Names Can Hurt You (by Cueing your Response to Incentives):
Experimental Evidence on Identity and Development

Karla Hoff* and Priyanka Pandey
The World Bank

This draft: April 1, 2011

We present an experiment that allows us to discriminate among several hypotheses about how identity affects behavior. In the experiment, boys from high and low castes in India solve mazes under piece rates and tournaments. Boys in castes that were traditionally Untouchable can solve mazes as well as high-caste boys and both can increase output under tournament incentives, compared to piece rate. Yet when caste identities are publicly revealed, performance generally declines but by a greater amount for the low caste and tournament incentives elicit zero or a perversely negative response. To explain these results, we argue that an individual has multiple worldviews (and of his place in the world) and that contextual changes cue a worldview that influences the preferences/behaviors he expresses. Our findings shed light on (a) why more salient caste boundaries appear to hurt school enrollment of both high- and low-caste children in South Asia, (b) why relative to rural air “[medieval] town air makes one free,” and (c) why groups fight over seemingly trivial rules of etiquette.

*Corresponding author: Hoff (email: khoff@worldbank.org). We thank members of the MacArthur Foundation Network on the Effects of Inequality, Jeffrey Carpenter, Paul DiMaggio, Ernst Fehr, Anjini Kochar, Vijayendra Rao, Tauhid Rahman, the late Ken Sokoloff, and Joe Stiglitz for helpful comments and discussions. We also thank seminar participants at Boston University, Brookings, Cornell, Delhi School of Economics, George Mason University, LSE, the Harvard-MIT development seminar, and the World Bank. We thank Muriel Niederle for advice at an early stage of this work. We owe a special debt to Anaka Narayanan and Ram Pratap for their assistance with data collection, and to Shweta Arya, Sonal Vats, and Sam Zhang for research assistance. This work was made possible by a grant from the World Bank-Netherlands Partnership Program, with additional support from the MacArthur Research Network on Inequality and Economic Performance.
I. Introduction

There are a number of hypotheses about how identity—an individual’s sense of the social categories to which he belongs—might affect behavior. We present an experiment that allows us to discriminate among some of these hypotheses. We argue that identity can have a large effect on the response to economic incentives by those at the top as well as at the bottom of a social hierarchy, but that the effect depends on context. Seemingly inconsequential contextual features appear to cue particular norms and meanings and in doing so to cause large changes in behavior.

Examples of the power of environmental cues to elicit changes in preferences/behavior are commonplace. Pavlov’s dog salivated at the sound of a bell because the sound was associated with consumption in the past. For a similar reason, cigarette smokers experience a heightened nicotine craving when they see an open box of cigarettes. Laibson (2001) presents a cue theory of consumption to explain why preferences change rapidly as cues change.

Our paper extends the literature on consumption-based cues by exploring the effect of identity-based cues in the context of a real effort experiment under incentives. Sociologists have argued that individuals have multiple, partly inconsistent worldviews and action frames (e.g., Swidler 1986, DiMaggio). Seemingly irrelevant changes in the environment can lead individuals to hold, at least momentarily, a different view of the world and of the “normal” way to behave.

The subjects in our study are 6th and 7th grade boys in rural North India (the state of Uttar Pradesh) drawn from the extreme ends of the caste hierarchy: individuals in the highest castes (subjects whom we will hereafter call H) and the lowest (hereafter, L). The latter, self-designated as Dalits, were subject to the traditional practice of untouchability. They make up 21 percent of the population of Uttar Pradesh. Caste is useful for studying the effects of identity for several reasons. First, caste is fixed by birth, so that there are no selection effects on identity.
Second, there are no physical markers associated with caste (Gupta 2000). Third, rural India is in what the sociologist Swidler calls a “culturally unsettled time,” in which widely divergent views of how society should be organized are in competition.

It will be useful to describe the traditional view. In this view, it is illegitimate for an L to acquire education, wealth, or the knowledge of the high castes. A fable from an epic dated around 1000 BC captures that idea (as reported in Jadhav, 2005, pp. 2-3). In the fable, a Dalit watches hidden from view a guru who is teaching archery to princes. One day the guru sees the Dalit shoot a deer bolting at lightning speed. Amazed at his precision, he questions him and discovers that the boy is a Dalit. The child wanted to offer the guru the traditional gift offered to a teacher in gratitude—so he offered himself as the guru’s slave. Instead, the guru asked the Dalit for his right thumb, in which rested all the prowess of archery. So he severed his thumb and laid the bleeding stub and the guru’s feet. According to Jadhav, every Indian child hears this myth, whose moral is that the outcaste is forever an outcaste. Untouchability continues to be practiced in almost 80 percent of Indian villages despite the constitutional abolition of Untouchability (Shah et al. 2006). Residential areas for L are still generally confined to the outskirts of villages, and in our survey of all households in a village near the experiment site, we find that half of both low-caste and high-caste respondents indicate that a Dalit would not sit in the presence of a high caste person. Children learn social structures through understanding such distinctions (Basso, ed.).

However, recent political movements in Uttar Pradesh have created vastly better conditions for Dalits (e.g. Narayan 2010 and Pritchett et al. 2010)—namely, that all people have equal rights. The Chief Minister of the state of Uttar Pradesh at the time of our experiment was a Dalit. Clearly, the traditional and modern views of the organization of society imply different
“normal” ways to behave.

In our experiment, H and L 6th and 7th graders solve mazes for 30 minutes (in two equal rounds) under piece rate and winner-take-all tournament incentives. We use three conditions to manipulate cues to caste. From all subjects, approximately one to two hours before the experiment, we asked the subject’s name, caste, father’s name, grandfather’s name, and village. (We asked more than caste to avoid experimenter demand effects.) In the control condition, we did not reveal this information, and each session had 3 H and 3 L. In the second condition, which we call “Caste Revealed,” we publicly revealed this information in the session, which again comprised 3 H and 3 L. In the third condition, “Caste Segregated,” we revealed the information in the same way as in Caste Revealed but now the session 6 H or 6 L.

Since a session was made up of children from several different villages, subjects in the control condition—sitting in an unfamiliar place with several unfamiliar peers--may reasonably believe that his peers do not know his caste. Subjects in “caste revealed” would lose their anonymity. Subjects in “segregated” would be cued for a caste-based organization of society.

We find that in the control condition, L solve puzzles as well as H under both piece rate and tournament incentives, and both groups respond to tournament incentives, solving significantly more mazes than under piece rate incentives.1 In contrast, in Caste Revealed, neither H nor L respond positively to tournament incentives. In Caste Segregated, H also do not respond to tournament incentives, while L respond negatively. We can calibrate the identity effect for L of Caste Segregated under a tournament: the effect more than reverses the incentive effect of the tournament in the control condition. Cueing caste and a caste-based organization of society generally impairs the performance of both H and L but the impairment is larger for L (as we

1 A brief summary of the results for piece rate, but not tournament, treatments appears in Hoff and Pandey (2006).
would expect from the literature on stereotype threat, surveyed below). Thus our identity treatments tend to reproduce caste inequality.

Although there is a long tradition of exoticizing caste, caste is just one of many examples of category systems characteristic of virtually all societies that represent achievements by groups that are low in the hierarchy as illegitimate (Gupta 2000). Other examples include, for particular societies in the past or present, gender identities, racial identities and class identities. When in 1607 the danger of shipwreck appeared so great to a sailor in the British Royal Navy that he warned the admiral of the error and thereby revealed his knowledge of navigation, the admiral “had the man hanged for mutiny on the spot” (Sobel, 1995, p. 13).² The words “uppity” and “social climber” capture the idea that certain people are viewed as having a “natural” place, departures from which are “unnatural.”

How much do social identities, e.g., of race or gender or medieval vassal, shape preferences/behavior? Historians such as Sabaens (1984, p. 59) and Dirks (2001) and more recently, economists such as Akerlof and Kranton (2000) and Austen-Smith and Fryer (2003) have posed this question. Our experiment with caste contributes to this literature.

Five models of the role of identity in preferences are described in Section II. Section III reviews the literature in psychology and sociology on the effect of cues on behavior. In Section IV we present the experimental design. Section V discusses our results in broad terms, and Section VI estimates treatment effects. Section VII discusses two studies that suggest that an increase in the salience of caste boundaries hurts educational outcomes of both Dalits and higher castes, which is consistent with our experimental findings. Section VII discusses implications of our findings for development.

---

²The sailor’s warning was well-founded. No course correction was made, and four of five warships were shipwrecked.
II. Models of Preferences, Behavior, and Identity

In this section we outline five hypotheses about how identity affects behavior.

*Identity has no effect.* The standard model of economics assumes the existence of an individual with fixed preferences in which any sense of identity with others influences neither values nor behavior; contrary to John Donne’s poem, each man “is an island entire of itself.”

This hypothesis is one of the fundamental differences between the standard model of economics and the conception of the individual that has increasingly been found useful in other social sciences, in which socially defined variables such as conformity affect economic behavior.

*Identity is an element of preferences.* The hypothesis that an individual has, at any moment in time, a well-defined set of preferences and that they are always salient is maintained in recent work that substantially broadens the notion of preferences by incorporating one’s sense of identity (Akerlof and Kranton 2000, Durlauf 2006). In Akerlof and Kranton’s model of preferences, an individual views himself as a member of a social category (or categories) that are an element of his identity. The categories affect choices because an individual likes conforming to the ideals of that category and dislikes actions by others that deviate from those ideals.

*Identity is a choice.* An individual chooses what his relevant identities are and also their relative importance when they pull in divergent ways; see, e.g., Laibson). This choice problem makes sense only under the assumption that an individual has a meta-utility function. This approach can give rise problems of welfare analysis but, just as in the two models above, an individual has well-defined preferences that provide all the information that is relevant for describing his behavior, *i.e.*, his choices.

*An individual has multiple, inconsistent worldviews and action frames.* An individual has a deep utility function that expresses itself automatically in one way or another if stimulated

---

3 Sen (2006, p. 20) cites this poem in his critique of the standard conception of man in economics.
appropriately, for example, by a cue to some way of seeing himself or the world. Since culture is contested and a given identity has many meanings, the effect of cues to identity will depend on context, that is, on what other cues are present in the situation. One way that other cues can matter is that they influence the subjective meaning of a given act because individuals may care about this meaning.

We can make an analogy to DNA. DNA are the set of instructions for making an individual, but poorly understood features of the environment determine which genes express themselves.

Where the idea of a deep utility function becomes interesting is that it leads to the observation of inconsistent choices. Of course, if we knew all the stimuli to the individual, then the hypothesis of rationality (i.e. consistency) would be trivial. Since we do not observe all stimuli, it becomes a useful construct to posit multiple preferences, one for each self-concept.4

Useful for what purpose? We will argue that it is useful for understanding long-run social change, which entails both changes in the salience of particular identities and changes in the set of possible identities and in the meanings those identities can give to actions.

The multiple identity concept is different from that of state-dependent utility and becomes useful when the stimuli to which an individual is exposed change in a way that leads to the expression of one set of characteristics rather than another, not under the control of the individual. That is, behavior depends on context. Of course, if we could solve the signal extraction problem and infer that an individual behaves one way under one set of stimuli and

4 In contrast, Laibson’s model of cue-based consumption posits a unitary self. Changes in preferences arise from complementarities between those cues and consumption; e.g. I want to a scotch when I hear the sound of ice tinkling in a glass.
another way under another set, then we would have discovered that person’s state-dependent utility function (as in Romer 2000). But by assumption, one cannot solve this problem because the world is too rich in signals. This hypothesis has been investigated largely outside the economics literature, as we discuss in the next Section. The psychologists Ross and Nisbett call this hypothesis “situationalism.”

*The social construction of identities.* This last hypothesis is not about the influence on preferences of a sense of identity with others, but instead about the social construction of identities. Attributes have a material existence, but the activation of attributes as identities depends on the cultural schemas that inform their use. Thus the eight-year old Claude Steele (2010, p.1) discovered his black identity when he learned that “black kids couldn’t swim at our local area park, except on Wednesday afternoons.” The categorization of individuals into socially recognized groups and the meanings assigned to those categories are outcomes of a social process. The outcomes are an equilibrium if they lead to beliefs that support the relevance of, and the distinctions among, those social identities. (Examples are in Glaeser 2005 and Hoff and Stiglitz, 2010, 2011.)

**III. Related Literature on “Situationalism”**

In this section, we discuss some of the work in psychology and sociology that identifies effects on behavior brought about by situational features that cue identities.

*Stereotype susceptibility.* Hundreds of studies since the seminal article of Steele and Aronson (1999) on “stereotype threat” have demonstrated that if a group’s ability is negatively stereotyped in some domain, then priming either that identity or that domain may impair the group’s performance in that domain (add review). For example, asking black and white college students to check a box for race before they take a test drawn from GREs leads to a black-white
difference that is just what one would expect based on the back-white difference in SAT scores, whereas in the presence of stereotype threat, the difference is larger than would be expected based on the difference in SAT scores. Racial categorization automatically activates the race stereotypes and prime them to affect judgment (see e.g. Blari and Banerji 1996 or Ridgeay Correll p. 516. Ambady et al. show that that making salient an identity positively stereotyped in some domain has the reverse effect, leading to improved performance by some measures (see also Ambady, Pittinsky and Shih, 1999; Stone et al. 1999). The stereotype susceptibility results have been observed in students in college, high school, middle school and elementary school (e.g. Ambady et al 2001; a review is Steele 2010).

The explanation for stereotype susceptibility might be that there is a norm about how difficult a task is to perform. An agent chooses a level of effort to exert to accomplish a task. A belief that one is low ability has a substitution effect and an income effect, and the experimental findings show that the substitution effect of a stereotype dominates.

However, even when there is no norm about how difficult a task, is making stereotypes salient affects performance. Stereotype threat has been demonstrated in experiments with unfamiliar tasks such as peculiar memory tests, Stroop tests, and miniature golf (Schmader and Johns, Darling et al). One way to explain this finding is by adopting a model of individuals with multiple identities whose salience can be manipulated. The manipulation changes an individual’s prior belief in the difficulty of the task—or of the payoffs from success or failure.

*Each identity has its own priors, its own heuristics.* A novel task that does not evoke anxiety when race is not primed, may provoke anxiety when race is primed. Experimental results on stereotype threat, by showing how race/caste/gender salience affects performance in novel tasks, provides evidence against the unitary self model. In these results, the expression in behavior of
identity depends not on fixed internalized beliefs that are always salient to the individual, but instead on the adaptation to particular situational cues (as emphasized explicitly by several authors, e.g., Steele 2010, p. 87 and Ridgeway and Correll, p. 518). A survey suggests that the declines in performance in situations of stereotype threat arise because of impaired working memory (Schmader-Johns).

The enactment of power. A recent set of experiments in psychology (Smith et al. 2008) finds that invoking in a participant the feeling that he has little power reduces the level of his performance in complex cognitive tasks. As part of that study, subjects were asked to unscramble a set of sentences cuing different feelings. In one treatment, the scrambled sentences included words related to lack of power (e.g., subordinate, obey). In another treatment, the sentences included words relating to having power (e.g, authority, dominant). In the control treatment, the sentences included no words related to power. Cueing feelings of lack of power in subjects reduced their ability to plan tasks, hold information in working memory, and inhibit goal-irrelevant information. Cueing feelings of power had the opposite effect. This finding is related to Bandura’s (1977) self-efficacy theory, in which mentally representing oneself as inefficacious is self-fulfilling.

Some identities, such as those based on race or caste, can be viewed as “enactments of power” because one way that power expresses itself is as social practices that define categories and then that subject people in different categories to different rules (Bourdieu).

The activation of “repertoires” of behavior. Many sociologists believe that individuals have the capacity to maintain “inconsistent action frames.” (e.g., DiMaggio). One reason for this is that

“All real cultures contain diverse, often conflicting symbols, rituals, stories, and guides to action…a culture is not a unified system that pushes action in a consistent direction.
Rather it is more like a ‘toolkit’…from which actors select differing pieces…individuals know how to do different things in different circumstances.” (Swidler, 276-77)

Situations shape the publicly available meanings that make some patterns of action natural while discouraging others.\(^5\)

Another way that cuing may shape behavior is that it influences individuals’ sense of what is normal behavior in a given situation (Bichierri). In turn, this matters because there is compelling evidence that people like to be viewed as normal for their group. Electricity use falls by an amount equivalent to a 13-34 percent increase in price when individuals learn that they are using more electricity than the average for their group (see Mullainathan). We are more likely to vote if we are told that most others in our group do so, even though when more vote, the less likely our own vote is to matter. When we are primed for ethnicity, some of our preferences shift in the direction of those of the group (Benjamin et al. 2010). The psychologists Dana et al. (2005) find in a cleverly designed Dictator Game that many individuals place positive value on meeting the expectations of others even when those others are anonymous and when no encounter between the actors and those with knowledge of the action will ever occur.

We conclude this discussion a speculative comment. In the real world, unlike the artificial world of an experiment, it may be adaptive to be aware of “one’s place” and to avoid situations in which one would feel “out of place.” This conjecture gains support from experiments that investigate the mechanisms underlying stereotype threat. Blacks in a task characterized as relevant to intellectual ability are more likely in a word completion task to fill in the missing letters in _ _ MB to spell DUMB rather than NUMB, whereas whites were not.

---

\(^5\) The power of framing and heuristics to affect judgment is illustrated by the following experiment (Kahneman et al.). Subjects are asked to estimate (a) how many 7-letter words of which the 6\(^{th}\) letter is “n” appear in the first 10 pages of *War and Peace* and later are asked to estimate (b) the number of 7-letter words of which the last 3 letters are “ing.” The average response to (a) is less than (b), which violates logic.
(Steele and Aronson). In conditions of stereotype threat, women performing a math test did not recruit the regions of the brain associated with mathematical learning, but instead showed heightened activation in a neural region associated with social and emotional processing (Krendl et al. 2008). It may be adaptive for individuals to respond to cues that they are not the dominant group in a situation by adopting a frame of mind in which others’ expectations of them are front and center.

IV. Experimental Design

In order to study the effect of identity on output under piece rate and tournament incentives, we undertook an experiment in which high- and low-caste junior high school boys in rural India were paid to solve mazes. About two hours before the maze-solving sessions began, a research assistant asked every participant his name, village name, father’s name, grandfather’s name, and caste. This interview occurred in private and a few minutes before the subject boarded the car that transported him to the experiment site. Three conditions varied the salience of this identity information:

Control. A session was composed of 3 H and 3 L subjects. The experimenter revealed no information about the subjects.

Caste Revealed. A session was composed of 3 H and 3 L subjects. The experimenter began the session by saying that she wanted to confirm some information with each subject, and that he should nod if it was correct. Then she turned to each subject and stated his name, village name, father’s name, grandfather’s name, and caste.

Caste Segregated. A session was composed of 6 H or 6 L subjects. The experimenter revealed the same information, in the same way, as in Caste Revealed.

The identity information was not limited to caste in order to reduce experimenter demand effects.

---

6 The mazes are level 2 mazes at [http://games.yahoo.com/games/maze.html](http://games.yahoo.com/games/maze.html). We provided subjects in each session with a packet of 15 mazes, enlarged to facilitate grading. Gneezy et al. (2002) and Carpenter et al. (2010) showed that people respond to tournament incentives in solving mazes.
The distinction between the control and experimental conditions is that subjects’ identities are publicly revealed, and so we call the latter the *identity treatments*. We interpret the last treatment as a cue to a caste-based organization of society because participants would have plausibly understood that segregation was not accidental but deliberate. Subjects from a single village were transported in the same car and contained an equal mix of H and L. The probability of six H (respectively, L) children from a random draw of six subjects from all 6th and 7th grade boys in this area was less than 0.002.

We now explain the incentive system. Subjects experienced one of two regimes:

- **Piece Rate in Both Rounds**: Participants play two identical rounds. At the beginning of the first round and again at the beginning of the second round, they are told that their reward is one rupee for every maze solved.

- **Piece Rate in Round 1 and Tournament in Round 2**:
  - In the first round, participants are told that will receive a reward (one rupee) for every maze solved
  - In the second round, participants are told that only the participant who solves the most mazes in the session (which includes six persons) will receive a reward. The reward is six rupees for every maze solved. The other participants in the group will not receive payment for their output.

The advantage of this design, which contrasts with the standard approach of only piece rate or only tournament incentives within a single treatment, is to permit within-subject as well as across subject comparisons by incentive, and also to give subjects unfamiliar with mazes a sure reward if they learned how to solve them. Note that incentive conditions faced by all subjects in the first round were identical, and so we will pool these observations in presenting the round-level results. Table 1 summarizes the design and sample size, 582 subjects in total. Each subject participated in only one treatment.
We recruited subjects from randomly chosen villages within an hour’s drive from Hardoi, in the state of Uttar Pradesh. Our goal was to find subjects who were representative of their age group and who would feel comfortable in a classroom setting. We chose as subjects 6th and 7th grade boys because we found that 95 percent of children of the lowest caste (Chamar) age 11-12 were in school, whereas the fraction of rural children attending high school in Uttar Pradesh according to the National Sample survey (1995-96) was only 59 percent. In the schools we visited, children appeared to be automatically promoted to the next grade level each year. We are not aware of any selection factors that make the children who were drawn from the 6th and 7th grades unrepresentative of all 11-12 year olds.

To recruit children, we visited homes each evening to ask parents’ permission to pick up their children the next day to drive them to the nearby junior high school that served as the site of the experiment. We also recruited from schools. We did not have selection problems because everyone was interested in participating. We told the parents that participants would be paid for

<table>
<thead>
<tr>
<th>Treatments</th>
<th>Revealed caste?</th>
<th>Caste composition of a session</th>
<th>Incentive per maze solved</th>
<th>Number of subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Round 1</td>
<td>Round 2</td>
<td></td>
</tr>
<tr>
<td>Piece Rate in both rounds under:</td>
<td></td>
<td>3 H and 3 L</td>
<td>1 rupee</td>
<td>1 rupee</td>
</tr>
<tr>
<td>Control</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caste Revealed</td>
<td>Yes</td>
<td>3 H and 3 L</td>
<td>1 rupee</td>
<td>1 rupee</td>
</tr>
<tr>
<td>Caste Segregated</td>
<td>Yes</td>
<td>6 H or 6 L</td>
<td>1 rupee</td>
<td>1 rupee</td>
</tr>
<tr>
<td>Piece Rate in round 1 and tournament in round 2 under:</td>
<td></td>
<td>3 H and 3 L</td>
<td>1 rupee</td>
<td>6 rupees, winner takes all</td>
</tr>
<tr>
<td>Control</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caste Revealed</td>
<td>Yes</td>
<td>3 H and 3 L</td>
<td>1 rupee</td>
<td>6 rupees, winner takes all</td>
</tr>
<tr>
<td>Caste Segregated</td>
<td>Yes</td>
<td>6 H or 6 L</td>
<td>1 rupee</td>
<td>6 rupees, winner takes all</td>
</tr>
</tbody>
</table>
showing up and paid additional rewards depending on performance. We explained that our purpose was to study children in India.

The experiment was conducted in unused classrooms in one of two schools. Upon arrival at the school, participants were randomly assigned to a treatment and were directed to a classroom where they waited in silence as a staff person entertained them. To the extent possible, we formed sessions of children from six different villages, but that was not always possible, and so we control for the number of other children in a session that a subject knows. When enough cars had arrived so that a session of six in the same treatment could be formed, participants were directed into a different classroom where they remained for the entire experimental session, which lasted about 70 minutes.

Each session proceeded as follows. The Experimenter gave participants the show-up fee of 10 rupees (a significant amount compared to the 6 rupee unskilled adult hourly wage) to drive home the fact that the children were playing for money. (The instructions are in the Web Appendix.) She illustrated how to solve a maze by first solving a trivially easy maze on a wall poster with an erasable surface, and then solving on another wall poster a maze of the same difficulty as the ones used in the experiment. The Experimenter gave participants five minutes to solve one practice maze. She explained the reward system just before round 1. She asked each of the six subjects in the session a question about the reward system using a hypothetical example of scores, and did not proceed until the subject had answered a question correctly. She explained that the performance and earnings of each subject would be revealed only to him. After the first round, the experimenter gave no feedback on performance, explained that there would be one more 15-minute period of solving mazes, and then explained the incentive system
and again tested every subject on his understanding of the incentive. After the session, participants provided personal information in an exit survey, and mazes were graded blind.

To promote trust, participants received a piece of fruit on entering the room of the experiment, a “consolation prize” of two rupees if they solved no mazes, and full awards to all winners who tied in a tournament. Average earnings including the show-up fee were slightly less than one-half a day’s adult wages: 18 rupees for subjects in the treatments with only Piece Rate, and 21 rupees for the treatments with Tournament in round 2. However, the top performers in the treatments with Tournament earned 110 rupees.

Based on our description of the treatments, the predictions of the literature on identity, discussed in Sections II and III, are fairly clear. If preferences are fixed and always salient, then reminding an individual of his caste or revealing it to his peers should have no effect. If stereotype susceptibility influences preferences, then making caste salient should reduce L output and, if anything, increase H output, regardless of the incentive.

If, however, the mechanisms in play include the activation of one out of multiple, inconsistent worldviews, then for H the predicted effect of the identity treatments is ambiguous. On the one hand, under the caste system it is an ideal for H to be superior. The incentive to achieve that ideal should boost H performance. On the other hand, the identity treatments reduce the incentive for effort if they cue a worldview in which society is organized by caste and so mobility occurs primarily at the level of the caste, not the individual.

Consider finally the predictions of cue theory for L. In the caste system, efforts towards upward mobility by L are illegitimate. In the Laws of Manu, it is written that “A man of low caste who through covetousness lives by the occupations of a higher one, the king shall deprive of his property and banish.” The attitudes underlying these rules are still present in villages. For
instance, Gupta (2000) writes that high-caste men regularly beat low-caste men to keep them “in their place.” Given this, we believe that this theory predicts for L an interaction between the identity treatments and Tournament. The identity treatments change the meaning of excelling. Excelling in a meritocratic society has a positive affective response (see the discussion in Carpenter et al. 2010). In contrast, excelling in a caste system might elicit a negative affective response for an L. Table 2 summarizes the theoretical predictions that we will test.

**Table 2. Predicted Effects of Identity Treatments**

<table>
<thead>
<tr>
<th>Theory</th>
<th>Predicted effect of Identity Treatments on output of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High caste</td>
</tr>
<tr>
<td>1. Standard neoclassical model and models with identity as a fixed element of preferences</td>
<td>None: preferences are fixed and salient</td>
</tr>
<tr>
<td>2. Stereotype susceptibility and enactment of power</td>
<td>Output rises: self-confidence and focus increase</td>
</tr>
<tr>
<td>3. Individuals have multiple, inconsistent action frames and worldviews</td>
<td>Ambiguous change: Cues an identity whose norm is to be superior, which increases effort, but cues a worldview in which rewards depend on status and power rather than output, which decreases effort</td>
</tr>
</tbody>
</table>

**V. Descriptive Statistics**

We begin by describing the subjects and the broad patterns in the results. We conducted experimental sessions in January and March 2003 and March 2005. In each time period, at least 36 subjects participated in the pure Piece Rate-Control treatment. As shown in Table A-1 in the Appendix, there were no significant differences by time period. Therefore we pool the data across the three time periods.
We collected data on parents’ education and occupation since that might be expected to influence children’s cognitive development and social status. As shown in Table 3, across all treatments 45 per cent of H subjects have a mother with at least primary schooling, but only 12 per cent of L subjects do. A similar caste gap in the education of subjects’ fathers is found. The fraction of subjects both of whose parents are illiterate is less than 5 percent for H but 28 percent for L. Only 8 percent of H subjects have fathers who are wage laborers, but 18 percent of L subjects do. For both caste groups only about 5 percent of mothers work outside the home.

We collected data on two other control variables in the post-play survey. We were interested in the effect on performance of prior exposure to mazes. 12 percent of H subjects and 11 percent of L subjects in our experiment had not been exposed to mazes before. We were also interested in the effect of variations in the degree of anonymity. Therefore we collected data on the number of other children in the sessions that a subject knew.

The randomization of subjects between the control and identity treatments was largely successful as measured by $t$-tests. Columns (3) and (6) report the $p$-values for differences by caste in subjects assigned to control or identity treatments.\(^7\) Significant differences appear only with respect to the exposure to mazes (exposure is higher in the identity treatments), and with respect to the fraction of mothers of L subjects who have at least a primary education (the fraction is higher in the identity treatments). These differences, as well as the caste differences in observable characteristics, highlight the need to examine whether any of these characteristics influence our treatment effects. For example, low socioeconomic class is a source of stigma, and stigma affects intellectual functioning in stereotype threat conditions (see Claire and Croizet 20**). In that case, the effect of our identity treatments could be mediated by the effect of

\(^7\) A test of the randomization across all six sets of experimental conditions is reported in the Web Appendix.
priming *class* rather than *caste*. To test that, it would be important to the extent possible to control for correlates of class.

**Figure 1. Average Output by Experimental Conditions and Caste**

Figure 1 shows the mean performance by caste and round under each set of experimental conditions. The data is reported in groups: round 1 data are the first group, round 2-Piece Rate is the second, and round 2-Tournament is the third. Across all three groups, H and L are equally good at solving mazes when caste is not revealed, and output is higher under Tournament. For example, in round 2 the mean H output was 5.86 under Piece Rate and 6.83 under Tournament. The comparable L outputs were 5.53 and 6.80. A *t*-test indicates that the caste differences are in all cases insignificant (*p > 0.10*) under the control conditions and that the responses to Tournament are significant.
Figure 1 suggests that the experiment was successful in eliciting from L many differences in output that make sense in light of earlier results on stereotype threat. Across the three groups in the figure, L subjects perform worse in both identity treatments than in the control. The effects in round 2 are large: for example, L output in Piece Rate-Caste Revealed is 23 percent lower than under the control ($p<.05$). The comparable treatment effect on H is a small and statistically insignificant increase in output. The net effect of these two changes is a significant caste gap in round 2, indicated by the dotted line.

For H, the effect of shifting from the control to Caste Segregated reduces output in all three groups in the figure. The declines are statistically significant in round 2. Stereotype susceptibility does not predict this.

We conclude our discussion of descriptive statistics by showing in a more disaggregated way the ability of the identity treatments to differentiate H from L subjects. Figure 2 shows the ratio of L subjects to all subjects whose second round score ranked them at or above a given decile. If L subjects were equally represented throughout the achievement distribution and if the experimental conditions affected both L and H castes in the same way, then we would expect all graphs in the figure to be straight lines at 0.5 (i.e., any cut of the distribution would have a proportion of L subjects equal to about 0.5). The figure shows that in the control condition, under Piece Rate and Tournament, respectively, L subjects comprise 0.56 and 0.66 of the top decile. In contrast, these figures fall to 0.16 and 0.42, respectively, in Caste Revealed. This suggests that Caste Revealed creates a social context that eliminates the top achievers among L subjects, whereas it does not do so for H. Recalling from Figure 1 that under piece rate incentives, Caste Segregated depresses average output about the same amount for H and L, the
graph for that treatment, lying almost at the halfway line, suggests that this same result holds true across the ability spectrum.

Figure 2. Proportion of Low-Caste Subjects with Round 2 Output at or above Each Decile (Cumulative)

Note. There is, in general, more than one subject whose performance ranks him at the border between two deciles. In that case, we calculated the proportion of the low caste among subjects whose performance was exactly the decile performance, and allocated low caste subjects in this proportion to both sides of the boundary.

VI. Econometric Analysis

We estimate the following model of an individual’s output in a round:

\[
\text{Output} = \alpha + \omega \cdot \text{(round is 2)} + \beta \cdot \text{(subject is H)} + \gamma \cdot \text{(session cues identity)} + \\
\delta \cdot \text{(subject is H * session cues identity)} + \tau \cdot \text{(Tournament)} + \lambda \cdot \text{(Tournament * subject is H)} + \theta \cdot \text{(Tournament * subject is H * session cues identity)} + \mu \cdot Z + \text{error}
\]
where \( Z \) is a vector of variables measuring individual and family characteristics of the participant. The omitted category in the regressions is an L in round 1 under the Piece Rate-Control condition, and thus the constant \( \alpha \) measures the output level for an L under those experimental conditions, controlling for other variables. The next seven coefficients are measures of the round, caste and treatment effects:

- \( \omega \) measures the difference between output in rounds 1 and 2.
- \( \beta \) measures the difference between an H and L subject’s output in Piece Rate-Control.
- \( \gamma \) is a vector that measures the difference between an identity treatment (Caste Revealed or Caste Segregated) and the control. Thus, letting \( R \) and \( S \) correspond to the elements of the vector for Caste Revealed and Caste Segregated, respectively, \( \alpha + \omega + \gamma_S \) indicates the output level of L in round 2 of Caste Segregated-Piece Rate when we control for individual characteristics.
- \( \delta \) is a vector that measures the difference between an identity treatment for an H compared to an L. Thus, \( \alpha + \omega + \beta + \gamma_S + \delta_S \) indicates H output level in round 2 in Caste Segregated-Piece Rate, controlling for individual and family characteristics.
- \( \tau \) measures the difference in output between the Tournament and round 2 of Piece Rate for an L.
- \( \lambda \) measures the caste difference in the response to Tournament.
- Finally, \( \theta \) is a vector that measures the interaction between Tournament and the identity conditions. This means that H output in Tournament-Caste Revealed is \( \alpha + \omega + \beta + \gamma_R + \delta_R + \tau + \lambda + \theta_R \), controlling for other characteristics.

Table 3 reports OLS coefficients and robust standard errors clustered at the level of the individual. Regression (1) uses only treatment and caste indicators. Regression (2) adds controls for individual characteristics: grade in school, previous exposure to mazes, and the number of others in a session known to a participant. The addition of these controls reduces the estimated treatment effects under Piece Rate. Between regressions (1) and (2), the only change in the set of significant treatment effects is that the effect of Caste Revealed on L is no longer significant.
Since the interaction terms make it difficult to assess the treatment effects and caste gaps, we report these in Table 4, columns (1) and (2). We can see from this table that the estimated caste gap in Piece Rate-Caste Revealed is still significant (as it was in Figure 2 for round 2), but now the treatment effects of Caste Revealed on each caste—both the movement up by H and the movement down by L—are individually insignificant. In contrast, Caste Segregated always reduces output by a large and significant amount—for both H and L under both Piece Rate and Tournament ($p<.05$).

Consider next the effects of shifting from Piece Rate to Tournament. These are shown in the last block of the table and in Figure 3. It is easy to see that both castes respond positively and significantly to Tournament under the control conditions, but neither does so under the identity conditions. The response by L subjects becomes perversely negative under Caste Segregated, with output falling by 1.6 compared to Piece Rate ($p=.001$).

**Figure 3. Predicted Output under Piece Rate and Tournament by Caste and Condition**

![Predicted output graph](image)

*Notes.* Based on Table 4, column (2). Treatment effect based on this regression are in the last three rows of Table 5, columns (1) and (2).
Does class, rather than caste, underlie the effects of the identity treatments. To check this we consider indicators of class. Examples of possibly relevant class variables, reported in column (3) of Table 3, are the education of parents and their dependence on earning a precarious living through daily wage labor. If class explains our results, then we would expect education and a non-wage laborer to have positive and significant coefficients. In the post-play survey, we asked about parents’ occupation and then formed a binary variable for wage labor, recording values as Yes or No. We find in column (3) that the only parental variables that have explanatory power after controlling for caste are father’s education, but that contrary to the hypothesis that class stigma mediates our results, the regression results mean that having a father with some primary education rather than none reduces output by one-fifth of a standard deviation. The additional contribution of parental variables over and above caste and treatment effects is not different from zero by an $F$-test ($F(6,486)=1.58$, $p=0.11$). On the basis of this regression, we cannot reject the hypothesis that parental variables have no effect. It might be, however, that parental variables matter for the low caste but not the high caste because parental education alleviates caste stigma. Therefore in unreported regressions, we rerun specification (3) separately for H and L subjects. We still find, however, that parental variables have little explanatory power. We also checked for the effect of having both parents illiterate. We find that this is not significant (result not shown).

In a village, the most important source of power and status is generally landownership. We could not ask our generally 12-14 year olds subjects the landholdings of their parents, but for a sample of 278 subjects we obtained landownership information directly from parents and in case of doubt checked with the land records’ office. We find that that the effect in all treatments
is insignificant, very small, and sometimes has the “wrong” (i.e. negative) sign (see Hoff and Pandey, 2004, p.51)

Up to now we have considered treatment effects on output. In Table 4, column (4), we reestimate the model but use as the dependent variable the change in output of a subject between rounds. The treatment effects based on this model are reported in Table 5, columns (3) and (4). Since almost 90 percent of subjects had not seen mazes before participating in our experiment, there is a lot of scope for learning by doing. The differences between rounds are positive in all treatments. For H, Table 5 shows that Caste Segregated significantly reduces the change in output between rounds when the second round is Piece Rate, but not when the second round is Tournament. In contrast, for L, the treatment effect is about twice as large when the second round is Tournament rather than Piece Rate (-1.22 vs. –2.33; p<.01 in each case). We emphasize this because it shows a pattern in which the interaction between Tournament and Caste Segregated is very different for H than for L. Figure 4 above show this for levels and Figure 5 shows it for the differences in levels between rounds.

We can decompose our effects into a discrete effect—does a subject learn how to solve a maze?—and a marginal effect on performance. (Logit table to be added here, showing large effect of Caste Revealed on L’s failure rate.) In Table 6 we report the treatment effects, controlling for individual variables, using only those subjects that solved at least one maze during the 30-minutes of maze-solving (the underlying regression is in the Web Appendix). Table 5 shows that the treatment effects under Piece Rate-Caste Revealed are no longer significant at p<.05; this is consistent with our finding in the logit regression of a large treatment effect on the failure rate. But the effects of the identity treatments under Tournament are little changed and remain significant: $p < .05$ for Caste Revealed and $p < .01$ for Caste Segregated.
Figure 4. Predicted Change in Output Between Rounds by Caste and Condition

Notes. Based on Table 4, column (4). Treatment effects based on that regression are in the last three rows of Table 5, columns (3) and (4).

We conclude this section with a comparison of our findings with the predictions in Table 2. The hypothesis that identity does not matter for preferences, or that it does matter but that preferences are fixed and salient, is refuted. Stereotype susceptibility can explain many of our results, but not the reductions in output by H under Caste Segregated nor the interaction between caste, Caste Segregated, and Tournament. We conjecture that competition over rank does not trigger an increase in effort in the identity treatments because of the restraining effect of the
ascriptive system, in which one’s social status in life is pre-determined. This affects both H and L. But the effect is larger for an L. For L, Tournament gives rise to a more than offsetting influence from caste identity, since not distinguishing oneself is the prescribed role in life for a Dalit. It may be that a cue to a caste-wise organization of society leads to the expression of preferences to conform to the prescribed social role under that system.

VII. Discussion

The experimental results imply that identity cues can erect invisible restraints on achievement by those at the top of the social hierarchy and even more by those at the bottom. Our results are more credible if observational studies support them. Jacoby and Mansuri (2010) study settlement-level data of population and school enrollment in Pakistan. They find that low caste children are less likely to enroll in school than the average child if the children live in a predominantly high caste hamlet and there is a local school in the settlement. The opposite is true if the low caste children live in a hamlet that is not predominantly high caste. The implication is that low caste children’s schooling is depressed not by an inherent problem of caste, but rather by the situation that they find themselves in when they live in a predominantly high caste hamlet.

There is also evidence from India that investment in education by the high caste responds to changes in the salience of the caste divide. Kochar (2009) finds that an Indian government program that boosted low caste schooling by constructing schools in low-caste hamlets unexpectedly also caused significant increases in schooling by other castes. The result was that aid that was successfully targeted to only the low caste did not narrow the educational achievement gap between that group and others. One interpretation of this outcome, based on our experimental findings for high caste subjects in Caste Segregated, is that a situation that
affirms the superiority of high castes reduces their incentive for effort. Policies that narrow the caste gap in achievement would plausibly weaken the high caste’s sense of security in their ascriptive position and thus increase their incentive for effort.⁸

Our experimental results help explain an idea among historians called a “cultural technology of rule.”⁹ For instance, Dirks (2001, p. 9) writes that the caste system was such a technology, that the British strengthened it, and that it sustained British power in India as much as the power of superior arms did. In *The Invention of Tradition*, Ranger (1983) describes how the British in Kenya taught the indigenous population how to behave as the subordinate class by selecting and modifying ideas from local culture. (Members of the British colonial service who contributed to inventing the rituals that taught this ideology could thereby advance their careers.) In light of the theory of Swidler (1986), we can characterize a “cultural technology of rule” as the molding not of a person’s preferences so much as the molding of the “toolkit” or “repertoire” of ways of behaving that are automatically cued by caste.

**VIII. Conclusion**

The broad question that this experiment addresses is the extent to which changes in behavior can be explained by changes in social contexts that cue particular aspects of identity. We find strong evidence that the interplay between categorization processes (identities) and situational cues has first-order importance for economic responsiveness, learning, and thus for output. We highlight four implications.

---

⁸ We thank Anjini Kochar for bringing this result to my attention.
⁹ Studies of the construction of social identities in discriminatory regimes—first, the elaboration of categories and second, the ranking of those categories of individuals with respect to inherent moral and social qualities—include Myrdal (1944), Lorcin (1995), Dirks (2001) and Benton (2002). The subordinate group, lacking equal access to education and law, cannot easily defend itself against a racial ideology. In emphasizing the role of social identities in the persistence of group inequality, we follow Loury (2002, see esp. pp. 20, 58), but the mechanisms he analyzes are different from the one we study here.
Nonseparability between incentives and preferences. Institutions affect not only rules but also preferences because they change the meanings attached to certain actions and individuals’ concept of themselves. This implies that adopting the rule of law has a larger effect on outcomes than a standard analysis would suggest. The change in incentives affects actions and social identities. And the change in social identities has an additional effect on actions, making individuals more responsive because it becomes legitimate for everyone to compete and to win.

Institutional overhang. Since worldviews cannot be changed at the stroke of a pen, the preceding result implies that institutions affect outcomes long after the rules have been abolished and the forces that gave rise to them have changed.

Identity as a causal factor in development. The linkage between identities and preferences suggests the possibility that changes in category systems defining the possible set of identities affect development. McCloskey and Mokyr (2010) argue that the invention of the identity of a gentleman as someone who does not cheat his partners in business was a causal factor in the Industrial Revolution. A striking recent example are the identities created as a byproduct of government efforts in China’s urban policies. In China, a class of second-class citizens emerged in response to the household registration system (hukou) set up by government to control population movement to urban areas. An individual’s hukou is generally based on the birthplace of his parents. Because of this system, many children in Beijing have a rural migrant hukou status. Such individuals have a stereotype as “hicks”—rural at least by background, poor, and unsophisticated. In a replication of our experiment, Afridi, Li, and Ren (2010) find that

---

10 That is, institutions operate both as symbolic systems and as rules and patterns of activity (e.g. Friedland and Alford 1991). Recent work in economics suggest several other reasons why the separability breaks down. Incentives affect social preferences; e.g., the desire to give to charity or to respect the commons is greater without any monetary incentives than with low monetary incentives (e.g. Gneezy et al.). A second example is that the incentive offered by a principal may signal his belief in the agent’s ability and thereby affect the agent’s own belief in himself (Benabou and Tirole). A survey and analysis of evidence for non-separability is Bowles (work in progress).
priming that stereotype and making individuals’ hukou status public reduces by 10 percent the number of mazes solved by 8-12 year old children with a rural migrant hukou, while it slightly improves the performance of those children with an urban local hukou. This study shows that a system of two classes of citizens creates inequality in a domain—solving mazes—in which being a hick should not matter.

*Changes in context as a causal factor in development.* As Swidler (2001, p. 280) argues, and our results suggest, in a culturally unsettled time context will determine which style of action takes root and thrives. This can explain a famous hypothesis of the medievalist Henri Pirenne: when towns emerged in the medieval period, two changes occurred: “It was not only that the serf, having escaped from the countryside, found legal freedom in the towns, but that the whole social atmosphere there was open to ambition and talent..” (Cipolla, 1976, p. 119). This idea is captured in the phrase, *stadtluft macht frei.* A modest policy implication is to make caste less salient in Indian schools so that children while in school escape the restraining influences of caste on performance.
References (not complete)


Benabou, Roland and Jean Tirole


McCloskey “Bourgeois Virtue.”


Romer, Paul. “Thinking and Feeling.”


