Implicit Gender-STEM Stereotypes and College Major Choice

Stephanie Owen (Colby College)
Derek Rury (Oregon State University)*

This version: December 1, 2025

Abstract

To study the role of implicit stereotypes in explaining the gender gap in college major choice, we administer a gender-science Implicit Association Test to a sample of undergraduates, and link results to survey and administrative transcript data. Women with a one standard deviation higher male-science association are 8-10 p.p. less likely to intend to major in STEM, 4 p.p. less likely to take STEM courses, and 6 p.p. less likely to declare a major. Men show the opposite. Results are robust to controls including explicit beliefs, preferences for major characteristics such as salary and job flexibility, and female role models.

Key Words: STEM, Stereotypes, College Major, Gender, Education
JEL Codes: I21, J16, D83, D91

*Email: sowen@colby.edu; ruryd@oregonstate.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, as well as Josh Hyman and Basit Zafar for valuable feedback on an early draft. Cristian Bancayan Navarro provided excellent research assistance for this project. This study was approved by Colby College's Institutional Review Board (#2023-080), the University of Michigan's Health Sciences and Behavioral Sciences Institutional Review Board (HUM00240215), and the University of Chicago's Social and Behavioral Sciences Institutional Review Board (IRB23-1230).

1 Introduction

Despite great strides in education and the labor market, women remain underrepresented in science, technology, engineering, and mathematics (STEM) college degrees and associated occupations, with implications for economic inequality and efficiency. STEM fields tend to be higher-paying, so differences in major contribute to the gender pay gap (Sloane et al., 2021). And, to the extent that different choices by gender reflect inefficient sorting, removing barriers would increase overall productivity (Hsieh et al., 2019).

In this paper, we marry two research literatures, from psychology and economics, to investigate one potential explanation for the gender gap in STEM: the role of implicit (unconscious) gender stereotypes. Social scientists have long argued that stereotypes–overgeneralized beliefs about certain groups–may affect behavior (Steele, 1997; Bordalo et al., 2016). Our analysis focuses on the role of *implicit* stereotypes in particular: those that shape individuals' thinking and behavior unconsciously, possibly in contrast to consciously-held beliefs or goals (Greenwald and Banaji, 1995; Bertrand et al., 2005). We study the stereotype that STEM fields are associated with men.

Psychologists have documented the existence of implicit gender stereotypes and simple correlations with academic performance (Nosek et al., 2002; Kiefer and Sekaquaptewa, 2007), self-concept (Cvencek et al., 2011), and STEM interest (Lane et al., 2007; Cundiff et al., 2013; De Gioannis, 2022). These researchers have offered compelling explanations for how stereotypes can shape decisions, highlighting factors such as self-concept (Ertl et al., 2017), identification with a subject (Steele and Aronson, 1995), and sense of belonging (Spencer et al., 2016). However, these studies tend to ignore the possibility of omitted variable bias and fail to control for any factors that might confound the relationship between implicit stereotypes and outcomes. They also rely on self-reported measures of STEM intent, making it hard to know how results translate to actual decisions.

Recent work in economics has found plausibly causal relationships between implicit stereotypes and real-world economic outcomes (Rooth, 2010; Glover et al., 2017; Carlana, 2019; Alesina et al., 2024; Martínez, 2025). However, this work has focused on the effects of stereotypes held by individuals other than members of 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 primarily first year, first semester college students to capture their unconscious beliefs about the relationship between gender and STEM. 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.

We find that 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 (s.d.) 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, beliefs about the gender of recent STEM and humanities graduates of the university, and the presence of female role models. Results using administrative data on STEM-course taking and official major declaration show similar patterns. For women, a one s.d. higher IAT score is associated with a 4 p.p. reduction in taking any STEM courses in the same semester, and a 6 p.p. lower likelihood of being declared as a STEM major. Men are 8 p.p. more likely to take any STEM, and 6 p.p. more likely to be a major (though this last result is not statistically significant).

Although we lack the exogenous variation in implicit stereotypes needed to definitively establish a causal relationship, our results (especially when considered alongside previous work on the causal effect of stereotypes about others, e.g., Carlana (2019)) 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, existing explanations (e.g., mathematical ability) leave much of it statistically unexplained (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. 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.

The paper proceeds as follows. Section 2 summarizes relevant literature and our contribution to it. Section 3 describes our setting and data sources. Section 4 describes our empirical approach and results, and Section 5 concludes.

2 Related Literature and Contribution

2.1 Implicit Stereotypes and STEM Outcomes

Social psychologists have long been interested in the role of implicit stereotypes in shaping behavior. Greenwald and Banaji (1995) introduced the concept of implicit social cognition to describe mental processes that operate without conscious awareness or control yet influence social judgments and behavior. Their seminal paper defines implicit stereotypes as "introspectively unidentified (or inaccurately identified) traces of past experience that mediate attributions of qualities to members of a social category," and provides a framework to understand behaviors such as unintended discrimination. Subsequent work has shown that implicit gender stereotypes (as measured by the IAT) negatively predict performance (quantitative SAT scores and college-level calculus grades) for women (Nosek et al., 2002; Kiefer and Sekaquaptewa, 2007), and predict math self-concept among elementary school children (Cvencek et al., 2011) as well as college students (Nosek et al., 2002). (We discuss the IAT in more detail in section 3.2.)

Several studies have shown that female students with stronger implicit gender stereotypes are less likely to be science majors (Smeding, 2012; Smyth et al., 2009) or be interested in a science career (Young et al., 2013). However, these studies examine students who are later in college and have already selected a major, making it hard to disentangle whether their implicit stereotypes formed their choices or the other way around. For this reason, we focus on a sample of primarily first-year, first-semester students.

In a handful of studies most closely related to our own, gender-science implicit stereotypes have been shown to predict interest in and intent to major in STEM. Lane et al.

(2012) and Cundiff et al. (2013) find that first year college women with stronger implicit stereotypes are less likely to state STEM interest; De Gioannis (2022) finds the same in a sample of Italian high school seniors. The literature is more mixed on whether the opposite relationship holds for men. Lane et al. (2012) find that men with stronger stereotypes are more likely to intend STEM, but Cundiff et al. (2013) and De Gioannis (2022) find no effect for men.

These studies document bivariate correlations only, and do not control for any potential confounding factors. The exception is De Gioannis (2022), who controls for maternal education, interest in STEM and humanities, identification with STEM and humanities, high school performance, and importance of job salary and social utility in choosing a major. However, the context of that paper–Italian high school students, who have already chosen a specialized track in sciences, humanities, or languages–is distinct from the U.S. context, and students' college specialization choices may follow a different process. Furthermore, to our knowledge, none of the work in psychology has confirmed findings about stated intentions with administrative data on observed behavior. Our rich survey data *and* our access to detailed administrative data allow us to control for many if not most of the previously unobservable factors that may have been responsible for previously observed correlations, and to supplement findings on subjective STEM intentions with actual behavior (course-taking and declared major).

2.2 Causal Evidence on Implicit Stereotypes

Economists have primarily studied implicit stereotypes as a driver of labor market discrimination. Rooth (2010) showed that employers with stronger implicit stereotypes regarding Arab-Muslim men relative to Swedish men display a larger ethnic gap in real-life hiring callback rates; *explicitly* stated attitudes and stereotypes were much more weakly correlated with hiring decisions. Glover et al. (2017) use plausibly exogenous variation in grocery store worker scheduling to show that more racially biased managers (measured by an IAT) cause higher worker absenteeism and lower performance.

More recent work by education economists has been interested in implicit stereotypes held by teachers and the consequences for students. Carlana (2019) shows that Italian girls who are conditionally randomly assigned to a middle school math teacher with higher implicit gender-science stereotypes perform worse in math, have lower self-confidence, and choose lower academic tracks. Martínez (2025) follows teachers and students in Peru and finds that teachers with higher gender stereotype IAT scores benefit boys and hurt girls on a variety of longer-term educational and labor market outcomes. Alesina et al. (2024) show that Italian teachers' implicit stereotypes regarding immigrant children predict observed bias in grading. In two experiments, revealing implicit bias to teachers reduced grading discrepancies.

Although economics research regarding implicit stereotypes has emphasized identifying causal relationships, it has thus far only considered how stereotypes held by *others* affect their behaviors towards and outcomes of the stereotyped group. We add to this literature by examining implicit stereotypes and behaviors of the stereotyped group themselves. Though the psychology literature has done this (see Section 2.1), our study brings an econometric emphasis on omitted variables, using rich survey and administrative data to control for other important inputs into major choice and provide the strongest evidence to date that implicit gender stereotypes predict–and may causally determine–major choice.

2.3 Gender Gaps in STEM

Finally, we contribute to a broader literature about gender gaps in STEM. Field choice is a complex decision, and there are many hypothesized mechanisms for why men and women choose different specializations. Potential sources of gender differences include: mathematical aptitude and comparative advantage (Breda and Napp, 2019; Aucejo and James, 2021; Speer, 2023; Goulas et al., 2024), risk aversion and willingness to compete (Niederle and Vesterlund, 2007; Buser et al., 2014, 2017), interest and relevance of the topics/curriculum (Jensen and Owen, 2000; Owen and Hagstrom, 2021), preference for certain types of jobs and job characteristics (Wiswall and Zafar, 2015, 2018; Kuhn and Wolter, 2022), discrimination and bias (Carlana, 2019; Lepage et al., 2025), toxic culture and harassment (Aycock et al., 2019; Minnotte and Pedersen, 2023), response to grades and academic feedback (Owen, 2010; Kugler et al., 2021), and method of assessment (Azmat et al., 2016; Iriberri and Rey-Biel, 2021; Griselda, 2024).

These explanations can be difficult to disentangle, as they may reinforce each other. We attempt to isolate the potential role of implicit stereotypes, but acknowledge the difficulty of doing so. For example, academic performance might reinforce stereotypes, or the threat of the stereotype could affect performance. A number of researchers have studied the importance of female instructors and role models (Bettinger and Long, 2005;

Carrell et al., 2010; Porter and Serra, 2020; Patnaik et al., 2024) and the gender composition of peers (Booth et al., 2018; Bostwick and Weinberg, 2022; Calkins et al., 2023). One reason that the gender of others might matter is because the lack (or presence) of other women and girls confirms (or contradicts) an implicit stereotype.

With the above body of research in mind, we carefully designed our survey to capture, to the extent possible, the many mechanisms that may contribute to gender differences in major choice. This allows us to control for them in our analysis and provide evidence on how much omitted variable bias may be driving the implicit stereotype-major correlation. This approach likely over-controls for factors that could themselves affect or be affected by implicit stereotypes; thus, we argue that our estimates of the relationship between implicit stereotypes and major choice are likely conservative ones.

3 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. 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.

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.[1] Of the 2,500 registrar students we invited

[1] Appendix B includes recruitment text.

to participate, 276 (11%) completed consent and 171 (6.8%) are in our final analysis sample.

3.1 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.

3.2 Implicit Association Test

To measure implicit 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 measures the strength of automatic association between concepts. It is premised on the assumptions that (1) the strength of an implicit stereotype is reflected in an association between, e.g., gender and STEM; and (2) it takes additional time and effort to make a decision that overrides an unconscious stereotype or automatic association. 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 D. 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]

[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 D.

The IAT score is based on the average difference in 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 (Blanton et al., 2009). However, a growing number of papers studying teachers (Carlana, 2019; Martínez, 2025), 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 result 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.

3.3 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, standardized test scores, 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

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

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. Although the administrative data contain students' official majors, a large portion of our sample (60%) have no official major as most 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.)

3.4 Final Sample

To be in our analysis sample, we require a survey response and a complete IAT.[6] Our final sample (Owen and Rury, 2025) consists of 540 students; descriptive statistics for our sample and how they compare to both the targeted 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 scored in the 93rd percentile of the math section of the SAT or ACT and had a 3.9 GPA in high school, and the majority took calculus. (Due to test-flexible policies, over a quarter are missing a test score.) The vast majority of the sample–87%–took any STEM

[6] Of the 508 (ECoach) + 276 (Registrar) = 784 students who started the survey, 87% made it to the end, and 75% completed the IAT, our requirement for sample inclusion. We drop 62 IAT responses with high error rates, and 1 with overly slow response times, as suggested by Greenwald et al. (2003).

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.

4 Results

4.1 IAT Scores

Figure 1(a) 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 versa, and therefore shows an automatic association for male with STEM. Figure 1(a) illustrates that students in our sample on average hold male-STEM stereotypes. Previous research has established that this is true in the wider population as well, as we would expect (Nosek et al., 2009, 2007). We find no statistical differences by gender in the mean or distribution of IAT scores; 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). However, looking to the tails of the distribution in Figure 1(a), there are slightly more women with strong female-STEM associations and more men with strong male-STEM associations.

Although we do not have causal evidence on the formation of stereotypes, we can investigate what predicts them, providing suggestive evidence. We run a series of single-variate regressions as well as a multivariate regression of standardized IAT score on the characteristics available in our data–and the same set of variables we condition on in our primary analysis. (From this point forward, we standardize IAT scores to have mean zero and standard deviation one.) In other words, for a given characteristic (X) such as gender, we estimate:

$$IAT_i = \alpha_0 + \alpha_1 X_i + \upsilon_i \tag{1}$$

and additionally estimate a multivariate version with all characteristics:

$$IAT_i = \gamma_0 + \gamma_1 X_{1i} + \gamma_2 X_{2i} + ... + \gamma_K X_{Ki} + \upsilon_i \tag{2}$$

Table 2 shows the results of estimating Equation 1 (column 1) and Equation 2 (column

2). Although the unconditional gender difference in IAT is not significant, conditional on all observables, women have (marginally significantly) weaker implicit stereotypes by 0.175 standard deviations ($p < 0.10$) In both the single- and multi-variate specifications, being a first generation college student predicts a lower level of male-STEM stereotype (by around a quarter of a standard deviation). In other words, children of college-educated parents have more stereotypically gendered views. The mechanism for this is not obvious from our available data; it may be related to maternal education (and occupation) specifically, but our data do not distinguish between parents and do not include parental occupation. We construct a measure of gender occupational segregation in the labor markets students grew up in, but that does not predict IAT scores.[7]

We also find that explicit beliefs predict implicit beliefs: the more female a student states they believe STEM to be, the less they implicitly associate men with STEM. Stating a 10 percentage point higher belief about the proportion of women in STEM (e.g., thinking STEM is 50 percent women vs. 40 percent) is associated with an IAT score that is a tenth of a standard deviation lower. It is unsurprising that explicit and implicit beliefs are correlated; however, in our subsequent analyses we investigate whether implicit beliefs have predictive power over and above explicitly stated ones.

Several other characteristics we might expect to predict implicit gender-STEM associations do not: race, international status, family income, or high school academics. We hypothesized that exposure to female STEM role models might soften stereotypical views (Porter and Serra, 2020); however, our measure of having a female (or non-binary) role model–gender of the student's favorite math or science teacher in high school–does not predict IAT score. Our measure of female role models in *college*–the proportion of STEM instructors who are female–does strongly predict implicit stereotypes: students with all female STEM instructors have nearly 0.4 s.d. lower implicit stereotypes than those with none. We interpret this as a correlation only, since while it's plausible that previous exposure to women in STEM could affect stereotypes, it is just as plausible that students with stronger stereotypes select out of classes with female faculty.

Overall, we can statistically explain little of the variation in IAT: the R^2 from the multivariate regression is 0.099. Though an interesting question, how implicit stereotypes are produced is not our primary one. We leave further investigation to future work.

[7]Specifically, we use zip-code-by-occupation-by-gender level data from the ACS to calculate a Duncan index of dissimilarity for the home zip code of students in our sample, during their childhood years. We include more detail in Appendix E.

4.2 Implicit Stereotypes and Intended Major

The focus of this study is the relationship between implicit stereotypes and major choice. In Figure 1b, we plot the relationship between raw IAT scores, intended major, and gender as a binned scatterplot. 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):

$$Y_i = \beta_0 + \beta_1 Female_i + \beta_2 IAT_i + \beta_3 Female_i \cdot IAT_i + \boldsymbol{X_i}\boldsymbol{\gamma} + \boldsymbol{Z_i}\boldsymbol{\lambda} + \boldsymbol{W_i}\boldsymbol{\varphi} + \varepsilon_i \qquad (3)$$

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. 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 3); 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.[8]

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

[8]The 17 p.p. gender gap in intended major mirrors the university's historical gender gap in STEM degrees, as well as the national gender gap of 17 p.p. (National Center for Education Statistics, 2021).

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: percentile on the math section of the SAT or ACT, high school GPA, and an indicator for taking calculus in high school. (We also include indicators for missing test score and GPA.) 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 13.0 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 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., 2025), 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 (7.6 p.p.).

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.

We are interested in *implicit* stereotypes in particular, meaning beliefs students are unaware of. Thus, in column 6 we add controls for students' *explicit* beliefs 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 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 high school 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 (see Appendix E) based on student zip-code, measuring how gender segregated their childhood local labor market was.

We note here that including these role model measures may 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 if somewhat weaker with the inclusion of role model controls (columns 8 and 9).

4.3 Robustness and Heterogeneity

As further evidence against remaining omitted variable bias, we point out that as we progressively add controls to Table 3, the R^2 increases substantially, from 0.065 to 0.291. Oster (2019) shows that coefficient stability itself does not rule out bias if the observables explain little of the outcome's variance. The stability of our coefficients of interest even as we add controls with significant explanatory power reassures us that remaining OVB is probably minimal. More formally, according to the procedure in Oster (2019), unobservables would need to be more than five times as important as observables for men, and more than four times as important for women in order for the true "effect" of IAT to be zero.[9]

The stability of our estimated coefficients to controlling for observables also provides some reassurance that the IAT-major relationship holds for various types of students. In Appendix Tables A1 and A2, we test for heterogeneity in the IAT-gender-intended major relationship by race/ethnicity (underrepresented minority status), parent education, family income, country of residence, and SAT/ACT math performance. The point estimates suggest that the negative IAT-major correlation may be stronger for underrepresented minority women and first-gen women, and for domestic relative to international women. However, the estimates are noisy and we cannot reject equality by subgroup.[10] In terms of external validity, we cannot rule out that our sample is unobservably different from

[9] We estimate Oster's δ with psacalc in Stata, using separate regressions for men and women; the coefficients of interest are very similar to those produced by the interacted model. We follow the benchmark suggested by Oster and set R_{max} at $1.3\tilde{R}^2$, where $\tilde{R}^2$is the R^2 from the regression with all controls.

[10] In Appendix Table A3 we estimate effects by recruitment sample (Registrar vs. ECoach). Results suggest it is the Registrar sample driving results. This could be due to differences in the sample (ECoach students are by construction much more likely to be in a STEM course and have higher baseline rates of STEM interest) or to slight differences in recruitment language and method.

college students as a whole. This is an inherent limitation of studies requiring participant opt-in (e.g., Wiswall and Zafar, 2015).

To further argue that that the IAT reflects *implicit* beliefs and not explicit ones, we run a robustness check where we estimate Equation 3 but replace IAT score with a measure of *explicit* gender-STEM beliefs, operationalized as the proportion of STEM graduates that a student believes are women. The results, in Appendix Table A4, show no relationship between this measure and intended major, for either men or women.

Another potential concern is that the observed relationship between implicit stereotypes and intended major choice may reflect causality in the reverse direction, whereby students' academic interests and experiences shape their stereotypes rather than the other way around. During college, experiences such as taking STEM courses and interacting with female peers or faculty (or a lack thereof) could change stereotypes. Although we cannot fully rule this out, we re-estimate our main results with first-year, first-semester students only, who have had limited college experiences that could move their implicit stereotypes. We do the same exercise with students who have not yet declared a major, and whose interest is thus less fixed. The results for both groups (Appendix Table A5) are nearly identical to those from the full sample.

It is also possible that once a student has identified an interest in STEM (or not), they update their implicit stereotypes to be more in line with their decision (e.g., a woman who decides she likes STEM starts to implicitly associate it more with her own gender). The nature and timing of our data do not allow us to test this; we highlight it as an important avenue for future research. We believe the weight of the existing theory and evidence in psychology and economics suggests a causal role for implicit stereotypes affecting behavior, but emphasize that our results are consistent with this hypothesis rather than definitive.

4.4 Implicit Stereotypes and Administrative Outcomes

We next 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 4 shows the results of estimating Equation 3 for six additional outcomes. From this point forward, we include the full set of control variables, i.e., those in column 9 of Table 3. 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 4 tell the same story as Table 3: 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 around 6 percentage points both semesters. Women with a one s.d. higher IAT score are 5.5 p.p. less likely to be declared as STEM in Fall 2023, and 6.1 p.p. less likely to be declared in the spring.[11,12]

In Table 5, we investigate the predictive power of implicit stereotypes for STEM performance, measured as GPA in Fall 2023 (column 1) and Spring 2024 (column 2). Only students who took at least one STEM course in the given semester are included. Column 1 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, possibly due to anticipated discrimination (Lepage et al., 2025). However, 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 4 showed that women with the strongest male-STEM stereotypes are more likely to not continue in STEM. But in Table 5, column 2, 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 may 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),

[11] In Table A6 of the 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 A6 are very similar to those in Tables 3 and 4.

[12] We test for robustness to different functional forms of our controls in Appendix Table A7. We further note that we control for every survey item in the full model, so are not selectively omitting covariates.

who find that receiving multiple signals that they don't belong, in the form of grades and gender composition, can decrease female STEM persistence.

5 Discussion and Conclusion

To summarize, we find a remarkably robust relationship between implicit stereotypes, gender, and STEM outcomes–both stated and observed. 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, are 8-10 p.p. more likely to take any STEM courses, and take more STEM credits. Women, on the other hand, are 8 to 10 p.p. *less* likely to be interested in STEM when they hold stronger implicit stereotypes. They are 4 p.p. less likely to take any STEM, take fewer STEM credits, and are 5-6 p.p. less likely to declare a STEM major. These results hold conditional on measures of demographic background, academic preparation, major- and job-related preferences, explicitly stated beliefs about STEM, and the presence of female role models. 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 put the magnitude of our findings in perspective, they are comparable to other well-known determinants of STEM pursuit. For example, Carrell et al. (2010) show that increasing the share of female introductory STEM instructors by one standard deviation raises the probability that women take higher level math by about 2 percentage points.[13] Porter and Serra (2020) find that exposure to role models increases subsequent enrollment in economics courses by about 8 percentage points for women. Conlon (2021) finds that providing students with salary information increases the likelihood of majoring in the corresponding field by 9 p.p. (he does not report differences by gender).

Although we have framed our paper around the lack of women pursuing STEM, our findings may well apply beyond STEM, speaking to a growing concern regarding the lack of men in fields like education and nursing (Reeves, 2022). Our results suggest that men are also influenced by implicit stereotypes. Another way to state our finding for men is that the less men associate male with STEM, the more likely they are to pursue

[13]This is based on the estimated effect on female students of increasing the proportion of female faculty by 1 (4.2 p.p., as shown in Table V) times the standard deviation of the independent variable, 0.27 (Table II).

non-STEM, less male-dominated fields. A version of the IAT that measured implicit stereotypes for men and women in STEM versus care fields (rather than humanities, as in our version) might reveal more.

The magnitudes of the associations we found between implicit stereotypes and major choice highlight their importance for future research and policy aimed at closing gender gaps. 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), even if the underlying implicit stereotypes don't change.

Changing implicit stereotypes directly is likely more difficult. 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. This is not impossible: one recent paper found that men who were born just after paternity leave policies were enacted grew up to have less stereotypical attitudes measured by an IAT (Fontenay and González Luna, 2024). A better understanding of the early production of implicit stereotypes is an important avenue for future research. In the shorter term, 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).

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.

Aucejo, E. and James, J. (2021). The path to college education: The role of math and verbal skills. *Journal of Political Economy*, 129(10):2905–2946.

Aycock, L. M., Hazari, Z., Brewe, E., Clancy, K. B., Hodapp, T., and Goertzen, R. M. (2019). Sexual harassment reported by undergraduate female physicists. *Physical Review Physics Education Research*, 15(1):010121.

Azmat, G., Calsamiglia, C., and Iriberri, N. (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6):1372–1400.

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.

Bertrand, M., Chugh, D., and Mullainathan, S. (2005). Implicit discrimination. *American Economic Review*, 95(2):94–98.

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.

Breda, T. and Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences*, 116(31):15435–15440.

Buser, T., Niederle, M., and Oosterbeek, H. (2014). Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447.

Buser, T., Peter, N., and Wolter, S. C. (2017). Gender, competitiveness, and study choices in high school: Evidence from switzerland. *American economic review*, 107(5):125–130.

Calkins, A., Binder, A. J., Shaat, D., and Timpe, B. (2023). When Sarah meets Lawrence: The effects of coeducation on women's college major choices. *American Economic Journal: Applied Economics*, 15(3):1–34.

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.

Cundiff, J. L., Vescio, T. K., Loken, E., and Lo, L. (2013). Do gender–science stereotypes predict science identification and science career aspirations among undergraduate science majors? *Social Psychology of Education*, 16(4):541–554.

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

De Gioannis, E. (2022). Implicit gender-science stereotypes and college-major intentions of italian adolescents. *Social Psychology of Education*, 25(5):1093–1112.

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

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.

Fontenay, S. and González Luna, L. (2024). Can public policies break the gender mold? Evidence from paternity leave reforms in six countries. Working paper.

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.

Goulas, S., Griselda, S., and Megalokonomou, R. (2024). Comparative advantage and gender gap in STEM. *Journal of Human Resources*, 59(6):1937–1980.

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

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.

Griselda, S. (2024). Gender gap in standardized tests: What are we measuring? *Journal of Economic Behavior & Organization*, 221:191–229.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and US economic growth. *Econometrica*, 87(5):1439–1474.

Iriberri, N. and Rey-Biel, P. (2021). Brave boys and play-it-safe girls: Gender differences in willingness to guess in a large scale natural field experiment. *European Economic Review*, 131:103603.

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.

Kuhn, A. and Wolter, S. C. (2022). Things versus people: Gender differences in vocational interests and in occupational preferences. *Journal of Economic Behavior & Organization*, 203:210–234.

Lane, K. A., Banaji, M. R., Nosek, B. A., and Greenwald, A. G. (2007). Understanding and using the implicit association test: IV: What we know (so far) about the method. In *Implicit Measures of Attitudes*, pages 59–102. The Guilford Press.

Lane, K. A., Goh, J. X., and Driver-Linn, E. (2012). Implicit science stereotypes mediate the relationship between gender and academic participation. *Sex Roles*, 66(3):220–234.

Lepage, L.-P., Li, X., and Zafar, B. (2025). Anticipated discrimination and major choice. National Bureau of Economic Research Working Paper no. 33680.

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. (2025). The long-term effects of teachers' gender stereotypes on labor outcomes. IZA Discussion Paper No. 17674.

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.

Minnotte, K. L. and Pedersen, D. E. (2023). Sexual harassment, sexual harassment climate, and the well-being of STEM faculty members. *Innovative Higher Education*, 48(4):601–618.

National Center for Education Statistics (2021). Number of persons 25 to 34 years old and percentage with a bachelor's or higher degree, by undergraduate field of study, sex, race/ethnicity, and U.S. nativity and citizenship status: 2021 (table 104.6). *Digest of Education Statistics*.

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., Banaji, M. R., and Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology*, 83(1):44.

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. (2010). Grades, gender, and encouragement: A regression discontinuity analysis. *The Journal of Economic Education*, 41(3):217–234.

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.

Owen, S. and Rury, D. (2025). Data for: "Implicit gender-STEM stereotypes and college major choice". Last accessed September 2025.

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.

Reeves, R. V. (2022). *Of boys and men: Why the modern male is struggling, why it matters, and what to do about it.* Brookings Institution Press.

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.

Sloane, C. M., Hurst, E. G., and Black, D. A. (2021). College majors, occupations, and the gender wage gap. *Journal of Economic Perspectives*, 35(4):223–248.

Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex Roles*, 67(11):617–629.

Smyth, F. L., Greenwald, A. G., and Nosek, B. A. (2009). Implicit gender–science stereotype outperforms math scholastic aptitude in identifying science majors. Unpublished manuscript.

Speer, J. D. (2023). Bye bye Ms. American sci: Women and the leaky STEM pipeline. *Economics of Education Review*, 93:102371.

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.

Steele, C. M. and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5):797.

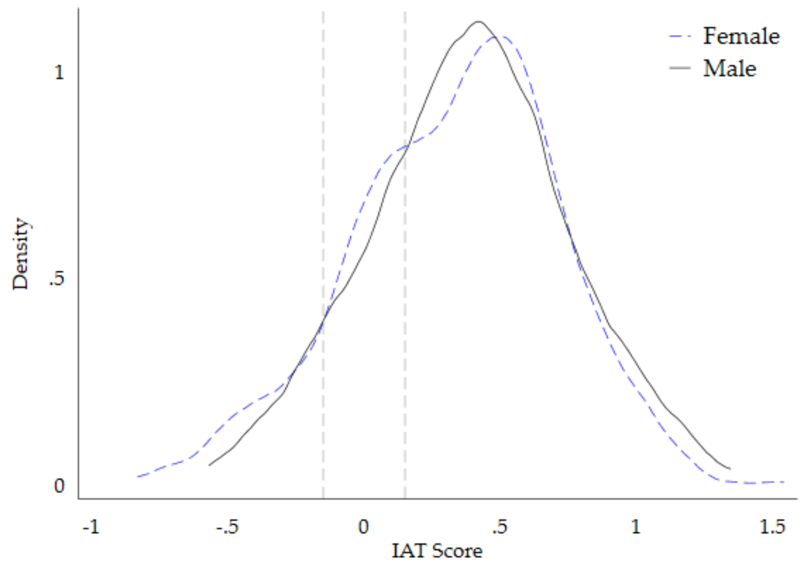
Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

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.

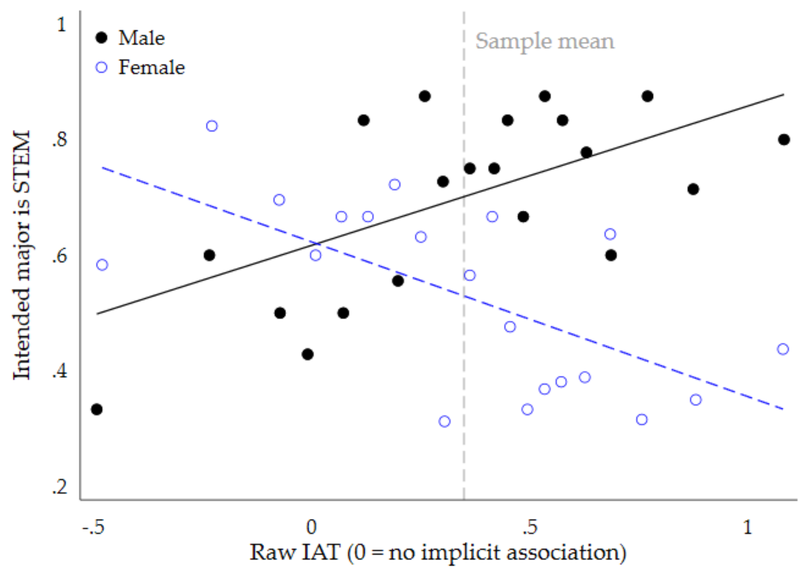
Young, D. M., Rudman, L. A., Buettner, H. M., and McLean, M. C. (2013). The influence of female role models on women's implicit science cognitions. *Psychology of Women Quarterly*, 37(3):283–292.

Exhibits

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 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). Positive scores indicate an automatic association for male with science and female with humanities (stereotypical association) while negative scores indicate the non-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 survey item and completed the IAT (N=539; 394 women and 145 men; one student submitted a survey and completed the IAT but did not answer the intended major item). Figure is a binned scatterplot using 20 equally sized bins for each gender.

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
SAT/ACT math percentile	93.49	93.30	93.11
Missing test score	0.261	0.311	0.263
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. SAT/ACT math percentile is in units of national percentiles, and is based on the max score submitted by the student. If students have both an SAT and ACT score, we take the average. 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.

Table 2: Single- and Multi-variate Associations of Student Characteristics with IAT Score

	Dependent variable: IAT score (standardized)	
	(1) Single-variate	(2) Multi-variate
Female	-0.127	-0.175+
	(0.095)	(0.103)
First year	-0.103	-0.103
	(0.095)	(0.104)
Underrepresented minority	0.035	-0.024
	(0.115)	(0.135)
International student	-0.200	-0.133
	(0.186)	(0.257)
First-generation college student	-0.238*	-0.259*
	(0.112)	(0.120)
Family income ≥ \$100K	0.010	-0.060
	(0.102)	(0.110)
SAT/ACT math percentile	-0.003	-0.007
	(0.005)	(0.006)
High school GPA	-0.568	-0.686
	(0.406)	(0.498)
Took calculus in HS	-0.030	0.028
	(0.089)	(0.098)
Important for major: good at subject	0.094	0.217+
	(0.101)	(0.112)
Important for major: engaged	-0.092	-0.170
	(0.106)	(0.118)
Important for major: having friends	-0.120	-0.121
	(0.106)	(0.112)
Important for major: salary	-0.144	-0.107
	(0.092)	(0.099)
Important for major: work flexibility	-0.111	-0.095
	(0.087)	(0.096)
Important for major: impact on society	-0.018	0.006
	(0.087)	(0.093)
Important for major: peers/culture	-0.008	-0.001
	(0.090)	(0.103)
Expected salary in top major (\$1000s)	-0.001+	-0.001
	(0.001)	(0.001)
Expected salary in 2nd major (\$1000s)	-0.001	-0.000
	(0.001)	(0.001)

Continued on next page

Table 2 – *Continued from previous page*

	Dependent variable: IAT score (standardized)	
	(1) Single-variate	(2) Multi-variate
Belief: % women in STEM (0-100)	-0.011*	-0.008+
	(0.004)	(0.004)
Belief: % women in humanities	0.006	0.009*
	(0.004)	(0.004)
Belief: HS GPA of STEM majors	0.029	0.085
	(0.148)	(0.178)
Belief: HS GPA of humanities majors	-0.109	-0.250
	(0.130)	(0.155)
Non-male STEM role model in HS	0.034	-0.016
	(0.088)	(0.093)
Non-male humanities role model in HS	-0.105	-0.057
	(0.087)	(0.090)
Prop. female STEM instructors, SU22-FA23	-0.356*	-0.396*
	(0.157)	(0.176)
Labor market gender segregation index	0.301	-0.468
	(14.021)	(14.931)
R^2		0.099
N		540

Notes: $+p < 0.1$, $*p < 0.05$, $**p < 0.01$, $***p < .001$. Robust standard errors in parentheses. Each coefficient in column (1) is from a separate regression of standardized IAT score on the characteristic. All coefficients in column (2) are from the same multivariate regression. More positive IAT scores indicate stronger male-STEM implicit association. Sample size for single-variate regressions varies based on level of missingness. Multivariate regressions include missing dummies for all variables and code missings as zeroes.

Table 3: Relationship between IAT Score, Gender, and Intended Major

	Dependent variable: student intends STEM major								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female (β_1)	-0.172***	-0.173***	-0.130**	-0.141**	-0.130**	-0.117*	-0.114*	-0.115*	-0.115*
	(0.045)	(0.046)	(0.045)	(0.046)	(0.046)	(0.049)	(0.049)	(0.049)	(0.049)
IAT score, std. (β_2)	0.090*	0.081*	0.089*	0.076*	0.080*	0.084*	0.082*	0.083*	0.083*
	(0.040)	(0.040)	(0.036)	(0.037)	(0.038)	(0.038)	(0.038)	(0.038)	(0.039)
Female x IAT (β_3)	-0.191***	-0.185***	-0.187***	-0.168***	-0.168***	-0.173***	-0.172***	-0.167***	-0.167***
	(0.047)	(0.047)	(0.043)	(0.044)	(0.045)	(0.045)	(0.045)	(0.045)	(0.045)
IAT effect for women	-0.100***	-0.104***	-0.098***	-0.092***	-0.089***	-0.089***	-0.090***	-0.084***	-0.084***
($\beta_2 + \beta_3$)	(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	0.179	0.236	0.254	0.267	0.276	0.289	0.291
N	539	539	539	539	539	539	539	539	539
Male outcome mean	0.710								
Female outcome mean	0.533								

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 implicit association. 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. Sample size is one fewer than the full sample in Table 1 due to one student who filled out the survey and completed an IAT but did not answer the intended major survey item.

Table 4: Relationship between IAT Score, Gender, and Observed 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 (β_1)	-0.011	-1.367**	-0.097*	0.039	-1.330*	-0.078
	(0.027)	(0.480)	(0.047)	(0.040)	(0.533)	(0.048)
IAT score, std. (β_2)	0.079***	1.309***	0.059	0.104***	1.735***	0.058
	(0.021)	(0.384)	(0.041)	(0.030)	(0.407)	(0.040)
Female x IAT (β_3)	-0.117***	-2.241***	-0.114*	-0.141***	-2.660***	-0.118*
	(0.024)	(0.432)	(0.047)	(0.036)	(0.460)	(0.046)
IAT effect for women	-0.038**	-0.932***	-0.055*	-0.037+	-0.925***	-0.061**
($\beta_2 + \beta_3$)	(0.013)	(0.197)	(0.023)	(0.019)	(0.216)	(0.022)
R^2	0.465	0.320	0.172	0.192	0.259	0.210
N	540	540	540	540	540	540
Male outcome mean	0.904	8.719	0.349	0.829	8.466	0.390
Female outcome mean	0.860	6.721	0.188	0.802	6.515	0.236

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 implicit association. 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 5: Relationship between IAT Score, Gender, and STEM Performance

	(1) STEM GPA FA23	(2) STEM GPA SP24
Female (β_1)	0.022	0.041
	(0.065)	(0.059)
IAT score, std. (β_2)	-0.043	0.069
	(0.049)	(0.045)
Female x IAT (β_3)	0.089	0.030
	(0.058)	(0.053)
IAT effect for women	0.046	0.099***
($\beta_2 + \beta_3$)	(0.030)	(0.030)
R^2	0.258	0.258
N	466	427
Male outcome mean	3.511	3.521
Female outcome mean	3.498	3.489

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 implicit association. 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. 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, Gender, and Intended Major: Heterogeneity by Race and Socioeconomic Status

	Dependent variable: student intends STEM major					
	Race/ethnicity		Max parent ed		Family income	
	Non-URM	URM	BA	1st gen	$\geq$100K	<100K
Female (β_1)	-0.154**	-0.279*	-0.178***	-0.040	-0.162**	-0.133
	(0.052)	(0.107)	(0.051)	(0.118)	(0.063)	(0.090)
IAT score, std. (β_2)	0.101*	0.091	0.110**	-0.013	0.110*	0.054
	(0.046)	(0.079)	(0.042)	(0.108)	(0.044)	(0.110)
Female x IAT (β_3)	-0.195***	-0.215*	-0.193***	-0.142	-0.221***	-0.168
	(0.054)	(0.093)	(0.051)	(0.119)	(0.057)	(0.118)
IAT effect for women	-0.094**	-0.123*	-0.082**	-0.155**	-0.111**	-0.115**
($\beta_2 + \beta_3$)	(0.028)	(0.048)	(0.028)	(0.050)	(0.036)	(0.042)
R^2	0.059	0.105	0.065	0.073	0.084	0.058
N	417	99	430	98	260	157
Male outcome mean	0.692	0.750	0.730	0.560	0.690	0.706
Female outcome mean	0.529	0.481	0.533	0.548	0.514	0.569

Notes: $+p < 0.1$, $*p < 0.05$, **p< 0.01, $***p < .001$. Robust standard errors in parentheses. Results in each column are from a regression of intended STEM major on gender, standardized IAT score, and their interaction, estimated separately by subgroup. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. Each subgroup analysis is limited to students who are not missing a recorded characteristic (i.e., known race, known parent education, etc.).

Table A2: Relationship between IAT Score, Gender, and Intended Major: Heterogeneity by Residency and Quantitative Background

	Dependent variable: student intends STEM major			
			SAT/ACT math	
	Country of residence		Above sample	Below sample
	International	Domestic	median	median
Female (β_1)	-0.138	-0.164***	-0.125+	-0.083
	(0.200)	(0.047)	(0.064)	(0.087)
IAT score, std. (β_2)	-0.030	0.108*	0.040	0.092
	(0.101)	(0.044)	(0.055)	(0.072)
Female x IAT (β_3)	0.044	-0.209***	-0.145*	-0.189*
	(0.231)	(0.050)	(0.068)	(0.081)
IAT effect for women	0.014	-0.102***	-0.105*	-0.098**
($\beta_2 + \beta_3$)	(0.208)	(0.024)	(0.041)	(0.037)
R^2	0.028	0.067	0.053	0.046
N	28	511	193	206
Male outcome mean	0.786	0.702	0.818	0.590
Female outcome mean	0.643	0.529	0.716	0.503

Notes: $+p < 0.1$, $*p < 0.05$, **p< 0.01, $***p < .001$. Robust standard errors in parentheses. Results in each column are from a regression of intended STEM major on gender, standardized IAT score, and their interaction, estimated separately by subgroup. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. SAT/ACT subgroup analysis is limited to students with a non-missing test score. SAT/ACT quantitative score is in units of national percentiles. If a student has submitted both scores, we use the average. The above/below median split is within the sample; the sample median corresponds to the 97th percentile nationally.

Table A3: Relationship between IAT Score, Gender, and Intended Major: Heterogeneity by Recruitment Sample

	Dependent variable: intends STEM		
	Registrar sample	ECoach sample (all)	ECoach sample (first years)
Female (β_1)	-0.102	-0.095	-0.132+
	(0.088)	(0.059)	(0.078)
IAT score, std. (β_2)	0.163**	0.040	-0.014
	(0.055)	(0.042)	(0.049)
Female x IAT (β_3)	-0.278***	-0.087+	-0.012
	(0.069)	(0.050)	(0.062)
IAT effect for women	-0.115**	-0.047+	-0.026
($\beta_2 + \beta_3$)	(0.040)	(0.028)	(0.039)
R^2	0.454	0.277	0.359
N	171	368	215
Male outcome mean	0.558	0.775	0.783
Female outcome mean	0.289	0.650	0.671

Notes: $+p < 0.1$, $*p < 0.05$, **p< 0.01, $***p < .001$. Robust standard errors in parentheses. Results in each column are from a regression of intended STEM major on gender, standardized IAT score, and their interaction, estimated separately by subgroup. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. By design, the Registrar sample is first year students only. 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 A4: Relationship between Explicit Gender-STEM Beliefs, Gender, and Intended Major

	Intends STEM Major
Female	-0.179***
	(0.046)
Explicit belief: Percent STEM degrees going to women (std.)	0.002
	(0.039)
Female x Explicit belief	0.028
	(0.047)
Explicit belief effect for women (main effect + interaction)	0.030
	(0.026)
R^2	0.028
N	537

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, 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. For this analysis, explicit belief is standardized to have mean 0 and s.d. 1, for comparison with Table 3. Analysis is limited to students who answered survey questions about intended major and explicit beliefs. Sample size differs from Table 3 due to two students who have an IAT score and stated intended major but did not answer the item about the proportion of women in STEM.

Table A5: Relationship between IAT Score, Gender, and Intended Major, Limiting to Undeclared and First Year Students

	Dependent variable: intends STEM	
	Undeclared students only	First year students only
Female (β_1)	-0.144*	-0.168**
	(0.066)	(0.058)
IAT score, std. (β_2)	0.052	0.115*
	(0.050)	(0.047)
Female x IAT (β_3)	-0.142*	-0.207***
	(0.059)	(0.055)
IAT effect for women ($\beta_2 + \beta_3$)	-0.090**	-0.092**
	(0.030)	(0.028)
R^2	0.366	0.288
N	324	385
Male outcome mean	0.676	0.686
Female outcome mean	0.500	0.498

Notes: $+p < 0.1$, $*p < 0.05$, **p< 0.01, $***p < .001$. Robust standard errors in parentheses. Results in each column are from a regression of intended STEM major on gender, standardized IAT score, and their interaction, estimated separately by subgroup. Intended STEM major is based on a survey item asking students the subject they're most likely to major in. 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 A6: Relationship between IAT Score, Gender, 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 or econ FA23	(5) Any STEM or econ courses SP24	(6) # STEM and econ credits SP24	(7) Declared STEM or econ SP24
Female (β_1)	-0.134**	-0.009	-1.526**	-0.092*	-0.001	-1.415**	-0.081+
	(0.045)	(0.027)	(0.471)	(0.047)	(0.034)	(0.501)	(0.048)
IAT score, std. (β_2)	0.099**	0.081***	1.263***	0.053	0.070*	1.503***	0.047
	(0.034)	(0.021)	(0.374)	(0.040)	(0.028)	(0.387)	(0.040)
Female x IAT (β_3)	-0.157***	-0.122***	-2.113***	-0.104*	-0.093**	-2.227***	-0.100*
	(0.042)	(0.024)	(0.420)	(0.047)	(0.033)	(0.439)	(0.047)
IAT effect for women	-0.058*	-0.041**	-0.850***	-0.052*	-0.023	-0.724***	-0.053*
($\beta_2 + \beta_3$)	(0.023)	(0.013)	(0.197)	(0.023)	(0.017)	(0.206)	(0.023)
R^2	0.306	0.450	0.335	0.172	0.202	0.276	0.214
N	539	540	540	540	540	540	540
Male outcome mean	0.766	0.911	9.226	0.349	0.890	9.185	0.397
Female outcome mean	0.571	0.865	7.025	0.198	0.853	7.157	0.251

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 implicit association. Intended STEM/econ major is based on a survey item asking students the subject they're most likely to major in. 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 A7: Relationship between IAT Score, Gender, and STEM Outcomes, with More Flexible Controls

	(1) Intend STEM major	(2) Any STEM courses FA23	(3) # STEM credits FA23	(4) Declared as STEM FA23	(5) Any STEM courses SP24	(6) # STEM credits SP24	(7) Declared as STEM SP24
Female (β_1)	-0.096+	-0.001	-1.092*	-0.086+	0.049	-1.047+	-0.075
	(0.050)	(0.027)	(0.489)	(0.047)	(0.042)	(0.546)	(0.050)
IAT score, std. (β_2)	0.076+	0.089***	1.241**	0.056	0.094**	1.596***	0.051
	(0.042)	(0.022)	(0.396)	(0.042)	(0.033)	(0.431)	(0.042)
Female x IAT (β_3)	-0.163***	-0.125***	-2.185***	-0.118*	-0.130***	-2.542***	-0.118*
	(0.048)	(0.025)	(0.445)	(0.048)	(0.039)	(0.486)	(0.048)
IAT effect for women	-0.087***	-0.036**	-0.944***	-0.062**	-0.036+	-0.946***	-0.067**
($\beta_2 + \beta_3$)	(0.024)	(0.013)	(0.204)	(0.023)	(0.020)	(0.223)	(0.023)
R^2	0.341	0.511	0.372	0.272	0.263	0.325	0.281
N	539	540	540	540	540	540	540
Male outcome mean	0.710	0.904	8.719	0.349	0.829	8.466	0.390
Female outcome mean	0.533	0.860	6.721	0.188	0.802	6.515	0.236

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, standardized IAT score, and their interaction. These regressions include the full set of controls from the main analysis, but with the following differences. For race, we include non-mutually exclusive dummies for white, Asian, Black, Hispanic, Native, multi-race, and unknown/missing (rather than a URM indicator). For parent education, we include 9 max parent education indicators, from elementary school to doctorate (rather than a first-gen indicator). For family income, we include 7 indicators, from <25K to >200K (rather than an above 100K indicator). For the 7 Likert scale questions about importance of factors for choosing major, we include indicators for each possible response on the 5-point scale (rather than dichotomizing as unimportant vs. important).

Appendix B. Recruitment Text

Recruitment message: Registrar sample

Dear [*First Name*],

You are invited to participate in a research study about how undergraduate students make academic decisions. Your participation will help inform strategies to support students as they choose their academic paths.

To learn more and participate in the study, please click here: [*Survey Link*]

If you participate by December 4, you have a chance of receiving a $50 Amazon gift card. We will randomly draw 40 participants who complete the study within this time, each of whom will receive a $50 Amazon gift card.

Completing the study should take 10-15 minutes. It can be completed on your phone, laptop, or tablet. Your participation is voluntary, and your responses will be kept confidential.

Sincerely,

Stephanie Owen, Principal Investigator (University of the Midwest and Colby College)

[Name redacted]*, Principal Investigator (University of the Midwest)

Derek Rury, Principal Investigator (University of Chicago)

*We have redacted the name and affiliation of our institutional partner in order to maintain anonymity of the institution.

Recruitment message: ECoach sample

Dear [*First Name*],

The ECoach team and [*ECoach Course*] are helping one of our research colleagues recruit people for a study. For the chance to win one of forty $50 Amazon gift cards, please help our colleagues understand how students make academic decisions by December 4.

Here's what the researchers want you to know:

- **Why we're doing this**: We want to learn how implicit stereotypes held by undergraduates impact their academic decisions. Your participation will help inform strategies to support students as they choose their academic paths.
- **What you could get**: If you participate by December 4, you have a chance of receiving one of forty $50 Amazon gift cards. We'll randomly draw 40 names from those who complete in time.
- **Survey details**:
  - The study should take 10 to 15 minutes.
  - It can be completed on your phone, laptop, or tablet.
  - Your participation is voluntary, and your responses will be kept confidential.
  - Even if you get more than one invitation (from another ECoach course), you'll only fill out the survey once.
- **The researchers**:
  - Stephanie Owen, Principal Investigator (University of the Midwest and Colby College)
  - *[Name redacted]*, Principal Investigator (University of the Midwest)
  - Derek Rury, Principal Investigator (University of Chicago)

Yes, I'll help [*Hyperlink to study*]

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	Slightly important	Moderately important	Very important	Extremely important
1	2	3	4	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)

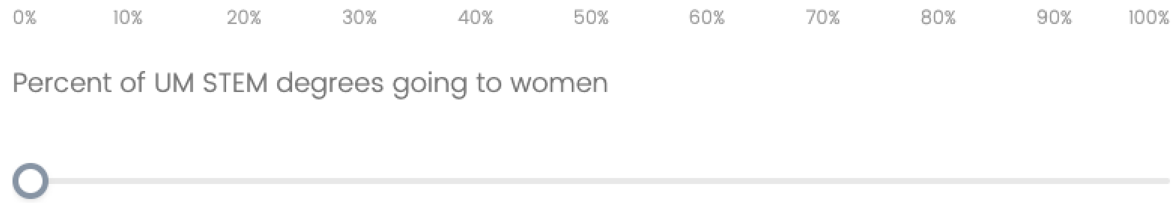
Having a positive impact on society (after graduation)

Work culture/peers (after graduation)

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

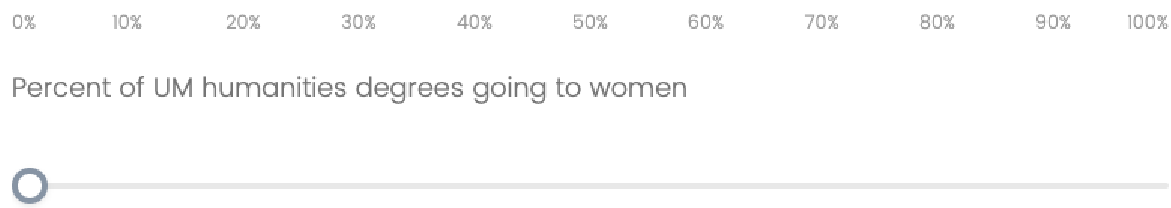
[*pct_stem_women*]

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



[*stem_hs_gpa*]

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

E/F	D	C	B	A
0	1	2	3	4

STEM majors' high school GPA

[*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?

E/F	D	C	B	A
0	1	2	3	4

Humanities majors' high school GPA

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. 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.

References: Appendix D

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

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.

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 implicit association. 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).

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:

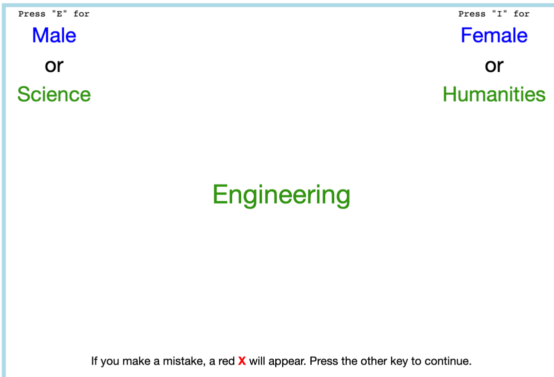
Category	Items
Male	Man, Son, Father, Boy, Uncle, Grandpa, Husband, Male
Female	Mother, Wife, Aunt, Woman, Girl, Female, Grandma, Daughter
Science	Astronomy, Math, Chemistry, Physics, Biology, Geology, Engineering
Humanities	History, Arts, Humanities, English, Philosophy, Music, Literature

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

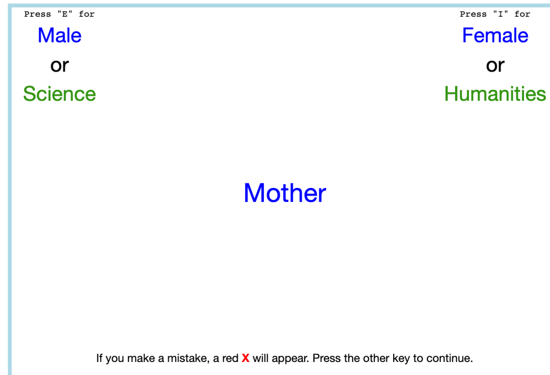
Continue

Notes: This is a screenshot from the keyboard device version of the IAT. The touchscreen version specifies the participant will touch the left or right side of the screen.

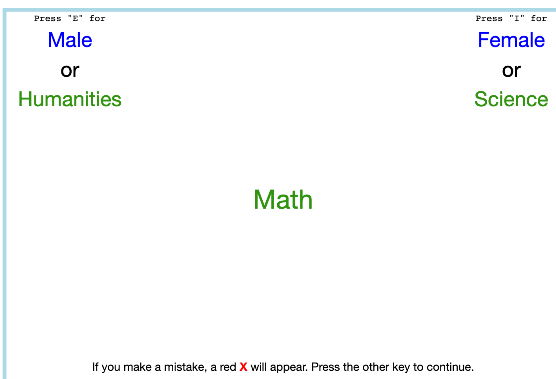
Figure B2: Sample IAT Tasks



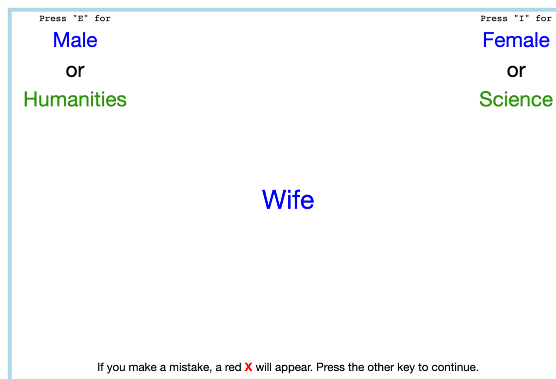
(a) Male-science/female-humanities (stereotypical)



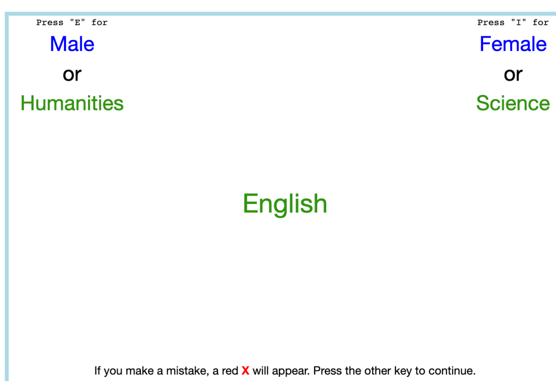
(b) Male-science/female-humanities (stereotypical)



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



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



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

Appendix E. Duncan Index

To measure local occupational gender segregation, we 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 (Cortes and Pan, 2019; Baker and Cornelson, 2018). 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 data from the five-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. The Duncan Index is calculated as:

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

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 ranged from 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.

References: Appendix E

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

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