Childhood confidence, schooling, and the labor market: Evidence from the PSID

Lucy Page and Hannah Ruebeck*

MIT

We link over- and under-confidence in math at ages 8-11 to education and employment outcomes 22 years later among the children of PSID households. About twenty percent of children have markedly biased beliefs about their math ability, and beliefs are strongly gendered. Conditional on measured ability, childhood over- and under-confidence predict adolescent test scores, high school and college graduation, majoring or working in STEM, earnings, and unemployment. Across all metrics, higher confidence predicts better outcomes. These biased beliefs persist into adulthood and could continue to affect outcomes as respondents age, since intermediate outcomes do not fully explain these long-run correlations.

* Lucy Page is a PhD candidate in economics at MIT. Hannah Ruebeck is a PhD candidate in economics at MIT (hruebeck@mit.edu). We are grateful to David Autor, Esther Duflo, Amy Finkelstein, Kartini Shastry, Rohini Pande, Frank Schilbach, and four anonymous referees for thoughtful and helpful comments. An online appendix is available with this paper on the JHR website. Most data used in this paper is publicly available on the website of the Panel Study of Income Dynamics (https://psidonline.isr.umich.edu/). Along with replication codes, the harmonized dataset used in this paper is available via the Harvard Dataverse: https://doi.org/10.7910/DVN/WBK5WA. Several control and outcome variables used in the analysis rely on the restricted PSID dataset, which can be obtained via the process described at https://simba.isr.umich.edu/restricted/ProcessReq.aspx. Both authors are supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. This project is also supported by the George and Obie Shultz Fund at MIT. The authors declare that they have no relevant or material financial interests that relate to the research described in this article.

JEL codes: D91, I20, J24
I. Introduction

Long-standing research in psychology finds that people have biased beliefs about their abilities in a range of domains.¹ Prior research has focused on “optimism bias,” or over-confidence about one’s performance, belief accuracy, or future outcomes (Moore and Healy 2008; Sharot, Korn, and Dolan 2011; Taylor and Brown 1988). In contrast, psychologists also document “imposter syndrome,” a form of systematic under-confidence in which people attribute their successes to luck or effort rather than skill (Langford and Clance 1993; Sakulku 2011). Recent lab-based work in behavioral economics has sought to microfound this empirical evidence of biased beliefs by documenting that people systematically under-weight or over-weight signals about the truth, especially in ego-relevant domains like intelligence and beauty (see Benjamin (2019) for a review).

Do these confidence gaps matter for economic decision-making in the real world? There are key reasons to expect that they might. For example, if adolescents or young adults perceive ability and educational investment to be complements, under-confident students might exert less effort in school or end their education earlier (Bénabou and Tirole 2002). Later, under-confident adults may be less likely to complete costly and uncertain job applications or may select away from jobs with higher returns to performance (Dohmen and Falk 2011). Individuals’ beliefs about their own ability could also affect outcomes by shaping how others perceive them. If parents or teachers mistake confidence for aptitude and expect the returns of education to increase with ability, they may invest more in more confident children (Papageorge, Gershenson, and Kang 2018; Dizon-Ross 2019). More confident applicants may appear more capable during job interviews, improving their employment prospects (Mobius and Rosenblat 2006; Schwardmann and van der Weele 2019).
As yet, there is limited evidence for how confidence affects economic outcomes in realistic settings and over the long term. In addition to the lab-based work on the short-term implications of confidence gaps cited above (Mobius and Rosenblat 2006; Dohmen and Falk 2011; Schwardmann and van der Weele 2019), a small parallel literature in economics and sociology examines longer-term outcomes and finds that those with higher self-esteem get more education, are more likely to be employed, and earn higher wages (Murnane et al. 2001; Waddell 2006; Drago 2011; de Araujo and Lagos 2013). However, this literature has struggled to demonstrate that these associations are not driven by omitted variables like unobserved ability. These papers typically control for IQ in an attempt to account for cognitive ability, but it is not feasible to control for subjects’ “ability” across all domains that affect generalized self-esteem.

In this paper, we address the limitations of both prior literatures by examining the real-world and long-term implications of a dimension of confidence in which we can observe and control for demonstrated ability: childhood over- and under-confidence in math.² We use unique data from the Panel Study of Income Dynamics (PSID) to identify biased beliefs in math in a sample of 2,985 children in core PSID households; we then relate their childhood over- and under-confidence to educational and employment outcomes up to 22 years later, controlling for test scores, general confidence, and other key confounders.

The PSID is an ideal setting in which to examine long-term links with childhood confidence. Our sample is based on child-focused PSID supplements that measure children’s performance on a standardized math test and their own reports of how “good” they are at math. We combine these measures to identify over-confident children as those who scored poorly on the math assessment and yet said they were good at math, and to identify under-confident children as those who scored well but said they were bad at math. The structure of the PSID also allows us to
observe much of respondents’ young adulthood: the child supplements and core survey followed our sample from 1997 through 2019, so we observe our oldest respondents from age 12 into their thirties.

Biased beliefs about math ability are prevalent in our sample: 5-20 percent of children are markedly over-confident and 7-16 percent are markedly under-confident (using several definitions of biased beliefs, described in more detail below). Over- and under-confidence in math are highly gendered: girls are 2.3 percentage points (pp) (20 percent) more likely to be under-confident and 2.7 pp (27 percent) less likely to be over-confident in math than boys. In contrast, girls are 30 percent less likely to be under-confident in reading than boys. This pattern is consistent with evidence that adults are more likely to be over-confident in stereotypically gender-congruent domains (Coffman 2014; Coffman, Collis, and Kulkarni 2019; Bordalo et al. 2019; Shastry, Shurchkov, and Xia 2020).

One key concern with our measures of over- and under-confidence is that they may just capture children’s private information about their own ability, driven by measurement error in the cognitive tests. We have several key pieces of evidence against this concern. First, the math assessment that we use has high test-retest reliability (Hicks and Bolen 1996). Second, over- and under-confidence persist between waves of the child survey among the 60 percent of our sample with multiple measurements, so our measures seem to capture a stable psychological trait. Next, as we’ve noted, our measures show gender variation that is consistent with prior work on gendered patterns in belief updating, and which we would not expect to see in random testing error. Finally, our results largely persist when we use alternate measures of childhood over- and under-confidence that are less vulnerable to measurement error; we calculate these measures based on test scores and self-reported ability averaged over two waves of the PSID child supplement.
Our main analysis is simple: we estimate the associations between biased beliefs about one’s math ability in childhood and later educational and employment outcomes, controlling for childhood math and reading score deciles, working memory, general confidence, and a host of information on respondents’ demographics and family backgrounds.

Children’s biased beliefs in math strongly predict many of their medium- and long-term educational and employment outcomes. First, confidence has large associations with educational achievement: over-confident children score higher than others with comparable prior scores on math assessments five years later, while under-confident children score lower. Biased beliefs in math also predict educational attainment: over-confident children are more likely to graduate from high school and under-confident children are less likely to graduate from college than others with comparable childhood scores. Under-confident children are also less likely to major in STEM during college and attend less selective colleges, though the latter result is imprecisely estimated. Finally, childhood math confidence predicts key employment outcomes at ages 26 and up. Under-confident children are less likely to work in STEM occupations as adults, and we find suggestive evidence that more confident children earn more and are less likely to be unemployed.

While we do not claim that these associations are causal, we do show that they are robust to several key potential confounders. First, children may form inaccurate beliefs about their ability in part because of how their parents or teachers perceive them, and these adult beliefs may themselves affect children’s later success (Papageorge, Gershenson, and Kang 2018; Jussim and Harber 2005; Wang, Rubie-Davies, and Meissel 2018). However, our main results are robust to controlling for parent and teacher expectations for children’s later educational attainment, teacher perceptions of children’s competence, and parent-reported measures of investment like often doing homework with their child. Second, children may assess their own ability relative to their
school or classroom, while we evaluate their demonstrated ability relative to a national sample. We are limited in our ability to measure school quality, but the measures we do have—proxies for school income, investment, and average achievement—do not correlate with over- and under-confidence, conditional on our other controls. Controlling for these measures of school quality does not change our results. Finally, our results are also robust to controlling for childhood “Big-Five” personality traits, suggesting that over- and under-confidence in math are distinct from these more commonly studied attributes.  

In addition to testing these confounders, we also show that our results hold when we use fourteen different formulations of over- and under-confidence—varying all the key decision points in constructing our main measures—as our primary independent variables.

Two dynamic patterns could underlie the associations we estimate. First, children’s over- and under-confidence could alter early patterns of educational investments by parents, teachers, or children themselves; these early investment patterns could then snowball forward into long-term gaps in education and employment. On the other hand, if children’s biased beliefs persist, they may have direct psychological effects on choices and performance at each stage in a young adult’s development, conditional on his or her performance up to that point. Our evidence suggests that this latter explanation may play a role in the associations we observe. Over- and under-confidence persist through adolescence and into young adulthood (ages 18-27), so biased beliefs could continue to directly affect young adults’ decision-making as they age. Childhood confidence also continues to substantially predict later-life education when we hold fixed all intermediate outcomes.

Our results suggest that over- and under-confidence merit study as psychological traits with key economic implications. While our results are not causally identified, they are consistent with childhood confidence having important effects on later-life outcomes. Our evidence is also
consistent with the idea that those with more confidence fare uniformly better: under-confident children have worse outcomes than their peers with comparable test scores, while over-confident children have better.\textsuperscript{6} Our results leave ample room for future work: to experimentally test the impacts of childhood biased beliefs, to clarify the mechanisms underlying the associations we observe, and to design and test interventions that build confidence in childhood and later life.

Our paper contributes to three literatures in economics.\textsuperscript{7} First, we add to a recent literature estimating the returns to psychological or social attributes in the labor market; we provide the first evidence on the returns to over- or under-confidence in the specific academic domain of math. In addition to the work on general self-esteem and long-term outcomes that we cite above, parallel literatures examine the associations between economic outcomes and the Big-Five personality traits (Almlund et al. 2011; Heckman, Jagelka, and Kautz 2019), competitiveness (Buser et al. 2021), and children’s time, risk, and social preferences (List, Petrie, and Samek 2021). While our data do not measure children’s competitiveness or time and risk preferences, our results are robust to controlling for measures of the Big-Five traits in childhood. Together with this prior work, our paper suggests that future work should disentangle the economic importance of these various traits.

Second, we extend the literature on asymmetric belief updating in adults by documenting over- and under-confidence in a large sample of children in a real-world setting. This heterogeneity matches the lab-based economics literature, which has found mixed patterns of asymmetric updating (Benjamin 2019; Zimmermann 2020). As we’ve noted, the gender gaps in math confidence that we observe are consistent with lab-based evidence that people over-weight positive ability signals in stereotypically gender-congruent domains (Coffman, Collis, and Kulkarni 2019).
Finally, the studies most relevant to our own examine the role of beliefs about ability in educational settings. Owen (2020) shows that male college students over-estimate their own ability in STEM and under-estimate the ability of others, while women are more likely to over-estimate others’ ability; giving students information about their ability then shrinks gender gaps in beliefs and STEM credits. We find that even children have biased beliefs about their own abilities, with similar gendered patterns. Since children’s beliefs may be more malleable than those of college students, our work suggests that interventions like Owen’s may be fruitful at younger ages. Owen does not assess whether the de-biasing intervention has effects beyond the same semester, but our results suggest that longer-term effects could be substantial.

While Owen (2020) intervenes specifically to change students’ beliefs about their ability, other interventions target self-perceptions more broadly. For example, several studies show that building children’s generalized self-efficacy and grit can narrow gender gaps in both confidence and willingness to compete in math (Falco, Summers, and Bauman 2010; Alan and Ertac 2019). Similarly, Carlana, La Ferrara, and Pinotti (2022) find that a multifaceted career-counseling intervention among high-achieving immigrant students in Italy increases self-efficacy and successfully closes native-immigrant gaps in pursuing a more academic high-school track. In contrast, we study math-specific confidence.\textsuperscript{8} We also show that math confidence predicts long-term outcomes even when controlling for general confidence, so interventions to close math confidence gaps may be important complements to interventions that build general self-efficacy or grit.

Finally, Diamond and Persson (2017), the only related paper that considers both biased academic beliefs and long-term outcomes, show that receiving an undeservedly marked-up grade on a test at ages 14-16 leads to higher later test scores, more likely high school and college
graduation, and higher earnings. Since marked-up scores in one subject raise later scores across all subjects, the authors argue that these effects arise in part by changing students’ beliefs about their own ability. However, they do not actually observe students’ beliefs about their own ability, as we do. Together, our papers strongly suggest that students’ biased beliefs about ability matter for later educational and employment outcomes.

The paper proceeds as follows. Section II lays out a conceptual framework for how childhood confidence might affect economic outcomes, and Section III describes our sample and our measures of biased beliefs. Section IV analyzes the prevalence and predictors of childhood over- and under-confidence in our sample, and Section V describes our strategy for estimating the links between biased beliefs in math and long-run economic outcomes. Section VI presents results, Section VII describes the stability of our results to potential confounders and alternate definitions of confidence, and Section VIII explores the dynamic patterns that these long-term associations might follow.

II. How might childhood math confidence affect economic outcomes?

Ability or skill is a primary independent variable in almost every economic model of student and worker decision-making. These include settings where agents are investing in their own futures, like deciding to continue with schooling, choosing a college major or career, or searching for a job (Becker 1964; Roy 1951; McCall 1970; Borjas 1987; Kirkeboen, Leuven, and Mogstad 2016), as well as settings where teachers or parents decide, for example, how to invest in or tailor their pedagogy to a child (Fryer 2018; Dizon-Ross 2019).

Over- and under-confidence would enter any of these models if ability is imperfectly observed: by parents, teachers, and even by the student or worker themself. Where ability and effort are complements, like college applications, over-confident agents may work harder.
Consistent with these cases, psychological theories of motivation, including Bandura’s (1986) Social Cognitive Theory or Expectancy-Value Theory (see Wigfield and Eccles (2000)), emphasize that individuals are more likely to attempt and succeed at tasks in which they feel competent. Where ability and effort are substitutes, like some school tests, over-confident agents may reduce their effort. Bénabou and Tirole (2002) model how over-confidence can persist in equilibrium in either setting.

Over- and under-confidence may also affect outcomes in any setting where teachers or parents decide how to invest time and resources into children based on their perceptions of each child’s ability. If adults interpret more confident children as more skilled, they may over-invest in over-confident children and under-invest in under-confident children. Dizon-Ross (2019) shows that parents have inaccurate beliefs about their children’s academic performance, and that correcting those beliefs causes them to adjust their investments. Similarly, Papageorge, Gershenson, and Kang (2018) show that having a teacher with higher expectations increases a student’s chance of completing college. The same forces could operate in job applications, where potential employers are uncertain about applicants’ skill: in lab experiments, Schwardmann and van der Weele (2019) show that interviewers rate more confident job applicants more favorably, and Mobius and Rosenblat (2006) show that employers offer higher wages to more confident workers.

**Our focus on confidence gaps in math, not reading**

Our data allows us to identify over- and under-confidence in both math and reading, but we focus the remainder of the paper on biased beliefs in math for several reasons. First, performance in math can be measured more objectively than performance in reading, so children’s beliefs about their math ability may be more precise. Next, past work suggests that math ability during
childhood and young adulthood more strongly predicts later achievement than does reading ability (Duncan et al. 2007; Castex and Kogan Dechter 2014; Goodman 2019). We find similar patterns in our data in Appendix Table A1, where we regress our main education and employment outcomes on childhood test scores and the set of controls that we will use throughout our main analysis. While both math and reading score percentiles predict later academic achievement, educational attainment, and working in high-education occupations, only math scores predict earnings and unemployment. Thus, children’s perceptions of their own ability in math may also link more strongly with later-life achievements than do their self-perceptions in reading. Finally, the Bureau of Labor Statistics predicts that employment in STEM occupations will continue to grow at faster rates than non-STEM occupations through 2030, so math ability may become an even more important predictor of success in the labor market.

That said, we conduct all of the subsequent analysis for reading confidence (Appendix Tables A2-A5). Reading confidence robustly predicts few educational or employment outcomes.

III. Measuring confidence and later-life outcomes in the PSID

III.A. Sample and survey design

We explore the links between biased childhood beliefs and outcomes in young adulthood using the rich data of the Panel Study of Income Dynamics (PSID). The PSID was first collected in 1968 among 5,000 nationally representative households from two independent samples: a national sample of low-income families from the Survey of Economic Opportunity (the “SEO sample”) and a national sample drawn by the Survey Research Center (the “SRC sample”). The PSID has since surveyed the descendant households of the original sample annually from 1968 to 1997 and biennially thereafter, adjusting the sample in 1997 to again make it nationally representative.
We combine the core PSID with two supplements that follow respondents from childhood into young adulthood: the Childhood Development Supplement (CDS) and the Transition into Adulthood Supplement (TAS). The CDS was introduced in 1997, sampling up to two children per PSID household who were then between the ages of 0 and 12 (3,563 children). The CDS collects detailed information from children themselves, from their primary caregivers, and from their elementary school teachers on areas including children’s cognitive and emotional development, health, and exposure to parenting practices. The original CDS sample was re-interviewed in 2002-2003, then aged 5-17, and those still below age 18 were included in a third CDS wave in 2007.

In 2005, the PSID introduced the TAS as a bridge between the CDS and the main PSID survey for CDS respondents, the oldest of whom had reached ages 18 to 20 by that year. The TAS has been collected biennially since 2005, with younger CDS respondents aging into the TAS sample at 18. Individuals participate in the TAS until they become economically-independent heads of their own households, at which point they enter the adult PSID sample and are surveyed every two years. The TAS is designed to capture respondents’ social and career development as they enter adulthood; we use its modules on education, employment, income, and personality.

The PSID-CDS-TAS data structure is uniquely suited to exploring the links between childhood confidence and long-term educational and employment outcomes. First, the CDS both administers a math test and asks children to evaluate their own math ability; we combine children’s test scores and self-assessments to identify over- or under-confidence in math. Section III.B below details the CDS tests, self-assessments, and our confidence measures. Second, following CDS children into the TAS and then the PSID allows us to observe detailed data on
educational and employment outcomes over 22 years, following our oldest respondents into their mid-thirties. Finally, the extensive data on parents’ employment and income in the PSID and on parenting practices and other child characteristics in the CDS allows us to control for many covariates that could confound the relationship between biased beliefs and long-run outcomes.

For example, the detailed child module in the CDS allows us to control for other forms of ability and confidence that are distinct from skill and confidence in math, but which may correlate with them. We construct a measure of general confidence as the mean of standardized variables capturing whether children see themselves as broadly competent (see Appendix B for details); we have no measure of true ability by which to normalize this general confidence scale, so we use it as a control for unobserved abilities and other dimensions of confidence that may correlate with biased beliefs in math and also affect later-life outcomes. We also control for children’s scores on the Digit Span subtest of the Wechsler Intelligence Scale for Children (Revised), a measure of short-term memory. Finally, the CDS and core PSID collect detailed household information on total family income, household heads’ education, primary caretakers’ values and mental health, household structure, and financial characteristics like whether the household receives food stamps. Section V will detail the family and child variables that we control for in our main analysis.

Our final sample consists of the 2,985 CDS respondents with at least one year of math cognitive tests and self-assessments in the CDS, about 84 percent of all CDS respondents.9 We report summary statistics for this sample in Appendix Table A6; all variables are observed in the same year in which we first observe childhood over- or under-confidence in math.

Our sample is non-randomly selected from the national population, both because the initial 1968 PSID sample oversampled low-income families and because there is unobserved selection in whether CDS participants report math test scores and self-assessments.10 This selection
appears in our sample statistics in notable ways. First, our sample is disproportionately Black: 45.8 percent are White, 41.7 percent are Black, and only 7.5 percent are Hispanic, while the U.S. Census Bureau reports that 69.1 percent of the US residents were White, 12.1 percent Black, and 12.5 percent Hispanic in 2000 (Greico and Cassidy 2001). While the Census Bureau reports median household income in 1997 of $55,336, our sample’s median taxable income is slightly lower, at $52,029 (both in 2016 USD). On the other hand, our study sample performs disproportionately well on the CDS standardized tests: we observe median CDS math and reading score percentiles of 60 and 54, respectively, relative to national norming samples.

While we do not weight our sample to be nationally representative in our main analysis, we include results that do so in Appendix Tables A7-A10. These weights are based on those published by the CDS, which capture the inverse probability of respondents’ inclusion in the CDS sample; we then recalibrate these CDS weights via iterative proportional fitting, or raking, to ensure that our sample matches marginal distributions of percentile CDS math scores, race in 2000, and total household income in 1997. Our main results are less precisely estimated when we use weights, though they remain qualitatively similar. We present all descriptive statistics both for the weighted and unweighted samples.

III.B. Measuring over- and under- confidence in math

Data on children’s self-reported and demonstrated ability in math

The CDS assesses children’s math skills using the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R), a test of academic achievement commonly used by school psychologists in the 1990s (Stinnett, Havey, and Oehler-Stinnett 1994; Hicks and Bolen 1996; Duffy and Sastry 2014). The CDS administers the Applied Problems subtest of the WJ-R, comprising 60 word problems of increasing difficulty that assess math reasoning and knowledge. Each child completes only a subset of the test, beginning at a “basal” level, where
they answer six consecutive questions correctly, and ending at a “ceiling” level, where they get six consecutive questions wrong. The CDS then reports each respondent’s percentile rank relative to the nationally representative WJ-R norming sample for their age group; we use these percentile ranks as our measure of each child’s demonstrated ability in math. Panel A of Figure 1 shows the distribution of these scores in our sample.

In addition to collecting this measure of performance in math, the CDS also asks all respondents ages 8 or older to assess their own ability in math, asking them to answer, “How good at math are you?” on a scale of 1 (not at all good) to 7 (very good). Children never receive their scores on the WJ-R math test, so these self-reports do not reflect feedback from the CDS. Panel B of Figure 1 shows the distribution of these self-assessments. Math self-perceptions are highly skewed towards positive responses, with over 89 percent of respondents ranking themselves as “Okay” or better at math. This skew may be partially explained by the distribution of percentile scores in Panel A, which skews heavily towards higher-performing children. While shifted upwards, children’s self-reports do contain information about objective ability: in Panel C of Figure 1, average math test percentiles rise almost linearly with self-reported ability in math.

We measure children’s over- and under-confidence in math in the first wave of the CDS in which they have non-missing cognitive test scores and self-assessments, leaving us with a sample of 2,985 children. We first measure confidence for the median child before age 12, and we observe confidence by age 13 for 83 percent of children. Thus, we will interpret our measures as childhood over- and under-confidence in math. Throughout, our analysis will control for both birth year and the age at which we first observe confidence.

**Defining binary measures of over- and under-confidence**
We first identify over- and under-confidence in math using large mismatches between children’s score percentiles and their self-assessments. In particular, we classify any respondent as under-confident in math if she scored above the 75th percentile nationally and ranked her own ability at 1 to 4, corresponding to the bottom 47 percent of the subjective-ability distribution in our sample, or if she scored above the 50th percentile nationally and ranked herself at 1 to 3, corresponding to the bottom 10 percent of the subjective-ability distribution. We define over-confidence among low-achievers using similar thresholds, but we account for the skewed self-assessment distribution by using stricter cut-offs to identify biased beliefs. In particular, we identify any respondent as over-confident in math if she scored below the 25th percentile nationally and rated her own ability at 6 or 7, corresponding to the top 39 percent of the subjective-ability distribution in our sample, or if she scored below the 50th percentile and rated herself at 7, corresponding to the top 22 percent of the subjective-ability distribution.

These measures of math over- and under-confidence have several key strengths: they are easy to define and observe, they refrain from putting too much stock in the cardinal value of children’s self-assessed ability, and they account for the upward skew in self-assessments, which we consider to be a form of response bias separate from over- or under-confidence.14

However, our measures also have several limitations. First, we can only identify over-confidence among children scoring below the 50th percentile and under-confidence among those scoring above the 50th percentile; note that this strategy matches the existing literature, which typically documents under-confidence (imposter syndrome) among high-achievers (Sakulku 2011). Another limitation is that these measures are not directly comparable to measures of over- and under-confidence from the lab-based literature, which can precisely quantify gaps between respondents’ true and self-assessed performance on a quiz or rank relative to a group (Coffman,
Collis, and Kulkarni 2019; Möbius et al. 2014; Eil and Rao 2011). Our measures of over- and under-confidence, in contrast, identify coarse categories of children with large gaps between their self-assessments and observed scores. Our second measure of biased beliefs, described below, aims to partially address these limitations.

**Defining a more continuous measure of biased beliefs**

Our second confidence measure identifies biased beliefs as the difference between children’s self-reports and their observed performance on the CDS math test. To transform these objects to the same scale, we split the distribution of children’s percentile scores uniformly into seven bins, where 1 includes the lowest 14 score percentiles and 7 includes the highest 14 score percentiles relative to the national norming sample. We then assume that students with full information about the national distribution of scores and their place in it would have self-reported their math ability as the bin from 1 to 7 in which their score percentile falls. We take the difference between their actual self-report and this bin as our measure of biased beliefs; this measure then takes on integer values from -6 to 6. For ease of interpretation, we standardize this variable to have mean 0 and standard deviation 1 throughout the rest of the paper.

This measure has three strengths relative to our main measure: it allows for more granularity in the extent of biased beliefs, aligns more closely with measurements of biased beliefs in the lab-based literature, and relies on fewer choices by the authors. However, by assuming that we can identify even small biases in beliefs about math ability, it is more likely to conflate actual biased beliefs with children’s private information about their math ability (described in more detail in the next section). It may also be confounded by forms of reporting bias other than over-confidence that generate the overall upward skew in self-reports (see footnote 14).
We present results for all outcomes using both the binary and more continuous formulations of biased beliefs, and in general the results are extremely consistent. To ensure that our main results do not arise just from our particular choice of confidence measures, we show that they are robust to a range of alternate definitions of both our indicators for over- and under-confidence and this more continuous measure of biased beliefs. We describe these alternate measures in Section VII.

III.C. Biased beliefs or measurement error?

One key concern with our measures of biased beliefs in math is that they may conflate over- and under-confidence with children’s private information about their math ability, perhaps driven by measurement error in the WJ-R assessment. Four key pieces of evidence support the claim that our measures truly capture biased beliefs in childhood.

First, prior work has shown that the WJ-R assessment is a reliable measure of children’s math skills, with test-retest reliability for the applied math problems of about 0.85 in large samples (Hicks and Bolen 1996). We can also verify WJ-R reliability across math domains in our sample using the 1997 wave of the CDS, which administered both the Calculation and Applied Problems subtests of the WJ-R. For the 1,450 children who took both tests, the correlation in percentile ranks on the two sections is 0.69. Our binary designations of children as over-confident, under-confident, or neither are also highly consistent whether we measure objective math ability using children’s percentile scores on the Calculation or Applied Problem subtest: 81 percent of children with both measures are classified in the same category regardless of which ability measure we use. Another ten (nine) percent switch from under-confident (over-confident) to neither or vice versa.
Second, our measures of childhood math confidence persist over time. About 60 percent of the children in our sample appear in two waves of the CDS, allowing us to construct two measures of over- and under-confidence taken five years apart. Children appear in a second CDS wave at ages 13 to 19, so these second-wave measures capture biased beliefs in adolescence.

Table 1 regresses our adolescent measures of biased beliefs on our childhood measure of the same variable, controlling for a set of demographics and parent characteristics that we will use throughout our later empirical analysis; we outline these specifications in detail in Section V. These regressions show substantial persistence: respondents who were over-confident in math as children are about 3 times as likely (12pp more likely) to be over-confident in math as adolescents, while under-confident children are about 1.7 times as likely (4pp more likely) to be under-confident as adolescents. Similarly, a one-standard-deviation increase in the degree of biased beliefs in childhood predicts 0.19sd more biased beliefs in adolescence. If our confidence measures just captured random testing variability, we would not expect to see such substantial persistence.

Third, our main results are largely robust to using measures of over- and under-confidence that reduce potential measurement error by combining observations of children’s test scores and self-reported ability across two waves of the CDS. We discuss these measures and results in more detail in Section VII. If measurement error is uncorrelated across tests taken 5 years apart, these average confidence measures will be less vulnerable to it than are our main measures.

Finally, we describe in the next section that we observe substantial gender gaps in math over- and under-confidence, with girls more likely to be under-confident and less likely to be over-confident. This pattern is consistent with gender stereotypes about math ability, which may shape children’s beliefs even at young ages, and mirrors results for adults in the lab (for instance,
Coffman, Collis, and Kulkarni (2019)). Our measures of over- and under-confidence could only
be entirely explained by measurement error if this error took a similar gendered pattern, beyond
its correlation with WJ-R Applied Problems scores and with the many other controls we outline
in Section V.\textsuperscript{20} We consider a few possible sources of non-random measurement error that could
generate these patterns: skill in some dimension of math that the test does not cover, test-taking
anxiety, and test-taking motivation.

First, the CDS data allow us to test for gender gaps in one central dimension of math skill that
our main test scores do not directly capture: calculation skills. Using the 1997 CDS sample,
when children took both the WJ-R Calculation subtest and the WJ-R Applied Problems subtest,
we find no evidence that boys have better calculation skills conditional on the applied problems
scores that we use in our main analysis.\textsuperscript{21}

Next, differential measurement error in the CDS math tests could arise if boys or girls are
more prone to testing anxiety that impairs performance. While past work finds that boys show
higher physiological stress during test-taking (Weekes et al. 2006; Stroud, Salovey, and Epel
2002), other research suggests that physiological stress only impairs performance when students
psychologically appraise it as an indicator of potential failure (Jamieson, Mendes, and Nock
2013; Mattarella-Micke et al. 2011). Girls tend to have higher psychological test anxiety and
math anxiety, and most commentary suggests that it is these psychological manifestations of
anxiety that pose first-order risks to test performance (Devine et al. 2012; Erturan and Jansen
2015; Ballen, Salehi, and Cotner 2017). Thus, we would expect girls’ test performance to
differentially lag their true skill, producing gender gaps in confidence that would conflict with
our empirical results.
Finally, we turn to test-taking motivation. Past work finds that girls are somewhat more motivated than boys to exert effort on low-stakes tests, so boys’ CDS math scores may be differentially low relative to their true skill in math (Segal 2012; DeMars, Bashkov, and Socha 2013; Gneezy et al. 2019). Then, boys may appear more over-confident by our measures. While it is hard to fully eliminate this possible confounder in our setting, our results are robust to controlling for agreeableness and conscientiousness, two Big-5 personality traits that are positively correlated with unincentivized test effort (DeMars, Bashkov, and Socha 2013; Segal 2012). (See Section VII for more details.)

Together, most evidence from our empirical setting and from past work on test-taking strongly suggests that our confidence measures capture a meaningful psychological trait. However, we cannot fully eliminate the risk that these measures capture children’s private information on some aspect of math ability that the test systematically excludes. Any such confounder could only explain our results if it is differentially weak among girls and affects outcomes beyond its correlation with demonstrated math ability, general confidence, digit span score, reading ability, and the many other controls we outline in Section V.

IV. Patterns of over- and under-confidence in the population

This section documents the prevalence and correlates of over- and under-confidence in our sample. Besides documenting biased beliefs in math in a real-world setting, these results are useful both to validate our measures of biased beliefs and to inform our strategy for estimating the links between childhood confidence and long-run outcomes, which we describe in Section V.

IV.A. Prevalence of biased beliefs

We find substantial over- and under-confidence among children in our sample: using our main binary measures, 8.5 percent of children are over-confident at their first measurement,
while 12 percent are under-confident.\textsuperscript{22} Since these measures identify large gaps between children’s self-assessed and objective performance, these shares are strikingly high. Turning to our more continuous measure of biased beliefs, 21 percent of children report the same ability bin as their percentile score would imply, 8.7 percent of children report ability levels that are at least 3 bins lower than that of their score, and 17 percent report ability levels that are at least 3 bins higher, where each bin spans 14 score percentiles. See Appendix Figure A1 for the full distribution of the continuous confidence measure. It is notable that over- and under-confidence are both prevalent in this large sample, given psychology’s focus on over-confidence (Moore and Healy 2008) and the mixed evidence from lab experiments on asymmetric belief updating (Benjamin 2019).

Next, older children have more accurate beliefs. Panel A of Appendix Figure A2 plots the share of children who are over- or under-confident in math by age; Panel B plots the cumulative density function for the continuous confidence measure for three age groups, pooling respondents’ observations across CDS waves. Both panels show that average belief accuracy increases almost monotonically as children age. We focus on the associations between confidence and later-life outcomes using first-observed confidence, so our confidence observations are drawn from young ages with more biased beliefs. We eliminate bias due to the timing of our confidence measurements by including fixed effects for the age at which confidence was measured in all regressions.

\textbf{IV.B. Biased beliefs and other child characteristics}

Over- and under-confidence correlate with other child characteristics in largely expected ways (Table 2 and Appendix Table A11). Unsurprisingly, children with higher general confidence are more likely to be over-confident and less likely to be under-confident in math,
and children with higher digit span scores are less likely to be under-confident. Math test score deciles strongly predict confidence gaps (though some of this correlation arises mechanically from how our measures are constructed), while reading test score deciles do not (Appendix Table A11). We will control for children’s general confidence, digit span scores, and test score decile fixed effects in math and reading in all regressions of later-life outcomes on childhood biased beliefs.

Conditional on these measures of ability, children who have ever been in a gifted program are 8.7pp less likely to be under-confident in math and 2.6pp more likely to be over-confident. These correlations could reflect that schools and children share private information on children’s ability conditional on CDS scores, that being in a gifted program alters children’s confidence, or that children’s confidence influences their treatment at school conditional on ability. To avoid controlling for mediators of the effects of confidence, our regressions will not control for this variable or other signals of ability from schools, like repeating a grade.23

Finally, gender is the strongest demographic predictor of math confidence. Girls are 2.3pp (20 percent) more likely to be under-confident and 2.7pp (27 percent) less likely to be over-confident in math than boys with the same score deciles, and on average, girls’ beliefs are 0.1 standard deviations (sd) less positively-biased than the average boy’s. Note that girls do not have more accurate beliefs, simply more negatively-biased ones.24,25 This finding is consistent with prior literature showing that adults are more over-confident in gender-congruent domains (Coffman, Collis, and Kulkarni 2019), but it is notable that we find it in children, the majority of whom have not yet entered puberty. These gender differences are present at almost all ages, but due to small sample sizes the patterns are imprecise (available upon request).26
Perhaps surprisingly, we find no significant links between children’s math confidence and their parents’ education or occupation, household income, or race, conditional on all other characteristics. However, noise in these estimates means we cannot reject potentially large correlations.

V. Confidence and long-term outcomes: Empirical strategy

Our empirical strategy is simple: we estimate the associations between biased ability beliefs in math and later education and work outcomes, holding fixed measured childhood ability. We use the PSID’s rich data on childhood environment and family characteristics to control for extensive pre-determined confounders, but we refrain from interpreting our estimates as the causal effects of confidence. We estimate the following specification:

$$Y_{it} = \alpha + \beta_1 Over_{i0} + \beta_2 Under_{i0} + A'_{i0} \mu + X_{i0}' \delta_1 + X_{i0}' \delta_2 + \gamma_s + \omega_t + \epsilon_{it}$$

where $Y_{it}$ is individual $i$’s outcome of interest in adolescence or adulthood, measured in wave $t$ of the TAS or PSID, and $Over_{i0}$ and $Under_{i0}$ are indicators for being over- or under-confident in math as a child, respectively. All of our main tables also include regressions in which we replace $Over_{i0}$ and $Under_{i0}$ with the single $Z_{conf_{i0}}$ variable, the more continuous measure of the degree to which a child is over- or under-confident, in standard deviations. Due to power limitations, we assume that $Z_{conf_{i0}}$ has a linear relationship with our outcomes of interest.27

Next, all of our regressions include $A_{i0}$, a vector of controls for childhood ability. In particular, $A_{i0}$ includes linear controls for childhood digit span score and general confidence, as well as fixed effects for test score deciles in both reading and math.28 Our basic specification also includes state fixed effects $\gamma_s$, TAS or PSID wave fixed effects $\omega_t$ when the outcome is observed multiple times for each individual,29 a set of child controls $X_{i0}^C$, and a set of parent controls $X_{i0}^P$. In
our first specification, $X_{t0}^C$ and $X_{t0}^P$ include only variables that are certainly unaffected by respondents’ childhood math confidence: $X_{t0}^C$ includes fixed effects for race, birth year, quarter of birth, gender, and age at which we observe confidence, and $X_{t0}^P$ includes family income, its square, and fixed effects for both parents’ levels of education. All variables indexed at $t = 0$ are from the first CDS wave in which a child had WJ-R scores and an ability self-assessment. Since about two-thirds of the children in our sample have a sibling in the sample, we cluster standard errors by family. Our coefficients of interest are $\beta_1$ and $\beta_2$.

Our second specification takes advantage of the detailed caregiver interviews in the CDS to add additional controls for child and family characteristics that may correlate with both confidence and long-run outcomes. In addition to expanding the set of child controls, $X_{t0}^C$, with the primary caregiver’s assessment of the child’s general health, this specification supplements $X_{t0}^P$ with additional parent and family controls: whether the family receives government transfers; whether the household includes the father or has two adults; parents’ beliefs about gender norms and the qualities that are most important for success; and parent mental health (see Appendix B for details). Finally, we add four indicators for whether the child’s parents work in STEM or another high-education occupation (based on Anaya, Stafford, and Zamarro (2021); see footnote 8). We focus on this specification throughout the text, but results are generally consistent across these two specifications.

VI. Confidence and long-term outcomes: Results

The following section presents our results, documenting strong associations between childhood under- and over-confidence in math and key later-life outcomes: adolescent test scores, graduation from high school and college, college major, career choice, earnings, and unemployment. We present these results in Tables 3, 4 and 5.
VI.A. Medium-term educational achievement

We first examine the links between childhood confidence and medium-term educational achievement, measured as adolescent scores on the CDS math assessments. We observe these scores at children’s second CDS observation, about 5 years after we first observe their confidence in math.

Children’s biased beliefs in math significantly predict adolescent math performance (Table 3, columns 1 and 2). Using our binary measures (Panel A), children who are over-confident in math score 2.7 percentiles (standard error = 1.5p) higher on the math assessment five years later than others with comparable baseline scores, while under-confident children score 5.9 percentiles (se = 1.5p) lower. Using our more continuous measure (Panel B), a child with 1 standard deviation (sd) higher math confidence in childhood scores 2.8 percentiles (se = 0.57p) higher on the math assessment 5 years later than others with comparable baseline scores. Children marked as over- or under-confident in our binary metrics differ from others by an average gap of 1.8sd and -1.6sd in continuous degrees of confidence, respectively, so our estimate magnitudes are remarkably consistent across the two panels. In contrast, there is no relationship between childhood math over- or under-confidence and adolescent reading scores using either measure of biased beliefs (Table 3, columns 3-4).

These associations are large relative to the links between raw math ability and later scores: increasing one’s childhood math score by 10 percentiles is associated with scoring on average 5.3 percentiles higher in adolescence (Column 1 of Appendix Table A12). Thus, being over-(under) confident in math predicts as large a gap in adolescent test scores as does increasing (decreasing) one’s childhood math test score by 5-11 percentiles.

VI.B. Educational attainment
Biased beliefs in math during childhood also predict important gaps in high school and college graduation. Children who are over-confident in math are 6.2 percentage points (se = 2.6pp) more likely to graduate from high school, and children who are under-confident in math are 5.8pp (se = 2.8pp) less likely to graduate from college (Table 3, columns 5-8, Panel A). Since only 30 percent of our sample graduates from college, being under-confident predicts a 20 percent drop in the likelihood of college graduation. We find very similar results using our more continuous measure in Panel B: a child with 1sd higher math confidence in childhood is 1.8pp (se = 1.0pp) more likely to graduate from high school and 3.3pp (se = 1.1pp) more likely to graduate from college, though the first is only marginally significant. Again, the magnitudes of these results are similar regardless of which confidence measure we use.

These gaps are large relative to the associations between childhood math scores and educational attainment in our data: childhood math scores do not substantively predict high school graduation, and increasing test scores by one decile is associated with being 2.9pp more likely to graduate from college on average (Appendix Table A12, columns 3-4).

**VI.C. College quality, college major choice, and graduate education**

Next, we consider later-education outcomes among those who went to college: college quality, college major choice, and whether respondents complete a graduate degree. Since we restrict to college graduates, these regressions use much smaller samples than for our previous outcomes.

First, we find imprecise links between childhood math confidence and the quality of colleges that children later attend. We consider two quality measures: first, an index of general college quality, and second, colleges’ 75th-percentile math SAT scores among incoming freshmen – a more specific measure of math quality. We focus our discussion on colleges’ 75th-percentile
math scores (Table 4, column 3 and 4), but our results are similar using the more general college quality index (columns 1 and 2). Under-confident children attend schools whose 75th-percentile math SAT scores are 11.3 points (se = 5.9 points) lower than others with the same childhood scores (p = 0.06); with 95 percent confidence, we can reject that under-confident children attend schools with math SAT scores that are over 0.3 points higher or 22.9 points lower. Over-confidence is not significantly associated with college quality among childhood low-scorers, but again we observe wide confidence intervals: we cannot reject that over-confident children attend colleges that have 20 points lower to 26 points higher SAT scores than their peers. Using our more continuous measure of biased beliefs in Panel B yields consistent, but imprecise, results.

Next, we find that childhood under-confidence in math is starkly associated with major choice among those who go to college (Table 4, columns 5 and 6). Among those with a 4-year college degree, students who were under-confident in math are 16.2pp (se = 3.6pp) less likely to earn a STEM major than their peers with comparable childhood scores, an 86 percent drop from the share of STEM majors across all college graduates in our sample. This large gap means that under-confident children who score above the 50th percentile on the CDS math test are only 1.3 times as likely to major in STEM, conditional on going to college, as the average child who scores below the 50th percentile; in contrast, other childhood high-scorers are 3.6 times as likely to major in STEM as low-scorers. We obtain very similar results using our more continuous measure of biased beliefs in Panel B: a 1sd increase in confidence is associated with a 7.8pp (se = 2.3pp) increase in the likelihood of majoring in STEM.

Finally, we find no significant relationships between biased beliefs and getting a graduate degree, though again our standard errors are large (Table 4, column 7 and 8).

**VI.D. Employment outcomes**
Next, we examine the links between childhood over- and under-confidence in math and employment outcomes in young adulthood: occupation type, earnings, and employment status. We follow respondents in the adult PSID when they age out of the TAS, so we observe our oldest CDS respondents through age 36 at the end of our sample period. Since respondents’ employment outcomes in their early twenties may not yet be representative of their long-term career trajectories, we restrict the sample to observations in which respondents are older than 25; we observe about 70 percent of our sample above this threshold at least once.\textsuperscript{35}

We first consider job choice. Under-confident children are about 4.9 pp (se = 1.6 pp) less likely to work in a STEM occupation\textsuperscript{36} than their peers (Table 5, columns 1 and 2), a gap that is approximately equal to the baseline rate at which respondents later work in STEM in our sample. We find a similar result with our measure of the degrees of over- and under-confidence, where a 1sd increase in childhood confidence is associated with a 1.8 pp (se = 0.6 pp) increase in the likelihood that one works in STEM. These confidence gaps are large relative to the link between childhood math scores and later STEM employment, which is precisely estimated but very close to zero (Appendix Table A12, column 9).

On the other hand, there are no gaps in the likelihood that over- or under-confident children work in non-STEM high-education occupations\textsuperscript{37} (Table 5, columns 3 and 4). These results are reassuring for our empirical design: the fact that math confidence matters for STEM employment, but not other high-education employment, helps to validate that we properly isolate long-term associations with children’s biased beliefs in math, rather than picking up correlations with unobserved self-esteem or other abilities. Taking these point estimates at face value, over half of the under-confident children who do not pursue STEM careers switch into other high-education occupations, while the rest pursue other work. However, our 95-percent confidence
intervals include estimates suggesting that under-confident children are up to 3.9pp less likely or up to 9.1pp more likely to work in other high-education occupations than their peers.

Next, we consider respondents’ earnings. Our regression results are imprecisely estimated, but they broadly suggest that higher math confidence is associated with higher earnings later in life (Table 5, columns 5 and 6). While our binary measures of over- and under-confidence are not significantly associated with earnings (Panel A), a 1sd increase in the degree of childhood confidence is associated with 5.9 percent (se = 2.9 percent) higher earnings in adulthood. This gap is large relative to the association between childhood math scores and adult earnings: increasing test scores by one decile is associated with 6 percent higher earnings on average (Appendix Table A12, column 11).

Finally, we consider unemployment. Our regressions suggest that higher confidence may be associated with lower unemployment risk (Table 5, columns 7 and 8). Again, our binary indicators for over- and under-confidence are not significantly associated with unemployment (Panel A), but a 1sd increase in childhood confidence is associated with a 2.3pp (se = 0.9pp) lower likelihood of having been unemployed in the previous year. This gap is large relative to the association between childhood math scores and unemployment: increasing test scores by 10 percentiles is associated with 1.2pp lower unemployment risk on average (Appendix Table A12, column 12).

While most of our results are quite stable – both in magnitude and precision – to the many robustness tests we run in Section VII, our results for earnings and unemployment should be interpreted with caution. They are only statistically significant when using our more continuous measure of biased beliefs, which is more vulnerable to measurement error, and we show in Section VII below that they are not robust to using measures of confidence that minimize
measurement error by using data from two waves of the CDS. That said, they are suggestive and are consistent with our other findings on the long-term links between childhood confidence and later-life outcomes.

**VII. Robustness**

In this section, we show that our main results are robust to controlling for a range of possible confounding variables and to many alternate definitions of our key measures of biased beliefs.

**VII.A. Key confounders: Personality, adult investment, and school quality**

First, we show that math over- and under-confidence predict long-run outcomes beyond their correlation with (1) more commonly studied personality traits, (2) parent and teacher beliefs and investment, and (3) and elementary-school quality. We do not control for these variables in our main specifications because they are likely jointly determined with math confidence, but they may confound the links we estimate. See Appendix E for more details on data used in this section.

Section 1 of Appendix Table A13 adds controls for children’s Big-Five personality traits: conscientiousness, agreeableness, neuroticism, openness, and extroversion. The CDS did not administer standard psychometric scales to identify the Big-Five traits among children, so we construct these measures from caregivers’ reports of child behavior (see Appendices B and E for details.) These traits could confound the long-term associations that we observe: other work shows that Big-Five personality traits correlate with contemporaneous educational and employment outcomes (Almlund et al. 2011; Heckman, Jagelka, and Kautz 2019), and we find some correlations between these common personality traits and measures of over- and under-confidence in our sample (Appendix Table A14). However, our estimates of the links between over- and under-confidence and long-run outcomes are broadly robust to controlling for the Big-Five traits.
Next, we add controls to our main specification for parent investments, like reading or doing homework with the child, teacher ratings of children’s academic, social, and physical competence, and the educational attainment that parents and teachers predict for the child. (Note that we only observe teacher perceptions for 20-34 percent of the sample.) Teacher and parent beliefs and investments do correlate with children’s beliefs in math in our sample (Appendix Table A15); if these adults’ investments affect children’s later-life success, they may drive the links between childhood math confidence and later outcomes that we observe (Papageorge, Gershenson, and Kang 2018; Dizon-Ross 2019). However, Section 2 of Appendix Table A13 shows that children’s over- and under-confidence continue to predict long-run outcomes in similar ways when we add controls for adult perceptions and investment to our main regressions.

Finally, our results are robust to controlling for the quality of school a child attended when we first observe their biased beliefs in math. If children assess their own ability relative to their peers, not the national distribution, school quality may shape children’s self-assessments in math; over-confident children could just be those with low-performing peers, for example. However, these patterns would tend to bias our results towards zero, since later-life outcomes may be worse for children from lower-achieving schools. We use restricted data from the CDS to match students with data on the percent of students at their school who qualified for free or reduced-price lunch (a proxy for income), the average student-teacher ratio at their school (a proxy for educational inputs), and levels and trends of their school’s mean achievement levels in math and reading. Reassuringly, our results do not change meaningfully when we control for school quality (Appendix Table A13, Section 3).

VII.B. Alternate definitions of biased beliefs

Next, we show that our results are robust to a range of alternate measures of biased childhood beliefs in math. Appendix F describes each of these alternate measures in more detail. None of
these changes affects our main conclusions: that over- and under-confidence strongly and meaningfully predict long-term education and working in STEM.

**Redefining over- and under-confidence:** We first redefine our binary measures of over- and under-confidence by altering the CDS math score and self-report cutoffs on which they rely, making those designations more or less strict than our main measures. Second, we construct more data-driven measures of over- and under-confidence—what we refer to as the relative confidence measures—that identify over- and under-confident children as those in the tails of the distribution of math scores at each self-reported ability level. Finally, a third class of binary over- and under-confidence measures marks a child as over-confident if the degrees-of-confidence measure is greater than 2 and under-confident if it is less than -2.

**Redefining degrees of confidence:** Next, we also test robustness to the key design choice in our more continuous measure of confidence: how we map self-assessed ability and observed scores to the same scale. Our main measure assumes that children with accurate beliefs would report the numbered bin from 1 to 7 in which their CDS score falls when test score percentiles are uniformly distributed across 7 bins (that is, each bin covers about 14 percentiles). We test robustness to two other transformations: the first assumes that children should have reported the bin from 1-7 in which their test score would fall if they had the CDS’ empirical self-assessment distribution in mind, and the second instead differences children’s percentiles of self-assessed ability and demonstrated ability. Each of these is converted to standard-deviation units to facilitate comparisons.

**Measurement error:** To reduce the likelihood that our results are driven by measurement error, we also construct alternate confidence measures using testing and self-assessment data from two waves of the CDS for the 60 percent of children with multiple measures. We take two
approaches to redefining our indicators for over- and under-confidence. In the first, we average children’s test scores and self-reported ability over two waves and then apply our standard cutoff rules to these average scores and self-reports. In the second, we calculate indicators for being over- or under-confident separately in each of two waves and then average these indicators. We use the same logic in defining multi-wave versions of the more continuous confidence measure.

Results: Appendix Figures A4-A16 present specification charts showing results for each of our main outcomes of interest using these alternate measures of biased beliefs. For simplicity, Appendix Table A16 presents a subset of these results: we iterate through alternate definitions of biased beliefs for each outcome, always using the control variables from our preferred specification. Panel A shows the results for over- and under-confidence, and Panel B shows the results for our more continuous measure of biased beliefs. Most coefficients that are statistically significant in our main results are remarkably stable, leaving our conclusions unchanged. The only exceptions are our results for earnings and unemployment, which disappear when we use the more continuous measure of biased beliefs based on two waves of the CDS.

VIII. Snowballing investment or persistent over- and under-confidence?

Childhood over- and under-confidence in math are associated with gaps in key educational and employment outcomes down the line, from adolescent math performance to career choices in young adulthood. As we outlined in Section II, these confidence gaps could arise if over- and under-confidence shape children’s own investment decisions or those of parents, teachers, or potential employers. In this section, we explore the dynamic patterns through which these confidence gaps open up and persist. On one hand, math confidence could produce investment gaps in childhood that in turn snowball through children’s later education and occupational choices. On the other hand,
childhood over- and under-confidence in math may persist into adulthood and directly affect choices and performance at each stage of life, conditional on past achievement.

This section explores whether biased beliefs persist into adulthood, and whether gaps in later-life outcomes can be fully accounted for by the links between confidence and intermediate investments that we observe.

**VIII.A. The persistence of childhood confidence in math**

While Table 1, discussed in Section III.B., shows that over- and under-confidence in math persist from childhood into adolescence, we also find that childhood biased beliefs persist even until we last observe respondents in the TAS at ages 18 through 27 (Table 6). This persistence is a necessary condition for children’s biased beliefs to have direct behavioral effects on their educational and career choices as they age. We use the wealth of questions in the TAS to construct measures of young adults’ confidence in math and reading, generalized academic confidence, career confidence, and general confidence. See Appendix B for more detail on each of these measures.

First, we calculate an index of adult math confidence as the mean of standardized ratings of how good respondents think they would be in a job requiring math or technology. By this metric, childhood math confidence strongly persists into adulthood. Respondents who were over-confident in math as children score about 0.26sd (se = 0.06sd) higher in math confidence as adults than others with comparable childhood scores, while under-confident children score about 0.25sd (se =0.05sd) lower (Table 6, columns 1 and 2, Panel A). Likewise, a 1sd increase in our more continuous measure of childhood math confidence predicts 0.17sd (se = 0.02sd) higher math confidence as an adult (Panel B). In contrast, children who were under-confident in math score about 0.16sd (se = 0.05sd) higher in adult reading confidence—measured by standardizing subjects’ ratings of how good they would be in a job requiring them to read and write a lot—than
others with comparable childhood math scores (Table 6, columns 3 and 4). This pattern may arise because under-confident children are less likely to work in STEM occupations, making them more likely to have a job requiring reading and writing.

Next, childhood over- and under-confidence in math predict gaps in general academic confidence and career confidence in adulthood (Table 6, columns 5-8). Generalized academic confidence captures respondents’ beliefs in their skill at solving problems, thinking logically, listening, and teaching others, and career confidence captures respondents’ belief that they can attain and succeed in their dream job. Children who are over-confident in math score about 0.09sd (se = 0.05sd) higher in adult academic confidence and 0.11sd (se = 0.05sd) higher in adult career confidence than peers with comparable childhood test scores. Similarly, a 1sd increase in childhood math confidence predicts a 0.04sd (se = 0.02) increase in adult academic confidence and a 0.06sd (se = 0.02) increase in adult career confidence. While it is unsurprising that adult math, academic, and career confidence are correlated, it is reassuring that the links between childhood and adult math confidence are several times larger than those between childhood math beliefs and these other forms of adult confidence.

On the other hand, there are no significant relationships between childhood math confidence and a measure of adult general confidence (Table 6, columns 9-10), which captures respondents’ conviction in their ability to lead and supervise, their independence and decisiveness, and their life’s direction. Since these regressions control for childhood and adolescent general confidence, they suggest that while general confidence correlates with math confidence in childhood, childhood math confidence is not significantly linked with the evolution of general confidence as respondents age.40
In sum, childhood over- and under-confidence in math persist through childhood and into young adulthood as confidence gaps across academic domains and in one’s career. If these biased beliefs directly affect respondents’ educational or employment success in adulthood, this persistence may be a key factor in the long-term economic associations that we observe.

**VIII.B. Gaps in intermediate outcomes do not fully explain results**

Despite the persistence of childhood confidence, the links we observe between childhood biased beliefs and later-life outcomes could still be fully explained by gaps in intermediate educational investments. In Figure 2, we explore the role of past investment by estimating the marginal relationships between childhood biased beliefs and later-life outcomes, conditional on all intermediate, observable outcomes along the chronological chain of education and entry into the labor market. We then compare these results to those from our baseline specification. If childhood biased beliefs continue to predict long-run gaps conditional on intermediate outcomes, these remaining gaps may be related to contemporaneous adult confidence. Of course, this analysis is imperfect, especially since we cannot control for all intermediate investments.

Figure 2 reproduces our baseline estimates (Tables 3, 4, and 5, even-numbered columns) for math over- and under-confidence in darker blue, while the lighter blue points present our estimates with controls for all outcomes that precede the outcome of interest. In particular, we re-examine educational outcomes through college holding fixed adolescent math and reading test scores, re-examine having a graduate degree and occupation choice holding fixed all previously observed educational outcomes, and re-examine log earnings and unemployment history with controls for all educational outcomes and past occupation choices.

Many of the large confidence gaps we’ve observed in educational and employment outcomes persist when we condition on observable intermediate outcomes. Controlling for adolescent
academic achievement does not change the relationship between childhood biased beliefs and any of our educational outcomes, and under-confidence remains half as predictive of working in STEM when we control for all educational outcomes, including whether respondents majored in STEM. Gaps in respondents’ earnings fall by up to 60 percent when we condition on intermediate outcomes, though our standard errors remain large. The unemployment coefficients are largely unaffected when we add intermediate outcomes as controls.

Together with the persistence of math confidence into adulthood, these results suggest that over- and under-confidence may continue to directly affect economic outcomes as respondents age.

IX. Conclusion

In this paper, we identify over- and under-confidence in math among a large sample of children. In doing so, we are the first to show that even children have markedly biased beliefs about their own math ability. These beliefs are distinct from Big-Five personality traits and general confidence. Girls are less confident in math than boys with the same test scores and general confidence, so gender stereotypes about math may shape ability perceptions even at young ages.

We then estimate striking associations between respondents’ childhood over- and under-confidence in math and their educational and employment outcomes up to 22 years later, including comprehensive controls for children’s demonstrated ability and family backgrounds. In the near term, under-confident children perform worse on the CDS math tests five years later, while over-confident children score higher. In the longer term, childhood math confidence significantly predicts key aspects of later education and work trajectories: whether respondents graduate from high school and college, their college major and occupation choices, their earnings, and whether they experience
unemployment. We do not observe similar associations with long-run outcomes for childhood confidence in reading, a puzzle that we leave for future work.

Our results suggest that biased beliefs about math ability in childhood may predict later-life outcomes both through accumulated differences in educational investments and by continuing to affect economic outcomes as respondents age. Childhood over- and under-confidence persist into adolescence and adulthood, and childhood confidence continues to broadly predict later-life outcomes, particularly in education, when we control for all observable educational and career investments along the chronological chain of education and labor-market entry.

While our results are not causal, they suggest that confidence in math may crucially shape the education we achieve and jobs we get, with effects possibly taking root as early as childhood. Our results provide key early evidence on the importance of math confidence, but they leave substantial room for future exploration. Besides re-examining the associations we estimate for math over- and under-confidence in an experimental setting, research should explore the mechanisms by which childhood math confidence affects later-life outcomes. For example, do less confident children perform worse later because they get less encouragement from teachers, or do they simply choose to exert less effort at school? Next, we’ve seen that high-achievers with low confidence are less likely to work in STEM jobs; do they fare worse in job interviews for those positions, or do they simply not apply? Finally, if future research verifies that confidence causally affects later-life outcomes, what interventions can close those gaps?
References


Fahle, Erin M., Belen Chavez, Demitra Kalogrides, Benjamin R. Shear, Sean F. Reardon, and Andrew D. Ho. 2021. “Stanford Education Data Archive: Technical Documentation (Version 4.1).”


Table 1: The persistence of math over- and under-confidence

<table>
<thead>
<tr>
<th>Panel A. Math over-confidence</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.118***</td>
<td>0.116***</td>
</tr>
<tr>
<td>(N)</td>
<td>(0.031)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>N</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td>Sample mean</td>
<td>0.041</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B. Math under-confidence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.042*</td>
<td>0.042*</td>
</tr>
<tr>
<td>(N)</td>
<td>(0.024)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>N</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td>Sample mean</td>
<td>0.063</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Math degrees of confidence</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.182***</td>
<td>0.185***</td>
</tr>
<tr>
<td>(N)</td>
<td>(0.025)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>N</td>
<td>1747</td>
<td></td>
</tr>
<tr>
<td>Sample mean</td>
<td>0.000</td>
<td></td>
</tr>
</tbody>
</table>

Basic controls: ✓ ✓
Added background controls: ✓

Notes: This table regresses adolescent confidence outcomes on childhood math confidence and controls. Adolescent confidence is measured five years after childhood confidence. In each row, the dependent variable is the adolescent measurement of the independent variable described. Panels A and B show persistence in binary over- and under-confidence, respectively, and Panel C shows persistence in degrees of confidence, standardized to have mean 0 and standard deviation 1. All controls that are time-variant are observed in the same year as the confidence measures. Basic controls include child gender, race, decile fixed effects for math and reading test percentile scores, digit span test scores, a general confidence index, family taxable income and its square, parent education, quarter-of-birth fixed effects, year-of-birth fixed effects, age at which confidence was measured fixed effects, and state fixed effects. We also include fixed effects for adolescent test score deciles in math and reading. Added background controls are parents’ rating of child health, indicators for receiving government transfers, household structure, parenting practices, parent occupation, and parent mental health and confidence. We include missing indicators and recode missing values to zero for all controls. Standard errors are clustered by family and are shown in parentheses below each estimate. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 percent level, respectively.
Table 2: Demographic predictors of over- and under-confidence

<table>
<thead>
<tr>
<th>Demographic Characteristics</th>
<th>Over-confidence</th>
<th>Under-confidence</th>
<th>Degrees of confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>-0.027***</td>
<td>0.023**</td>
<td>-0.097***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Black</td>
<td>0.014</td>
<td>0.015</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.038*</td>
<td>0.019</td>
<td>-0.037</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Asian or Native American</td>
<td>-0.021</td>
<td>0.024</td>
<td>-0.042</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.06)</td>
</tr>
<tr>
<td>Only child</td>
<td>0.005</td>
<td>0.016</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>First child</td>
<td>0.015</td>
<td>0.038**</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Second child</td>
<td>0.029*</td>
<td>0.004</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Father graduated high school</td>
<td>-0.011</td>
<td>-0.039</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Father has bachelors</td>
<td>0.014</td>
<td>-0.009</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Mother graduated high school</td>
<td>-0.019</td>
<td>-0.010</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mother has bachelors</td>
<td>-0.007</td>
<td>-0.025</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Father works in STEM</td>
<td>0.004</td>
<td>-0.036</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mother works in STEM</td>
<td>-0.010</td>
<td>-0.006</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Father works in non-STEM high-educ</td>
<td>0.000</td>
<td>-0.008</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.03)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Mother works in non-STEM high-educ</td>
<td>-0.017</td>
<td>0.019</td>
<td>-0.062*</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Family taxable income (thous 2016 USD)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.00)</td>
<td>(0.00)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Other Child Characteristics</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child ever in gifted prog</td>
<td>0.026**</td>
<td>-0.087***</td>
<td>0.146***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Child ever in special ed prog</td>
<td>0.007</td>
<td>-0.008</td>
<td>0.071</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Child has repeated grade</td>
<td>-0.016</td>
<td>-0.012</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Parent’s rating of child health</td>
<td>-0.001</td>
<td>-0.013**</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>
Table 2: Demographic predictors of over- and under-confidence (continued)

<table>
<thead>
<tr>
<th>School Quality Measures</th>
<th>Over-confidence</th>
<th>Under-confidence</th>
<th>Degrees of confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent FRPL</td>
<td>-0.027</td>
<td>0.058*</td>
<td>-0.074 (0.08)</td>
</tr>
<tr>
<td>Student-teacher ratio</td>
<td>0.000</td>
<td>-0.000</td>
<td>0.002** (0.00)</td>
</tr>
<tr>
<td>Average math and reading achievement</td>
<td>-0.003</td>
<td>0.003</td>
<td>-0.015 (0.01)</td>
</tr>
<tr>
<td>Difference btwn math and reading achievement</td>
<td>0.010</td>
<td>-0.004</td>
<td>0.030 (0.01)</td>
</tr>
<tr>
<td>Cohort slope of average achievement</td>
<td>-0.036</td>
<td>0.079</td>
<td>-0.228 (0.04)</td>
</tr>
<tr>
<td>Unable to link to NCES id</td>
<td>0.049*</td>
<td>0.013</td>
<td>0.109 (0.07)</td>
</tr>
</tbody>
</table>

| Other Child Ability Measures                 |                  |                  |                      |
| Digit span score                             | -0.000           | -0.004**         | 0.009** (0.00)       |
| General confidence                           | 0.038***         | -0.054***        | 0.211*** (0.02)      |
| Mean of dependent variable                   | 0.085            | 0.121            | 0.000                |

N: 2985  R-squared: 0.21  0.21  0.57

Notes: Each column regresses a measure of childhood biased beliefs in math on child characteristics. In columns 1 and 2, the dependent variables are our main indicators for over-confidence or under-confidence, respectively. In column 3, the dependent variable is the more continuous degrees of confidence, standardized to have mean 0 and standard deviation 1. All variables are taken from the first year in which we observe children’s confidence in math. Additional controls include fixed effects for math and reading test score deciles, birth year, birth quarter, state, and age at which confidence was measured. The coefficients on the test score deciles are shown in Appendix Table A11. We include missing indicators (not shown) and recode missing values to zero for all controls. All variables are either continuous or binary indicators, except for child race and birth order. The omitted category for race is non-Hispanic whites, and the omitted category for birth order is any birth order higher than two. Standard errors are clustered by family and are shown in parentheses below each estimate. *, **, and *** indicate significance at the 10, 5, and 1 percent level, respectively.
Table 3: Childhood math confidence and medium-term educational achievement and attainment

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Adolescent math scores</th>
<th>Adolescent reading scores</th>
<th>High school degree</th>
<th>College degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Over-confidence</td>
<td>2.637*</td>
<td>2.666*</td>
<td>-0.362</td>
<td>-0.286</td>
</tr>
<tr>
<td></td>
<td>(1.468)</td>
<td>(1.496)</td>
<td>(1.381)</td>
<td>(1.385)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-5.705***</td>
<td>-5.860***</td>
<td>0.353</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td>(1.482)</td>
<td>(1.497)</td>
<td>(1.439)</td>
<td>(1.452)</td>
</tr>
<tr>
<td>N</td>
<td>1747</td>
<td>1747</td>
<td>1745</td>
<td>1745</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.147</td>
<td>0.138</td>
<td>0.997</td>
<td>0.951</td>
</tr>
</tbody>
</table>

Panel A: Independent variables are binary measures of over- and under-confidence

Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units

| Confidence          | 2.806***                | 2.827***                  | 0.111             | 0.128         |
|                     | (0.566)                 | (0.569)                   | (0.587)           | (0.580)       |
| N                   | 1747                    | 1747                      | 1745              | 1745          |
| Sample mean of dep. var. | 50.808               | 48.231                    | 0.876             | 0.297         |

Basic controls: ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
Added background controls: ✓ ✓ ✓ ✓ ✓ ✓ ✓

Notes: This table regresses educational achievement and attainment outcomes on childhood biased beliefs and controls. Biased beliefs are measured in the earliest CDS wave in which a child has non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence, and basic controls (in odd-numbered columns) or extended controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on the more continuous measure of degrees of confidence, standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Standard errors are clustered at the family level and shown in parentheses below each estimate. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 percent level, respectively.
Table 4: Childhood math confidence and college quality, college major choice, and post-college schooling

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>College quality index</th>
<th>College's 75th pctile math SAT score</th>
<th>STEM major</th>
<th>Graduate degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Over-confidence</td>
<td>0.067</td>
<td>0.037</td>
<td>6.244</td>
<td>3.133</td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td>(0.148)</td>
<td>(12.112)</td>
<td>(11.829)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-0.095</td>
<td>-0.127</td>
<td>-9.312</td>
<td>-11.312*</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td>(0.082)</td>
<td>(5.958)</td>
<td>(5.925)</td>
</tr>
<tr>
<td>N</td>
<td>1107</td>
<td>1107</td>
<td>1117</td>
<td>1117</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.866</td>
<td>0.601</td>
<td>0.819</td>
<td>0.537</td>
</tr>
</tbody>
</table>

Panel A: Independent variables are binary measures of over- and under-confidence

| Confidence          | 0.044                 | 0.041                               | 4.198      | 3.631          | 0.077***     | 0.078***   | 0.023        | 0.022        |
|                     | (0.047)               | (0.046)                             | (3.460)    | (3.417)        | (0.023)      | (0.023)    | (0.025)      | (0.025)      |
| N                   | 1107                  | 1107                                | 1117       | 1117           | 736          | 736        | 810          | 810          |
| Sample mean of dep. var. | 0.053              | 594.172                             | 0.189      | 0.200          |             |            |              |              |
| Basic controls:     | ✓                     | ✓                                   | ✓          | ✓              | ✓            |            | ✓            | ✓            |
| Added background controls: | ✓                   | ✓                                   | ✓          | ✓              | ✓            |            | ✓            | ✓            |

Notes: This table regresses college outcomes on childhood biased beliefs and controls. Biased beliefs are measured in the earliest CDS wave in which a child has non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence, and basic controls (in odd-numbered columns) or extended controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on the more continuous measure of degrees of confidence, standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Standard errors are clustered at the family level and shown in parentheses below each estimate. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 percent level, respectively.
Table 5: Childhood math confidence and employment outcomes

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Works in STEM (1)</th>
<th>Non-STEM high-educ occ. (2)</th>
<th>Ln(earnings) (3)</th>
<th>Ln(earnings) (4)</th>
<th>Ln(earnings) (5)</th>
<th>Ln(earnings) (6)</th>
<th>Ln(earnings) (7)</th>
<th>Ln(earnings) (8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over-confidence</td>
<td>0.011 (0.016)</td>
<td>-0.019 (0.025)</td>
<td>0.045 (0.085)</td>
<td>-0.034 (0.030)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.014 (0.017)</td>
<td>-0.025 (0.026)</td>
<td>0.064 (0.085)</td>
<td>-0.035 (0.030)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-0.049*** (0.016)</td>
<td>0.029 (0.033)</td>
<td>-0.067 (0.056)</td>
<td>0.006 (0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.049*** (0.016)</td>
<td>0.026 (0.033)</td>
<td>-0.075 (0.057)</td>
<td>0.005 (0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>4592</td>
<td>4592</td>
<td>4592</td>
<td>4592</td>
<td>4423</td>
<td>4423</td>
<td>4975</td>
<td>4975</td>
</tr>
<tr>
<td>OC = -1*UC? p-value:</td>
<td>0.096</td>
<td>0.127</td>
<td>0.822</td>
<td>0.987</td>
<td>0.833</td>
<td>0.917</td>
<td>0.437</td>
<td>0.395</td>
</tr>
</tbody>
</table>

Panel A: Independent variables are binary measures of over- and under-confidence

Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units

| Confidence          | 0.018*** (0.006) | -0.001 (0.011)              | 0.049* (0.028)   | -0.023** (0.009) |                 |                 |                 |                 |
|                     | 0.018*** (0.006) | 0.001 (0.012)               | 0.059** (0.029)  | -0.023** (0.009) |                 |                 |                 |                 |
| N                   | 4592              | 4592                        | 4592             | 4592             | 4423            | 4423            | 4975            | 4975            |
| Sample mean of dep. var. | 0.046             | 0.163                       | 10.185           | 0.167            |                 |                 |                 |                 |

Basic controls: ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
Added background controls: ✓ ✓ ✓ ✓ ✓

Notes: This table regresses employment outcomes on childhood biased beliefs and controls. Biased beliefs are measured in the earliest CDS wave in which a child has non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence, and basic controls (in odd-numbered columns) or extended controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on the more continuous measure of degrees of confidence, standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1, minus the controls for adolescent test score deciles. Basic controls also include year fixed effects when the outcome is observed in a panel. Standard errors are clustered at the family level and shown in parentheses below each estimate. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 percent level, respectively.
Table 6: Childhood math confidence and young adult confidence outcomes

<table>
<thead>
<tr>
<th>Dependent variable:</th>
<th>Math confidence</th>
<th>Reading confidence</th>
<th>Academic confidence</th>
<th>Career confidence</th>
<th>General confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>Over-confidence</td>
<td>0.264***</td>
<td>0.260***</td>
<td>0.002</td>
<td>-0.002</td>
<td>0.090*</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.058)</td>
<td>(0.067)</td>
<td>(0.067)</td>
<td>(0.047)</td>
</tr>
<tr>
<td>Under-confidence</td>
<td>-0.250***</td>
<td>-0.251***</td>
<td>0.159***</td>
<td>0.163***</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.048)</td>
<td>(0.053)</td>
<td>(0.054)</td>
<td>(0.033)</td>
</tr>
<tr>
<td>N</td>
<td>6632</td>
<td>6632</td>
<td>6634</td>
<td>6634</td>
<td>8096</td>
</tr>
<tr>
<td>OC = -1*UC?</td>
<td>0.850</td>
<td>0.863</td>
<td>0.064</td>
<td>0.057</td>
<td>0.362</td>
</tr>
<tr>
<td>p-value:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.473</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.418</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.268</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.272</td>
</tr>
</tbody>
</table>

Panel A: Independent variables are binary measures of over- and under-confidence

Panel B: Independent variable is degrees of over- and under-confidence in standard deviation units

| Confidence          | 0.167***        | 0.168***           | -0.046*             | -0.048*          | 0.037**            |
|                     | (0.021)         | (0.021)            | (0.026)             | (0.026)          | (0.017)            |
| N                   | 6632            | 6632               | 6634                | 6634             | 8096               |
| Sample mean of dep. var. | -0.000        | -0.000             | -0.000              | 0.009            | 0.000              |
| Basic controls:     | ✓               | ✓                  | ✓                   | ✓                | ✓                  |
| Added background controls: | ✓             | ✓                  | ✓                   | ✓                | ✓                  |

Notes: This table regresses confidence in young adulthood on childhood biased beliefs and controls. Biased beliefs are measured in the earliest CDS wave in which a child has non-missing test scores and self-assessed ability. In Panel A, the outcome is regressed on an indicator for over-confidence, an indicator for under-confidence, and basic controls (in odd-numbered columns) or extended controls (in even-numbered columns). The p-value listed tests whether the coefficient on the over-confidence indicator is equal to -1 times the coefficient on the under-confidence indicator. In Panel B, the outcome is regressed on the more continuous measure of degrees of confidence, standardized to have mean zero and standard deviation one in our sample, and the same sets of controls. All controls are the same as described in Table 1. Basic controls also include year fixed effects when the outcome is observed in a panel. In this table, we add controls for adolescent test score deciles in math and reading, as well as adolescent general confidence and digit span scores in all specifications. Standard errors are clustered at the family level and shown in parentheses below each estimate. *, **, and *** indicate significance at the 0.1, 0.05, and 0.01 percent level, respectively.
Panel A: Demonstrated ability on CDS test

Panel B: Self-assessed ability

Panel C: Demonstrated vs. self-assessed ability

Figure 1. Distributions of self-assessed and demonstrated ability

Notes: We plot first-observed math test scores and self-assessments for the 2985 CDS respondents with at least one year of both measurements. Panel A plots the distribution of respondents’ percentile ranks relative to a national norming sample on a WJ-R math test. Panel B plots the distribution of children's responses when asked to answer “How good at math are you?” on a scale from 1 (not at all good) to 7 (very good). Finally, Panel C plots the average math percentile rank within each category from 1 to 7 of children's self-reported ability in math.
Figure 2. Main results controlling for intermediate outcomes

Notes: This figure plots the coefficients on over-confidence, under-confidence, and degrees of confidence in our baseline specification (2) and the same coefficients when we add controls for mediating factors to these specifications. When the outcome is high school or college graduation, majoring in STEM, or college quality, we control for adolescent math and reading test scores. When the outcome is earning a graduate degree, we control for all education outcomes through college. When the outcome is occupation choice, we control for all observed educational outcomes. When the outcome is earnings or unemployment, we control for all observed educational outcomes and occupational choice.
We refer throughout the paper to *ability* and *beliefs about ability*, but we do not mean to imply that ability or beliefs are innate or fixed. Rather, we are referring to someone’s ability or perceived ability to perform well in a certain domain or task at a particular time.

We report all of the following analysis for parallel measures of reading over- and under-confidence in Appendix Tables A2-A5. We discuss our focus on math confidence in Section II.

Using weights that adjust our sample to be nationally representative, these ranges are 6-30 percent and 6-15 percent, respectively.

While these four pieces of evidence strongly suggest that our measures of over- and under-confidence capture more than *random* measurement error on the cognitive test, they do not negate the possibility that children have private information on a form of math ability that the test systematically excludes. We discuss this possibility in detail in Section III.C.

The PSID child assessments do not include standard psychometric scales for the Big Five, so we construct proxies for these traits using parents’ reports of children’s behavior and personality. See Appendices B and E for details.

Since girls are more likely to be under-confident in math and less likely to be over-confident, these associations could help to explain key gender gaps in the labor market. Unfortunately, our results are too imprecise for us to conclude whether controlling for biased beliefs in math reduces the gender gaps in adolescent test scores, majoring or working in STEM, or earnings.

Psychology research on academic confidence studies how these beliefs develop as children age (Eccles et al. 1984) and depend on social constructs like gender and race (Herbert and Stipek 2005; Usher and Pajares 2006). This work relies on self-reported psychometric scales and does not compare self-reported ability to a measure of objective ability, as we do in this paper.

Contemporaneous work by Anaya, Stafford, and Zamarro (2021) uses the same data from the PSID and its child supplements to examine the relationship between majoring in STEM and early childhood achievement, self-assessed ability, and parent occupation, though they focus on including parent occupation as a novel explanatory variable in this regression. Like theirs, our main specifications include indicators for whether children’s parents work in STEM, but adding these controls does not change our results. Anaya, Stafford, and Zamarro also describe similar gender gaps in ability beliefs to those we document, but they do not specifically study over- and under-confidence or their relationships with long-term outcomes. In addition to this difference in our central research questions, we see our work as building on theirs in three ways: (1) We use a more comprehensive set of available data from the PSID and its child supplements; (2) We consider a larger set of outcomes observed over a much longer time frame; and (3) We define several new measures of over- and under-confidence to deal with complications with the raw data, an issue that Anaya, Stafford, and Zamarro do not discuss.

In both the full CDS sample and our final analysis sample, 53 (38) percent of children are descended from the SRC (SEO) sample and 9 percent of children are from the immigrant sample added to the PSID in 1997.

Most children who are missing test scores or self-assessments lack this data because they skipped the entire section of the CDS administered to the child, while completing the survey portions administered to the primary caregiver. These respondents largely have similar demographics to those for whom we observe confidence measures, but their mothers are less likely to have a high school degree, they have lower total family income, and they are about a
year younger. Students who take the math cognitive assessments but do not give self-assessments (about 25 percent of the children who are missing test scores or self-assessments) score much lower on both the math assessment and the Digit Span memory test (Appendix Table A17).

Since the error terms in our regressions are unrelated to the sampling criterion, conditional on our extensive controls for family income, race, and other characteristics, weighting may not improve the estimator’s consistency and may reduce its precision (Solon, Haider, and Wooldridge 2015). These recalibrated weights put less weight on children with high CDS math scores, in some cases leaving us underpowered to detect the correlations between under-confidence and long-term outcomes. A natural concern is that our unweighted regressions may estimate non-representative partial relationships between confidence and outcomes, if these associations are heterogeneous by race or income. However, weighted least squares estimates are not necessarily closer to the true population average partial relationship than ordinary least squares estimates (Solon, Haider, and Wooldridge 2015). Instead, we directly estimate heterogeneity by characteristics related to the sampling scheme, like family income, race, age, and being in an SRC-sample family. These results are imprecise and show no robust patterns of heterogeneity (results available upon request).

The 1997 CDS wave also included 58 WJ-R questions on calculation skills, and we use this test in the next section to assess the reliability of our over- and under-confidence measures.

We first observe confidence from the 1997 CDS wave for 1,075 children, from the 2002 CDS wave for 1,347 children, and from the 2007 CDS wave for 563 children.

For example, upward response bias could arise based on children’s interpretation of the qualitative labels on the scale (“Not at all good” at 1, “Okay” at 4, and “Very good” at 7) if, for example, they think that nearly everyone is at least “Okay” at math. Upward skew could also arise if self- or social-image concerns make children unwilling to tell a surveyor that they are worse than “Okay” at math. If, on the other hand, this upward skew does reflect true aggregate over-confidence, our estimates for long-term associations with over-confidence would simply reflect links with particularly over-confident beliefs in math.

Several studies find test-retest reliability of about 0.75 for certain ages, though these studies use small samples (Shull-Senn et al. 1995).

We find similar reliability using our more continuous measure of degrees of confidence, which takes on integer values from -6 to 6. There, 32 percent of children are assigned the same value regardless of which math test we use as the measure of demonstrated ability, 62 percent are within one integer, and 83 percent are within two integers. See Appendix Figure A3 for the full joint distribution of the more continuous confidence measures based on the two math subtests.

The regressions in Table 1 add controls for children’s adolescent test score deciles in math and reading to our main specification. We add these controls to purge any correlations induced by the effects of childhood confidence on adolescent test scores, since childhood over- and under-confidence predict later test scores (see Section VI) and higher-scoring (lower-scoring) children are mechanically more likely to be classed as over-confident (under-confident).

While our main measure of under-confidence persists only weakly into adolescence, several alternate definitions of under-confidence are strongly persistent (Appendix Figure A4). Our main measure’s limited persistence may relate to the fact that adolescent test scores are much less upward-skewed than childhood test scores, so fewer respondents can be classified as
under-confident in adolescence. The more persistent alternate definitions of under-confidence, in contrast, identify under-confidence among respondents with a wider set of test scores and thus are less affected by this distributional shift. Like our main measure, these alternate measures predict substantial gaps in long-run outcomes (Appendix Figures A5-A16).

19 Despite this benefit, we do not use these averages as our preferred measures of confidence for three reasons: (i) over- and under-confidence at older ages may be more likely to be confounded by unobserved variables; (ii) we are interested in adolescent test scores and confidence measures as outcome variables; and (iii) only 60 percent of our sample has confidence measurements over multiple waves of the CDS.

20 Differential random error by gender could not fully explain the gendered patterns of over- and under-confidence we observe, since the gender with more variable performance would be more likely to be both over- and under-confident. Nonetheless, comparing boys’ and girls’ performance on the Calculations and Applied Problems subtest in the 1997 CDS sample suggests that neither gender has differentially variable test performance. 81% of both boys and girls receive the same binary confidence designation when calculated using either the Calculations or Applied Problems percentile score as a measure of math skill, and the joint distributions of the more continuous measures are very similar for boys and girls (Appendix Figure A3).

21 We estimate the following regression: \( CALC_{ptilei} = \beta_0 + \beta_1 AP_{ptilei} + \beta_2 Female_{i} + \beta_3 AP_{ptilei} \times Female_{i} + \epsilon_{i} \). Coefficient \( \beta_3 \) is not significantly distinguishable from zero, and \( \beta_2 \) is significant and positive. Thus, girls have stronger calculation skills than boys conditional on their Applied Problems scores, which would tend to make girls look more over-confident by our measures, the opposite of what we find.

22 We find similar results when applying our raked weights to obtain nationally representative estimates: 9.2 percent of children are over-confident and 9.8 percent are under-confident.

23 Math over- and under-confidence also correlate with children’s other attitudes towards math and school in reasonable ways (Appendix Table A18), suggesting that our measures isolate over- and under-confidence in the particular domain of math. See Appendix C for more discussion.

24 In fact, there is no gender gap in the likelihood of having accurate or almost accurate beliefs (degrees of over- and under-confidence equal to zero, or between -1 and 1, respectively). Results are available upon request.

25 Appendix Figure A17 shows that this gender gap is extremely robust to using alternate definitions of over- and under-confidence and alternate ways of calculating the more continuous degrees-of-confidence measure. This figure plots the coefficient on the female indicator when we exchange the dependent variables in Table 2 with these alternate measures (discussed further in Section 7).

26 We also test whether childhood gender gaps in math under-confidence explain gender gaps in later education and employment outcomes: adolescent test scores, majoring in STEM, and earnings. Specifically, we estimate the change in the coefficient on gender when we estimate our preferred specification with and without the indicator for under-confidence (following Buser, Niederle, and Oosterbeek (2021)). The results (available upon request) are quite noisy, so we leave it to future research to determine whether confidence gaps in math can help explain these and other gender gaps.
Appendix Figures A18, A19, and A20 show our main results when we relax this assumption. We plot the coefficients on indicators for each integer value of the variable underlying $Z_{Conf_{z0}}$: $Conf_{z0} = -6$, $Conf_{z0} = -5$, ..., $Conf_{z0} = 6$. While these results are noisy, the point estimates suggest that this linearity assumption is reasonable. We also show in Appendix D that we cannot generally reject the null hypothesis that over- and under-confidence predict economic outcomes in similar (opposite-signed) ways, further supporting this linearity assumption.

One might worry that controlling for general confidence absorbs too much of the variation in math over- and under-confidence if over- and under-confidence in math are dimensions of confidence in general. While the economic impacts of general confidence are certainly of interest, we take the conservative approach of isolating math-specific over- and under-confidence as cleanly as possible by controlling for general confidence. That said, our results are remarkably similar with or without the control for general confidence (available upon request).

For some outcome variables, like earnings and unemployment, we have multiple years of outcomes across TAS and PSID waves for each respondent. In contrast, we observe our educational outcomes (for example, whether respondents ever majored in STEM) only once per respondent; we do not include survey wave fixed effects in regressions linking childhood confidence to these outcomes. Note that we do not include respondent fixed effects even in regressions with multiple outcome observations per respondent, since we only measure childhood confidence once. We cluster standard errors by family in all regressions.

Age at which confidence is first observed and birth year are not collinear. For example, children who had their confidence measured when they were 8 years old could have been born in 1989, 1995, or 1999 (and had their confidence measured in the 1997, 2003, or 2007 CDS, respectively).

While we estimate the relationships between confidence and later outcomes in regressions with test score decile fixed effects, here we run otherwise identical regressions replacing these fixed effects with linear controls for test scores. We benchmark the links between biased beliefs against the coefficients on these linear score controls throughout Section VI.

Math scores do significantly predict high school graduation, but the coefficient is precisely estimated and very small (increasing test scores by 10 percentiles is associated with being 0.8pp more likely to graduate from high school). The magnitude of this linear coefficient is half of the size of the coefficient on reading scores. Reading skills may be more important for high school graduation than math because, for example, fewer years of math study are required to graduate.

Using restricted data from the TAS, we link respondents with college quality data from the National Center for Education Statistics (NCES) for the first college they attended in the first year they attended that college. Following Cohodes and Goodman (2014), we construct an index of college quality as the first component from a principal component analysis of colleges’ 75th-percentile math SAT scores among incoming freshmen, graduation rates, and per-pupil instructional expenditures, separately by year. Details on variable construction are available in Appendix B. We then standardize this index to have mean 0 and standard deviation 1 in the full sample of four-year colleges in the US by year.

We define STEM fields as engineering, math and computer sciences, and natural sciences. We find similar results if we also include health fields.

Appendix Table A19 replicates these results using one observation per child, where the dependent variable is calculated as the average outcome observed over ages 28-33. The results are meaningfully the same.
We define STEM fields to include computer and mathematical occupations, architecture and engineering occupations, and life, physical, and social-science occupations. We find similar results if we include healthcare occupations as STEM.

We define non-STEM high-education occupations as management, business, and financial occupations, legal occupations, education, training, and library occupations, and occupations that focus on writing and communication (a subset of media, arts, and entertainment occupations). We exclude health fields, as they are STEM-adjacent.

Besides testing alternate confidence measures, the specification charts also show that our main results are robust to dropping children in the lowest and highest math score deciles from our sample. Across all confidence measures, children at the upper (lower) tail of the score distribution are mechanically most likely to be identified as under-confident (over-confident).

Unlike our measures of biased beliefs from the CDS, these TAS confidence variables are not paired with measures of demonstrated ability in adulthood. However, the ideal regressions would test the links between childhood over- and under-confidence and biases in adult confidence, so as to avoid conflating the persistence of biased beliefs with the links between childhood confidence and adult achievement. We approximate this ideal by controlling for adolescent math and reading scores, digit span scores, and general confidence as proxies for adult ability.

As additional evidence that our results capture links with math confidence, not general self-esteem or ability, we consider a set of placebo outcomes: individuals’ relationship status, general mental health, social anxiety, alcohol consumption, and dangerous behavior as young adults (all from the TAS). We expect each of these outcomes to be affected by general self-esteem, but not by math over- and under-confidence specifically. Reassuringly, we generally find no relationships between biased beliefs in math and any of these placebo outcomes, except that math over-confidence predicts a lower likelihood of being in a romantic relationship (Appendix Table A21).