Implicit Gender-STEM Stereotypes and College Major Choice

Stephanie Owen (Colby College)
Derek Rury (University of Chicago)*

This version: March 27, 2025

Abstract

Implicit stereotypes about gender and STEM may unconsciously shape students' academic choices and contribute to gender gaps in major choice, but there is limited economic evidence on this channel. To study this relationship, we administer a gender-science Implicit Association Test (IAT) to a sample of undergraduates, and link results to original survey data and administrative transcript data. We find that on average, women are no more likely than men to implicitly associate men with STEM relative to humanities. However, implicit stereotypes are strongly predictive of behavior. Male students with a one standard deviation higher male-science association are 7-9 ppt *more* likely to intend to major in STEM, while female students are 8-10 ppt *less* likely. We find similar relationships between implicit stereotypes and observed STEM course-taking and officially declared major. These patterns are robust to controls for expected earnings, preferences for major characteristics such as salary and job flexibility, presence of female role models, and explicit beliefs about women in STEM and humanities. Our results suggest that implicit stereotypes may be a promising focus for interventions targeting gender gaps.

Key Words: STEM, Stereotypes, College Major, Gender, Education
JEL Codes: I26, I29, D83

*Email: sowen@colby.edu; rury@uchicago.edu.
The authors thank participants at Opportunity Insights, the Harvard Graduate School of Education, the University of Chicago, the Boston University Wheelock Educational Policy Center, UC Davis, the Northeastern Economics of Education Conference, and the Association for Education Finance and Policy Conference for helpful feedback. Cristian Bancayan Navarro provided excellent research assistance for this project.

1 Introduction

Despite great strides in education and the labor market, women remain under-represented in science, technology, engineering and mathematics (STEM) college degrees and associated occupations. These gaps have implications for economic inequality (the gender pay gap) as well as efficiency (if barriers are preventing optimal sorting).

In this paper, we marry two research literatures, from psychology and economics, to investigate one potential explanation: the role of implicit gender stereotypes. Social scientists have long argued that stereotypes–overgeneralized beliefs about certain groups–may affect behavior (Steele, 1997; Bordalo et al., 2016). We are interested in the role of *implicit* stereotypes in particular: those that shape individuals' thinking and behavior unconsciously, possibly in contrast to explicit beliefs or goals.

Psychologists have carefully documented the existence of implicit gender stereotypes and their relationship with academic performance and self-concept (Kiefer and Sekaquaptewa, 2007; Cvencek et al., 2011; Ertl et al., 2017). However, they have yet to establish a link with long-term, consequential real-world outcomes, such as major choice. Recent work in economics has found plausibly causal relationships between implicit stereotypes and economic outcomes (Rooth, 2010; Glover et al., 2017; Carlana, 2019; Martínez, 2022; Alesina et al., 2024). However, this work has focused on the effects of stereotypes held by individuals other than the stereotyped group itself, including teachers, managers, and employers.

We combine and add to these two literatures by investigating the relationship between implicit stereotypes about a student's own gender and their real-world academic outcomes. Our primary research question is whether students' implicit stereotypes about women and science help explain the gender gap in STEM major choice.

We administer a gender-science implicit association test (IAT) to a sample of college students to capture their unconscious beliefs about the relationship between gender and science. The IAT has a long track record in social science as a way for researchers to measure subconscious stereotypes (Greenwald and Banaji, 1995; Greenwald et al., 2003). We pair the IAT with a survey to capture students' intended major, as well as a number of preferences and beliefs related to major that prior research has shown to influence major choice. We link these data to administrative data on course selection and major choice. Our rich survey and administrative data allow us to control for many factors that could potentially confound the relationship between IAT score, gender, and major choice.

The typical student in our sample implicitly associates male with STEM subjects and female with humanities, with no difference by the student's gender. We find that implicit stereotypes are strongly predictive of stated and observed behavior. Men with a one standard deviation stronger implicit stereotype are 7 to 9 percentage points (p.p.) *more* likely to state that they intend to major in a STEM subject, while women are 8 to 10 p.p. *less* likely. These relationships are robust to controlling for factors hypothesized to explain STEM gender gaps, including academic preparation, preferences for pecuniary and non-pecuniary features of post-college job prospects and in-college experiences, beliefs about major-specific ability, and the presence of same-gender role models. Results using administrative data on STEM-course taking and official major declaration show similar patterns.

Although we lack the exogenous variation in implicit stereotypes needed to definitively establish a causal relationship, our results are consistent with students' implicit stereotypes influencing their interest in and persistence through STEM majors. The robustness of our findings to the most likely confounding variables suggests that implicit stereotypes are capturing a distinct determinant of major choice, one that keeps men in and drives women out of STEM.

While the gender gap in STEM has been the subject of much research (see Delaney and Devereux (2021) and Patnaik et al. (2021) for recent reviews), the current study provides evidence for a channel largely ignored in the economics literature. To our knowledge, we are the first to document a relationship between a stereotyped group's implicit beliefs and an economically meaningful outcome. Future research should seek to confirm the causal nature of the correlation we've established, and investigate solutions to a mechanism that may be leading to sub-optimal investments.

2 Setting and Data

Our setting is a large public research university in the Midwestern U.S. Admission to this university is highly competitive, with an average high school GPA of 3.9 and an acceptance rate of 18% for the cohort entering in fall 2023.

We recruited undergraduate students from two populations during the fall 2023 semester. Students in both targeted populations were offered a financial incentive for completing the study (a chance at one of 40 $50 Amazon gift cards). The first population, which we refer to as our ECoach sample, consisted of 6,620 students in a set of nine introductory courses in biology, chemistry, computer science, economics, engineering, math, physics, and statistics.[1] Students in these courses interact with an online platform called ECoach, which is a communication tool designed to provide tailored information and advice to students in large courses. Through an existing partnership with ECoach, we sent recruitment messages to enrolled students via the ECoach platform. 508 (7.7%) of the ECoach population participated in some way (measured as completing the consent page of the survey), and 369 (5.6%) are in our final analysis sample.

[1] The courses were: Biology 171 (Introductory Biology: Ecology and Evolution), Chemistry 130 (General Chemistry: Macroscopic Investigations and Reaction Principles), Economics 101 (Principles of Economics I), Electrical Engineering and Computer Science (EECS) 183 (Elementary Programming Concepts), EECS 203 (Discrete Mathematics), Engineering 101 (Introduction to Computers and Programming), Math 105 (Data, Functions, and Graphs), Physics 140 (General Physics 1), and Statistics 250 (Introduction to Statistics and Data Analysis). Students could be enrolled in multiple of these courses, and around 18% were.

Our second source of participants is a random sample of 2,500 first-year students who were not enrolled in an ECoach course. The original list of students was provided by the university's office of the registrar, so we refer to this as our registrar sample. We sent recruitment emails directly to these students. Of the 2,500 registrar students we invited to participate, 276 (11%) completed consent and 171 (6.8%) are in our final analysis sample.

2.1 Student survey

We administered a survey designed to elicit students' beliefs and preferences about different majors. First, the survey asked students to select their top and second choice of major. Specifically, we asked students which major they were most (and second most) likely to graduate with a degree in, from a list of 16 majors (plus a write-in option).[2] Our primary outcome is based on the top/primary major item. We classify students as intending a STEM major if they selected (or wrote in) any of the following as their most likely major: engineering, biology, computer science, math, statistics, data science, neuroscience, environmental science, and other natural sciences.

The remainder of the survey captured other beliefs and preferences relevant to major choice. We asked students to rate how important various factors were in their choice of major on a Likert scale, ranging from from 1 ("Not at all important) to 5 ("Extremely Important"). These factors included: "Feeling like I'm good at the subject," "Being engaged with the coursework (while in school)," "Making/having friends or study partners in the major," "Expected salary (after graduation)," "Work flexibility (after graduation)," "Having a positive impact on society (after graduation), and "Work culture/peers (after graduation)." We also asked them what they expected their salary to be if they graduated with a degree in (separately) their first and second choice of major.

To capture students' explicitly held beliefs about gender and STEM, we asked students to report the proportion of STEM graduates and humanities graduates from the university that they believed were female. To measure beliefs about relative ability across fields, we asked them to estimate the high school GPA of graduates who completed a STEM degree, and of those who completed a humanities degree. Lastly, to proxy for the presence of pre-college female role models, we asked students to provide the name and the title (e.g. Ms./Mr./Mx.) of their favorite math or science teacher and of their favorite English or social studies teacher in high school; the title allows us to classify role models by gender. The full survey instrument is included in Appendix C.

[2]For potential double majors, the survey asked students to distinguish between their "primary" and "secondary" major.

2.2 Implicit Association Test

To measure stereotypes, we presented students with a link to a gender-science IAT at the end of the survey.[3] The IAT is a tool that was developed by social psychologists to capture implicit attitudes and stereotypes, meaning those that might be hidden or subconscious to the individual but nevertheless affect judgment and behavior (Greenwald and Banaji, 1995; Greenwald et al., 1998). The IAT has been used in hundreds of studies to document implicit attitudes regarding different racial and ethnic groups (McConnell and Leibold, 2001), genders (Salles et al., 2019), and religions (Rowatt et al., 2005), among other topics. The IAT is premised on the idea that it takes additional time and effort to make a decision that overrides an unconscious stereotype. Mechanically, the IAT measures how quickly someone completes a set of word categorization tasks, and scores reflect relative response time for stereotypical vs. non-stereotypical tasks.

The IAT we administer for this paper, the gender-science IAT, contains four categories of words: male (boy, uncle, etc.); female (daughter, woman, etc.); science (engineering, geology, etc.); and humanities (literature, history, etc.); the full set of words appear in Appendix B. The test consists of seven short modules, each of which asks participants to sort a series of words to the left or right of a screen using keystrokes or touch (the IAT can be completed on a computer, tablet, or mobile phone). Three of the modules are practice rounds to familiarize participants with the categories and the tasks. In the four modules used for scoring, participants are asked to either sort male and science words to one side and female and humanities words to the other (stereotypical pairing) or male and humanities to one side and female and science to the other (non-stereotypical). In total, participants perform 60 stereotypical and 60 non-stereotypical sorting tasks. Screenshots from the gender-science IAT can be found in Appendix B.[4]

The IAT score is based on the average difference in raw response times between stereotypical versus non-stereotypical sorting tasks. We use the scoring algorithm developed by Greenwald et al. (2003). Though we report raw scores in summarizing IAT results, for our primary analysis we standardize within our sample, for more easily interpretable magnitudes.

Though the IAT has been widely used in research for decades, there is a debate within the social science literature about its validity.[5] First, some have argued that the test is only weakly predictive of the discriminatory behavior it should theoretically influence

[3]The IAT we used was administered by Project Implicit (PI). Housed at Harvard University, PI is a repository of research and education about implicit biases, and programs custom IATs for organizations and researchers.

[4]Most studies using an IAT use an instrument that is "counterbalanced," meaning the order in which participants are asked to do the stereotype-conforming versus non-conforming tasks is randomly assigned. In our setting, due to a programming error, this was not the case: all participants did the stereotype conforming task (male-science) first. Although this has been shown to affect IAT score levels (Greenwald et al., 1998), it should not affect comparisons across groups, which is the focus of this study. We discuss this further and provide evidence in Appendix B.

[5]Carlana (2019) and Alesina et al. (2024) have excellent summaries of critiques of the IAT, which informed our discussion here.

(Blanton et al., 2009). However, a growing number of papers studying teachers (Carlana, 2019; Martínez, 2022), employers (Rooth, 2010), and managers (Glover et al., 2017) have found meaningful, plausibly causal links between IAT scores and real world behavior. Second, IAT scores may be unstable over time, and reflect temporary personal and environmental factors. We view the IAT as a noisy measure of a latent characteristic; any noise would attenuate the correlations we find. To the extent that we find that the IAT has predictive power relating to college major choice, we view the current study as providing evidence regarding these first two critiques. Finally, the IAT may be manipulable, with participants able to target a particular results once they learn how it works. One study found that participants were able to fake results when instructed to do so (Fiedler and Bluemke, 2005). However, it is not clear to what extent a participant would attempt to fake the IAT on their own, and how easily they could do so without prior knowledge of how the test works. Although we cannot completely rule this out, we follow recommended scoring procedures (Greenwald et al., 2003) and drop IAT scores from participants with unusually slow reaction times, which can be a sign of manipulation.

2.3 University records

We link our survey and IAT results to university administrative records. These data contain all baseline demographic and academic characteristics for the sample such as gender, race, class standing, high school GPA, socioeconomic status, and residency status. The data also contain students' full academic trajectories while at the university: course-taking, major declaration, and official grades. Because these are administrative data, they contain full information on academic outcomes for all students. Some students are missing information on pre-college characteristics such as high school GPA and parental education, which is collected as part of the application process. This is because some information, such as parental education, is self-reported, and some applicants, such as international and transfer students, do not submit certain information. Note that although the administrative data contain a student's official major, a large portion of our sample (60%) have no official major as many are first year students. We therefore use stated major choice from the survey as our primary outcome. We examine officially declared major and STEM course-taking as secondary outcomes.

We also use the transcript data to construct a measure of the presence of female STEM role models in college. We classified instructor first names as male or female using the Genderize package in R, which provides access to a comprehensive database of names and associated genders. We then calculated what percentage of a given student's STEM instructors from summer 2022 to fall 2023 were female. (A student who didn't take any STEM would be assigned a zero.)

2.4 American Community Survey

Finally, we supplement our internal data with data from the American Community Survey (ACS). We use ACS to calculate a Duncan index of gender dissimilarity, commonly used in research on gender, racial, and occupational segregation to measure how evenly groups are distributed across different categories. In our application, the Duncan index represents the proportion of workers who would need to switch occupations within a geographic region to achieve an even distribution of men and women across occupations. We use the five-year ACS samples between 2011 and 2023, at the occupation-by-ZIP-code level. We calculate a Duncan index for each zip code each year, then take the average index for each zip code over all years. Details of the Duncan index can be found in Appendix D.

2.5 Final sample

To be in our analysis sample, we require a survey response and a complete IAT response.[6] Our final sample consists of 540 students; descriptive statistics for our sample and how they compare to both the sampled population and the university as a whole are in Table 1.

Women are strongly overrepresented in our sample: 73% compared to 54% in the targeted population and 53% university-wide. (The university only records two genders.) On all other observable characteristics, our analysis sample closely resemble the target population and the university as a whole. By design, most (71.5%) of our sample are in their first year. The majority are white (61%) or Asian (32%). Seven percent are Black and 11% are Hispanic (note we report non-mutually exclusive race categories). Five percent are international. Only 18% are first-generation; the majority of our sample has a parent with a bachelor's degree. Roughly half have family income above $100,000 (though over a fifth are missing information on family income). The average student had a 3.9 GPA in high school, and the majority took calculus. The vast majority of the sample–87%–took any STEM courses in the semester that the study took place (unsurprising given we oversampled from introductory STEM courses); however, only 23% were officially declared as STEM majors. 58% indicated in the survey that their most likely major was a STEM field.

3 Results

3.1 IAT scores

Figure 1a shows the distribution of raw IAT scores. A positive score means a participant was quicker at sorting male with science and female with humanities words than vice

[6] We drop 62 IAT responses with high error rates, and 1 with overly slow response times, as suggested by Greenwald et al. (2003).

versa, and therefore shows an automatic association for male with STEM. Students in our sample on average hold male-STEM stereotypes. This is true in the wider population as well, as we would expect (Nosek et al., 2009, 2007). We find no differences by gender in the mean or distribution of IAT scores, though visually there are slightly more women with strong female-STEM associations and more men with strong male-STEM associations. This is consistent with patterns documented using the largest available database of IAT scores (collected and maintained by Project Implicit), wherein men and women have similar gender-science IAT scores on average (Nosek et al., 2007).

3.2 IAT scores and intended major

Our primary research question is how implicit stereotypes, as measured by the IAT, predict stated and observed academic choices for men and women. To start, we plot the relationship between raw IAT scores, intended major, and gender as a binned scatterplot (Figure 1b). It is clear from this visual that implicit stereotypes are strongly related to major choice, but in opposite ways for men and women. Men with stronger automatic associations between male and STEM are more likely to indicate they plan to major in a STEM field. For women, the opposite holds: the stronger their implicit stereotype, the *less* likely they are to intend a STEM major. In other words, the more a student implicitly associates STEM with their own gender, the more likely they are to intend to major in it.

Figure 1(b) represents a simple correlation between IAT score and intended major. This does not necessarily imply that implicit stereotypes causally influence major choice, as there may be omitted variables that correlate with both IAT and intended major, and which vary by gender. Though we do not have exogenous variation in implicit stereotypes, our rich survey and administrative data allow us to test the robustness of the relationship to a number of potential confounders. Formally, we estimate the following using ordinary least squares (OLS):

$$Y_i = \beta_0 + \beta_1 Female_i + \beta_2 IAT_i + \beta_3 Female_i \cdot IAT_i + \boldsymbol{X_i \gamma} + \boldsymbol{Z_i \lambda} + \boldsymbol{W_i \varphi} + \varepsilon_i \qquad (1)$$

where Y_i is a STEM-related outcome for student i. Our primary outcome of interest is stated intention to major in STEM (i.e., listing a STEM field as their top choice of major on the survey). We also examine observed STEM course-taking and officially declared major using transcript data. For both the contemporaneous (Fall 2023) and subsequent (Spring 2024) terms, we create variables measuring any STEM course-taking, the number of STEM credits, and declaration of a STEM major.

$Female_i$ is an indicator for the student being female (as recorded in the student's official record) and IAT_i is standardized IAT score, with higher values indicating stronger male-STEM implicit stereotypes. (From this point forward, we standardize IAT scores to have mean 0 and standard deviation 1.) β_2, and β_3 are the parameters of interest, telling us how implicit stereotypes predict outcomes for men and women. β_2 is the increase in

Y associated with a one standard deviation increase in IAT score for men; $\beta_2 + \beta_3$ is the equivalent increase for women.

$\boldsymbol{X_i}$ is a vector of student demographic and academic characteristics, including race, parental education and income, international status, and pre-college academic preparation. $\boldsymbol{Z_i}$ is a vector of beliefs and preferences related to different majors. Finally, $\boldsymbol{W_i}$ are environmental factors including the presence of female role models and the local labor market a student grew up in. We describe our control variables in detail below, as we progressively add them to the model.

We begin with the raw correlation between gender, IAT score and STEM major choice, without any controls (Column 1 of Table 2); this is equivalent to Figure 1(b) (though on a different scale since we have standardized IAT). The estimated $Female$ coefficient ($\hat{\beta_1}$) simply confirms the well-documented STEM gender gap, with the average female student (with an average IAT score) 17 percentage points (p.p.) less likely than a male one to intend a STEM major. The coefficient for IAT ($\hat{\beta_2}$) implies that male students with one standard deviation higher IAT score (i.e., stronger male-STEM association) are 9 p.p. *more* likely to major in STEM than men with less strong stereotypes ($p < 0.05$). The sum of the base IAT coefficient and the gender interaction imply that the relationship for women is the exact opposite: a woman with a 1 s.d. higher IAT score is 10 p.p. *less* likely to major in STEM ($p < 0.001$). The interaction term, 19.1 p.p., is also statistically significant, implying a statistically different relationship between IAT and major choice by gender.

Even with similar distributions of IAT scores (as shown in Figure 1a), the different relationship between IAT and academic choices by gender could explain the STEM gender gap. If β_2 and β_3 are in fact causal parameters, then reducing all students' average IAT score by a standard deviation (which is equivalent to the average student having no association between STEM and gender) would decrease men's rate of STEM intent by 9 p.p. and increase women's by 10 p.p., thus closing–and in fact slightly reversing–the gender gap.

Of course, interpreting these correlations causally requires a stronger assumption about ε_i. Determinants of major choice–and gender differences in those determinants–are the subject of a deep research literature (Delaney and Devereux, 2021; Patnaik et al., 2021; Altonji et al., 2012). Many of the factors in the error term could be correlated with both IAT and gender, meaning the results in Column 1 may be picking up omitted factors other than the implicit stereotypes that students hold. To test this, and to argue that the IAT is picking up a distinct determinant of major, we progressively add to our model a rich set of covariates that previous research has suggested may explain gender gaps.

We first add in basic demographic variables: race (indicator for underrepresented minority), parent education (indicator for first-generation college student), family income (indicator for above \$100,000), and international status. It is possible that students from different socioeconomic or cultural backgrounds have different gender views, which may correlate with other determinants of major choice. However, our results with demographic controls (column 2), are almost identical to those of column 1.

We next add in academic covariates: high school GPA and an indicator for taking calculus in high school.[7] Prior research has documented gender gaps in quantitative skill (Lohman and Lakin, 2009; Andreescu et al., 2008), which itself predicts STEM entry and persistence (Altonji et al., 2012; Arcidiacono, 2004); however, some have argued that mathematical aptitude explains little of the gender gaps in field choice (Cheryan et al., 2017). Although adding in academic preparation in column 3 shrinks the gender gap in major choice from 17.3 to to 14.7 p.p., the predictive power of IAT score is largely unchanged.

A number of studies have suggested that men and women hold different preferences for academic subjects and their associated careers. For example, Wiswall and Zafar (2018) found that female students rank job flexibility high and therefore shy away from STEM fields as a result. Other work posits that women are less interested in (or less informed about) the range of content and career opportunities of certain fields (Owen and Hagstrom, 2021; Jensen and Owen, 2000), while still other work suggests gender differences in peer effects (Bostwick and Weinberg, 2022; Booth et al., 2018), anticipated discrimination (Lepage et al., 2022), and confidence in academic ability (Owen, 2023; Niederle and Vesterlund, 2007). To capture these preferences, we control for a series of seven survey items asking students to rate the importance of various factors in the choice of major. These include three in-college factors (perceived ability in the subject, engagement with coursework, and having friends in the major) and four post-graduation factors (expected salary, job flexibility, work culture/peers, and having a positive impact on society). Students rated each item on a 1-5 Likert scale; we dichotomize each as important (4 or 5) or not (1 to 3). Column 4 adds these controls. The relationship between implicit stereotypes, gender, and STEM intent is very similar to the previous columns, though the IAT-major coefficient for men is somewhat smaller (7 p.p.) and only significant at the 10% level.

Similarly, in column 5 we add covariates for students' expected earnings in their top and second choice of major, motivated by work showing that expected earnings are a key driver of major choice (Conlon, 2021; Baker et al., 2018), and the possibility that women and men may be differentially motivated by earnings, and/or differently informed about earnings potential (Wiswall and Zafar, 2018; Reuben et al., 2017). The IAT-major intent correlations remain very similar to the previous columns.

Although stereotypes are inaccurate, exaggerated beliefs, they are based on a kernel of truth (Bordalo et al., 2016); it is true (and in fact motivates the current study) that more men than women pursue STEM fields. Furthermore, we are interested in *implicit* stereotypes in particular, meaning beliefs students are unaware of. Thus, in column 6 we add controls for students' *explicit* stereotypes about gender and major choice, operationalized as the proportion of STEM (and humanities) graduates of their university who they estimate are female. (Our measure is similar to the gender stereotype

[7]Students in our sample largely applied to the university under a "test flexible" policy precipitated by Covid, and many do not have ACT or SAT scores.

measure used by Kugler et al. (2021).) Although adding these two beliefs shrinks the gender gap (female coefficient) slightly, the association between IAT and intended major is unchanged for both genders.

Owen (2023) found that women and men have different beliefs, on average, about their relative performance in STEM courses, with women more likely to believe the median STEM major has higher grades than they do in reality. If these inaccurate beliefs correlate with implicit stereotypes, they may explain the IAT-intended major relationship. In column 7, we add in controls intended to capture beliefs about ability required for different majors, in the form of two survey items about what a student estimates to be the average high school GPAs of STEM and humanities majors at their university. The estimated coefficients closely resemble the previous set.

Another commonly cited explanation for (and potential solution to) gender gaps in STEM is the presence–or lack–of female role models (Patnaik et al., 2024; Porter and Serra, 2020; Carrell et al., 2010; Bettinger and Long, 2005). Perhaps gender-STEM stereotypes are related to–and possibly formed by–the gender of role models available to students. If this were true, it may be that our stereotype measure is picking up a lack of role models, which is the true cause of women not pursuing STEM. We include three measures of female role models. The first is based on a survey question asking students to name their favorite math or science teacher, with a dropdown for title (Ms., Mr., or Mx.); we use the selection of "Ms." or "Mx." as a proxy for non-male STEM role models, and the equivalent for English/social studies teachers. To measure the presence of female STEM role models in college, we use transcript data and instructor names to calculate the proportion of each student's STEM instructors who are women. Third, we construct a Duncan dissimilarity index (described in more detail in Section 2.4 and in Appendix D) based on the zip-code level labor market a student grew up in, which measures how gender segregated the local labor market is.

We note here that including these role model measures in our regression may in fact be over-controlling, if role models affect the formation of implicit stereotypes. In the case of college-level role models, this may be jointly determined with IAT, with students with stronger stereotypes choosing to take fewer courses with women. Thus, adding these controls would likely lead to a conservative estimate of our coefficients of interest. However, our coefficients of interest remain stable with the inclusion of role model controls (columns 8 and 9).

To summarize, we find a remarkably robust relationship between implicit stereotypes, gender, and intended major. Men with a one standard deviation stronger male-STEM association as measured on the IAT are 7 to 9 p.p. more likely to state an interest in STEM. Women, on the other hand, are 8 to 10 p.p. *less* likely to be interested in STEM when they hold stronger implicit stereotypes. Although we do not have exogenous variation in stereotypes, the robustness of this relationship to many of the commonly cited determinants of major choice and the gender gap therein leads us to conclude that the IAT is picking up something distinct.

To further argue that that IAT reflects *implicit* beliefs and not explicit ones, we run a robustness check where we estimate Equation 1 but replace IAT score with a measure of *explicit* gender-STEM stereotypes, operationalized as the proportion of STEM graduates that a student believes are women. The results, in Online Appendix Table A2, show no relationship between this measure and intended major, for either men or women.

It is also important to point out that as we progressively added controls to Table 2, the conditional gender gap in STEM intent shrank (from 17.2 to 13.4 p.p.) and the R^2 increased (from 0.065 to 0.272). So it is not the case that these other factors don't matter for major choice, or for the gender gap. Rather, they matter but the IAT maintains distinct predictive power.

3.3 IAT scores and other STEM outcomes

Because the majority of our sample is comprised of first year students, most of whom have not yet officially declared a major, we have thus far focused on stated major intent as captured by our survey. Of course, students' observed behavior might differ from what they state on a survey, and this divergence may differ by gender. In this section, we examine a number of observed STEM-related outcomes using university administrative data. We look at the extensive margin of STEM (taking any courses), the intensive margin (number of credits), and officially declared major.

Table 3 shows the results of estimating Equation 1 for six additional outcomes. From this point forward, we include the full set of control variables, i.e., those in column 9 of Table 2. Columns 1 through 3 measure STEM choices in Fall 2023, the same semester that students took the survey and IAT; columns 4 through 6 measure the same outcomes in the subsequent semester, Spring 2024. The estimated coefficients in Table 3 tell the same story as Table 2: men with stronger implicit gender-STEM stereotypes are more likely to take STEM courses (8 p.p. for fall, and 10 p.p. for the subsequent spring), and take more STEM credits (1.3 in fall, 1.7 in spring), than men with weaker stereotypes. Women with stronger implicit stereotypes are 4 p.p. less likely to take any STEM in both semesters, and take around 1 fewer STEM credit. The relationship between IAT and officially declaring a STEM major is not statistically significant for men, though the coefficient is between 5 and 6 percentage points both semesters. Women with a one s.d. higher IAT score are 6.1 p.p. less likely to be declared as STEM in Fall 2023, and 6.8 p.p. less likely to be declared in the spring.[8]

In Table 4, we investigate the predictive power of implicit stereotypes for three final outcomes. Column 1 is a measure of STEM persistence, where the outcome is taking any STEM courses in spring 2024, conditional on stating an intent to major in STEM on the fall survey. In other words, Column 1 of Table 4 is equivalent to column 4 of Table 3, but with the sample limited to students with an initial interest in STEM. The signs

[8]In Table A1 of the online appendix, we test the sensitivity of our results to including economics in the definition of STEM. Economics, like many STEM fields, is male-dominated and may be subject to the influence of stereotypes. However, the gender-science IAT does not include any social science words. The results in Table A1 are very similar to those in Tables 2 and 3.

of the coefficients are similar to those in Table 3, but smaller in magnitude. Conditional on initial STEM interest, men with a one s.d. higher IAT score are 3 p.p. more likely to continue on to take STEM courses in the spring; women are 3 p.p. less likely. This implies two facts: first, implicit stereotypes appear to matter not just for initial STEM interest, but for persisting through more coursework. Second, the fact that the coefficients are smaller in magnitude than they are for the full sample implies that implicit stereotypes matter more for students *without* an initial intent to major in STEM, especially for men.

The final two columns of Table 4 look at STEM performance, measured as GPA in Fall 2023 (column 2) and Spring 2024 (column 3). Only students who took at least one STEM course in the given semester are included. Column 2 reveals that in the concurrent semester, IAT score is not associated with STEM grades. However, by Spring 2024, women with higher IAT scores have higher STEM GPAs, by around a tenth of a grade point per standard deviation increase in IAT score. This could mean that stronger implicit stereotypes cause women to perform better. This would be at odds with the theory of stereotype threat, in which the existence of a stereotype can cause the disadvantaged group to perform worse out of fear of confirming the stereotype (Steele, 1997; Spencer et al., 1999, 2016). The result could alternatively be consistent with a positive selection story for women. Recall that Table 3 and column 1 of Table 4 showed that women with the strongest male-STEM stereotypes are more likely to drop out of STEM. But in Table 4, column 3, the remaining women with stronger male-STEM associations have higher GPA. This could mean that among women with strong male-STEM stereotypes, only the higher-performing ones remain. In other words, strong aptitude for STEM can help some women overcome implicit stereotypes, while women with both mediocre performance and strong implicit stereotypes are receiving multiple signals (implicit and explicit) that push them out of STEM. This result is consistent with Kugler et al. (2021), who find that receiving multiple signals that they don't belong, in the form of grades and gender composition, can decrease female STEM persistence.

4 Discussion and Conclusion

If implicit stereotypes are unconsciously keeping women out of STEM, several types of interventions might prove useful targets for future research. In some settings, simply making individuals aware of implicit biases or stereotypes has been enough to change behavior (Alesina et al., 2024; Boring and Philippe, 2021). Presenting students with counter-stereotypical examples might help counteract stereotypes; indeed, this might help explain the success of several recent interventions featuring female role models (Porter and Serra, 2020; Patnaik et al., 2024). However, gender stereotypes have been documented in children as young as six (Bian et al., 2017), suggesting that changing them may require early and sustained intervention.

References

Alesina, A., Carlana, M., La Ferrara, E., and Pinotti, P. (2024). Revealing stereotypes: Evidence from immigrants in schools. *American Economic Review*, 114(7):1916–1948.

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1):185–223.

Andreescu, T., Gallian, J. A., Kane, J. M., and Mertz, J. E. (2008). Cross-cultural analysis of students with exceptional talent in mathematical problem solving. *Notices of the AMS*, 55(10):1248–1260.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Baker, M. and Cornelson, K. (2018). Gender-based occupational segregation and sex differences in sensory, motor, and spatial aptitudes. *Demography*, 55:1749–1775.

Baker, R., Bettinger, E., Jacob, B., and Marinescu, I. (2018). The effect of labor market information on community college students' major choice. *Economics of Education Review*, 65:18–30.

Bettinger, E. P. and Long, B. T. (2005). Do faculty serve as role models? The impact of instructor gender on female students. *American Economic Review*, 95(2):152–157.

Bian, L., Leslie, S.-J., and Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, 355(6323):389–391.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., and Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology*, 94(3):567.

Booth, A. L., Cardona-Sosa, L., and Nolen, P. (2018). Do single-sex classes affect academic achievement? An experiment in a coeducational university. *Journal of Public Economics*, 168:109–126.

Bordalo, P., Coffman, K., Gennaioli, N., and Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794.

Boring, A. and Philippe, A. (2021). Reducing discrimination in the field: Evidence from an awareness raising intervention targeting gender biases in student evaluations of teaching. *Journal of Public Economics*, 193:104323.

Bostwick, V. K. and Weinberg, B. A. (2022). Nevertheless she persisted? Gender peer effects in doctoral STEM programs. *Journal of Labor Economics*, 40(2):397–436.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *The Quarterly Journal of Economics*, 134(3):1163–1224.

Carrell, S. E., Page, M. E., and West, J. E. (2010). Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Cheryan, S., Ziegler, S. A., Montoya, A. K., and Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143(1):1.

Conlon, J. J. (2021). Major malfunction: A field experiment correcting undergraduates' beliefs about salaries. *Journal of Human Resources*, 56(3):922–939.

Cortes, P. and Pan, J. (2019). Gender, occupational segregation, and automation. Economics Studies at Brookings report.

Cvencek, D., Meltzoff, A. N., and Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child Development*, 82(3):766–779.

Delaney, J. and Devereux, P. J. (2021). Gender and educational achievement: Stylized facts and causal evidence. CEPR Discussion Paper No. DP15753.

Duncan, O. D. and Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20(2):210–217.

Ertl, B., Luttenberger, S., and Paechter, M. (2017). The impact of gender stereotypes on the self-concept of female students in STEM subjects with an under-representation of females. *Frontiers in Psychology*, 8:703.

Fiedler, K. and Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology*, 27(4):307–316.

Glover, D., Pallais, A., and Pariente, W. (2017). Discrimination as a self-fulfilling prophecy: Evidence from French grocery stores. *The Quarterly Journal of Economics*, 132(3):1219–1260.

Greenwald, A. G. and Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1).

Greenwald, A. G., Brendl, M., Cai, H., Cvencek, D., Dovidio, J. F., Friese, M., Hahn, A., Hehman, E., Hofmann, W., Hughes, S., et al. (2022). Best research practices for using the implicit association test. *Behavior Research Methods*, pages 1–20.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6):1464–1480.

Greenwald, A. G., Nosek, B. A., and Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2):197.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., and Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1):17.

Jensen, E. J. and Owen, A. L. (2000). Why are women such reluctant economists? Evidence from liberal arts colleges. *American Economic Review*, 90(2):466–470.

Kiefer, A. K. and Sekaquaptewa, D. (2007). Implicit stereotypes and women's math performance: How implicit gender-math stereotypes influence women's susceptibility to stereotype threat. *Journal of Experimental Social Psychology*, 43(5):825–832.

Kugler, A. D., Tinsley, C. H., and Ukhaneva, O. (2021). Choice of majors: Are women really different from men? *Economics of Education Review*, 81:102079.

Lepage, L.-P., Li, X., and Zafar, B. (2022). Anticipated gender discrimination and grade disclosure. National Bureau of Economic Research Working Paper no. 30765.

Lohman, D. F. and Lakin, J. M. (2009). Consistencies in sex differences on the cognitive abilities test across countries, grades, test forms, and cohorts. *British Journal of Educational Psychology*, 79(2):389–407.

Martínez, J. (2022). The long-term effects of teachers' gender stereotypes on labor outcomes. Working Paper.

McConnell, A. R. and Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37(5):435–442.

Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122(3):1067–1101.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., Greenwald, A. G., et al. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European review of social psychology*, 18(1):36–88.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., et al. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, 106(26):10593–10597.

Owen, A. L. and Hagstrom, P. (2021). Broadening perceptions of economics in a new introductory economics sequence. *The Journal of Economic Education*, 52(3):175–191.

Owen, S. (2023). College major choice and beliefs about relative performance: An experimental intervention to understand gender gaps in STEM. *Economics of Education Review*, 97:102479.

Patnaik, A., Pauley, G., Venator, J., and Wiswall, M. (2024). The impacts of same and opposite gender alumni speakers on interest in economics. *Economics of Education Review*, 102:102579.

Patnaik, A., Wiswall, M., and Zafar, B. (2021). College majors 1. *The Routledge Handbook of the Economics of Education*, pages 415–457.

Porter, C. and Serra, D. (2020). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–254.

Reuben, E., Wiswall, M., and Zafar, B. (2017). Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. *The Economic Journal*, 127(604):2153–2186.

Rooth, D.-O. (2010). Automatic associations and discrimination in hiring: Real world evidence. *Labour Economics*, 17(3):523–534.

Rowatt, W. C., Franklin, L. M., and Cotton, M. (2005). Patterns and personality correlates of implicit and explicit attitudes toward Christians and Muslims. *Journal for the Scientific Study of Religion*, 44(1):29–43.

Salles, A., Awad, M., Goldin, L., Krus, K., Lee, J. V., Schwabe, M. T., and Lai, C. K. (2019). Estimating implicit and explicit gender bias among health care professionals and surgeons. *JAMA Network Open*, 2(7):e196545–e196545.

Spencer, S. J., Logel, C., and Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology*, 67(1):415–437.

Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35(1):4–28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52(6):613.

Wiswall, M. and Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

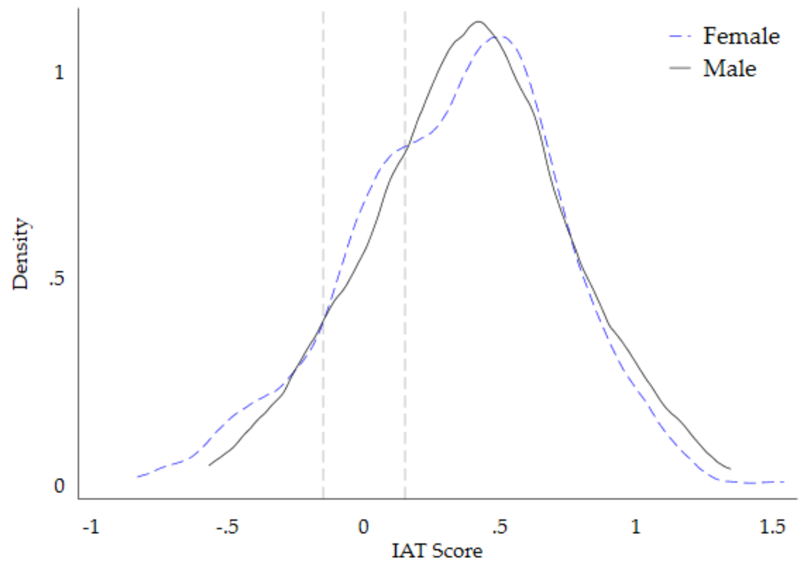
Exhibits

Table 1: Descriptive Statistics for Analysis Sample and Targeted Population

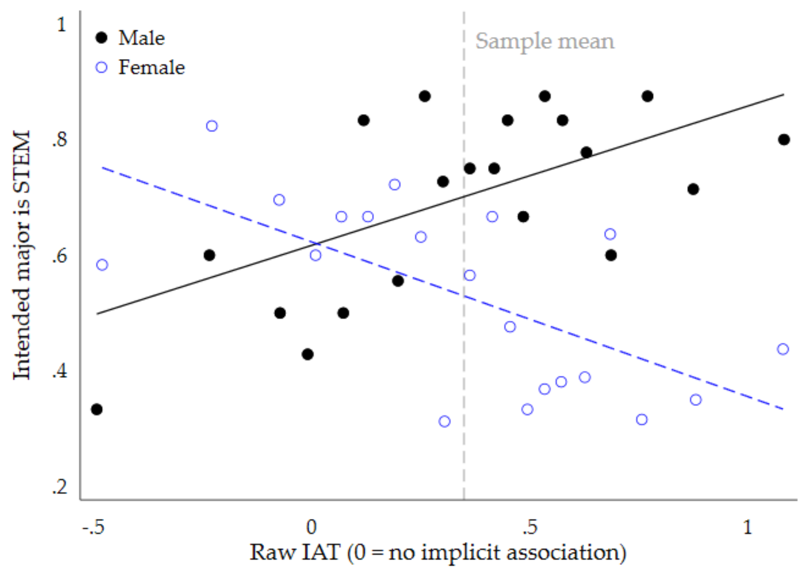
	Analysis sample	Target population	All university undergrads
Female	0.730	0.537	0.529
First year	0.715	0.707	0.264
White	0.611	0.598	0.615
Asian	0.317	0.301	0.294
Black	0.074	0.082	0.065
Hispanic	0.106	0.138	0.103
International	0.052	0.065	0.082
First-generation college	0.181	0.183	0.154
Missing parent ed	0.020	0.012	0.030
Family income >100K	0.483	0.509	0.510
Missing family income	0.226	0.225	0.232
High school GPA	3.923	3.901	3.886
Missing HS GPA	0.091	0.105	0.212
Took calculus in HS	0.648	0.610	0.591
Took any STEM courses FA23	0.872	0.888	0.753
Declared as STEM major FA23	0.231	0.262	0.354
Top intended major STEM	0.581	-	-
N	540	9,041	34,082

Notes: Analysis sample is limited to students who completed the survey and IAT. Target population includes students enrolled in participating ECoach courses, as well as the random sample of 2,500 first year students provided by the registrar's office. Race categories are not mutually exclusive; multi-racial and ethnic students have values of 1 for multiple categories. STEM courses and declared major are classified by 2-digit CIP code, and include natural resources and environmental sciences, computer science, engineering, math and statistics, biological sciences, and physical sciences. Intended STEM major is based on a survey item asking students the field they are most likely to major in, and STEM includes engineering, biology, computer science, math, statistics, data science, neuroscience, environmental science, and other natural sciences.

Figure 1: IAT Score Raw Distribution and Correlation with Major Choice, by Gender



(a) Raw Distribution of IAT Scores by Gender



(b) Relationship between IAT Score and Intended Major, by Gender

Notes: (a) Sample includes students who answered the student survey and completed the IAT (N=540; 394 women and 146 men). IAT is scored following the algorithm in Greenwald et al. (2003). Negative scores indicate an automatic association for female with science and male with humanities (non-stereotypical association) while positive scores indicate the stereotypical association. Scores with an absolute value less than 0.15 (region between gray lines) are considered to have little to no association either way (Greenwald et al., 2009).
(b) Sample includes students who answered the intended major item on the survey and completed the IAT (N=539; 394 women and 145 men). Figure is a binned scatterplot using 20 equally sized bins for each gender.

Table 2: Relationship between IAT Score, Gender, and Intended Major

	Dependent variable: student intends STEM major								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female	-0.172***	-0.173***	-0.147**	-0.161***	-0.148**	-0.137**	-0.133**	-0.134**	-0.134**
	(0.045)	(0.046)	(0.045)	(0.046)	(0.046)	(0.048)	(0.048)	(0.049)	(0.049)
IAT score (std.)	0.090*	0.081*	0.083*	0.070+	0.074+	0.078*	0.075+	0.078*	0.080*
	(0.040)	(0.040)	(0.038)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)	(0.039)
Female x IAT	-0.191***	-0.185***	-0.184***	-0.165***	-0.165***	-0.170***	-0.168***	-0.167***	-0.169***
	(0.047)	(0.047)	(0.045)	(0.045)	(0.046)	(0.046)	(0.046)	(0.046)	(0.046)
IAT effect for women	-0.100***	-0.104***	-0.102***	-0.095***	-0.092***	-0.092***	-0.093***	-0.089***	-0.089***
(main + interaction)	(0.024)	(0.024)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Demographics?		x	x	x	x	x	x	x	x
Academic preparation?			x	x	x	x	x	x	x
Major importance factors?				x	x	x	x	x	x
Salary beliefs?					x	x	x	x	x
Explicit gender-major beliefs?						x	x	x	x
Major ability beliefs?							x	x	x
Role models?								x	x
Local gender segregation?									x
R^2	0.065	0.099	0.156	0.216	0.236	0.248	0.258	0.271	0.272
N	539	539	539	539	539	539	539	539	539

Notes: +$p < 0.1$, * $p < 0.05$, ** p< 0.01, ***$p < .001$. Robust standard errors in parentheses. Results are from a regression of intended STEM major on gender, standardized IAT score, and their interaction. More positive IAT scores indicate stronger male-STEM bias. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. Regressions also include missing indicators for all included control variables.

Table 3: Relationship between IAT Score, Gender, and STEM Outcomes

	(1) Any STEM courses FA23	(2) # STEM credits FA23	(3) Declared as STEM FA23	(4) Any STEM courses SP24	(5) # STEM credits SP24	(6) Declared as STEM SP24
Female	-0.018	-1.584**	-0.117*	0.023	-1.599**	-0.104*
	(0.027)	(0.485)	(0.047)	(0.040)	(0.533)	(0.049)
IAT score (std.)	0.079***	1.318***	0.058	0.102**	1.710***	0.056
	(0.021)	(0.393)	(0.042)	(0.032)	(0.423)	(0.041)
Female x IAT	-0.119***	-2.311***	-0.119*	-0.142***	-2.704***	-0.124**
	(0.024)	(0.441)	(0.049)	(0.037)	(0.477)	(0.047)
IAT effect for women	-0.040**	-0.993***	-0.061**	-0.041*	-0.994***	-0.068**
(main effect + interaction)	(0.013)	(0.200)	(0.023)	(0.020)	(0.223)	(0.023)
R^2	0.461	0.296	0.145	0.166	0.222	0.168
N	540	540	540	540	540	540

Notes: $+p < 0.1$, * $p < 0.05$, ** p< 0.01, ***$p < .001$. Robust standard errors in parentheses. Results are from a regression of the outcome on gender, standardized IAT score, and their interaction, as well as the full set of controls. STEM courses and declared major are classified by 2-digit CIP code, and include natural resources and environmental sciences, computer science, engineering, math and statistics, biological sciences, and physical sciences. More positive IAT scores indicate stronger male-STEM bias. Controls include the full set of demographics, academic preparation, major importance factors, salary beliefs, major ability beliefs, explicit gender-major beliefs, role models, and local labor market occupational gender segregation. Regressions also include indicators for missing any of the above.

Table 4: Relationship of IAT Score with STEM Performance and Persistence

	(1) Persisted in STEM SP24	(2) STEM GPA FA23	(3) STEM GPA SP24
Female	0.000	-0.005	0.018
	(0.023)	(0.065)	(0.060)
IAT score (std.)	0.031*	-0.048	0.062
	(0.015)	(0.052)	(0.046)
Female x IAT	-0.061**	0.089	0.034
	(0.021)	(0.059)	(0.053)
IAT effect for women	-0.030*	0.040	0.096**
(main effect + interaction)	(0.014)	(0.030)	(0.030)
R^2	0.177	0.227	0.238
N	313	466	427

Notes: $+p < 0.1$, * $p < 0.05$, ** p< 0.01, ***$p < .001$. Robust standard errors in parentheses. Results are from a regression of the outcome on gender, standardized IAT score, and their interaction, as well as the full set of controls. More positive IAT scores indicate stronger male-STEM bias. STEM GPA is calculated using STEM courses only, where STEM is classified by 2-digit CIP code, and include natural resources and environmental sciences, computer science, engineering, math and statistics, biological sciences, and physical sciences. STEM GPA regressions are limited to students who took any STEM courses in the specified semester. Persistence is defined as taking any STEM courses in Spring 2024; persistence regression only includes students who listed a STEM major as their most likely major on the survey. Controls include the full set of demographics, academic preparation, major importance factors, salary beliefs, major ability beliefs, explicit gender-major beliefs, role models, and local labor market occupational gender segregation. Regressions also include indicators for missing any of the above.

Appendix A. Supplementary Exhibits

Table A1: Relationship between IAT Score and STEM Outcomes, Including Economics as STEM

	(1) Top major STEM or econ	(2) Any STEM or econ courses FA23	(3) # STEM and econ credits FA23	(4) Declared STEM/econ FA23	(5) Any STEM or econ courses SP24	(6) # STEM and econ credits SP24	(7) Declared STEM/econ SP24
Female	-0.154***	-0.015	-1.762***	-0.114*	-0.018	-1.685***	-0.108*
	(0.045)	(0.027)	(0.477)	(0.047)	(0.034)	(0.501)	(0.049)
IAT score (std.)	0.095**	0.080***	1.244**	0.050	0.067*	1.464***	0.044
	(0.035)	(0.021)	(0.384)	(0.042)	(0.028)	(0.402)	(0.041)
Female x IAT	-0.158***	-0.123***	-2.155***	-0.107*	-0.094**	-2.253***	-0.104*
	(0.042)	(0.025)	(0.431)	(0.048)	(0.033)	(0.455)	(0.048)
IAT effect for women	-0.063**	-0.043***	-0.911***	-0.057*	-0.027	-0.789***	-0.060*
(main effect + interaction)	(0.023)	(0.013)	(0.202)	(0.023)	(0.018)	(0.213)	(0.023)
R^2	0.287	0.446	0.305	0.144	0.168	0.238	0.170
N	539	540	540	540	540	540	540

Notes: Robust standard errors in parentheses. $+p < 0.1$, * $p < 0.05$, ** $p < 0.01$, ***$p < .001$. Results from a regression of outcome on gender, standardized IAT score, and their interaction, as well as the full set of controls. More positive IAT scores indicate stronger male-STEM bias. STEM courses and declared major are classified by 2-digit CIP code, and include natural resources and environmental sciences, computer science, engineering, math and statistics, biological sciences, and physical sciences. Here, economics (CIP 45.06) is also included as part of STEM/econ. Controls include the full set of demographics, academic preparation, major importance factors, salary beliefs, major ability beliefs, explicit gender-major beliefs, role models, and local labor market occupational gender segregation. Regressions also include indicators for missing any of the above.

Table A2: Relationship between Explicit Gender-STEM Beliefs, Gender, and Intended Major

	Intends STEM Major
Female	-0.287
	(0.185)
Explicit belief: Percent	0.000
STEM degrees going to women	(0.004)
Female x Explicit belief	0.003
	(0.005)
Explicit belief effect for women	0.003
(main effect + interaction)	(0.003)
R^2	0.028
N	537

Notes: $+p < 0.1$, * $p < 0.05$, ** p< 0.01, ***$p < .001$. Robust standard errors in parentheses. Results from a regression of intended STEM major on gender, explicit gender-STEM beliefs, and their interaction. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. Explicit STEM belief is the proportion of STEM graduates the student believes are female, from 0 to 100.

Appendix B. IAT Details

The IAT consists of seven blocks, with between 20 and 40 tasks per block, summarized in Table B1. Each task requires a participant to sort a word to the left or right of a computer, tablet, or phone screen, using keystrokes or touch. The words fall into four categories: male, female, science, and humanities. Figure B1 lists all possible words. The correct sorting depends on the block. There are three practice rounds that familiarize participants with the procedure (e.g., sorting science words to one side and humanities to another; and the same for male and female). In the four rounds used for scoring, participants must either sort male and science words to one side and female and humanities to the other (stereotypical pairing), or sort male and humanities to one side and female and science to the other (non-stereotypical). Within a block, words are presented randomly. For example, in block 3, a participant might see the first six words: Engineering, Math, Uncle, English, Woman, Son; another might see Mother, Chemistry, Literature, Wife, Father, Daughter. Figure B2 shows screenshots of sample tasks.

Studies using the IAT typically use an instrument that is "counterbalanced," meaning the order in which participants are asked to do the stereotype-conforming versus non-conforming tasks is randomly assigned. In our setting, due to a programming error, this was not the case: all participants did the stereotype conforming task (male-science) first. In other words, all of our participants saw the blocks in the order listed in Table B1, whereas a counterbalanced instrument would randomly assign some participants to see them in the order of 1, 2, 6, 7, 5, 3, 4.

There are well-documented order effects, where participants show stronger bias or stereotypes when they see the stereotypical task first (Greenwald et al., 1998). However, there is some disagreement among experts on whether it is best-practice to counterbalance instruments if the research question regards the correlation between the IAT and other variables, as varying order can introduce additional noise (Greenwald et al., 2022).

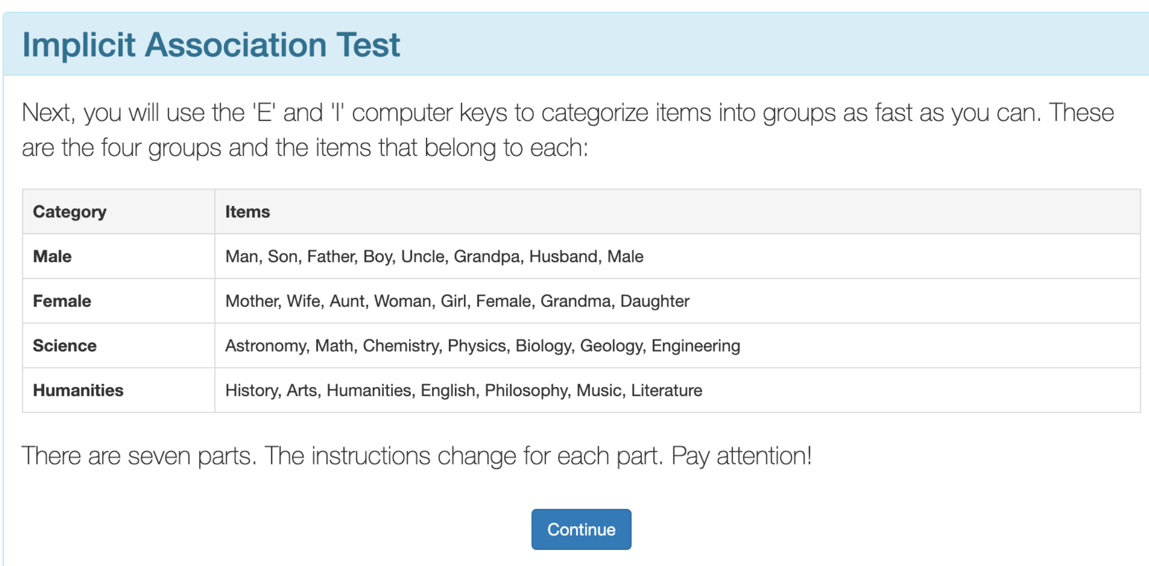
While we cannot directly test for how much this matters with our data, we bring in data from a related study to assuage any concerns about block order. In Fall 2024, a year after the focal study took place, we conducted a second round of data collection, paired with an experimental intervention (the analysis for this experiment is currently in progress). In this second round, the IAT was counterbalanced. We use the pre-intervention data from Fall 2024, which included a nearly identical survey and IAT, to show that our key finding is not sensitive to order. Table B2 shows that in the full sample, the relationship between gender, IAT, and intended major is qualitatively similar to our main result here (column 1). In the second column, we add a control for order; in columns 3 and 4, we split the sample by the block order of their IAT. Though the point estimates vary somewhat by order, they are not statistically distinguishable from each other, reassuring us that the counterbalancing error is not driving our results.

Table B1: Summary of IAT Blocks

Block	Left Categories	Right Categories	Number of Tasks
1	Science	Humanities	20
2	Male	Female	20
3*	Male Science	Female Humanities	20
4*	Male Science	Female Humanities	40
5	Humanities	Science	28
6*	Male Humanities	Female Science	20
7*	Male Humanities	Female Science	40

Notes: Blocks 3, 4, 6, and 7 (*) are used in scoring; the remaining blocks are practice rounds.

Figure B1: Categories and Words in Gender-Science IAT



Implicit Association Test

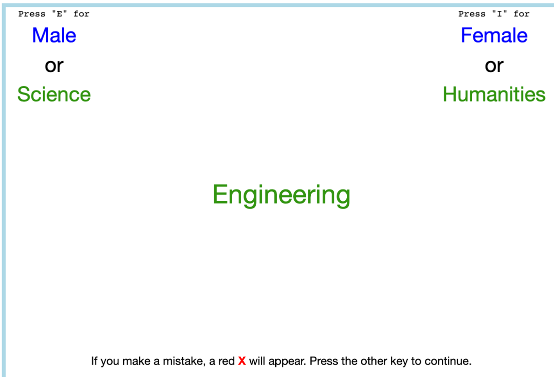
Next, you will use the 'E' and 'I' computer keys to categorize items into groups as fast as you can. These are the four groups and the items that belong to each:

Category	Items
Male	Man, Son, Father, Boy, Uncle, Grandpa, Husband, Male
Female	Mother, Wife, Aunt, Woman, Girl, Female, Grandma, Daughter
Science	Astronomy, Math, Chemistry, Physics, Biology, Geology, Engineering
Humanities	History, Arts, Humanities, English, Philosophy, Music, Literature

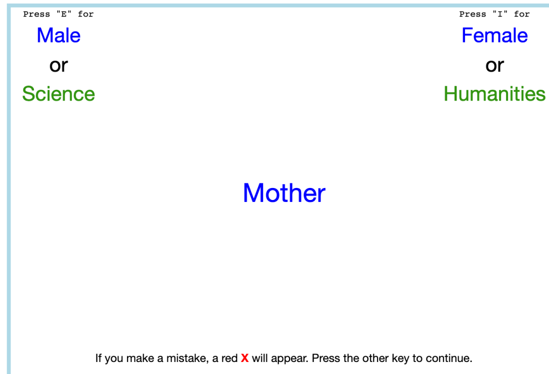
There are seven parts. The instructions change for each part. Pay attention!

Continue

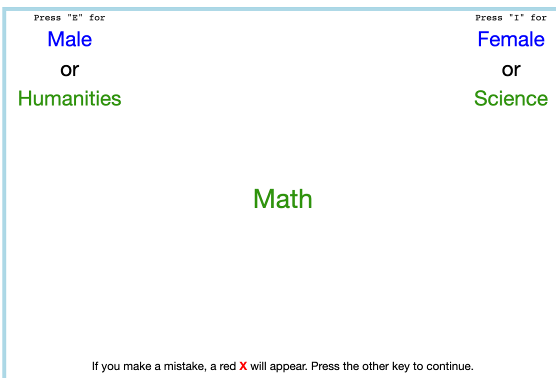
Figure B2: Sample IAT Tasks



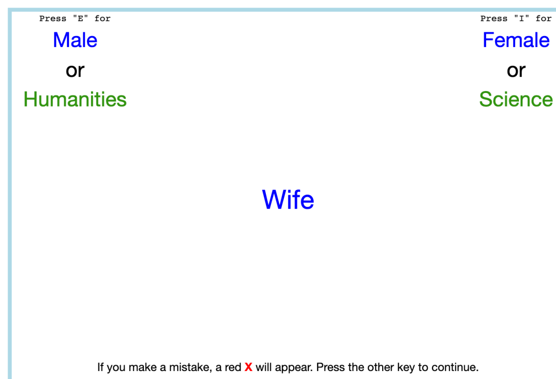
(a) Male-science/female-humanities (stereotypical)



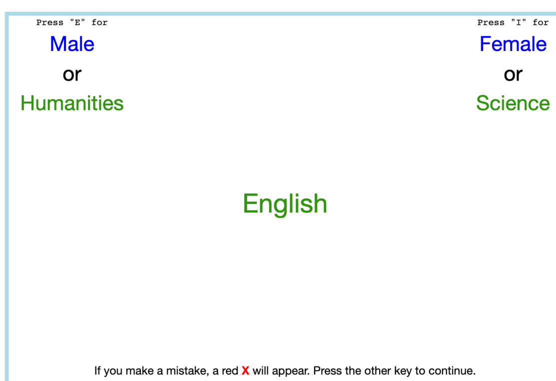
(b) Male-science/female-humanities (stereotypical)



(c) Male-humanities/female-science (non-stereotypical)



(d) Male-humanities/female-science (non-stereotypical)



(e) Male-humanities/female-science (non-stereotypical)

Table B2: Sensitivity of IAT-Intended Major Relationship to IAT Block Order

	Dependent variable: student intends STEM major			
	(1)	(2)	(3)	(4)
	No	Order	By order	
	controls	control	Order 1	Order 2
Female	-0.173***	-0.175***	-0.177**	-0.160**
	(0.040)	(0.040)	(0.060)	(0.058)
IAT score (std.)	0.059+	0.052	0.074	0.031
	(0.035)	(0.035)	(0.055)	(0.048)
Female x IAT	-0.177***	-0.183***	-0.214**	-0.151**
	(0.041)	(0.041)	(0.065)	(0.057)
IAT effect for women	-0.117***	-0.130***	-0.140***	-0.120***
(main effect + interaction)	(0.022)	(0.023)	(0.035)	(0.031)
R^2	0.070	0.076	0.102	0.052
N	689	689	336	353

Notes: Robust standard errors in parentheses. $+p < 0.1$, * $p < 0.05$, ** p< 0.01, ***$p < .001$. This analysis is based on participants in a Fall 2024 study and experimental intervention. IAT score and intended major were collected pre-intervention. Results are from a regression of intended STEM major on gender, standardized IAT score, and their interaction. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. More positive IAT scores indicate stronger male-STEM bias. Order refers to whether the IAT asked the participant to do the order-compatible (stereotypical) task first (Order 1), or the incompatible (non-stereotypical) task first (Order 2).

Appendix C. Survey Instrument

First, we would like to ask you some questions about your academic plans at the University of the Midwest.

[*expected_major_1*]

At UM, there are dozens of academic majors to choose from. What major do you think you're **most likely** to graduate with a degree in? If you plan to double major, please select what you consider to be your primary major.

Science, Technology, Engineering, and Math (STEM)

- Engineering
- Biology
- Natural science other than biology (astronomy, chemistry, earth science, geology, physics, etc.)
- Computer science
- Math
- Statistics
- Neuroscience

Humanities

- Arts (visual, performing, etc.)
- Languages (Spanish, French, Chinese, etc.)
- Other humanities (English, philosophy, history, etc.)

Social Sciences

- Economics
- Psychology
- Social science other than econ or psych (political science, sociology, etc.)

Other

- Business
- Health-related (e.g., kinesiology/movement science, pharmacy, public health, etc.)
- Public policy
- Other (please fill in)

[*expected_major_2*]

What major do you think you're **second most likely** to graduate with a degree in? If you plan to double major, this would be your second major.

Science, Technology, Engineering, and Math (STEM)

- Engineering
- Biology

- Natural science other than biology (astronomy, chemistry, earth science, geology, physics, etc.)
- Computer science
- Math
- Statistics
- Neuroscience

Humanities

- Arts (visual, performing, etc.)
- Languages (Spanish, French, Chinese, etc.)
- Other humanities (English, philosophy, history, etc.)

Social Sciences

- Economics
- Psychology
- Social science other than econ or psych (political science, sociology, etc.)

Other

- Business
- Health-related (e.g., kinesiology/movement science, pharmacy, public health, etc.)
- Public policy
- Other (please fill in)

[*expected_salary_1*]

Assume you graduate with a bachelor's degree in **[*expected_major_1*]** (and no double major). How much money do you predict you would make per year, at age 40, with that major? (If you expect to make more than $300,000, please select 300 below.)

$ 0 K $ 50 K $ 100 K $ 150 K $ 200 K $ 250 K $300 K

Expected salary (thousands of $)

[*expected_salary_2*]

Assume you graduate with a bachelor's degree in **[*expected_major_2*]** (and no double major). How much money do you predict you would make per year, at age 40, with that major? (If you expect to make more than $300,000, please select 300 below.)

$ 0 K $ 50 K $ 100 K $ 150 K $ 200 K $ 250 K $300 K

Expected salary (thousands of $)

[*major_factors*]

How important are/were each of the following in choosing your major?

Not at all important 1	Slightly important 2	Moderately important 3	Very important 4	Extremely important 5

Feeling like I'm good at the subject

Being engaged with the coursework (while in school)

Making/having friends or study partners in the major

Expected salary (after graduation)

Work flexibility (after graduation)

Having a positive impact on society (after graduation)

Work culture/peers (after graduation)

Next, we have a few questions about students and majors at UM.

[*pct_stem_women*]

Think about all of the undergraduate students who graduated from UM last year with a Bachelor's degree in a **math, technology, science, or engineering (STEM)** subject. What proportion of those students would you estimate are women?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Percent of UM STEM degrees going to women

[*pct_humanities_women*]

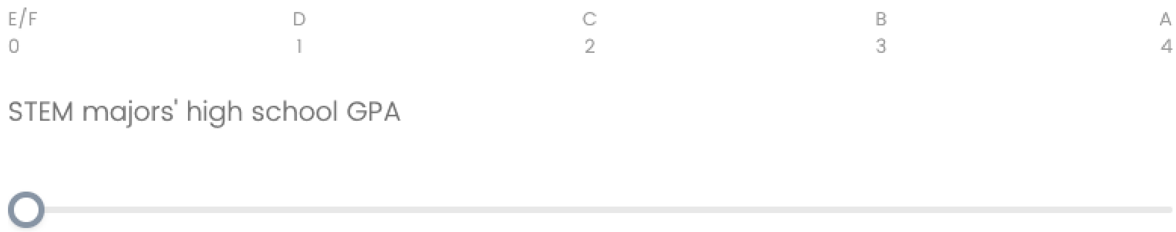
Think about all of the undergraduate students who graduated from UM last year with a Bachelor's degree in a **humanities** subject (includes English, languages, arts, history, philosophy, etc.). What proportion of those students would you estimate are women?

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Percent of UM humanities degrees going to women

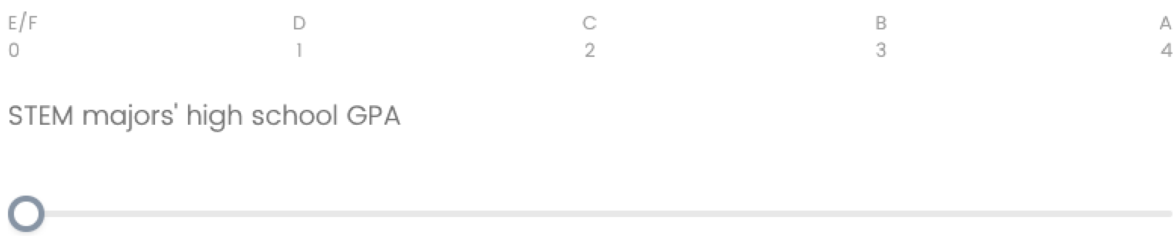
[*stem_hs_gpa*]

Again, think about all of the undergraduate students who graduated from UM last year with a Bachelor's degree in a **math, technology, science, or engineering (STEM)** subject. What do you think was the **average high school GPA** of these STEM majors?



[*hum_hs_gpa*]

Again, think about all of the undergraduate students who graduated from UM last year with a Bachelor's degree in a **humanities** subject (includes English, languages, arts, history, philosophy, etc.). What do you think was the **average high school GPA** of these humanities majors?



Now we'd like to ask you a couple of questions about your high school experience.

[*fav_teacher_hum*]

Think back to all of the **English and social studies** courses you took in high school. Who was your favorite high school teacher in those subjects?

	Title	Last name (fill in)
Favorite English/social studies teacher:		

[*fav_teacher_stem*]

Think back to all of the **math and science courses** you took in high school. Who was your favorite high school teacher in those subjects?

	Title	Last name (fill in)
Favorite math/science teacher:		

Thank you very much for answering the previous questions. As the second part of the study, we will now direct you to a link to take an implicit association test, or IAT. Completing the IAT should take about 5 minutes.

Please click the link below to take the Implicit Association Test. After you complete the IAT, you will be automatically entered into a lottery to win a $50 Amazon gift card.

Click here to take the IAT and complete the study.

Appendix D. Duncan Index

To measure local occupational gender segregation, we use data from the 5-year American Community Survey (ACS) files, available from the Census Bureau's website. These data are provided at the ZIP code level, which allows us to merge them with our student data. We use the 2011 to 2023 files; this range corresponds to the years when our sample were school-age.

Our analysis uses estimates of occupational distribution by gender across the 25 major occupations commonly classified by the Census as encompassing the primary jobs in the economy. We calculate the Duncan Index of Dissimilarity (Duncan and Duncan, 1955) which is commonly used in different settings studying segregration across various characteristics (Cortes and Pan, 2019; Baker and Cornelson, 2018):

$$D = \frac{1}{2} \sum_{i} \left| \frac{a_i}{A} - \frac{b_i}{B} \right|$$

Where: D is the segregation index, ranging from 0 (perfect integration) to 1 (complete segregation), a_i is the number of women in occupation i, b_i is the number of men in occupation i, A is the total number of women in the workforce and B is the total number of men in the workforce. This index provides a straightforward interpretation: it represents the proportion of individuals who would need to change occupations to achieve gender parity in the labor market.

We calculate D for each zip code and year. There are missing data elements at the zip code level. For example, the number of men in the Education occupation might be missing for some zip code-years due to data suppression rules. In all years we consider, the number of occupations with valid measures averaged between 19 to 22 out of 25, depending on the year. Occupations with missing counts in a given zip code and year are excluded from the calculation of the average within each year. Our final measure is at the ZIP code level, averaged over all years between 2011 and 2023. This captures the annual average level of occupational segregation in the ZIP code a student grew up in.