Stereotype Threat and the Student-Athlete

August 2012

Abstract - Achievement gaps may reflect the cognitive impairment thought to occur in evaluative settings (e.g., classrooms) where a stereotyped identity is salient (i.e., stereotype threat). This study presents an economic model of stereotype threat that reconciles prior evidence on how student effort and performance are influenced by this social-identity phenomenon. This study also presents empirical evidence from a framed field experiment in which students at a selective college were randomly assigned to a treatment that primed their awareness of a negatively stereotyped identity (i.e., student-athlete). This social-identity manipulation reduced the test-score performance of athletes relative to non-athletes by 12 percent. These negative performance effects were concentrated among male student-athletes who also responded to the social-identity manipulation by attempting to answer more questions.

Keywords: Social Identity, Stereotype Threat, Education, and Athletics
JEL Classifications: I2, C9, D0

1

I. INTRODUCTION

The recognition that "nonpecuniary motivations" play an important role in economic decision-making extends back to the early and influential research on discrimination by Becker (1957). However, economists have only recently begun to explore explicitly the behavioral and welfare implications of social identities (i.e., a person's "sense of self" with respect to membership in a particular social category). One prominent example is the study by Akerlof and Kranton (2000), which incorporates social identity into a general model of behavior and demonstrates its implications for a variety of economic outcomes. They argue, for example, that the stark patterns of occupational segregation by gender persist because individuals experience "anxiety and discomfort" when their occupation is inconsistent with the pre-existing "behavioral prescriptions" of their gender identity (e.g., male nurse, female trial lawyer). Co-workers may also experience discomfort from (and even seek to retaliate against) peers in gender-atypical occupations, encouraging firms to reinforce pre-existing gender-job associations (Akerlof and Kranton 2000). Similarly motivated economic models (i.e., recognizing individuals' interest in choosing behaviors that affirm their social identity) have been applied to a diverse set of other topics as well including health behaviors (e.g., women and smoking), racial inequality in labor-market outcomes (e.g., concerns about "acting white"), economic cooperation, and human-capital investments (Akerlof and Kranton 2002, 2010; Bénabou and Tirole 2006; Benjamin, Choi, and Strickland 2007; McLeish and Oxoby 2011).

However, this emerging social-identity literature has paid relatively little attention to the prominent social-psychology literature on the phenomenon known as "stereotype threat" (Steele and Aronson 1995). Stereotype threat refers to the perceived risk of confirming, through one's behavior or outcomes, negative stereotypes that are held about one's social identity. More

specifically, its key conjecture is that the threat of being viewed through the lens of a negative stereotype can create an anxiety that disrupts cognitive performance and influences outcomes and behaviors. In this study, I present a simple one-period economic model of stereotype threat adapted from the social-identity model introduced by Akerlof and Kranton (2002). This model reconciles seemingly contradictory results in the extant empirical literature by illustrating how the effects of stereotype threat on effort and performance depend on the complementarity of ability and effort as well as other context-specific factors. This model of stereotype threat differs from recent economic models of social identity in a straightforward but conceptually important detail. Previous economic models of identity have viewed individuals as choosing behaviors that correspond to the social norms for an identity (or set of identities) that they unambiguously view as their own. However, individuals experiencing stereotype threat do not necessarily feel that they personally do (or should) subscribe to the stereotyped traits of a social identity. Rather, it is the apprehension that *others* view them through the lens of a negative stereotype that is conjectured to create anxiety that compromises cognitive functioning and, eventually, identification with the stereotyped domain.

This study also presents the results of a "framed field experiment" (Harrison and List 2004) that focuses on manipulating the stereotype threat associated with a particular social identity: that of a student-athlete at a selective post-secondary institution. Most of the empirical evidence in support of the stereotype-threat phenomenon comes from laboratory experiments in which student-participants are randomly assigned to receive a treatment that "primes" their awareness of a stereotype prior to completing a test or some other task.[1] For example, in the seminal laboratory study by Steele and Aronson (1995), participating students were randomly

---

[1] See Aronson and Steele (2005) for a discussion of this growing literature. Interestingly, there is also a small but growing body of field evidence that finds effects from interventions designed to buffer students from the effects of stereotype threat.

assigned to be told that the test they were about to take was diagnostic of their ability (i.e., the stereotype-threat prime) or that the test was non-evaluative (i.e., the control condition). They found that black students in the "ability-diagnostic" condition performed significantly worse on tests than those in the control condition while the performance of white students was not significantly affected by how the test was framed. In another widely used variant of this study design, participants would first complete a brief questionnaire that included questions designed to prime their awareness of a racial or gender identity (e.g., Steele and Aronson 1995, Shih, Pittinsky, and Ambady 1999, Benjamin, Choi, and Strickland 2007).

The experiment presented in this study adapted this design to evaluate whether priming student awareness of their athletic status leads to achievement gaps. Two other recent studies (Yopyk and Prentice 2005; Harrison et al. 2009) present similar evidence based, respectively, on male students at Princeton University (i.e., 67 athletes and a cappella singers) and on student athletes at two large state universities (n = 88). In general, these studies suggest that manipulating awareness of an athletic social identity reduces cognitive performance, though there is also evidence that the test performance of male student-athletes *improved* on a more difficult test as a result of an identity manipulation (Harrison et al. 2009). The empirical evidence presented here contributes to this limited and somewhat contradictory evidence in several important ways. Most notably, an active literature in economics has recently engaged the question of whether (and under what conditions) the inferences drawn from laboratory experiments generalize to real-world settings (Levitt and List 2007a, 2007b; Falk and Heckman 2009; Levitt, List, and Reiley, forthcoming). The experimental design in this study reflects such concerns about "external validity" in two distinctive and intentional ways. First, students were recruited into the study without any explicit screening or indications that student athletes were

the focal point of the study. Second, the priming mechanism used in this study more closely resembles how an athletic social identity is manifested in field settings at selective institutions (e.g., questions about scheduling conflicts with seminars and labs). In contrast, the priming mechanism in the Yopyk and Prentice (2005) study (i.e., writing detailed comments on athletic participation) is unusually strong and does not necessarily parallel how a social identity as an athlete is likely to be manifested in real-world settings. Furthermore, the student athletes in the Yopyk and Prentice (2005) study were explicitly recruited from specific teams (i.e., ice hockey and football). The knowledge of this targeted study recruitment may have made the subjects exceptionally susceptible to the threat prime. Similarly, Harrison et al. (2009) recruited only student athletes and informed participants *ex ante* that the study was designed to improve their classroom performance, possibly allowing identity awareness to be primed in an unnaturally strong manner.

The experimental component of this study also contributes to the literature by addressing heterogeneity by gender in the effects of an identity threat. How the effects of an athletic-identity threat might vary by gender is a distinctly empirical question. If female student-athletes have strong academic self-identities but are still subject to the "dumb jock" stereotype, the performance implications of stereotype threat may be particularly large for them (Harrison et al. 2009). However, the "dumb jock" stereotype may be less relevant for female-athletes because they have stronger academic identities as well as because they do not play "high profile" sports for which a strong admissions advantage is thought to exist. In this case, the hypothesized effects of stereotype threat on both performance and effort would be larger among male student athletes. Another important design feature in this experiment concerns the random-assignment procedure. The earlier study by Yopyk and Prentice (2005) used simple randomization over a small number

5

of subjects and found a "failure of randomization" such that participants in a control condition had significantly higher SAT scores than other participants. This suggests the possibility of bias that would confound that study's main finding. Those who did not receive the stereotype-threat prime may have outperformed those who did simply because they had unobserved traits that predisposed them to do so. This study eschews simple randomization in favor of a "block randomization" procedure that leverages baseline traits to reduce the likelihood of imbalance across experimental conditions.

II. STEREOTYPE THREAT AND COLLEGE ATHLETICS

Akerlof and Kranton (2002), in an extension of their seminal economic analysis of identity (Akerlof and Kranton 2000) to schooling, present models in which students choose a particular social identity (e.g., leading crowd, nerd or burnout). Their subsequent utility is determined by the social status of one's chosen identity and by how well an endowment of traits (e.g., appearance, intelligence) and a chosen level of effort allow one to approximate the ideal of that chosen social identity. However, models of this type do not correspond exactly with how social psychologists conceptualize the interaction of stereotype threat and social identity. In particular, stereotype threat is not about conforming to the ideals of what an individual actor perceives as their salient social identity. Instead, the salient feature of stereotype threat is the apprehension and diminished cognitive performance that may be created by the suspicion about how one is viewed by *others*.

A simple extension of their baseline model illustrates how stereotype threat may influence student effort and outcomes. Specifically, an individual's utility, $wk(n,e) - c(e)$, reflects the return to performance, w, a performance level, $k(n, e)$, that is a function of ability, n,

6

and effort, *e*, and the disutility of expending effort, *c(e)*.[2] This model can be extended to capture the influence of stereotype threat on student performance by making the ability term, *n*, a decreasing function of situational threats, *t*, that create this anxiety (i.e., $n_t < 0$). A simple model of stereotype threat would then assume that an individual chooses a level of effort to maximize $wk(n(t),e) - c(e)$. In this model, stereotype threat influences student performance (1) because it reduces the efficacy of effort and (2) through its effects on the chosen level of effort, e^*. So, what does this model imply about the effect of stereotype threat on e^*? The relevant comparative static based on this model of student effort can be shown to take the following form:

$$\frac{\partial e^*}{\partial t} = \frac{wk_{en}n_t}{-(wk_{ee} - c_{ee})} \underset{<}{\overset{>}{=}} 0$$

The denominator of this expression is positive by the second-order condition. Therefore, given the defining assumption that stereotype threat decreases the productivity of effort (i.e., $n_t < 0$), the chosen level of effort will fall only if effort and ability are complements in the production of performance (i.e., $k_{en} > 0$). The intuition for this insight is straightforward: stereotype threat is, in effect, a negative ability shock that simultaneously compromises the return to complementary inputs like effort and thereby reduces the amount of effort chosen. However, this result also illustrates how the conjectured effect of stereotype threat on effort can depend critically on how subjects view the nature of the task under consideration. In particular, if an increase in effort is seen as a substitute for a negative ability shock (i.e., $k_{en} < 0$), then the introduction of stereotype threat should unambiguously increase the optimally chosen level of effort.

---

[2] Akerlof and Kranton (2002) extend this simple model by introducing additional arguments that reflect the returns and costs associated with particular social identities that are available to choose. However, stereotype threat is not directly about the consequences of choosing an identity. Instead, the key feature of stereotype threat is the cognitive disruption from situational threats due to concern about how one is viewed by others (e.g., being an athlete in classroom at a highly selective institution).

7

This insight provides a way to reconcile the small number of seemingly incongruous results observed in some stereotype-threat studies. In particular, three gender-based studies suggest that stereotype threat increased effort and motivation even by an amount sufficient to improve overall performance despite the reduced productivity of effort. Oswald and Harvey (2000) found that female undergraduates exposed to a hostile cartoon prior to taking a math test actually performed better when there was no attempt to reduce stereotype threat. Similarly, in a study of a visual-attention task, Jamieson and Harkins (2007) found that gender-related stereotype threat increased both the effort and performance of women but only when the experimental setting facilitated the opportunity to correct mistakes through increased effort (i.e., by allowing additional time). Third, a recent study by three economists (Fryer, Levitt, and List 2008) found that, despite reporting higher levels of stress, females performed at their best when primed to be aware of math-related gender stereotypes. A rational substitution of effort in response to the ability shock of stereotype threat provides one possible explanation for the small number of seemingly anomalous findings. For example, subjects who view the presence of increased stereotype threat as particularly illegitimate may be more likely to see increased effort as an attractive substitute (e.g., an "I'll show them" attitude).

Such contextual factors could have particular relevance for stereotype-threat effects related to athletic status at elite post-secondary institutions. In particular, this model implies that the relevant question is whether student-athletes *perceive* their intellectual effort as substitute or complement to their native ability. We cannot assess the character of such student perceptions directly. However, the "threatened" student-athletes at selective institutions typically have high-performing academic backgrounds (relative to their high school peers) and are likely to have a strong identification with academics. Their prior academic success, which was sustained while

8

participating in extracurricular activities, could plausibly lead them to view additional effort as a readily available substitute for enhancing their performance.[3]

A second limitation of this one-period model as well as the corresponding experimental evidence is that they do not capture the dynamic feedback mechanisms by which an identity threat could have amplified (or attenuated) performance implications over time. The social-psychology literature has stressed the importance of such "recursive" processes with respect to stereotype threat (e.g., Yeager and Walton 2011). For example, students subject to identity threats and diminished performance may choose to align themselves with a social identity that places less value on academic achievement (e.g., fraternities). In contrast, both institutional practices and social norms may be able to block this sort of preference formation by "buffering" students with a culture of unequivocally high expectations. Capturing such dynamic and contextual elements of identity threats in a more general model would be a potentially insightful extension of this basic framework presented here.

III. A STEREOTYPE THREAT EXPERIMENT

The experiment described here tests whether stereotype threat appears to contribute to the academic underperformance of college student-athletes by implementing a conventional stereotype-threat experiment but with the distinction that an athletic identity, rather than identities related to race or gender, is primed. As with any experiment, the external validity of these results for other populations and real-world settings is an important issue. For example, there are at least two reasons to suspect that a study of this type may have unique power when

---

[3] The experimental results presented in this study are weakly suggestive of this in that student-athletes assigned to the threat condition attempted to answer more test questions, a crude proxy for overall effort. Interestingly, this effect is concentrated among males for whom the stereotype of the "dumb jock" may be more powerful and for whom the academic underperformance of student athletes is large. However, it should be noted that there are other plausible mechanisms by which stereotype threat may influence effort and performance that are not captured by this simple model (e.g., a direct disutility from experiencing stereotypes). It is straightforward to show that this disutility provides an additional mechanism through which stereotype threats could increase effort.

based on students at Swarthmore College. First, the small size of the College suggests that student-athletes are particularly likely to view their athletic status as well known. Second, a fairly long history of animus in the College community with respect to the relationship between athletics and the core academic mission of the College also suggests that a definite athletic stigma exists in the community.[4] In other words, this setting may provide a uniquely powerful test of the relevance of stereotype threat related to athletic status. However, relative to conventional lab-based experiments, this experiment may have a somewhat stronger claim to external validity because it is a "framed field experiment" in the taxonomy of Harrison and List (2004). Specifically, this experiments uses field-relevant subjects (i.e., students) doing field-relevant tasks (i.e., test performance). This characterization is further supported by the fact that the study procedures took place in conventional College classrooms rather than in an unfamiliar laboratory environment. It should also be noted that the unusual level of scrutiny that characterizes participation in experimental studies also happens, in this application, to correspond uniquely with the relevant field context (i.e., the performance of college students in the highly evaluative context of academically selective classrooms).

A. Recruitment

At the beginning of the Spring 2008 semester, all students at Swarthmore College received emails inviting them to participate in a 1-hour experiment whose goal was "to examine the determinants of cognitive functioning." Students were also told that they would receive $15 for their hour of participation. In order to promote statistical power and generalizability,

---

[4] In general, student athletes both perceive that there are negative cultural stereotypes about their academic preparation and intelligence but do not believe that these stereotypes apply to them (Saliles 1996, Jackson et al. 2002). The anecdotes around the high-profile debates specific to Swarthmore College athletics (e.g., Longman 2000) suggest the perception of stigma may be uniquely strong there.

additional emails and a study-recruitment letter were sent to student-athletes.[5] All recruited students were directed to a secure web page where they could register for the study by completing a baseline questionnaire and indicating their scheduling availability. Ninety-one students completed this registration and were randomized according to the procedure described below. These participating students could then select into one of five scheduled sessions that were held in the 5th and 6th weeks of the semester. Seven students who had registered for the study did not ultimately attend a session leaving a final sample of 84.[6] Roughly a quarter of students at the College were athletes. However, 44% of the study participants (i.e., 37 out of 84) were athletes, which reflected the success of the differential recruitment strategy (Table 1).

In any experiment such as this one, an important question involves the other ways in which the participants do or do not resemble the larger population from which they were drawn. Table 1 provides just such a comparison. The study population actually resembled the overall student body quite closely with respect to gender and SAT scores. However, study participants were both more likely to be freshman instead of seniors and less likely to be black or Hispanic.

B. Randomization

In order to increase the likelihood that the unobserved participant traits were unrelated to treatment status, the randomization procedure used in this study exploited the baseline traits available from the initial questionnaire and the publicly available athletics rosters. Participants were matched to other participants with respect to traits thought to be relevant to the study (e.g., athletic status, math SAT scores). Randomization then occurred within these matched pairs. Students who were identified as current athletes from the roster data were first sorted into cells

---

[5] Current student-athletes were identified through the rosters that were publicly available on the College athletics web site. However, none of the recruitment materials indicated the athletic focus of the study. Furthermore, students could not directly infer from the emails or the mailer this selective recruiting strategy.
[6] OLS models of attrition indicate that it was unrelated to treatment status.

based on the gender-specific sport that they played (e.g., women's soccer, men's basketball).

Within each of these sport-gender cells, each participant was matched to another participant with

similar math SAT scores. In cases where there were an odd number of participants within cells,

participants were matched to someone of the same gender and similar math SAT score but a

different sport. Any remaining participant was assigned a treatment status through simple

randomization. The participants who did not appear on the current athletics rosters were assigned

a treatment status by a similar procedure. Specifically, they were first sorted by gender and by

graduating class. Then participants were paired within each class-gender cell to another

participant with a similar math SAT score. In cases where there were odd remainders in a

gender-by-cohort cell, they were matched with a similar residual from a neighboring gender-

specific cohort if available or randomized as a singleton otherwise.

The fundamental goal of randomization is to balance outcome-relevant participant traits

across treatment and control states so that any post-treatment differences observed across these

conditions can be attributed to the treatment. Whether randomization succeeded in balancing

unobserved, outcome-relevant traits cannot be definitively established. However, auxiliary

regressions in which treatment status is the dependent variable are uniformly consistent with a

"successful" randomization in that treatment status is unrelated to athletic status, gender,

graduating class, race-ethnicity, and SAT scores (see Dee 2009).

C. Experimental Procedures

An administrator who was blind to the treatment status of the individuals conducted each

of the 1-hour experimental sessions in normal classrooms. After completing the informed-

consent procedures, each participant received a folder containing their experimental materials

(i.e., questionnaires, a test, etc.). The administrator guided the students through the sequenced

completion of these materials beginning with a 1-page questionnaire. For students in both the treatment and the control states, the questionnaire elicited information on the student's graduating class, whether they lived in College housing and whether they had a roommate. For the students in the treatment condition (both athletes and non-athletes), the questionnaire then asked "Are you (or have you been) a member of a National Collegiate Athletics Association (NCAA) sports team at the College?". They were then asked to identify the sport(s) they played and to respond to three questions about the frequency with which (on a scale of 1 to 7) they experienced scheduling conflicts between athletics and, respectively, course/seminar meetings, laboratory sessions, and other academic lectures (e.g., evening lectures by outside speakers). For students in the control condition, the questionnaire continued instead with similarly structured questions related to the dining services on campus. The basic structure of these treatment and control questionnaires parallels those used in the stereotype-salience study by Shih, Pittinsky, and Ambady (1999) and a recent study by Benjamin, Choi, and Strickland (2007).

Following the completion of this brief questionnaire, the participants were instructed that they would have 30 minutes to complete a 39-question test. The administrator explained to the participants that they might not be able to finish the test in the allotted time but that they should try to answer correctly as many questions as possible. In other words, these instructions deliberately encouraged the subjects to simultaneously value both accuracy on an answered question and answering more questions. The test consisted of 30 quantitative questions and 9 verbal questions from a Graduate Record Examination (GRE). As in prior studies of stereotype threat, this study reports the effect of random assignment to the stereotype-threat prime on participants' test accuracy (i.e., the percent correct of answered questions) and on the number of questions answered. The stereotype-threat prime could influence the test accuracy of participants

through its effects on both cognitive functioning and test effort (i.e., respectively, the n and e terms in the theoretical model). The number of questions answered is commonly used as a less ambiguous proxy for participant effort (e.g., Fryer, Levitt, and List 2008).

This 30-minute assessment was explicitly designed to be difficult both to avoid ceiling effects in the question accuracy of the high-ability subjects (Table 1) and to provide meaningful variation in the number of questions answered. The items from the GRE are commonly characterized as more difficult (e.g., Harrison et al. 2009). This assessment amplified the difficulty of the GRE items by adding 9 GRE questions to a 30-question quantitative GRE section that was designed to be taken in 30 minutes by itself. This design was effective in both limiting ceiling effects and providing the intended variation in the number of questions answered. No subject answered all 39 questions accurately in the time allotted. Furthermore, though 80 percent of subjects answered 30 or more questions, only slightly more than a third of the subjects answered all 39 questions. This pattern suggests that the subjects responded to the instructions to prioritize both accuracy and the number of questions answered (e.g., there was not extensive guessing). However, the potential limitations of using the number of questions answered as a proxy for subject effort should be explicitly noted. In particular, subjects could direct their effort towards answering a given question accurately as well as to answering more questions.

At the conclusion of the 30 minutes allotted for the test, the students were then directed to a word completion exercise designed to test the cognitive activation of the stereotype. Specifically, this exercise consisted of 30 word fragments, 12 of which were designed with the possibility that they could be completed as sports-themed words (e.g., "GO _ _" which could be completed as "GOAL"). This list also contained word fragments that could be completed in a

way that suggested self-doubt (e.g., "DU _ _" as "DUMB") and 11 filler words. After this 10-minute exercise, the students were directed to a short questionnaire that consisted of the 7 questions that constitute the academic sub-scale of the self-regard survey (Fleming and Courtney 1984). The experiment then concluded with a short exit questionnaire where participants could indicate the extent to which they enjoyed the study and what they thought the study's purpose was. All students were also asked at this point to identify their race-ethnicity. For students assigned to the control condition, the exit questionnaire then contained the questions about athletics that had been in the opening questionnaire for students in the treatment condition.

IV. RESULTS

On average, the participating students answered 35 of the 39 available questions with 27 questions answered correctly. The primary measure of test performance, the percent of answered questions that were correct, averaged 78.4 percent with a minimum value of 43 percent, a maximum value of 97 percent and a standard deviation of 0.11. Figure 1 presents graphical evidence on the effects of the intervention by showing kernel-density estimates of the test-performance distributions by treatment status and athletic status. The top panel of Figure 1 indicates that, for the non-athletes, the distributions of test-score performance are remarkably similar by treatment status.[7] In contrast, the bottom panel of Figure 1 indicates that, for the athletes participating in the study, assignment to the threat condition led to a quite large leftward shift in the test-performance distribution, an effect consistent with the hypothesis of stereotype threat.

---

[7] These comparative distributions may appear to suggest that the treatment modestly increased test-score performance (e.g., a possible and plausible stereotype "lift"). However, this interpretation reflects just two or three outlying observations. Regression-adjusted comparisons indicate that the treatment did not have a statistically significant effect on the non-athletes.

The regression specification used to estimate the effect of the intervention on test performance (i.e., y_i) takes the following form:

$$y_i = \alpha + \beta(T_i A_i) + \gamma T_i + \delta A_i + \theta X_i + \varepsilon_i$$

where T_i and A_i are binary indicators that identify, respectively, whether student i was assigned to the treatment and was an athlete and ε_i is a mean-zero random error term. The coefficient of interest, β, reflects the unique effect the stereotype-threat intervention had on athletes. The variable, X_i, represents various other determinants of test performance, including fixed effects for gender and race, math and verbal SAT scores, and fixed effects for the student's graduating class and the session they attended. Given the random assignment, none of these control variables should have a substantive influence on the estimated value of β. However, these controls can improve the precision of this point estimate. The fixed effects for the session attended provide a control for unintended determinants that may have been unique to each session (e.g., administrator behavior, classroom setting, peer traits, etc.).

The key results from estimating this model are reported in Table 2. Interestingly, the estimated main effect of the treatment (i.e., the estimate of γ) was positive but small and statistically insignificant in all specifications. The key estimate of interest is the unique effect of the treatment on athletes (i.e., the estimate of β). The results in Table 2 indicate that this effect was uniformly negative and implied large and statistically significant reductions in test performance of student athletes. These point estimates ranged from 8.1 to 9.4 percentage points, which is equivalent to as much as a 12 percent reduction in the mean performance or 0.84 of a standard deviation. Treatment effects of this magnitude are common in laboratory tests related to stereotype threat (e.g., Steele and Aronson 1995, Shih, Pittinsky, and Ambady 1999, Yopyk and Prentice 2005). However, effects of this magnitude are also consistent with the achievement gaps

observed in field settings. For example, Shulman and Bowen (2001, Table 3.1) find that among males from the most recent "College and Beyond" cohort who attended cohort liberal arts colleges, the regression-adjusted effect associated with being an athlete in a high-profile sport is a 8.8 percentile point reduction in class rank. The corresponding effect for females was 6.1 percentile points. The key experimental results presented here suggest that stereotype threat could make a substantive contribution to the academic underperformance of student-athletes documented by Shulman and Bowen (2001).

The results in Table 2 also indicate that SAT scores strongly predicted test performance and that the female participants tended to perform somewhat worse. However, estimates based on the full sample provided only weakly suggestive evidence that athletes responded to stereotype threat with increased effort (Dee 2009). More specifically, the treatment increased the number of test questions answered by athletes by roughly 5 percent but reduced the number of correctly answered questions by 8 percent (Dee 2009). There were also no statistically significant treatment effects on the post-test measures of stereotype activation, self-doubt or academic self-regard (Dee 2009). This may reflect the unusual length of the test relative to other studies (i.e., 30 minutes instead of 5) and the general fatigue this appeared to create among the subjects.[8] However, it should also be noted that this could be due to the effects of the threat manipulation fading out over this study time period.

One potentially important type of treatment heterogeneity suggested by earlier research (e.g., Harrison et al. 2009) involves whether the effects of stereotype threat vary by gender. Female student-athletes would be particularly harmed by stereotype threat if they have strong academic identities that are susceptible to negative stereotypes. In contrast, the effort and

---

[8] One empirical result consistent with this interpretation is that virtually none of the observed student traits (e.g., race, gender, SAT scores) were significant predictors of the data collected in the last 20 minutes of the experiment (i.e., academic self-regard, self-doubt or sports-themed word completions).

performance implications of stereotype threat would be particularly salient for male student-athletes if, for example, the "dumb jock" stereotype applied differentially to them. Table 3 provides evidence on this question by estimating, separately for males and females, the effects of the treatment-athlete interaction on test performance and questions answered. The results indicate that the pejorative performance implications of the identity manipulation were over twice as large among males (i.e., 13 percentage-point reduction with a p-value of 0.057) relative to females (a statistically insignificant 6 percentage-point reduction). Furthermore, the identity manipulation increased the number of questions answered and the probability that they answered all of the questions, suggesting that male student-athletes responded to the identity threat with increased effort.[9] Given the small sample sizes (i.e., 45 females and 39 males), it is not surprising that the gender-specific treatment heterogeneity documented in Table 3 is not statistically significant. Nonetheless, the patterns in Table 3 are consistent with the hypothesis that athletic stereotype threat is more pronounced among males and that it contributes to the larger achievement gaps observed among male student athletes in field settings

V. CONCLUSIONS

The prominent role that social identity may play in influencing a broad array of economic and education-related outcomes is receiving an increasing amount of attention. This study makes two broad contributions to this literature. One is to adapt an economic model of social identity to reflect more accurately one of the most prominent conjectures in the literature on social identity: the phenomenon of stereotype threat and the role that endogenously chosen effort can play in mediating its effects. The second contribution of this study is to examine through a tightly

---

[9] The marginal effect from a probit specification is similar. The large size of this treatment-athlete interaction reflects the fact that non-athletes exposed to the identity prime were substantially less likely to answer all the questions, which is consistent with a threat reduction (i.e., a stereotype "lift" from affirming that one does not have the stereotyped identity) resulting in reduced effort.

controlled experiment whether stereotype threat contributes to a large, controversial, and policy-relevant achievement gap observed at many selective colleges and universities: the academic underperformance of student-athletes. The results of this framed field experiment are consistent with the hypothesis that the academic stigma associated with being a student-athlete at a highly selective college or university makes a substantial contribution to their academic underperformance, particularly for males. The relevance of this sort of experimental evidence for "natural" field settings (as well as for other institutions) is always an open question. This study is no exception and a compelling next step for research in this area would involve well-designed natural field experiments that assess the effectiveness of institutional strategies to ameliorate the consequences of this social-identity phenomenon.

REFERENCES

Akerlof, George and Rachel Kranton. "Economics and Identity," Quarterly Journal of Economics 115(3), August 2000, pages 715-753.

Akerlof, George and Rachel Kranton. "Identity and Schooling: Some Lessons for the Economics of Education," Journal of Economic Literature 40, December 2002, pages 1167-1201.

Akerlof, George and Rachel Kranton. Identity Economics: How Our Identities Affect Our Work, Wages, and Well-being. Princeton: Princeton University Press, 2010.

Aronson, Joshua and Claude M. Steele. "Stereotypes and the Fragility of Academic Competence, Motivation, and Self-Concept," in Handbook of Competence and Motivation, Andrew J. Elliot and Carol S. Dweck, editors, New York: Guilford Press, 2005.

Becker, Gary S. The Economics of Discrimination. Chicago: University of Chicago Press, 1957 .

Bénabou, Roland and Jean Tirole "Identity, Dignity and Taboos: Beliefs as Assets," working paper, December 2006

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. "Social Identity and Preferences," NBER Working Paper No. 13309, August 2007.

Bowen, William G. and Sarah A. Levin. Reclaiming the Game: College Sports and Educational Values. Princeton University Press, 2003.

Bowen, W.G. and D. Bok. The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions. Princeton, NJ: Princeton University Press.

Dee, Thomas S. "Stereotype Threat and the Student-Athlete," NBER Working Paper No. 14705, February 2009.

Falk, Armin and James J. Heckman (2009). Lab Experiments are a Major Source of Knowledge in the Social Sciences. Science 23(326), pp. 535-538.

Fleming, James S. and Barbara E. Courtney. "The Dimensionality of Self-Esteem II:
Hierarchical Facet Model for Revised Measurement Scales," Journal of Personality and
Social Psychology 46(2), 1984, pages 404-421.

Fryer, Roland G., Levitt, Steven D. and John A. List. "Exploring the Impact of Financial
Incentives on Stereotype Threat: Evidence from a Pilot Study," American Economic
Review, May 2008.

Harrison, Glenn W. and John A. List. "Field Experiments," Journal of Economic Literature Vol.
XLII, December 2004, pages 1009-1055.

Harrison, C. Keith; Stone, Jeff; Shapiro, Jenessa; Yee, Sharon; Boyd, Jean A.; Rullan, Vashti.
The Role of Gender Identities and Stereotype Salience With the Academic Performance of
Male and Female College Athletes. Journal of Sport and Social Issues 38(1), February
2009, 78-96

Jackson, J. S., Keiper, S., Brown, K. T., Brown, T. N., & Manuel, W. (2002). Athletic identity,
racial attitudes, and aggression in first-year Black and White intercollegiate athletes. In M.
Gatz, M. A. Messner, & S. J. Ball-Rokeach (Eds.), Paradoxes of youth and sport (pp. 159-
172). Albany: State University of New York Press.

Jamieson, J. P., & Harkins, S. G. (2007). "Mere effort and stereotype threat performance
effects," Journal of Personality and Social Psychology, 93, 544-564.

Levitt, Steven D. and John A. List (2007a). "Viewpoint: On the Generalizability of Lab Behavior
in the Field." *Canadian Journal of Economics*, 2007, 40(2): 347-370.

Levitt, Steven D. and John A. List (2007b)."What Do Laboratory Experiments Measuring Social
Preferences Reveal about the Real World?"*Journal of Economic Perspectives*, 2007,
21(2), pp.153-174.

Levitt, Steven D., John A. List and David Reiley (forthcoming). "What Happens in the Field
Stays in the Field: Professionals Do Not Play Minimax in Laboratory Experiments"
Econometrica.

Longman, Jere. "No More Football, Lots of Questions," The New York Times, December 5,
2000.

McLeish, Kendra and Robert Oxoby. "Social Interactions and the salience of social identity,"
Journal of Economic Psychology 32(1), 172-178, February 2011.

Oswald, D. L., & Harvey, R. D. (2000). Hostile environments, stereotype threat, and math
performance among undergraduate women. Current Psychology 19(4): 338-356.

Sailes, G. A. (1996). An investigation of campus stereotypes: The myth of Black athletic
superiority and the dumb jock stereotype. In R. E. Lapchick (Ed.), Sport in society: Equal
opportunity or business as usual? (pp. 193-202). Thousand Oaks, CA: Sage.

Shih, Margaret, Todd L. Pittinsky, and Nalini Ambady. "Stereotype Susceptibility: Identity
Salience and Shifts in Quantitative Performance," Psychological Science 10(1), January
1999, pages 80-83.

Shulman, James L. and William G. Bowen. The Game of Life: College Sports and Educational
Values. Princeton University Press, 2001.

Steele, Claude M. and Joshua Aronson. "Stereotype Threat and the Intellectual Test Performance
of African Americans," Journal of Personality and Social Psychology 69(5), 1995, pages
797-811.

Yeager, David S. and Gregory M. Walton. Social-Psychological Interventions in Education:
They're Not Magic. Review of Educational Research 81 (2), 267-301.

20

Yopick, Darren J.A. and Deborah A. Prentice. "Am I an Athlete or a Student? Identity Salience and Stereotype Threat in Student-Athletes," Basic and Applied Social Psychology 27(4), 2005, pages 329-336.

Table 1 - Average Traits of College Students and Study Participants

Variable	College	Study Participants			
		Total	Treatment	Control	p-value
Athlete	0.244	0.440	0.463	0.419	0.68
Female	0.520	0.527	0.537	0.535	0.99
White Non-Hispanic	0.435	0.615	0.732	0.605	0.22
Asian	0.173	0.220	0.195	0.279	0.37
SAT (Math)	736	724	729	721	0.51
SAT (Reading)	716	718	720	714	0.68
Class of 2008	0.272	0.187	0.195	0.186	0.92
Class of 2009	0.247	0.242	0.220	0.233	0.89
Class of 2010	0.225	0.220	0.171	0.256	0.35
Class of 2011	0.245	0.352	0.415	0.326	0.4
Enrollment	1,491	84	41	43	

Source: The Fact Book, Institutional Research Office, Swarthmore College
(http://www.swarthmore.edu/factbook.xml). College-specific athletic participation based on
gender-specific 2005-06 data weighted by male and female enrollment. SAT scores are based on
the matriculants from 2005, 2006, and 2007. The p-value refers to a test of the hypothesis that the
prevalence of the observed trait is the same across the treatment and control groups.

Figure 1 – Kernel Density Estimates of Test Performance by Treatment and Athletic Status
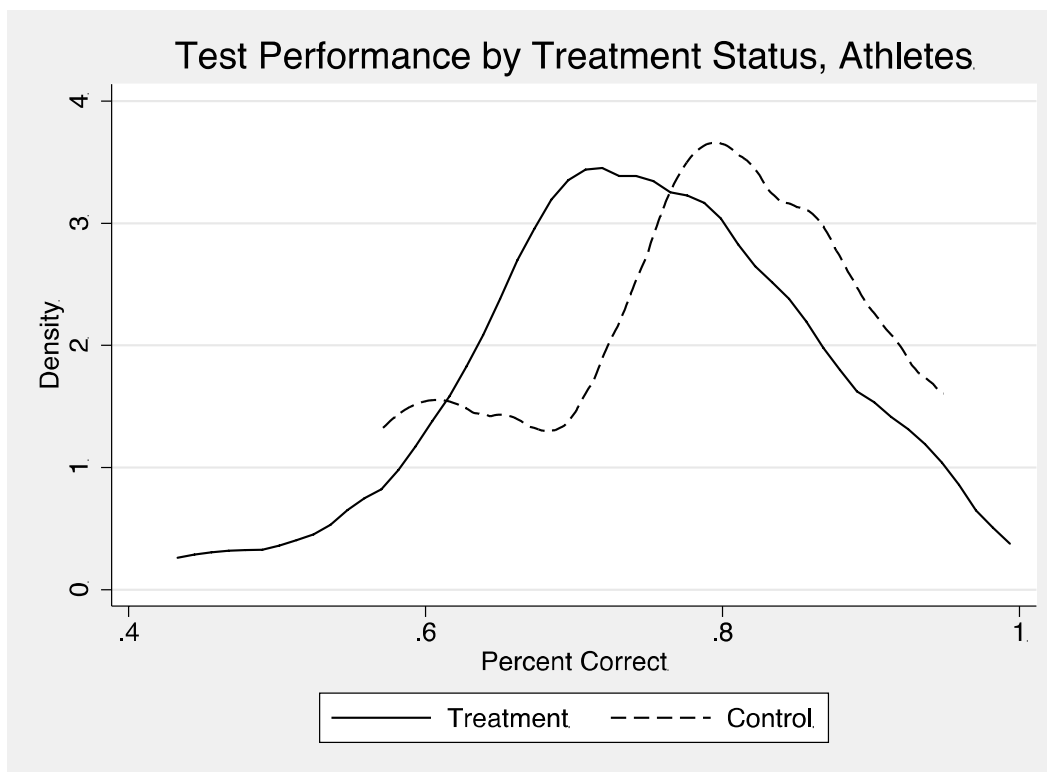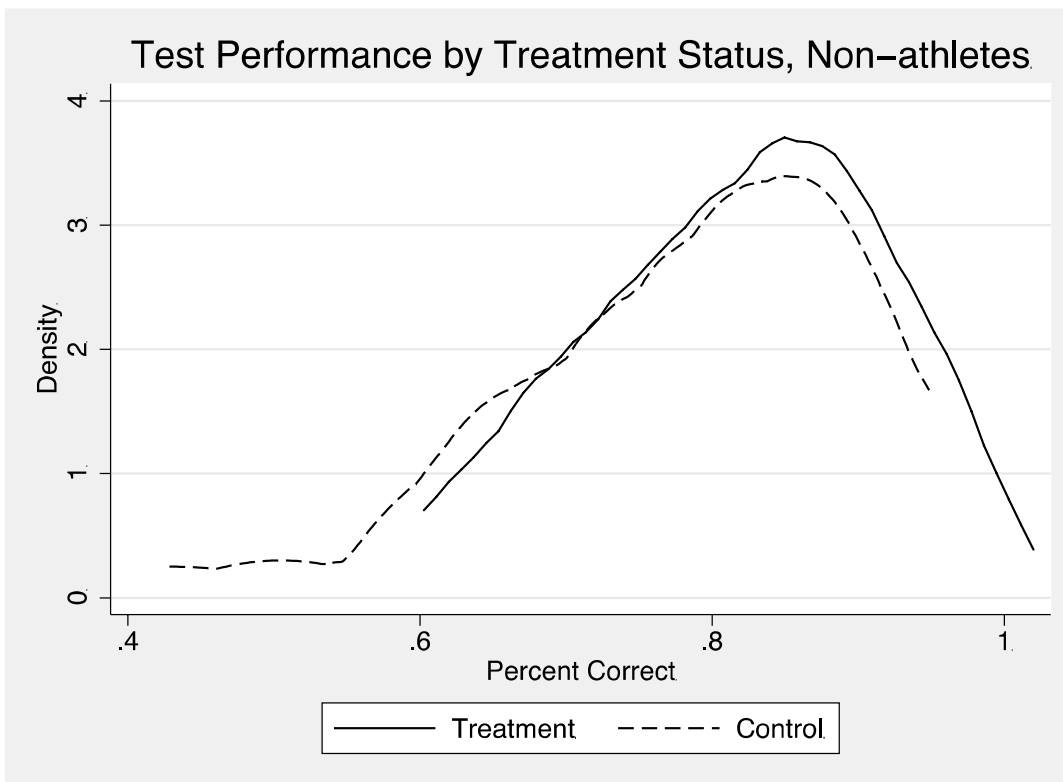
Table 2 - OLS Estimates of the Determinants of Test Performance

Independent Variable	(1)	(2)	(3)	(4)
Treatment x Athlete	-0.0808*	-0.0907**	-0.0859**	-0.0939**
	(0.048)	(0.041)	(0.041)	(0.042)
Treatment	0.0408	0.0356	0.0355	0.0372
	(0.032)	(0.027)	(0.028)	(0.029)
Athlete	0.0173	0.0766**	0.0665**	0.0634*
	(0.035)	(0.032)	(0.032)	(0.033)
Female	-0.0496**	-0.0483**	-0.0446**	-0.0440*
	(0.024)	(0.020)	(0.021)	(0.023)
White Non-Hispanic	0.0414	-0.0376	-0.0473	-0.0512
	(0.042)	(0.038)	(0.039)	(0.040)
Asian	0.0797*	-0.0030	-0.0037	-0.0110
	(0.047)	(0.043)	(0.044)	(0.044)
SAT (Math)		0.0008***	0.0008***	0.0008***
		(0.000)	(0.000)	(0.000)
SAT (Reading)		0.0006***	0.0006***	0.0006***
		(0.000)	(0.000)	(0.000)
Class Fixed Effects	No	No	Yes	Yes
Session Fixed Effects	No	No	No	Yes
R^2	0.125	0.383	0.407	0.434

Notes: The dependent variable is the percent of answered questions that were correct (sample mean = 0.784). ***$p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3 - Estimated Treatment Effects on Test Performance and Questions Attempted by Sex

Dependent Variable	Females		Males	
	Estimated Effect of Treatment x Athlete	Dependent Mean	Estimated Effect of Treatment x Athlete	Dependent Mean
Test Performance	-0.0598 (0.067)	0.764	-0.1289* (0.065)	0.807
Questions Attempted	-0.1706 (2.659)	34.8	6.8287* (3.771)	34.3
All Questions Attempted	-0.0993 (0.304)	0.356	0.9820*** (0.341)	0.359

Notes: All models condition on the student observables, class, and session fixed effects (e.g., model (4) in Table 2). ***$p<0.01$, ** $p<0.05$, * $p<0.1$.