Helsinki GSE

HELSINKI GSE DISCUSSION PAPERS 39 · 2025

Childhood Gender Nonconformity and Gender Gaps in Life Outcomes

Abigail R. Banan
Torsten Santavirta
Miguel Sarzosa

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HANKEN

Aalto University

Helsinki GSE Discussion Papers

Helsinki GSE Discussion Papers 39 · 2025

Abigail R. Banan, Torsten Santavirta, and Miguel Sarzosa
Childhood Gender Nonconformity and Gender Gaps in Life Outcomes

ISBN 978-952-7543-38-2 (PDF)
ISSN 2954-1492

Helsinki GSE Discussion Papers:
https://www.helsinkigse.fi/discussion-papers

Helsinki Graduate School of Economics
PO BOX 21210
FI-00076 AALTO
FINLAND

Helsinki, May 2025

Childhood Gender Nonconformity and Gender Gaps in Life Outcomes*

Abigail R. Banan† Torsten Santavirta‡ Miguel Sarzosa§

March 19, 2025

Abstract

We study the role of childhood gender conformity in generating gender gaps. We present a conceptual framework that uses gender norms to explain why some women make less profitable choices than comparable men. Using unique longitudinal survey and register data, we show that gender-nonconforming girls have substantially better education and labor market outcomes than gender-conforming girls. In contrast, gender-nonconforming boys perform substantially worse at school, sort into lower-paying occupations, earn less, and have a greater incidence of mental health disorders and substance abuse during adulthood than gender-conforming boys. Our analyses suggest that such divergence develops from an early age.

JEL Classification: I21, J13, J15, J16, J24
Keywords: gender nonconformity, gender norms, gender gaps

*This paper was supported by the Handelsbanken Research Foundation grant number P2016-0093 and the Swedish Research Council, Grant No. 2017-00878. We thank Ylva Almquist and Jannike Kjellström at Center for Health Equity Studies, Stockholm University for data provision and excellent support and Carina Mood and Simon Hjalmarsson at the Swedish Institute for Social Research (SOFI) for advice on how to access the auxiliary data. We are grateful to Daniele Paserman, Miriam Gensowski, Pia Pinger, Matti Sarvimäki, Colin Sullivan, and Seth Sanders for helpful comments and suggestions and to participants at the Purdue Labor Lunch (West Lafayette, 2021), the CSQIEP Seminar on Economics (2021), the AEA-ASSA Meetings (2022), the ALP-POP Conference (La Thuile, Italy, 2022), EDW (2022), SoLE Meetings (Minneapolis, 2022), Helsinki GSE lunch seminar (Helsinki, Finland, 2022), ESPE (Arcavacata, Italy, 2022), Purdue Summer Seminar (West Lafayette, 2022), Universidad del Rosario (Bogota, Colombia, 2022), Universidad de los Andes (Bogota, Colombia, 2022), WOLFE (York, UK, 2022), APPAM (DC, 2022), RIDGE (Montevideo, 2023), SEHO (Copenhagen, 2023).

†Department of Economics, Purdue University. 403 W. State St, West Lafayette, IN 47907. Email: abanan@purdue.edu.

‡Swedish Institute for Social Research (SOFI), Stockholm University, Sweden, and Department of Economics, University of Helsinki, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, Finland. Email: torsten.santavirta@helsinki.fi

§Department of Economics, Purdue University. 403 W. State St, West Lafayette, IN 47907. Email: msarzosa@purdue.edu.

1 Introduction

Numerous studies in the social sciences seek to understand why women, despite having surpassed men in terms of school outcomes for the past five decades in most Western countries, make educational and career choices that depress their labor market outcomes (Bertrand, 2011; DiPrete and Buchmann, 2013; Olivetti and Petrongolo, 2016; Bertrand, 2020; Lundberg, 2020). They tend to opt out of math-intensive STEM fields in high school and tertiary education, are underrepresented in the higher-paying STEM occupations (where the gender gap in earnings is relatively small), and bear most of the earnings losses due to childbirth (Goldin, 2014; Kleven et al., 2019). In an effort to solve this puzzle, much recent work on gender economics explores between-gender differences in preferences and the impact of societal gender norms and sticky stereotypes (West and Zimmerman, 1987; Akerlof and Kranton, 2000, 2002; Bertrand, 2011; Zafar, 2013; Wiswall and Zafar, 2014; Bertrand et al., 2015; Nollenberger et al., 2016). Prevailing social norms and gender stereotypes impose certain life choice expectations that differ between the sexes (e.g., women should be modest caretakers, men agentic breadwinners), in turn affecting career decisions, ultimately translating into the observed persistent gender gaps (Bertrand, 2020).

The burgeoning literature that attributes gender gaps to differences in male and female prescriptive stereotypes has thus far employed a binary concept of gender. Such an approach makes sense in that binary sex is a convenient proxy for gender norm identity. Perhaps more importantly, societal prescriptions largely police individuals into one of these two social categories: man or woman (Akerlof and Kranton, 2000). Yet, *within*-gender differences in cognition, traits, and preferences have been shown to be larger than the *between*-gender differences (Hyde, 2014; Hyde et al., 2019). Sizeable within-gender heterogeneity is partly the result of a non-trivial variation in the degree of conformity to societal prescriptions. Closing gender gaps thus necessarily requires a more nuanced understanding of the latter—our paper represents a first step in this direction. We exploit variation in preferences and behaviors during childhood to study how nonconformity to gender norms predicts lifetime outcomes.

The study of gender norms and their impact has become ever more important. Indeed, as Generation Z's formative years draw to a close, an increasing number of young individuals are contesting the prescriptions of binary gender norm identities (Pew Research Center, 2020). Approaches able to capture the heterogeneity in men's and women's outcomes across degrees of conformity to gendered expectations are indubitably needed (Lundberg, 2023). To add to this, gender norms themselves, though shown to be sticky (Alesina et al., 2013; Giuliano, 2018; Baranov et al., 2023), are not set in stone. Research demonstrates that they can change considerably over a generation (Goldin and Katz, 2002; Fernández et al., 2004; Fogli and Veldkamp, 2011; Bursztyn et al., 2020; Miho et al., 2023). Generally, gender norms have become more fluid (Diamond, 2020; Hyde et al., 2019) and the categorical notion of gender has been increasingly challenged across disciplines (Hyde et al., 2019; Yavorsky and Buchmann, 2019; Burn and Martell, 2022; Mittleman, 2022; Hernandez et al., 2024; Hsu, 2024).

We begin by presenting a conceptual framework that uses gender norms to explain why some women make less profitable choices than do men with similar preferences. In our framework, men and women have idiosyncratic endowments of skills and "male typicality" and choose how much "masculinity" to reveal (this can differ from the true endowment, at a cost) based on societal incentives. On the one hand, the market rewards masculinity. On the other, society punishes deviations from the stereotypes. Based on this framework, we derive implications for men and women with high and low innate masculinity levels. We show that gender norms lead to a misallocation of time between marketable and domestic activities.

We underpin these arguments using a Swedish longitudinal data set that links a 1966 survey of 10,154 13-year-olds to administrative registers spanning five decades. We measure gender conformity by re-purposing a battery of survey questions on general preferences and choices. While these do not directly ask about gender conformity, they do generate *empirically* gendered answers—especially given that the respondents were growing up in the 1960s. The questions explore preferences in leisure time interests. As these interests are measured in childhood, many are rather toy and play preferences (e.g., playing with trains, fixing bikes and baking) which literature has documented to be gendered (Adelson, 2012; Davis and Hines, 2020). They relate however to home production, market skills, and action-oriented activities.[1] We also include information on gender homophily and preferred subjects in school. We identify the variation common across all measures to create a one-dimensional con-

[1]The areas of interest map nicely onto two fundamental dimensions of Social Role Theory regarding how gender-typical behavior results from the imposition of different social expectations on men and women: *communion* and *agency*. Men are generally perceived as oriented towards agentic (motivated to master, receive recognition) goals and women towards communal (selfless and concerned with others) goals (Eagly and Steffen, 1984). Further, being good at sports is associated with high social status at school for adolescent boys, while it is less of a status marker for girls (Coleman, 1961). Preferences for interests is one area in which Hyde's (2005) gender similarity hypothesis is rejected by empirical research (Su et al., 2009). As Hyde (2014) notes, these differences are cultural artifacts that are not immutable.

tinuous index of gender typicality. Our metric does not discriminate between girls and boys. Rather, it reveals gender-typical and atypical behaviors and preferences for both genders. We define children with gender-atypical behaviors and preferences as "nonconformers" and document how childhood gender nonconformity (as opposed to conformity with gender-typical preferences associated with one's own gender) relates to later life outcomes. We cover early education outcomes and choices and then follow these individuals throughout life, observing their occupational and fertility decisions, health indicators, labor market success, and lifetime earnings.

Our results indicate that childhood gender nonconformity is strongly associated with life outcomes. Consistent with the implications of our conceptual framework, we observe asymmetric consequences for men and women. In general, childhood gender nonconformity is associated with better life outcomes for women and worse life outcomes for men. Gender-nonconforming women perform better at school, are more likely to continue into higher education, and to choose a STEM track in upper secondary school than are their gender-conforming female peers. Compared to gender-typical women, gender-nonconforming women are furthermore substantially more likely to work full time, sort into STEM occupations, postpone fertility, and earn more, especially if they go into male-dominated fields. Conversely, relative to gender-typical men, male nonconformers perform worse at school, are less likely to choose a STEM track in upper secondary or sort into STEM occupations, and earn roughly 10% less during their working career. To analyze earnings gaps in the context of endogenous occupational choices, we implement a Roy model of self-selection into occupations and counterfactual earnings. We confirm our reduced-form findings indicating that STEM occupations significantly reward gender-nonconformity in women. Gender-nonconforming women in STEM occupations earn 49% more than they do in non-STEM ones. When exploring possible mechanisms, we find that school experiences differ significantly for gender-nonconforming boys and girls. While gender-nonconforming girls enjoy school and befriend smarter peers, gender-nonconforming boys feel unsafe, have little interest in schoolwork and experience isolation. The latter translates to a higher incidence of behavioral problems, mental health issues, and substance abuse during adolescence. Our results are robust to potentially important confounding factors such as cognitive ability, parental socioeconomic status (SES), and having opposite-sex siblings.

The results in our Swedish context of a cohort who grew up in the 1960s highlight the matter that average gender gaps mask significant variation within both sexes. The progress made towards achieving gender equality has not benefited all women uniformly. A subgroup of women, namely those who challenged gender norms early in life, face significantly smaller gender gaps compared to other women. These results are in line with the findings of unpublished parallel work to ours in economics on female childhood gender conformity and schooling and labor market outcomes by Ayyar et al. (2024) and Brenøe et al. (2024) in various settings.[2] As to men, we find that their educational and occupational choices which much of the recent literature considers to be among the key remaining explanations for the gender gaps in labor market outcomes (Bertrand, 2020), seem to be driven by the more gender-conforming men while gender-nonconforming men choose careers with lower monetary returns. This is consistent with what Brenøe et al. (2024) find in the contemporary U.S. context of adult men. For men, along the masculinity dimension alone, De Haas et al. (2024) further show that conformity to masculinity norms correlates positively with economic outcomes such as labor supply. To our knowledge, ours is the only study that measures gender conformity in childhood and is able to observe the study persons from early on in adolescence and throughout their careers across a wide range of surveys and administrative data. We contribute to the literature by exploring the pathways through which gaps in schooling and labor market outcomes between gender non-conformers and their conforming counterparts open up. This, combined with sociometric data for complete classroom rosters and students' self reports on perceptions of the school social environment allows us to carefully describe when and how the documented performance gaps emerge and how they evolve throughout the life course.

We also directly contribute to the young and budding literature of the measurement of gender complexity (Fleming et al., 2017; Brenøe et al., 2022, 2024; Ayyar et al., 2024) by conceptualizing a novel data-driven way of measuring gendered childhood behaviors and preferences for which prescriptions of social gender categories provide salient guidance. In particular, ones that are grounded in Economics and Psychology literature as being associated with feminine and masculine prescriptions (Eagly and Steffen, 1984; Su et al., 2009; Ceci et al., 2014). Measuring gender conformity as early as in sixth grade allows us to follow their outcomes throughout their life while alleviating concerns of reverse causality or justification bias (Black et al., 2017). Another

[2]Somewhat at odds with these results however, two recently published studies on adulthood gender nonconformity using U.S. data find a negative correlation between self-assessed appearance-based gender nonconformity among respondents being in their late 30s and economic status, a result holding for both assigned genders and heterosexuals as well as sexual minorities (Hernandez et al., 2024; Hsu, 2024).

advantage of our preference-based measure in comparison to much of the literature is that its construction abstracts from binary sex; it picks up and consolidates the common variation of preferences across both assigned sexes.

The remainder of this article is structured as follows. Section 2 introduces a conceptual framework that rationalizes our empirical findings based on gender norms. Section 3 describes the data and the construction of our measure of gender conformity. Section 4 presents our regression analysis and the main results. Section 5 explores earnings gender gaps under the prism of endogenous occupational choices. Section 6 discusses potential mechanisms underlying our findings, and Section 7 concludes.

2 Framework

To fix ideas, we begin by presenting a simple model inspired by Akerlof and Kranton (2002) where gender conformity is salient. As in Bertrand (2011), we use gender-based differences in traits, preferences, and social norms, to rationalize why gender nonconformity should be associated with life outcomes. Unlike Akerlof and Kranton (2002), our simple framework does not explore identity choices. Rather, it describes how gender-(non)conforming traits and tastes influence later outcomes. In our model, individual i belongs to one of two categories $\mathbf{C} = \{F, M\}$, men and women, which are fixed. But following Akerlof and Kranton (2002), society holds prescriptions $\mathbf{P}$ for each category. That is, $\mathbf{P}$ collects social norms or, as Bertrand (2020) puts it, prescriptive stereotypes that dictate the expected behaviors for each category.[3] In our model, those norms prescribe that men's (women's) actions and choices should reveal high (low) levels of masculinity. Individuals reveal their level of masculinity through behaviors that society considers to be gender-specific. For instance, that boys (girls) should play with other boys (girls) and spend more time in athletic (domestic) activities.

Each person has two exogenously given traits: ability, n_i, and masculinity m_i. Ability and masculinity are both independently distributed bounded between $[0, 1]$. In particular, people with m_i close to one will find it easier to comply with the societal norms imposed on men. Conversely, people with m_i close to zero will find it easier to comply with the societal norms imposed on women. m_i's distribution could be

[3] Just like most of the literature on gender gaps, we assume that the social norms are exogenous to the individual and in this sense, we abstract from any general equilibrium effects. Endogenous social norms are widely studied in the literature on linear social interaction models (see, e.g., Blume et al., 2015; Boucher and Fortin, 2016; Ushchev and Zenou, 2020).

bimodal, where members of each gender tend to cluster around a different level of m_i. More women's m_i are closer to zero and more men's m_i are closer to one. That is, $\mathbb{E}[m_i|c_i = M] > \mathbb{E}[m_i|c_i = F]$.

Agents split their time endowment T into the production of marketable and domestic goods.[4] Let $\alpha(m_i)$ be the fraction of time agent i devotes to marketable activities $e_i = \alpha(m_i)T$. Then, $1 - \alpha(m_i)$ is the fraction of the time endowment devoted to domestic production, $h_i = (1 - \alpha(m_i))T$. We assume $\alpha'_{\hat{m}} > 0$ to reflect that time allocation, to a significant degree, is the product of societal norms: women are expected to spend more time in domestic activities, men are expected to devote more of their time in marketable ones (Eagly, 1987).[5] Society considers time allocation as a signal of the agent's masculinity endowment and it enforces the prescriptive stereotypes $\mathbf{P}$ by punishing deviations from its ideal of masculine men ($\hat{m}_i = 1$) and feminine women ($\hat{m}_j = 0$) (Moss-Racusin et al., 2010; Bertrand, 2020).[6] We materialize the enforcement mechanism through society levying a punishment $t > 0$ proportional to the deviations from the ideal time allocation dictated by social norms for men and women (Fortin, 2005; Booth and Van Ours, 2009). That is $t(1 - \alpha(\hat{m}))$ for men and $t\alpha(\hat{m})$ for women, where $\hat{m}_i$ refers to the level of masculinity agent i decides to *reveal*—our model's choice variable.

People derive utility from three payoffs. The first one, $w(c_i, m_i)n_i e_i$, is an age-specific payoff from their marketable activities (e.g., higher grades at school while adolescents, earnings when adult), which is increasing in ability n_i and the time devoted to it e_i. The age-specific payoffs rely on an outcome-specific societal reward function, $w(c_i, m_i)$, that varies by gender—includes gender based discrimination—and values masculinity (i.e., $w(M, \bar{m}) > w(F, \bar{m})$ and $w'_m > 0$). The latter characteristic could reflect the differences in traits that have been shown to contribute to gender productivity gaps in favor of men (Bertrand, 2011). For instance, risk aversion, willingness to compete, and time preferences are traits that have been shown to differ on average between men and women (Croson and Gneezy, 2009; Eckel and Grossman, 2008; Dohmen et al., 2010, 2011). These are potentially important determinants of the

[4] To keep interpretation simple, we consider as marketable any activity that is conducive to a public societal reward (e.g., spending time working in the labor market, or investing time in preparing for an exam). We also include indirect activities, e.g., playing with tools during childhood.

[5] Becker (1985), Albanesi and Olivetti (2009), and others leverage those societal norms to explain gender occupational and wage gaps.

[6] This closely relates to the concept of *Backlash effect* in social psychology that recognizes that people pay a social cost when they show gender atypical behaviors (Rudman, 1998; Rudman and Phelan, 2008; Babcock et al., 2017).

payoffs obtained from marketable activities holding skills and effort constant (Buser et al., 2014; Blau and Kahn, 2017; Shurchkov and Eckel, 2018; Cortes et al., 2021). These trait-related regularities are not limited to the labor market. Evidence on test taking behavior, for instance, shows that women tend to follow strategies that are less risky but lead them to misplace effort in questions that have lower marginal reward (Borges et al., 2022), and that women tend to score less than what they should in high-stakes exams (Ors et al., 2013). In adult outcomes like the labor market, $w'_m > 0$ can also reflect the evidence showing that incumbent men prefer a masculine environment in their work place and like to interact with, schmooze, and promote other men instead of women (Cullen and Perez-Truglia, 2021).[7] We assume these masculine traits are more easily deployed by people with high masculinity endowment. That is, although the reward function values revealed masculinity, the reward decreases as the revealed masculinity deviates from the true endowment. Therefore, $w(c_i, \hat{m}_i; \mathbf{P}, m_i) = w(c_i, \hat{m}_i - \frac{\gamma}{2}(\hat{m}_i - m_i)^2)$, where $\gamma > 0$.

The second payoff from which people derive utility is the consumption of a domestic good $G(h_i)$ that is produced devoting time to domestic activities and $G'_h > 0$. The third payoff comes from being true to oneself. That is, choosing to reveal a masculinity level different from the true masculinity endowments is personally costly.[8]

By normalizing $T = 1$, we can write the utility functions for people belonging to each of the two gender categories as:[9]

[7] For instance, in the case of adults $w(M, m_i)n_i e_i$ would entail a gender-specific wage rate that rewards more male dominated occupations than female dominated ones. People with higher ability and those who spend more time in marketable activities will earn more, given a wage rate. For the case of adolescents, consider for instance STEM enrollment as outcome. Then, the payoff entails a reward function that embeds a behavior that the society tends to attach to men (i.e., men are engineers, women are caregivers) (Bertrand, 2020), more skilled children and those who spend more time in STEM-related activities are more likely to enroll in STEM fields.

[8] This is a recurring argument for replacing the gender binary with a more complex conception of gender that stresses diversity. Maintaining the binary is associated with a number of negative consequences of gender stereotyping (Hyde et al., 2019). Further, stereotyping (in this case, gender-typing) may be particularly costly for those not fitting in as they cause belief distortions towards the individuals most representative to the group (gender) (Bordalo et al., 2016). Genicot (2022) incorporates a similar concept of adopting an identity that differs from one's own inherent identity into individual utility and fittingly calls it *compromising*.

[9] As in Albanesi and Olivetti (2009), we could add a term collecting the disutility of producing marketable and domestic goods $V(h, e)$. But as in this case, the individual will always exhaust her time endowment T, disutility $V(h, e)$ will only reflect the benefits of specialization. To see this, consider for instance, $V(h, e) = he$, which we can write as $V(\hat{m}) = (1 - \alpha(\hat{m}))\alpha(\hat{m})T^2$. Thus, the disutility will be maximum if $\alpha(\hat{m}) = 1/2$. That is, when the person splits her time in the production of domestic and marketable good equally. Adding that term does not change the model's predictions.

$$U_i(M; m_i, n_i, \mathbf{P}) = w(M, \hat{m}_i; m_i) n_i \alpha(\hat{m}_i) + \theta G(1 - \alpha(\hat{m}_i)) - \frac{(\hat{m}_i - m_i)^2}{2} - t(1 - \alpha(\hat{m}_i))$$
$$U_j(F; m_j, n_j, \mathbf{P}) = w(F, \hat{m}_j; m_j) n_j \alpha(\hat{m}_j) + \theta G(1 - \alpha(\hat{m}_j)) - \frac{(\hat{m}_j - m_j)^2}{2} - t\alpha(\hat{m}_j)$$

To maximize utility, people choose how much masculinity to reveal. The first order condition with respect to revealed masculinity can be written as:

$$\hat{m}_i = m_i + \frac{n_i \left[\alpha(\hat{m}_i) w'_{\hat{m}}(c_i) + \alpha'_{\hat{m}} w(c_i)\right] + \alpha'_{\hat{m}} \left\{[t - \mathbf{1}(c_i = F) 2t] - \theta G'_h\right\}}{1 + \gamma n_i \alpha(\hat{m}_i) w'_{\hat{m}}(c_i)} \tag{1}$$

where $\mathbf{1}[c_i = F]$ takes on value 1 if i is female and zero otherwise. Equation (1) indicates that optimal revealed masculinity depends positively on true masculinity, but that, as indicated in the social psychology literature, society's normative stereotypes become self-fulfilling as they push people to disclose a masculinity level away from their true endowment and closer to what is expected from their gender category (Bertrand, 2020). First, the societal punishment for deviating from the norms pushes revealed masculinity up for men and down for women. Second, the optimal deviation from true masculinity is mediated by how transferable "faked" masculinity is into the reward function. Low transferability implies a high γ, making the bias smaller by discounting the forces that incentivize the inflation and deflation of the revealed masculinity relative to the true one. Third, optimal revealed masculinity also depends positively on the marginal benefit of increased masculinity in the age-specific payoffs on the marketable activities, which materializes in two ways. The marginal increase in the societal reward function (the more the "market" rewards masculinity, the greater the incentive to reveal a higher masculinity endowment) and the marginal increase in the fraction of time devoted to the marketable activities. This is in line with De Haas et al. (2024) who show using international survey data that tracks adherence to dominance masculinity norms (i.e., norms that promote male dominance at the cost of women and non-conforming men) worldwide that men, who adhere strongly to dominance masculinity norms, supply more labor at the intensive margin and are more competitive . Note that $w'_{\hat{m}}(M, \hat{m})$ need not be equal to $w'_{\hat{m}}(F, \hat{m})$. In fact, a very plausible scenario is one in which $w'_{\hat{m}}(M, \hat{m}) > w'_{\hat{m}}(F, \hat{m})$, where masculinity is valued more in the production of the marketable good among men than among women. Thus, how societies reward masculinity in marketable activities can affect the reported masculinity differently by gender. Finally, optimal revealed masculinity will be lower if there is a higher marginal value of the domestic good.

Equation (1) also shows an important attribute of the model: $\alpha'_{\hat{m}}$, the extent to which society infers masculinity from a marginal increase in e_i, is critical in determining $\hat{m}_i$. A large $\alpha'_{\hat{m}}$ means that large changes in e_i imply small changes in masculinity. Thus, larger $\alpha'_{\hat{m}}$ in equation (1) implies that an agent who wants to inflate their masculinity by a given amount needs to increase the time spent at the marketable activity by more.

The first order condition shows that gender norms produce an interesting asymmetry across genders in the difference between the true and the revealed masculinity. We summarize this asymmetry in Implication 1.

Implication 1 (*Asymmetry*) *If $t > 0$ then $\hat{m}_i \geq m_i$ for $c_i = M$ and $\hat{m}_i \lesseqgtr m_i$ for $c_i = F$.*
Given a sizable enough societal punishment for deviating from the gender norm, men will have an incentive to inflate their masculinity, while women can reveal more or less masculinity than their true endowment depending on the size of the marginal ability-mediated gains relative to the marginal utility produced by the domestic good.

It is easy to see that men with high m_i have a strong incentive to reveal a high $\hat{m}_i$ and enjoy the benefits of higher payoffs, while not paying the cost of deviating from societal norms for men. Men with low m_i face a trade-off. The benefit of being true to their traits (disclose a low $\hat{m}_i$) is countered by the societal punishment t and the losses in the age-specific payoffs on the marketable activities. The higher the societal punishment t and the premium for masculinity are, the higher the disclosed masculinity will be, despite having a low true m_i. In fact, all men will inflate their masculinity unless they have a marginal utility of the domestic good that exceeds the marginal increases in the age-specific payoffs from their marketable activities plus the societal punishment from deviating. This formalizes the notion held in social psychology that men need to project more masculinity than what they really have because the societal normative prescriptions for men relate highly with status (Moss-Racusin et al., 2010).

For women, the direction of the bias $\hat{m}_j - m_j$ could go either way. Even though they face the possibility of greater rewards in the marketable activity if they choose an $\hat{m}_j > m_j$, the societal punishment for deviating from social norms and the gains to low masculinity through the increased consumption of domestic good push $\hat{m}_j$ to be less than m_j. Women with high m_j are incentivized to reveal their true high masculinity by facing higher rewards from marketable activities, especially if they are high-skilled. At the same time, they are disincentivized to reveal their high m_j by society's punishment for not complying with gender norms and the value given to the domestic good. An example of this type of behavior is documented by Bursztyn et al. (2017) who find that unmarried female MBA students under-report professional ambition when they are aware male peers will see their responses. Another example is Exley and Kessler (2022), who find that women are less likely to self-promote themselves when it comes to male-typical tasks than their equally performing male counterparts. When it comes to stereotypically female tasks, the gender gap in self promotion disappears.

Implication 2 (*The role of discrimination*) *Gender discrimination (i.e.,* $w(M, \bar{m}) > w(F, \bar{m})$*) leads to men choosing a higher revealed masculinity than women, even if they both have the same true masculinity endowment.*

Gender discrimination makes women's payment from the marketable activity more easily offset by the societal punishment and the payoff from the domestic good, which relate negatively with $\hat{m}$. As social norms incentivize people to reveal a masculinity level that is different from their true endowment, they also have an incidence on the misallocation of time in marketable *versus* domestic activities—relative to the time allocation we would observe in the absence of gender norms and punishments for deviations (Ashraf et al., 2022). We summarize this concept in Implication 3.

Implication 3 (*Misallocation*) *Time misallocation between marketable and domestic activities is increasing in the size of the societal punishment* t*.*

Communities with greater societal punishment t will have more outcome (e.g., occupation) sorting along gender lines as men (women) are pushed to inflate (deflate) their revealed masculinity further away from their true endowment. As t increases, the average true masculinity among men who sort into gender-conforming roles falls, and only those whose true masculinity is very low sort into gender-nonconforming roles. Thus there is an over-provision of marketable time and under-provision of domestic time among men. Conversely, for women, increasing t implies that the average true masculinity among those who sort into gender-conforming roles increases. Thus, there is an under-provision of time in marketable activities and an over-provision of domestic time among women. There is ample evidence attesting to that. More sexist societies (in the sense that more people hold anti-egalitarian views of the role of women) have lower female employment and labor force participation rates and more economic inequality (Fortin, 2005; Bertrand, 2020; De Haas et al., 2024). Placing people in roles they would not have chosen in the absence of gender norms is costly for themselves and for the society as a whole (Hsieh et al., 2019; Jiang, 2021). Further, De Haas et al. (2024) show that societies with stricter adherence to dominance masculinity norms have higher male suicide rates.

The asymmetry described in Implications 1-3 uncovers how heterogeneity in revealed masculinity $\hat{m}_i$ links to variation in life outcomes that, in turn, are heavily influenced by gender norms. On the one hand, gender-nonconforming men are those who choose a relatively low $\hat{m}_i$ even after being unequivocally incentivized to inflate it. Thus, they are men with low m_i who do not anticipate the net gains from further inflation to be worth it, due to, for instance, low productivity of 'faked' masculinity relative to true masculinity, γ, or high returns to home production. In this sense, revealed masculinity has a biological basis in our model through m_i.[10] We therefore expect to find a negative empirical relationship between the outcomes of marketable activities and gender nonconformity among men. On the other hand, the incentives to inflate $\hat{m}_i$ are ambiguous for women. In particular, holding all other parameters constant, gender-nonconforming women will be those with high m_i who find that the benefits in the marketable activity of relatively high $\hat{m}_i$ exceed the societal punishment of not deflating it and the payoff provided by domestic activities. Hence, we expect a positive empirical relationship between the outcomes of marketable activities and gender nonconformity among women.

3 Data

3.1 The Stockholm Birth Cohort

We use data from the Stockholm Birth Cohort Study (SBC) which follows the cohort of children born in 1953 who were living in the Stockholm metropolitan area in November 1963. This cohort study links individuals across two longitudinal data sets. The first is the Stockholm Metropolitan Study 1953–85, which consists of birth

[10] For example, males tend to outperform females on spatial tasks (Hines, 2004; Halpern, 2011). There is also evidence of neuroanatomic and hormonal differences between gender-typical and gender-atypical individuals (Bussey and Bandura, 1999; Davis and Risman, 2015; Folkierska-Zukowska et al., 2020). While acknowledging this biological basis, the development of a child's cognitive understanding of gender—for example, whether competitiveness and self confidence can be feminine, or whether empathic, nurturing activities can be masculine—is still deeply rooted in nurture and societal norms (West and Zimmerman, 1987; Eagly, 1987; Halpern, 2011).

records, Census data and a comprehensive in-class school survey for the complete cohort of children born in 1953 who were still living in the Stockholm metropolitan ten years later. The second is The Swedish Work and Mortality Database, an administrative data set which follows up individuals' education, earnings, employment history and morbidity through 2009 at its longest for all cohort members still living in Sweden by 1980 or 1990.[11] Since the SBC relies on a number of different ethical approvals not all administrative data are available through 2009. For example, exact mother's age at birth is observed until 1981 from birth records and employment statistics are observed in 1980 from the Census and from 1992-2009 on from the Longitudinal integrated database for health insurance and labour market studies provided by Statistics Sweden.

The SBC study includes an in-class school survey that was conducted in 1966 when the cohort members were in sixth grade (age 13). During one school day, all sixth graders in the county of Stockholm filled out two questionnaires, including cognitive tests (verbal, numeric and spatial), friendship nominations of three best friends, questions on whether they hang out with mixed- or single-sex reference groups, favorite school subject and preferences for leisure interests (e.g., domestic, mechanical and sports). Importantly, the survey took place at school which gave it a mandatory character. As a result, the non-response rate is only 9% (the percentage of pupils absent on that particular school day). The school survey was combined to data backwards in time, such as delivery records and Census data and forward in time to data on the cohort members' subsequent schooling careers, occupational choices, earnings over three decades, life events and demographic outcomes. Table 1 provides descriptive statistics on the individuals in the study split by assigned sex and Appendix Table A.1 split by assigned sex and gender conformity. See Appendix A and Appendix C for additional summary statistics and variable definitions.

[11]The two data sets were matched based on variables which are available in both data sets. For 96% of the original cohort, data were matched. See (Stenberg and Vågerö, 2006) for a description of the data set and the matching procedure. Codebooks are available online at: https://www.stockholmbirthcohort.su.se.

Table 1: Descriptive Statistics Split by Assigned Sex

	Obs.	All	Male	Female	Difference	S.E.
Sociodemographic Background						
Older brother	10,154	0.351	0.341	0.36	-0.018*	0.009
Older sister	10,154	0.331	0.326	0.335	-0.009	0.009
Younger brother	10,154	0.339	0.336	0.343	-0.007	0.009
Younger sister	10,154	0.32	0.313	0.326	-0.013	0.009
Professional mother	10,154	0.04	0.041	0.038	0.003	0.004
Working mother	10,154	0.187	0.187	0.186	0.001	0.008
Female head of house	10,154	0.074	0.07	0.079	-0.009*	0.005
Mother any college	10,154	0.018	0.019	0.017	0.002	0.003
Father any college	10,154	0.089	0.093	0.085	0.008	0.006
Home-ownership	10,154	0.184	0.186	0.182	0.004	0.008
Educational Outcomes						
GPA in grade 9*100 (scale 1-5)	9,701	321.96	320.01	323.84	-3.832**	1.545
Upper secondary dropout	9,396	0.423	0.416	0.43	-0.014	0.010
Any post secondary	9,396	0.429	0.408	0.448	-0.040***	0.010
STEM secondary track	6,843	0.383	0.593	0.18	0.413***	0.011
Any college	9,396	0.244	0.245	0.243	0.002	0.009
Labor Market Outcomes						
Log earnings age 37	9,596	1.626	1.847	1.413	0.434***	0.013
Log average earnings age 37-47	9,638	1.757	1.921	1.598	0.323***	0.012
Full time in 1980	9,763	0.657	0.797	0.522	0.275***	0.009
Not employed in 1980	9,763	0.167	0.118	0.214	-0.096***	0.007
Unemployed in 2000	9,373	0.060	0.057	0.063	-0.006	0.005
Professional	7,789	0.118	0.152	0.082	0.070***	0.007
Legal or business	10,085	0.18	0.174	0.185	-0.012	0.008
STEM Occupation	10,085	0.1	0.148	0.053	0.095***	0.006
Blue collar	10,085	0.244	0.392	0.101	0.291***	0.008
Clerical support	10,085	0.122	0.047	0.195	-0.148***	0.006
Teacher-other health	10,085	0.15	0.072	0.225	-0.153***	0.007
Service and sales	10,085	0.118	0.092	0.142	-0.050***	0.006
Marriage, Fertility, and Mental Health Outcomes						
Married by 1980	9,816	0.344	0.252	0.433	-0.180***	0.009
Married by 1990	9,816	0.671	0.616	0.724	-0.108***	0.009
Married by 2000	9,816	0.732	0.695	0.767	-0.072***	0.009
Divorced by 1980 $\mid$ married	3,376	0.099	0.077	0.111	-0.033**	0.011
Divorced by 1990 $\mid$ married	6,588	0.158	0.133	0.179	-0.046***	0.009
Divorced by 2000 $\mid$ married	7,182	0.288	0.260	0.312	-0.052***	0.011
Teenage childbearing	5,171			0.021		
Age at first birth	2,749			23.87		
Childlessness in 1980	5,171			0.468		
Childlessness in 1990	4,914			0.169		
Total fertility in 1980	5,171			0.81		
Total fertility in 1990	4,880			1.749		
Mental health disorders	9,817	0.079	0.088	0.071	0.016***	0.005
Substance abuse	9,817	0.039	0.053	0.025	0.028***	0.004
Leadership ability	3,612		0.000			
Ability to function under stress	4,492		0.000			

Note: Statistics are shown for the full sample and for the male and female subsamples. Variables Leadership ability and Ability to function under stress are only available for men and were standardized to have mean zero and standard deviation one. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Additional descriptive statistics can be found in this Appendix A

3.2 The Gender Conformity Index

To study the association between revealed masculinity $\hat{m}_i$ and life outcomes, we set out to construct a context-specific measure of revealed masculinity in childhood. Based on it, we consider men with low $\hat{m}_i$ and women with high $\hat{m}_i$ to be childhood gender nonconformers. Unlike the contributions of the personality trait-based approach to gender identity that impose prespecified scales of "femininity" to "masculinity" that are thought to distinguish men from women (Bem, 1974; Spence et al., 1975; Magliozzi et al., 2016; Brenøe et al., 2022, 2024), we develop a measure of childhood gender nonconformity based on seemingly innocuous reports on interests and choices during childhood. We use reports on preferences for leisure interests, subject at school and peer group preferences because, as we will show below, those preferences and choices *empirically* differ across genders at that time (late 1960s) and age (i.e., 13 year-olds). By not imposing any gendered structure on the data, we take an *unsupervised* learning approach in which we simply collect the common latent variation shifting the manifest responses on preferences and choices and generate a one-dimensional continuous measure that abstracts completely from gender categories. It is only *ex-post* that we plot the gender-specific distributions of the resulting continuous measure and confirm that it contains relevant information about childhood gender conformity as it clearly discriminates between most boys and girls.

Existing approaches predict individuals' assigned sex using survey items that discriminate men from women based on clearly defined statistical significance criteria (Lippa and Connelly, 1990; Cleveland et al., 2001; Fleming et al., 2017; Yavorsky and Buchmann, 2019; Burn and Martell, 2022), or *supervised* learning algorithms (Mittleman, 2022). The resulting prediction thus informs us about the extent to which an individual's reported behaviors are aligned with the so-called sex-typed behaviors of other male or female research participants. Ours shares the advantage with the gender diagnosticity approach of generating a social- and culture-specific measure of gender identity, but unlike the supervised gender prediction approach ours is completely unsupervised in the sense that we only consider childhood features but do not associated them with assigned sex. Instead, we collect the common variation from our selected variables for every individual in the data, men and women pooled.

3.2.1 Observed Measures as Inputs for the GCI

We use the following variables to construct our gender conformity index. See further details in Appendix B.

Gender homophily. We use the survey question "with whom do you spend most of your time?" Extensive literature in Psychology and Sociology documents the important role that gender homophily plays in building friendship links, especially for school-aged children (Maccoby, 1998; McPherson et al., 2001; Stehlé et al., 2013). Appendix Figure B.1a shows that students in our sample tend to spend most of their time with other same-sex students.

Favorite school subject. We find a clear difference between boys and girls on their preferred school subject. Boys more frequently reported subjects like history and math, while girls preferred subjects like home-economics, foreign languages, Swedish language arts, music and religion (see Appendix Figure B.1b). This difference in class preferences that we find along gender lines has been established by earlier research in psychology, education and economics (Fryer Jr and Levitt, 2010; Buccheri et al., 2011; Buser et al., 2014; Joensen and Nielsen, 2016; Justman and Méndez, 2016).

Preferences for leisure time interests. The school survey in sixth grade included a battery of scrambled items on preferences for the three following areas of leisure interests that map onto the dimensions of communal and agentic behavior (Wood and Eagly, 2015), and towards which boys and girls on average tend to have different preferences: domestic interests, mechanical interests, and sports (Lippa and Connelly, 1990; Lippa, 2005). Even though the items were scrambled in the questionnaire, the areas were sets of ten items each and the students rated their interest preferences for each individual item using a 4-step scale that ranged from "would be very much fun" to "would be very boring" practicing the mentioned interest. In this sense, the battery of items inquires about hypothetical scenarios. Thus, it does not inform about the child's proficiency in each particular interests.[12]

Domestic interests include items such as making clothes, cooking foreign dishes and interior decoration. Mechanical interests include items such as playing with model railways, repairing a bike or reading about space ships. Finally, the area of interests that deal with sports includes items such as bike racing, playing basketball for a club, high jump and coaching athletes.[13] Appendix Figure B.2 shows that the preferences collected by these three areas of interests differ greatly along gender lines. Appendix

[12]These responses translate to corresponding numerical values of 5, 4, 2, or 1 points and only the sum of scores for each set of interests made it to the coded data set (not the scores on the single items). Specifically, a score of 10 means the student would find practicing the area of interests (e.g., mechanical interests) very boring whereas a score of 50 instead means that the student thinks it would be very much fun.

[13]See Appendix Table B.1 for all the individual items.

Figures B.2(c) and B.2(a) show that boys had a higher average preference for sports, while girls have higher preference for domestic activities—have in mind that all children in the sample were born in 1953. Finally, Appendix Figure B.2(b) shows that most boys have a preference for mechanical interests, while most girls find them unappealing. We excluded commercial interests from our measure as nine out of ten items were referring to professional activities beginning with the words "working as" or "selling", e.g., "working as head of a department in an office" and verbal interests as the individual items could be conflated with verbal achievement, e.g., "learning a foreign language".

The differences in preferences between boys and girls are well established in social psychology and education literatures (Ashmore et al., 1986; Gibbons et al., 1997; Aros et al., 1998; Lippa, 2010). When allowed to choose, boys are more likely to select conventionally masculine toys (e.g., cars, trains), whereas girls are more likely to choose conventionally feminine toys (e.g., jewelry, cooking and nurturing games) (Adelson, 2012). Further, as was mentioned before, sports achievements are higher valued among boys than among girls (Coleman, 1961). Displaying gender-atypicality on many of our measured behaviors and preferences (e.g., opposite-sex toy and friend preferences) belong to the most common manifestations of childhood gender nonconformity (Adelson, 2012).

3.2.2 The Construction of the Index

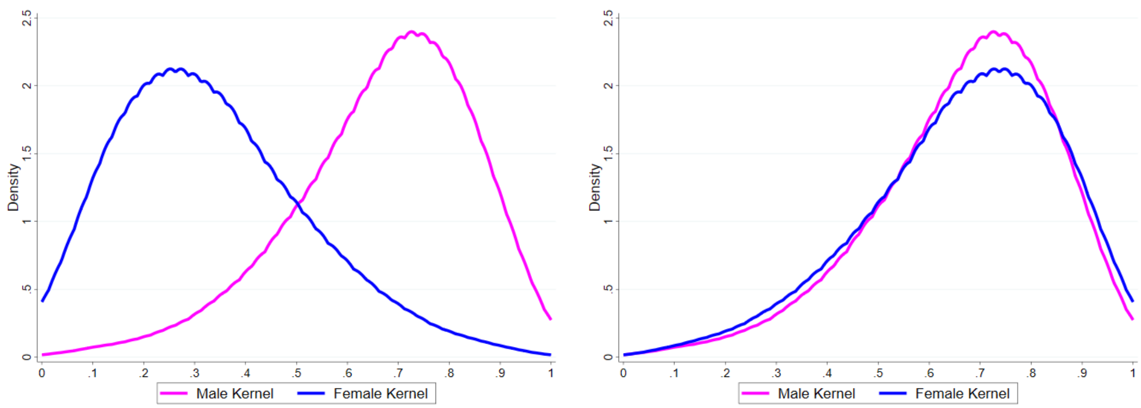
For our main empirical approach, we collect the common variation across the measures described above using principal component analysis (PCA).[14,15] We keep the variation in the first component—which amounts to 42% of the total variation in the manifest measures—to build our measure of interest. We normalize the values of the factor produced by that first component to the $[0, 1]$ interval. Figure 1(a) shows the distribution of component scores for boys and girls separately. This figure can be interpreted as a *revealed* masculinity scale since girls are clustered near zero and boys are clustered near one. Therefore, we can think of the boys who are near one and the girls who are near zero as expressing more gender-typical behaviors. To keep boys and girls in the same scale, we create a gender conformity index by reversing the gender-typical masculine factor for girls. By doing so, we have all of the girls who

[14]The school subject input variable is categorical. To use it in the PCA, we coded it to reflect if they are female dominated (-1), male dominated (1), or neither (0).

[15]In Section 5 we only use the scores of leisure interest (domestic, mechanical, sports) as the manifest test scores are required to be continuous.

Figure 1: Gender Conformity Index

(a) Gender Typicality by Gender (b) Gender Conformity Index



Note: Figure 1(a) plots the distribution of the component scores for boys and girls in our study sample of 4,983 boys and 5,171 girls. The mean (SD) for boys is 0.670 (0.174) and 0.333 (0.187) for girls. As in Figure 1(a), the score for girls is subtracted from one in order to flip the girls' distribution, so that the bulk of the girls' data is clustered near one in Figure 1(b). This allows us to make the assumption that the boys and girls near zero are gender-nonconforming since they are different than the norm set by their peers. To create this figure an in-built kernel density command was used on the modified first component variable while also splitting the data by gender. Data from Stockholm Birth Cohort.

reveal feminine characteristics and the boys who reveal masculine characteristics near one, and the boys who reveal feminine characteristics and the girls who reveal masculine characteristics near zero. Put another way, the children who are near one would be considered "extremely gender-conforming," and the children who are near zero are considered "extremely gender-nonconforming" regardless of their gender. Figure 1(b) shows the distribution of the gender conformity index by gender.

We define the gender-nonconforming individuals in relation to their gender's typical individuals. To operationalize childhood gender nonconformity (CGN), we set the threshold for conformity at the twentieth percentile of the GCI distribution. Thus, we define a binary variable that takes on value one for the bottom 20% of the GCI distribution and zero otherwise. We test the robustness of our results to the use of other thresholds, and we also provide results using the continuous GCI index.

3.2.3 Proof of Concept

In this subsection, we test the internal validity of our novel metric of gender nonconformity by providing evidence that it captures intrinsic preferences as opposed to parental influences in our analytic sample. We also test its external validity by replicating our metric in contemporary school surveys for three different European countries: Sweden, Germany, and the United Kingdom. We document the same bimodal GCI distributions in all three auxiliary surveys. We further repeat the internal validity test in each of these three surveys. We confirm that our metric is not sensitive to a) our choice of country; and b) our choice of cohort.

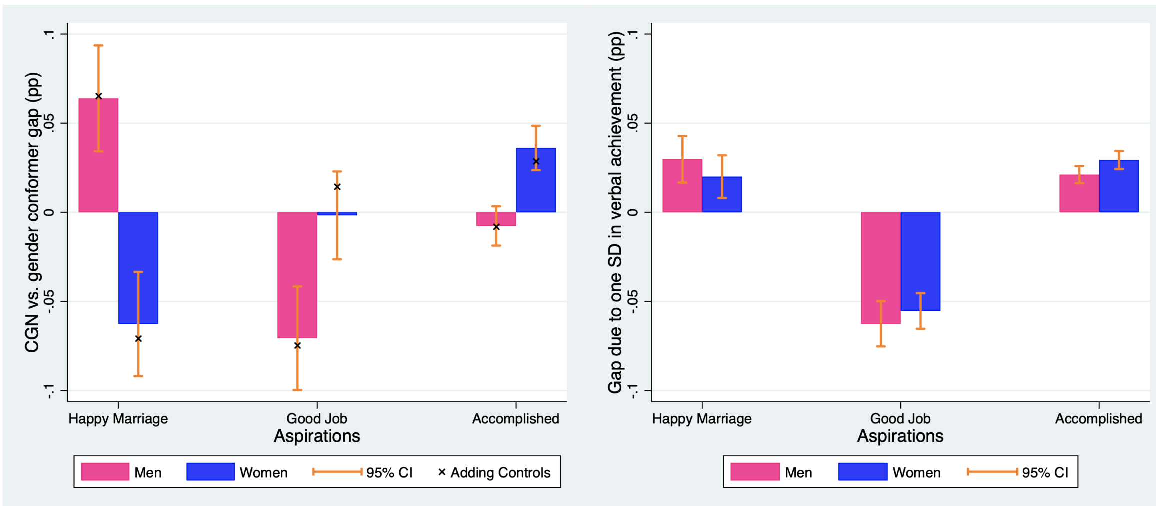
We first validate our measure using data on the marriage and career expectations of the cohort members in sixth grade. As mentioned, heavily influenced by gender norms, life goals is an area where men and women display vast differences from an early age (Lippa and Connelly, 1990). In Figure 2(a), we relate our CGN measure to self-reported marriage (communal) and career (agentic) aspirations ascertained in the same school survey as all questions underlying our CGN measure.[16] It would be reassuring to find strong and asymmetric correlations between our CGN measure and life goals. In congruence with the framework we presented in Section 2, our working hypothesis is that gender-nonconforming women (men) are more career-oriented (family-oriented) than their gender-typical counterparts.

Our results reassuringly show that female (male) gender nonconformers are less (more) likely to believe that having a "happy marriage" will be the most important factor determining their happiness during adulthood than their gender-typical counterparts. In addition, gender-nonconforming men value less getting a "good job" as a grown up as compared to their gender-conforming male peers, while female gender nonconformers aspire more for "accomplishing things in life" than female gender-conformers. Taken together, these associations suggest that the gender nonconformers in our data have clearly different life goals than their gender-conforming counterparts, and go against the gendered life goals postulated by the Social Role Theory (Eagly, 1987). The relationships between CGN and life goals remain unaltered, and even become stronger, when we control for cognitive achievement (both measured in the same survey). This result is relevant for two important reasons. First, cognitive achievement and life goals are correlated even after controlling for school fixed-effects to deal with neighborhood sorting, as we do throughout out the paper (Figure 2(b)). Second, cognitive achievement tests—especially ones gauging verbal and numeric achievement—are, to a great extent, the product of parental practices and investments (Agostinelli and Wiswall, 2016; Agostinelli et al., 2020). The answers to the questions we use to construct our GCI might also reflect parental practices and investments. Thus, if our GCI did collect information mainly inherent to differential parental investments, the

[16] According to Alice Eagly's (1987) Social Role Theory shared gender stereotypes develop from the gender division of labor and lead to differentiated skills and life goals.

Figure 2: Which will be the strongest driver of your happiness and life satisfaction as a grown up?

(a) By CGN (b) By Cognitive Achievement



Note: Data from Stockholm Birth Cohort. The question inquires about which aspect of life the child believes will be most important for her happiness as a grown up. Ninety-five percent of the students chose one of the following (exclusive) four categories: i.) to be happily married (51.4%), ii.) to have a good job (31.1%), iii.) to accomplish things (4%), and iv.) to have good friends (7.8%). The bars represent the coefficients of a regression of each aspiration on a gender-nonconformity dummy separately for men and women. We control for observable sociodemographic characteristics like household composition, parental education and home-ownership. We also include school fixed-effects. Black x symbols in Figure 2(a) indicate the value of the coefficient when the list of controls is expanded to include cognitive achievement. The variable in the vertical axis in 2(b) measures verbal achievement at age 13.

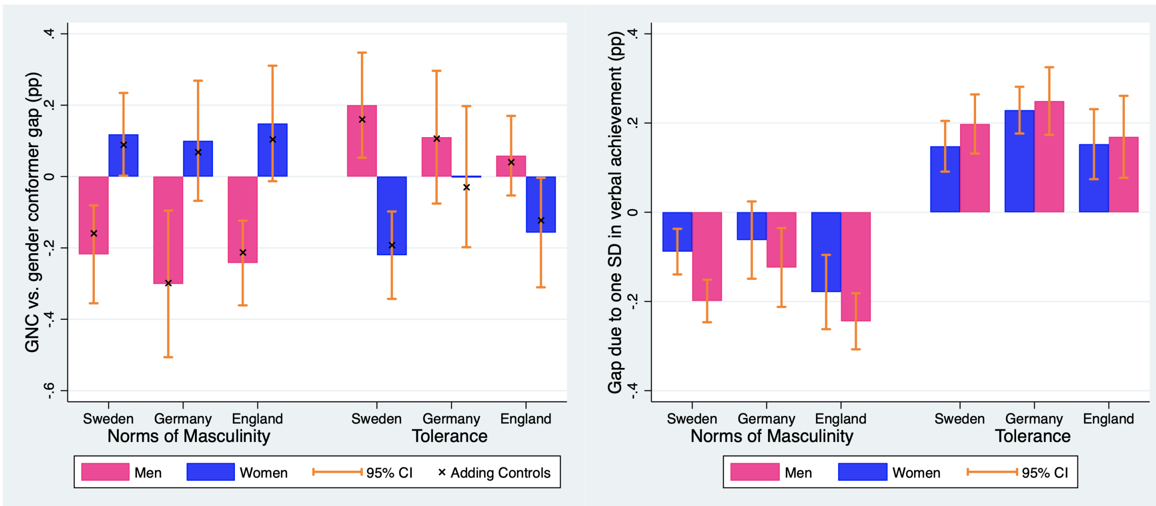
estimates in Figure 2(a) would mirror those that present the relationship between aspirations and cognitive achievement in Figure 2(b) and would significantly change once one controls for cognitive achievement. Reassuringly, this is not the case. These results lead us to confirm that the GCI we construct captures intrinsic preferences.

In order to test the external validity of our proposed metric of gender nonconformity, we use the longitudinal Youth in Europe Study (YES!) for three European countries representing school systems with different degrees of selectivity and extent of tracking: Germany (high selectivity), Sweden (moderate) and England (low) (Dollmann, 2021; Kalter et al., 2016). The study collects nationally representative data on 8^{th} graders in 2010. The survey collects information covering student performance, achievement, school environment, attitudes towards school, friendship and classmate networks, attitudes and norms and leisure time activities (see detailed descriptions of the data and instruments in Appendix B.1).

Using the same inputs as we did in our primary study sample (i.e., gender homophily, favorite subject at school and leisure time activities), we calculate our gender confor-

Figure 3: CGN and Gender Norms in Contemporary School Surveys

(a) By CGN (b) By Cognitive Achievement



Note: Data from YES! (Kalter et al., 2016). Samples exclude migrant children. Masculine norms collects information on how much the respondent agrees with men having to use violence in different contexts. Tolerance collects information on how much the respondent is at ease with a couple cohabiting without being married, divorce, homosexuality and abortion. For symmetry across Figures 2(a) and 3(a), the y-axis displays the percentage point difference between gender nonconformers (bottom 20% of the GCI by gender) and gender conformers. We include school fixed-effects. Black x symbols in Figure 3(a) indicate the value of the coefficient when the list of controls is expanded to include cognitive achievement. The variable in the vertical axis in 3(b) measures verbal achievement at age 13.

mity index in the contemporaneous data sets.[17] Reassuringly, Appendix Figure B.5 shows that the gender-specific distributions of the GCI in the three auxiliary surveys present similar gender-skewed patterns to the ones obtained from our primary data in Sub-figure 1(a). We then correlate the auxiliary surveys' GCI with the students' level of agreement to statements indicating that *men* should use violence in particular situations ("Norms of Masculinity"), and the students' level of agreement with cohabitation, divorce, abortion and homosexuality ("Tolerance"). A clear pattern emerges from Figure 3(a) across all three countries: female (male) gender-nonconformers are more (less) likely to sympathize with the statements gauging norms of masculinity than their gender-conforming counterparts. When it comes to tolerance, male (female) gender-nonconformers are more (less) tolerant than their gender-conforming counterparts. As all survey respondents in our auxiliary data were tested for verbal achievement we can contrast the correlations presented in Figure 3(a) with correlations of verbal achievement and norms of masculinity as well as tolerance (Figure 3(b)). As expected, both *a priori* and based on the results in our primary data (Figure 2), the latter correlations are symmetric for both genders, the higher the verbal

[17]Appendix B.1 describes the survey design and variables used as inputs for the gender nonconformity index. All codebooks and technical reports are available online at https://www.cils4.eu/.

achievement score at age 14, the less prone one is to sympathize with the norms of masculinity and the more tolerant one tends to be.[18] Thus, these patters serve as proof of concept and give us grounds to believe that our novel theory-based metric of gender nonconformity captures intrinsic preferences in multiple settings and cohorts.

3.2.4 Unsupervised vs. Supervised Learning

To further validate our measure, we compare our gender typicality index with the prediction whether Youth in Europe Study (YES!) survey respondents reported their assigned sex was male using as predictors the full set of response items of the 8th-grade survey (excluding the set of items on masculinity norms and the one on tolerance, all administrative variables, demographic variables such as birth country and individual and group identifiers). We follow the latest advances of gender prediction and automate the search for the best predictors through supervised learning by employing Least Absolute Shrinkage and Selection Operator (lasso) regressions. Of the 669, 676 and 675 answer choices available across 157 questions in the three auxiliary surveys for England, Sweden, Germany respectively, the lasso selected a total of 190, 187 and 153 distinct answer choices to include in the model. In each country sub-sample, at least five of the ten answer choices that lasso identified as most predictive of the 8th-grader's sex are answer choices of leisure time activities set of questions. In all three auxiliary surveys the most predictive answer choice is the "no time at all" answer choice on the video game question eliciting the time spent playing on a regular week day. Appendix Figure B.7 plots the distributions of the predicted probability of being boy separately by sex.[19] Again, reassuringly, we see that employing more than 150 survey questions and treating gender as a prediction task yields very similar bimodal distributions as we showed in Appendix Figure B.5 for our own GCI measure in the three auxiliary surveys. Table B.5 correlates separately our own GCI and the predicted probability with norms of masculinity and tolerance.

Three things stand out from the results of this validation exercise. First, both unsupervised and supervised learning work intuitively. Gender-nonconforming boys ad-

[18]Appendix Figure B.6 confirms that the results remain qualitatively identical for the same validation exercise when using the continuous GCI instead of the derived binary metric of childhood gender nonconformity (CGN).

[19]We estimate linear probability models and normalize the values of the predicted probability to the $[0, 1]$ interval. We invert the predicted probability by subtracting it from one for the female sub-sample to get both sexes on the same scale, i.e., 0 refers to the least gender conforming and 1 to the most gender conforming.

here less to norms of masculinity and are more tolerant than gender conforming boys while the gender-nonconforming girls adhere more to norms of masculinity than their same-sex gender conforming counterparts but are less tolerant (i.e., behave more like gender conforming men in this sense). Second, leisure interests seem to have the most predictive value even in the gender diagnosticity approach of predicting assigned sex. Third, starting off with the complete survey in the gender prediction task and narrowing down the predictors to a few hundred with lasso does not seem to outperform our own preference-based GCI measure precisely because leisure time preferences seem to be the most predictive school survey items of assigned sex. To summarize: We provide evidence that our data-driven measure picks up intrinsic preferences that form a continuum and a two-peaked distribution with significant overlap when both gender distributions are displayed separately. We operationalize a childhood gender conformity index based on it and validate it. A common feature to most data-driven metrics on gender conformity shared by ours is that it does not explicitly elucidate the prevailing societal gender norms that individuals either adhere to or not. It rather lets the data describe the stereotypical (or normative) behavior of both assigned sexes and deviations from it. In this sense it considers gender in the spirit of West and Zimmerman (1987) as a the collection of behaviors—what an individual "does"—and the degree to which an individual's behavior reflects identification with expectations about what it means to be male or female.

4 Regression Analysis

This section presents our main empirical analyses. Given that our aim is to explore the relationship between gender nonconformity and life outcomes, we regress the outcome y_{is} of student i in school s (e.g., education, career choices, labor market outcomes, marital status, mental health, and fertility outcomes) against a CGN_{is} dummy variable that takes the value of one if an individual was categorized as being a gender-nonconforming according to our metric (i.e., belonging to the bottom 20% of our gender conformity index), and zero otherwise. The reference group are thus the same-sex individuals who conform to the gender norms. We estimate the following equation:

$$\mathrm{y}_{is} = \alpha_0 + \omega_s + \mathrm{CGN}_{is}\beta + \mathbf{x}_{is}\alpha_1 + \epsilon_{is}, \tag{2}$$

for each gender separately, where $\mathbf{x}_{is}$ includes the following sociodemographic background covariates: single-parenthood, parental level of education, home-ownership status, whether the mother worked or had a professional position (Olivetti et al., 2018), and the number of younger and older siblings (Brenøe, 2021). The regressions also include ω_s, a school-level fixed effect that captures potential neighborhood-level confounders like those driving income and socioeconomic segregation. Thus, all our results refer to differences across students *within* the same school. These socio-economic controls capture predetermined factors that are known to determine parental investments—which can correlate with GCI—and later life outcomes (OECD, 2014). Recent literature has shown that parental investments respond to beliefs about their returns and to the child characteristics (Boneva and Rauh, 2018; Attanasio et al., 2020; Agostinelli et al., 2020; Walker ⓡ al., 2024). Given that the child's CGN status can be one of those characteristics, we do not include in $\mathbf{x}_{is}$ controls that are not predetermined. For instance, $\mathbf{x}_{is}$ should not include cognitive achievement scores because they inevitably capture acquired knowledge, personality and educational opportunity up until the time of taking the test (Borghans et al., 2016). All of which can be the direct or indirect—through parental investment—result of the child's CGN status. Appendix Table D.2 documents that our benchmark results are remarkably robust to adjusting for spatial ability, i.e., the ability measure that is weighted most towards fluid intelligence (Cattell, 1971). We adjust specifically for spatial ability as it is considered to be the most innate dimension of ability and hence the dimension that much of the literature on gender differences in cognitive ability has focused on (Hines, 2004; Halpern, 2011). Further, Section 5 takes a semi-structural approach and estimates the relation between life outcomes (i.e., earnings and occupation choices) and latent factors of gender nonconformity and cognitive ability. The latter identified from the common variation in cognitive ability test scores (verbal, numeric and spatial components). This allows us to analyze the association between each factor and life outcomes independently of each other.

Finally, to aid interpretation, we use the binary variable for whether a child is gender-nonconforming. In Section 4.4, we test the robustness of our results to using the continuous gender conformity metric (GCI) and alternative GCI metrics. Section 5 also takes a purely continuous approach.

4.1 Student Performance and Career Choice

We start by examining the link between gender nonconformity and student performance, education choices and career outcomes. To conserve space we only present estimates for our main variable of interest. The first panel in Table 2 reveals that

Table 2: Gender Nonconformity and Student Performance and Career Choices

	Explanatory variable: *Gender nonconformity*					
	Sample: Boys/Men			Sample: Girls/Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Student performance						
GPA in grade 9 (in hundredths)	-6.104**	(2.761)	4,746	21.328***	(2.474)	4,955
Upper secondary dropout	0.036**	(0.017)	4,588	-0.077***	(0.016)	4,808
Educational choice						
Any post secondary	-0.008	(0.018)	4,588	0.074***	(0.017)	4,808
STEM secondary track	-0.213***	(0.021)	3,374	0.170***	(0.018)	3,469
Any college	0.014	(0.015)	4,588	0.085***	(0.016)	4,808
Occupational choice†						
Legal or business	0.018	(0.013)	4,947	-0.001	(0.013)	5,138
STEM	-0.090***	(0.015)	4,947	0.023***	(0.007)	5,138
Blue collar	-0.011	(0.016)	4,947	-0.004	(0.011)	5,138
Clerical support	0.026***	(0.007)	4,947	-0.016	(0.014)	5,138
Teacher-other health	0.012	(0.009)	4,947	0.026*	(0.014)	5,138
Service and sales	0.027***	(0.009)	4,947	-0.024*	(0.013)	5,138
Did not work	0.018**	(0.008)	4,947	-0.004	(0.010)	5,138

Note: The entries in the top-two panels of the table represent the coefficient β from separate regressions of student performance and career choice on our binary variable for gender nonconformity as in equation (2). All regressions control for school fixed effects and sociodemographic background covariates such as parents' presence in the household, parents' level of education, a dummy for family living in an owner-occupied home, having older and younger brothers and sisters, and whether the mother worked or had a professional position. See top panel of Table A.2 for the complete set of the sociodemographic controls included. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. † Marginal effects from a multinomial logit (excluding school fixed effects) reported in the bottom panel. Multinomial logit's coefficient estimates provided by the authors

gender-nonconformity is positively associated to student performance for girls while negatively so for boys. The magnitude of the coefficients are sizable. Our regression results show that CGN girls' academic GPA in 9^{th} grade is 0.21 points greater than gender-conforming girls. In contrast, we find that CGN boys' academic 9^{th} grade GPA is 0.06 points *lower* than their gender-conforming male peers. Table 2 also shows that CGN girls are less likely to drop out from school before completing upper secondary than gender-conforming girls. A CGN girl is 7.7 percentage points (14% in relative terms) less likely to drop out of school after completion of compulsory school in 9^{th} grade than gender-conforming girls. We find the opposite for CGN boys. They are 3.6 percentage points (6.2%) *more* likely to drop out of school after completion of compulsory school as compared to gender-conforming boys.

After documenting that the link between gender nonconformity and student performance holds after controlling for potential confounders, and importantly, that opposite patterns for men and women remain robust, we assess its relation with educational and occupational choices in the bottom two panels of Table 2. We focus on educational track choices for upper secondary education, a pivotal junction that determines which prerequisites for tertiary education the student will acquire. We further explore occupational choices by age 27, which is late enough in life for most individuals to have completed their education and to be several years into their labor market history. Gender nonconformity is strongly associated with educational and occupational choices. We find that CGN girls were 7.4 percentage points (16.5%) more likely to take any post-secondary study and 17.0 percentage points (94.4%) more likely to apply to a STEM track than their gender-conforming female peers. The last gap is remarkable given that women have historically been vastly under-represented in STEM fields (Carrell et al., 2010; Kahn and Ginther, 2018; Bertrand, 2020). In contrast, CGN boys were 21.3 percentage points (35.9%) less likely to opt into the STEM track than their gender-conforming counterparts. However, even though CGN boys are less likely to choose STEM tracks, we do not find statistically significant evidence that CGN boys are less likely to reach post-secondary education than their gender-typical counterparts.

We find notable differences in occupational choices across gender conformity types. While women who were gender nonconformers during childhood (henceforth, CGN women) tend to sort into STEM occupations, CGN men tend to sort away from them and into female-dominated occupations like clerical support and services and sales. Given the general backdrop of women's underrepresentation in STEM occupations, which are on average high-paying and with relatively small gender gaps,[20] these occupational differences by CGN type must not be taken lightly. The differences are remarkable. While CGN women are 49% (2.3 percentage points relative to a 4.66% mean) more likely to work in a STEM occupation than gender-conforming women, CGN men are 55% less likely to sort into a STEM occupation than their gender-conforming male peers. Instead, CGN men are 63% and 31% *more* likely to choose clerical or service and sales occupations than comparable gender-conforming men.

4.2 Employment and Earnings

Thus far, we have documented a consistent pattern of female gender nonconformers outperforming gender-conforming women (and men, for that matter) at school and making career choices that should narrow the gender gap on the labor market even

[20]See, for instance, Goldin (2014); Bertrand (2020). In fact, in our data, STEM occupations pay on average 30% more, and the gender gap within them is 40% smaller than in the rest of the occupations.

Table 3: Gender Nonconformity and Labor Market Outcomes

	Explanatory variable: *Gender nonconformity* Sample: Men			Sample: Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Employment						
Full time in 1980	-0.036**	(0.015)	4,803	0.028	(0.017)	4,960
Not employed in 1980	0.021*	(0.012)	4,803	0.001	(0.015)	4,960
Unemployed in 2000	0.016*	(0.009)	4,585	0.012	(0.009)	4,788
Professional	-0.014	(0.013)	4,090	0.032***	(0.012)	3,699
Earnings						
Log earnings age 37	-0.094***	(0.023)	4,716	0.031	(0.021)	4,880
Log average earnings age 37-47	-0.096***	(0.024)	4,735	0.028	(0.019)	4,903
Log av. earnings 37-47 in STEM$^{\gamma}$	0.000	(0.075)	723	0.112	(0.070)	270

Note: The entries in the table represent the coefficient β from separate regressions of labor market outcomes on our binary variable for gender nonconformity as in equation (2). See top panel of Table A.2 for all the sociodemographic controls included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. $^{\gamma}$ Refers to earnings of those working in STEM occupations, estimates do not include school fixed effects due to the small number of observations.

in terms of earnings. The next question to ask is thus, did gender-nonconforming women capitalize on those choices in terms of better labor market outcomes and higher financial reward? Similarly, we ask whether gender-nonconforming men earn less due to making career choices that resemble those of gender-conforming women more than gender-conforming men. We report the results in Table 3.

We find that CGN men are 17.8% less likely to be employed and 4.5% less likely to have a full time job than gender typical men by age 27. Meanwhile, CGN women participate in the labor force as much as gender typical women, but they are more likely (5.4%) to hold a full time job as opposed to a part time one. Importantly, that job is 39% more likely to be a professional post.

Given this divergence in labor market participation, full-time employment, and occupational choices along gender typicality lines, we inquire if they have any bearing on earnings in the bottom panel of Table 3. We find that CGN men earn about 10% less than gender-conforming men during their prime labor market years (ages 37-47). This gap contrasts with the result that, during the same years, CGN women earn about 3% more than gender-conforming women—albeit not statistically significant at standard levels. The equivalent coefficient on the correlation between our continuous gender conformity index (GCI) and log average earnings (ages 37-47) is statistically significant though at the 5% level and implies that a one standard deviation less gender conforming woman earns on average 2% more (Appendix Table D.1). This coefficient is somewhat smaller and more imprecise than what unpublished parallel work to ours finds in the British context documents (5%) for a similarly aged cohort of women born in 1958 (Ayyar et al., 2024). This difference could potentially be explained by lower returns to education and a more compressed wage distribution in Sweden as compared to the UK (Tables 18-21 in the Deaton Review country reports Karimi et al., 2024; Cribb, 2024).[21]

The documented asymmetry in the way the labor market rewards male typicality goes in line with the predictions of our framework in Section 2, as it seems to matter whether *masculine* traits are displayed by a man or a woman and what type of occupation they are displayed in (Pan, 2015). Although imprecise, the results reported on last row of Table 3 suggest that CGN men who work in STEM occupations see no earnings *punishment*, CGN women in STEM see returns to their childhood gender nonconformity in adulthood. They earn 11% more than their gender-conforming female classmates who sort into STEM. Of course, the latter estimates may be biased due to sample selection. Namely, as we showed in Table 2, sorting into STEM occupations is itself influenced by childhood gender nonconformity. For this reason, in Section 5, we revisit the occupation-specific associations between earning and gender nonconformity using a Roy model in which we allow for endogenous occupational choices. The results remain: STEM occupations reward childhood gender nonconformity among women about four times more.

4.3 Marriage, Fertility, and Behavior

Lastly, we look at demographic and health outcomes in Table 4. Our results show that CGN girls were not less likely to get married but were nevertheless substantially more likely to get divorced in their 20s, their 30s and their 40s than their gender-conforming peers. These findings align with our framework in Section 2. Female gender atypicality would imply a lower production of the domestic good and increased production of the market good. Increased labor force participation and income would provide gender atypical women an outside option to staying in an unhappy marriage and increase their potential for divorce. These results are consistent

[21] In both ours (Appendix Table D.2) and their studies, the size of the coefficient shrinks substantially when adjusting for ability. Our directly comparable ability-adjusted coefficient of gender conformity index (GCI) for girls is no longer statistically significant (t=1.44) and implies an only 1.2% change in log average earnings (ages 37-47) as a result of a one standard deviation change in gender conformity.

with recent research stating that gender norms make marriages in which women are high earners less stable (Bertrand et al., 2015; Folke and Rickne, 2020). Furthermore, CGN girls are roughly 45% less likely to become mothers as teenagers and overall postpone childbearing relative to gender-conforming women by roughly four months. We further see that CGN girls are 5 percentage points more likely to remain childless by age 37, i.e., almost through completed fertility. The evidence suggests that they have substantially (14%) fewer children by age 27 than their gender-conforming counterparts but that they catch up and end up having only 4.7% fewer children by age 37, a difference which is marginally significant. This discrepancy in the results of fertility gaps of female gender-nonconformers vs. conformers of women in their twenties and of gaps in completed fertility is strikingly similar to what Ayyar et al. (2024) find in unpublished parallel work to ours.

This shrinking fertility gap when approaching completed fertility may help explain why the magnitude of the gain in earnings for CGN women is smaller than that of the loss for CGN men. It seems as if, just like their gender-conforming counterparts, CGN women end up paying the "child penalty", which research has shown to persist throughout a woman's career (Angelov et al., 2016; Doepke and Kindermann, 2019; Kleven et al., 2019). Taken together, these results suggest that CGN women, despite attaining higher and more profitable degrees and demonstrating similar preferences to the typical man, are yet penalized in the labor market for childbearing. This relates to the results documented by Goldin (2004, 2021) that show that only around 15% of college educated female 'baby boomers' in the US managed to combine career after marriage *and* childbearing.

Regarding mental health outcomes, we find that CGN boys are more likely to have mental health conditions and suffer from addiction during adulthood than gender-conforming boys. According to inpatient records, CGN men were 24% more likely to be hospitalized for mental health disorders during their life up until age 56 than gender-conforming men. Similarly, CGN men were 43.5% more likely to be hospitalized for substance abuse than gender-conforming men. As our framework describes, these adult outcomes can result from not fitting into the prescriptive stereotypes society imposes. They could be the long-term manifestation of i.) the emotional cost society puts on deviants from the norm in the form of social rejection (Landolt et al., 2004; Lippa, 2008a; MacMullin et al., 2021; Sarzosa and Urzúa, 2021), or ii.) the personal cost of having to fit into a masculine role (i.e., $\hat{m}_i > m_i$). Social psychology literature has established that even though gender norms have clear advantages for

Table 4: Gender Nonconformity and Demographic, Health & Socio-Emotional Outcomes

	Explanatory variable: *Gender nonconformity*					
	Sample: Men			Sample: Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Marriage						
Married by 1980	-0.021	(0.016)	4,825	-0.022	(0.018)	4,991
Married by 1990	-0.053***	(0.018)	4,825	-0.017	(0.016)	4,991
Married by 2000	-0.041**	(0.017)	4,825	-0.012	(0.016)	4,991
Divorced by 1980 | married	0.016	(0.023)	1,217	0.055***	(0.019)	2,159
Divorced by 1990 | married	0.011	(0.016)	2,974	0.037**	(0.016)	3,614
Divorced by 2000 | married	0.044**	(0.020)	3,353	0.033*	(0.020)	3,829
Fertility						
Teenage childbearing				-0.009*	(0.005)	5,171
Age (years) at first birth*				0.301**	(0.130)	2,749
Childlessness in 1980				0.060***	(0.017)	5,171
Childlessness in 1990				0.049***	(0.014)	4,914
Total fertility by 1980				-0.104***	(0.030)	5,171
Total fertility by 1990				-0.079*	(0.043)	4,880
Mental health						
Mental health disorders	0.021**	(0.011)	4,826	0.002	(0.009)	4,991
Substance abuse	0.023***	(0.009)	4,826	0.011*	(0.006)	4,991
Socio-emotional						
Leadership ability	-0.195***	(0.042)	3,612			
Ability to function under stress	-0.167***	(0.038)	4,492			

Note: The entries in the table represent the coefficient β from separate regressions of marriage, fertility, mental health and socioemotional outcomes on our binary variable for gender nonconformity as in equation (2). See top panel of Table A.2 for the complete set of sociodemographic controls included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. * Age at first birth is measured by 1981 as date of first birth is only available until 1981.

men, the pressure to conform to them may be harmful (Moss-Racusin et al., 2010). Failing to be as agentic, self-reliant, and stoic as society expects can undermine a person's self-worth (Carver et al., 2003; Yunger et al., 2004; Lippa, 2008a; DiFulvio, 2011). Table 4 provides some evidence on that matter. The last panel uses detailed military psychological evaluations of men at the time of conscription (end of high school), assessing leadership abilities and ability to function under stress, qualities related to agency. The results show that CGN men had a fifth of a standard deviation lower leadership ability and a sixth of a standard deviation less capacity to perform under stress than gender-conforming men in their cohort.

4.4 Summary of Robustness Checks

This sub-section summarizes the results of our robustness checks, which are presented in Appendix D. To support the validity of our results we run four variants of our main regressions. First, we replace our binary CGN variable with the underlying continuous gender conformity index as the independent variable of interest. We perform this analysis to confirm that our results are not influenced by our binary definition of gender nonconformity versus conformity. We report the results of this exercise in Appendix Table D.1. Second, Appendix Table D.2 shows that almost all our results remain qualitatively the same when adjusting for cognitive ability, the notable exception being GPA in grade 9 for male gender nonconformers. Third, in order not to conflate our nonconformity with general apathy towards leisure interests, we run our original analyses with a restricted sample. The restricted sample excludes individuals in the bottom fifth percentile of all reported interest in domestic interests, mechanical interests, and sports. The results for this exercise is reported in Appendix Table D.3. To ensure that favorite subject in sixth grade is not driving the results, Appendix Table D.4 removes this variable from our GCI index and picks the common variation from our three leisure interest scores (sports, domestic and mechanical) and gender homophily. Our key results remain qualitatively unchanged except for GPA in grade 9 for men and some of the marriage and fertility outcomes for women.

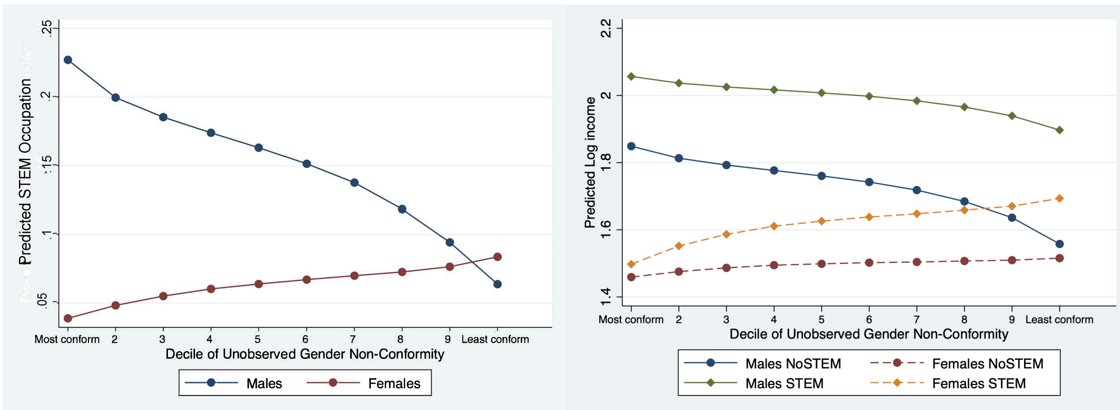
Overall, our results remain robust to these alternative specifications. CGN boys enroll less in STEM fields, earn significantly less than their gender-conforming male peers, are less likely to work full-time, and tend to forgo STEM occupations for clerical, service, and sales jobs. On the contrary, CGN girls earn higher grades, go on to enroll in STEM fields, and further their education more than their gender-conforming female peers. We find that CGN girls are more likely to work in STEM, to divorce, and to postpone childbirth.

5 Endogenous Occupational Choice & Earnings Gaps

The earnings gap is arguably the most studied gender gap in economics. Works like Bertrand (2020) and Goldin (2021) show the importance of analyzing it in the context of a sequence of career choices that end up defining the feasible set of monetary rewards for a given worker. In this study, we have already shown that the gender atypical individuals opt for substantially different careers relative to their gender-typical counterparts. Our *raw* earnings gaps reported in Table 3 embed the different

Figure 4: Gender Nonconformity and Career Outcomes

(a) Occupational Choices (b) Log earnings by Occupational choice



Note: Figure 4(a) presents the $E[\text{STEM}|\theta^{CGN}]$ in the vertical axis product of 20,000 simulations based on the findings of the choice equation (in system of equations (4)) of the Roy model presented in Section F.1. Figure 4(b) presents $E[y_0|\theta^{CGN}] = E[\mathbf{x}\beta_0] + \alpha^{Y_0,G}\theta^{CGN}$ and $E[y_1|\theta^{CGN}] = E[\mathbf{x}\beta_1] + \alpha^{Y_1,G}\theta^{CGN}$ as estimated by the outcome equations of the Roy model (again, in system of equations (4)) and simulated using 20,000 simulations. The horizontal axes in all panels displays the deciles of the gender-nonconforming factor. Data from Stockholm Birth Cohort.

career choices made up until that point in life. Here, to inquire about the counterfactual earnings had the individual chosen a male-dominated occupation as opposed to a female-dominated one (or vice versa), we present and estimate an extended Roy Model of potential outcomes (Heckman and Honoré, 1990). Specifically, we explore this counterfactual contrast across the entire distribution of gender conformity. Our Roy model endogenizes occupational selection while allowing earnings to be a function of observable and unobservable characteristics that depends on the occupational category (Heckman et al., 2006). In order to estimate our Roy model we consider a two-factor model in which both gender nonconformity and cognitive ability are latent factors The former is identified from the common variation in manifest scores of the reported preferences for leisure time interests (domestic, mechanical and sports) and the latter is identified from the common variation in cognitive ability test scores (verbal, numeric and spatial). This allows us to analyze the association between each factor and life outcomes independently of each other. In this section, we present the results of the model and refer the reader to Appendix F for the underlying model and the assumptions required for identification. See Appendix Table F.1 for the estimates of measurement system of manifest test scores.

Figure 4 presents the results. Figure 4(a) shows that CGN men are more likely to sort into (out of) lower (higher) paying occupations. Twenty-three percent of the most gender-conforming men work in a STEM occupation, while only 6.4% of the least gender-conforming do. That is a relative difference of about 73%. Although the relation between occupation choice and gender-conformity is not as steep among women as it is in men, we do find some interesting differences. Figure 4(a) indicates that virtually none of the gender-conforming women choose a STEM occupation, while about 8.3% of the least gender-conforming women do. Thus, CGN women are more likely to sort into STEM occupations than CGN men. That is remarkable given the vast under-representation of women in STEM occupations.[22]

Figure 4(b) presents the estimates of the outcome equations. It shows that gender nonconformity is punished among men regardless of the occupation. The punishment is larger in non-STEM occupations than in STEM ones. Relative to the most gender-typical men, CGN men in STEM occupations earn 15% less. The tantamount differential is 29% in non-STEM occupations. CGN women experience the opposite. While gender-typical women earn on average roughly the same in STEM occupations as outside of STEM occupations, CGN women in STEM occupations earn 49% more than CGN women in non-STEM fields. Thus, gender nonconformity pays off for women in terms of earnings as long as they sort into high-paying male-dominated occupations like the ones in STEM fields.

6 Gender Norms and Divergent Youth Paths

In Section 4.1, we show that, on average, CGN boys do poorly in school while CGN girls do exceptionally well. We then show that these gaps persist in many respects into adulthood. But what are the underlying mechanisms opening up these performance gaps at school? This section shows evidence attributing these divergent school paths to a combination of preferences (e.g., taste for school) and social interactions, all fueled by gender norms.

CGN girls thrive at school. Table 5 shows that CGN girls are 27% of a standard deviation *more* interested in school than gender typical girls that go *to the same school*. Likewise, CGN girls feel much safer in school than their gender-conforming female peers. In addition, CGN girls' social networks at school have different characteristics.

[22] In Appendix Figure F.1 we show the opposite regarding the take-up of lower paying occupations like service and sales. CGN men are twice more likely to sort into service and sales (12.9%) than into STEM (6.4%) occupations. On the contrary, gender-conforming men are about three times *more* likely to sort into STEM occupations (23%) than into service and sales (8%). In contrast, CGN women are 4 percentage points (18%) *less* likely to sort into a services or sales occupation than gender-conforming women.

Table 5: Gender Nonconformity and Social Interactions, Attitudes and Behaviors

	Explanatory variable: *Gender nonconformity*					
	Sample: Boys/Men			Sample: Girls/Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Panel A: Social Network Characteristics						
Nominating only one friend	0.016*	(0.009)	4,776	-0.003	(0.008)	4,985
At least one friends is CGN	0.099***	(0.017)	4,983	0.145***	(0.016)	5,171
Average verbal score of friends	-0.013	(0.025)	4,765	0.162***	(0.026)	4,968
Average numeric score of friends	0.011	(0.025)	4,765	0.126***	(0.024)	4,967
Average spatial score of friends	-0.037	(0.026)	4,765	0.098***	(0.023)	4,968
Panel B: Attitude toward school						
Student's feeling safe at school	-0.157**	(0.080)	4,961	0.349***	(0.082)	5,152
Student's interest in school work	-0.660***	(0.080)	4,967	0.674***	(0.084)	5,139
Panel C: Risky behavior at age 14-18						
Misbehavior	0.013**	(0.007)	4,983	0.006	(0.006)	5,171
Stealing	0.012	(0.012)	4,983	0.002	(0.006)	5,171
Crimes of violence	0.011	(0.008)	4,983	-0.002	(0.002)	5,171
Abuse of alcohol and narcotics	0.021***	(0.009)	4,983	-0.001	(0.004)	5,171
Drunkenness and abuse of solvents	0.021**	(0.009)	4,983	-0.004	(0.004)	5,171

Note: The entries in the table represent the coefficient β from separate regressions of measures of social interaction, attitudes and behaviors on our binary variable for gender nonconformity as in equation (2). See top panel of Table A.2 for the complete set of the sociodemographic controls included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

CGN girls tend to befriend substantially smarter peers relative to gender typical girls in the same school. Relative the to friends of gender typical girls, the average friend of CGN girls score 16%, 12% and 10% of a standard deviation higher on verbal, numeric and spacial cognitive tests, respectively. Those gaps are sizeable, and as the extensive literature on peer effects has shown, they are very likely to create positive spillovers that boost scholastic performance of the CGN girls (Sacerdote, 2011; List et al., 2020). Although, they report having a similar number of close friends as their gender-conforming classmates, Table E.1 of Appendix E shows that CGN girls increase the social capital of the classroom by linking female and male networks together. A higher share of CGN girls in the classroom reduces its clustering (i.e., the degree to which students in the network tend to group in cliques) and diameter (i.e., the shortest distance between the two most distant nodes in the network), both standard measures of social cohesion. That is, CGN girls connect social cliques that otherwise would not be connected. Recent literature highlights the importance social capital has in determining life outcomes ranging from education to intergenerational mobility (Chetty et al., 2022).

What about boys? Much of the literature on gender gaps in student performance attributes at least part of the male underperformance at school to school cultures or toxic masculinity that may adversely affect educational investments (Coleman, 1961; Akerlof and Kranton, 2002; Goldin et al., 2006; Bursztyn and Jensen, 2017; Bursztyn et al., 2018). Our results challenge the hypothesis of gender nonconformity leading to greater human capital accumulation for adolescent boys through shielding them from male-typical risky behaviors (Yavorsky and Buchmann, 2019; Mittleman, 2022). Using data from the Child Protection and Social Welfare Services for the high school years (ages 14-18), we observe the incidence of risky behavior during adolescence. Panel C of Table 5 shows that, on average, adolescent CGN men are *not* less likely than gender typical boys to be involved in delinquent or violent behaviors. They are roughly 50% *more* likely to misbehave as reported by either the school teachers or parents, and 31% more likely to have problems with substance abuse. Thus, we reject the 'toxic male school culture' hypothesis, at least for Sweden in the late 1960s.

Rather, our framework and results indicate that deviating from the norm can be harmful for CGN boys. Discomfort and anxiety of being forced to participate in gender-typical activities and fear of social rejection may predispose CGN adolescent boys to high-risk behavior and substance abuse (Adelson, 2012). According to Panel A of Table 5, CGN boys tend to be less socially connected than their gender-conforming peers. Relative to their gender-conforming peers are 27% more likely to have only one close friend, and the classmates that they do befriend are 28% more likely to be CGNs themselves.[23] Thus, unlike CGN girls, CGN boys face greater risk of social rejection, which is known to have negative effects on grades (Eriksen et al., 2014), skill accumulation (Sarzosa, 2021), and adult outcomes (Sarzosa and Urzúa, 2021). In fact, CGN boys are 15.7% of a standard deviation less likely to feel safe at school than their gender typical male classmates. In line with the literature showing that social rejection in the classroom increases the distaste for school (NAS, 2016; Sarzosa and Urzúa, 2021), CGN boys are 27% of a standard deviation less interested in schoolwork than gender typical boys.

Worth noting here is that we do not observe youth sexual orientation. Gender non-conformity is found to correlate positively with sexual minority representation (Lippa, 2008b; Hernandez et al., 2024; Hsu, 2024), which in turn is positively associated with

[23]The relative changes come from the following calculations. Overall, 5.7% of gender typical boys report having only one close friend. Thus, 0.016/0.057 yields the 27% increase in the probability of a CGN boy reporting having only one close friend relative to gender typical boys. Likewise, 35.4% of gender typical boys report befriending at least one CGN peer. CGN boys are 9.9% more likely to do so. Thus the relative change amounts to 28%.

school-based victimization, in particular among boys (Toomey and Russell, 2016; Humphries et al., 2021). Even though much of our analysis considers conformity along the whole distribution and our simplified binary definition of nonconformity (categorizing 20% of individuals as nonconformers) encompasses by necessity individuals of all sexual orientations, it is possible, that individuals identifying with youth sexual minority is partly driving our results on social rejection.[24] Some recent work further suggests that sexual minority youth has lower performance in high school (Badgett et al., 2024, and references therein), and face higher risk of substance abuse (Goldbach et al., 2014).

Taken together, these results indicate that the consequences of gender nonconformity differ greatly along gender lines even early in life. There is a return to childhood nonconformity for girls and a steep cost for CGN boys. The social interactions seem to play an important role in mediating those rewards and costs. While CGN girls thrive at school and make more and smarter friends, CGN boys feel isolated and unsafe, which in turn may contribute to the observed behavioral problems. These differences enroute CGN boys and girls in the divergent paths that we document throughout the paper.

7 Conclusions

Our study provides new insight on the relationship between childhood gender nonconformity and life outcomes. We use a novel approach to measure the former, facilitated by the availability of unique data on preferences in early adolescence and outcomes throughout individuals' professional careers. Among women, gender nonconformers fare better compared to their gender-conforming peers, while among men, the opposite is true. Female gender nonconformers outperform other women at school and are more likely to choose a STEM track and sort into higher-paying STEM jobs. They also delay fertility and earn more throughout their lives. Male gender nonconformers, on the contrary, do worse at school, achieve lower levels of education, sort into less well paid occupations, and have substantially lower earnings than comparable gender-conforming men. Furthermore, male gender nonconformers present a higher incidence of troubling behaviors and mental health issues.

[24] Individuals identifying as LGBT are typically below 10% of the population (Badgett et al., 2021). Further, recent evidence from the U.S. suggests that sexual minorities are a clear minority among gender nonconformers. In the fifth wave of the AddHealth study, roughly 17% of self-assessed gender nonconforming women in their late 30s identify as sexual minority while roughly 11% of gender nonconforming men do so (Hernandez et al., 2024).

Our study is descriptive and does not evaluate a specific policy, nor do our results provide direct policy implications. They, however, highlight the role of gender norms and their rewards to male typicality in determining individuals' career pathways and drive gender gaps. Our findings are in line with the recent literature in economics showing that gender norms and gender gaps in preferences hinder progress towards achieving equality in career choices and the labor market. We observe that gender gaps are narrower for girls who challenge gender norms as early as adolescence. In particular, when it comes to student performance gender gaps, our results suggest that the current state of the literature paints a vexingly incomplete picture. Our analysis suggests that the student performance gap in favor of girls may be driven by the gender-nonconforming girls. From a simplified static perspective, if they were absent from the classroom, girls would not be outperforming boys in our data. Our results also support the idea that the labor market rewards masculine traits more so than feminine ones, and that men who do not display the former pay a significant cost. Furthermore, that such men display a greater incidence of troubling behaviors and mental health issues suggests that the pressure to conform to the agentic male ideal during adolescence can translate into profound emotional harm for boys who go against the grain.

To our knowledge, this is among the first studies in the Economics literature to explore the role that societal prescriptions play in perpetuating gender gaps by analyzing *within*-gender complexity. Our evidence on drastic within-gender differences in life outcomes across the distribution of conformity to stereotypical gender norms and preferences shows just how critical it is to consider more complex gender norm identities. In this, we respond to Hyde et al.'s (2019) multidisciplinary call to move "beyond" a simple male-female gender binary and to develop new, pioneering research methods that allow to study the complexity of gender norm identity (Lundberg, 2023). Hopefully, increased availability of data about the complex notions of gender will spur more economics research on the emergence of gender gaps from early school years on.

References

Adelson, Stewart L. (2012) "Practice Parameter on Gay, Lesbian, or Bisexual Sexual Orientation, Gender Nonconformity, and Gender Discordance in Children and Adolescents," *Journal of the American Academy of Child & Adolescent Psychiatry*, 51 (9), P957–974, https://doi.org/10.1037/0033-295X.106.4.676. 1, 3.2.1, 6

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti (2020) "It Takes a Village: The Economics of Parenting with Neighborhood and Peer Effects," Working Paper 27050, National Bureau of Economic Research, 10.3386/w27050. 3.2.3, 4

Agostinelli, Francesco and Matthew Wiswall (2016) "Estimating the Technology of Children's Skill Formation," Working Paper 22442, National Bureau of Economic Research, 10.3386/w22442. 3.2.3

Akerlof, George A. and Rachel E. Kranton (2000) "Economics and Identity," *The Quarterly Journal of Economics*, 115 (3), 715–753, 10.1162/003355300554881. 1

——— (2002) "Identity and Schooling: Some Lessons for the Economics of Education," *Journal of Economic Literature*, 40 (4), 1167–1201, 10.1257/002205102762203585. 1, 2, 6

Albanesi, Stefania and Claudia Olivetti (2009) "Home production, market production and the gender wage gap: Incentives and expectations," *Review of Economic Dynamics*, 12 (1), 80–107, https://doi.org/10.1016/j.red.2008.08.001. 5, 9

Alesina, Alberto, Paola Giuliano, and Nathan Nunn (2013) " On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, 128 (2), 469–530, 10.1093/qje/qjt005. 1

Almlund, Mathilde, Angela Lee Duckworth, James Heckman, and Tim Kautz (2011) "Personality Psychology and Economics," *Handbook of the Economics of Education*, 4, 1–181, 10.1016/B978-0-444-53444-6.00001-8. C

Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016) "Parenthood and the gender gap in pay," *Journal of labor economics*, 34 (3), 545–579. 4.3

Aros, Jesse R, George A Henly, and Nicholas T Curtis (1998) "Occupational Sextype and Sex Differences in Vocational Preference-Measured Interest Relationships," *Journal of vocational behavior*, 53 (2), 227–242. 3.2.1

Ashmore, Richard D, Frances K Del Boca, and Arthur J Wohlers (1986) "Three - Gender Stereotypes," in *The Social Psychology of Female-Male Relations*, 69–119: Elsevier Inc. 3.2.1

Ashraf, Nava, Oriana Bandiera, Virginia Minni, and Victor Quintas-Martinez (2022) "Gender Roles and the Misallocation of Labour Across Countries," Unpublished Manuscript. 2

Attanasio, Orazio, Raquel Bernal, Michele Giannola, and Milagros Nores (2020) "Child Development in the Early Years: Parental Investment and the Changing Dynamics of Different Dimensions," Working Paper 27812, National Bureau of Economic Research, 10.3386/w27812. 4

Ayyar, Sreevidya, Uta Bolt, Eric French, and Cormac O'Dea (2024) "Imagine your Life at 25: Gender Conformity and Later-Life Outcomes," Working Paper 32789, National Bureau of Economic Research, 10.3386/w32789. 1, 4.2, 4.3

Babcock, Linda, Maria P. Recalde, and Lise Vesterlund (2017) "Gender Differences in the Allocation of Low-Promotability Tasks: The Role of Backlash," *American Economic Review*, 107 (5), 131–35, 10.1257/aer.p20171018. 6

Badgett, M. V. Lee, Christopher S. Carpenter, Maxine J. Lee, and Dario Sansone (2024) "A Review of the Economics of Sexual Orientation and Gender Identity," *Journal of Economic Literature*, 62 (3), 948–94, 10.1257/jel.20231668. 6

Badgett, M. V. Lee, Christopher S. Carpenter, and Dario Sansone (2021) "LGBTQ Economics," *Journal of Economic Perspectives*, 35 (2), 141–70, 10.1257/jep.35.2.141. 24

Baranov, Victoria, Ralph De Haas, and Pauline Grosjean (2023) "Men. Male-biased Sex Ratios and Masculinity Norms: Evidence from Australia's Colonial Past," *Journal of Economic Growth*, 28, 339–396, `https://doi.org/10.1007/s10887-023-09223-x`. 1

Bartholomew, D, M Knott, and I Moustaki (2011) *Latent Variable Models and Factor Analysis: A Unified Approach*: Wiley Series in Probability and Statistics. West Sussex, United Kingdom. F.2, F.3

Becker, Gary S. (1985) "Human Capital, Effort, and the Sexual Division of Labor," *Journal of Labor Economics*, 3 (1), S33–S58, `http://www.jstor.org/stable/2534997`. 5

Bem, S. L. (1974) "The measurement of psychological androgyny," *Journal of Consulting and Clinical Psychology*, 42 (2), 155–162. 3.2

Bertrand, Marianne (2011) "Chapter 17 - New Perspectives on Gender," in Card, David and Orley Ashenfelter eds. *Handbook of Labor Economics*, 4 of Handbook of Labor Economics, 1543–1590: Elsevier, https://doi.org/10.1016/S0169-7218(11)02415-4. 1, 2

——— (2020) "Gender in the Twenty-First Century," *AEA Papers and Proceedings*, 110, 1–24, 10.1257/pandp.20201126. 1, 2, 7, 2, 2, 4.1, 20, 5

Bertrand, Marianne, Emir Kamenica, and Jessica Pan (2015) " Gender Identity and Relative Income within Households," *The Quarterly Journal of Economics*, 130 (2), 571–614, 10.1093/qje/qjv001. 1, 4.3

Black, Nicole, David W. Johnston, and Agne Suziedelyte (2017) "Justification bias in self-reported disability: New evidence from panel data," *Journal of Health Economics*, 54, 124–134, https://doi.org/10.1016/j.jhealeco.2017.05.001. 1

Blau, Francine D and Lawrence M Kahn (2017) "The gender wage gap: Extent, trends, and explanations," *Journal of economic literature*, 55 (3), 789–865. 2

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Rajshri Jayaraman (2015) "Linear Social Interactions Models," *Journal of Political Economy*, 123 (2), 444–496, 10.1086/679496. 3

Boneva, Teodora and Christopher Rauh (2018) "Parental Beliefs about Returns to Educational Investments—The Later the Better?" *Journal of the European Economic Association*, 16 (6), 1669–1711, `https://doi.org/10.1093/jeea/jvy006`. 4

Booth, AlisonL. and Jan C. Van Ours (2009) "Hours of Work and Gender Identity: Does Part-time Work Make the Family Happier?" *Economica*, 76 (301), 176–196, https://doi.org/10.1111/j.1468-0335.2007.00670.x. 2

Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer (2016) "Stereotypes," *The Quarterly Journal of Economics*, 131 (4), 1753–1794, 10.1093/qje/qjw029. 8

Borges, Bruna, Fernanda Estevan, and Louis-Philippe Morin (2022) "Gender Differences in Prioritizing Rewarding Tasks," Working Paper. 2

Borghans, Lex, Angela Lee Duckworth, James J Heckman, and Bas Ter Weel (2008) "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43 (4), 972–1059, 10.1353/jhr.2008.0017. C

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries (2016) "What grades and achievement tests measure," *Proceedings of the National Academy of Sciences*, 113 (47), 13354–13359, 10.1073/pnas.1601135113. 4, C

Boucher, Vincent and Bernard Fortin (2016) "Some Challenges in the Empirics of the Effects of Networks," in Bramoulle, Yann, Andrea Galeotti, and Brian W. Rogers eds. *The Oxford Handbook of Economics of Networks*, Chap. 10, 277–302, Oxford: Oxford University Press. 3

Brenøe, Anne Ardila (2021) "Brothers increase women's gender conformity," *Journal of Population Economics*, 1–38. 4

Brenøe, Anne Ardila, Lea Heursen, Eva Ranehill, and Roberto A. Weber (2022) "Continuous Gender Identity and Economics," *AEA Papers and Proceedings*, 112, 573–77, 10.1257/pandp.20221083. 1, 3.2

——— (2024) "Continuous Gender Identity and Economic Decision Making," SSRN Working Paper, University of Zurich, University of Zurich, `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4889851`. 1, 3.2

Buccheri, Grazia, Nadja Abt Gürber, and Christian Brühwiler (2011) "The Impact of Gender on Interest in Science Topics and the Choice of Scientific and Technical Vocations," *International journal of science education*, 33 (1), 159–178. 3.2.1

Burn, Ian and Michael E. Martell (2022) "Gender typicality and sexual minority labour market differentials," *British Journal of Industrial Relations*, https://doi.org/10.1111/bjir.12671. 1, 3.2

Bursztyn, Leonardo, Georgy Egorov, and Robert Jensen (2018) "Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education," *The Review of Economic Studies*, 86 (4), 1487–1526, 10.1093/restud/rdy026. 6

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais (2017) "'Acting Wife': Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 107 (11), 3288–3319, 10.1257/aer.20170029. 2

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott (2020) "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, 110 (10), 2997–3029, 10.1257/aer.20180975. 1

Bursztyn, Leonardo and Robert Jensen (2017) "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure," *Annual Review of Economics*, 9 (1), 131–153, 10.1146/annurev-economics-063016-103625. 6

Buser, Thomas, Muriel Niederle, and Hessel Oosterbeek (2014) "Gender, competitiveness, and career choices," *The Quarterly journal of economics*, 129 (3), 1409–1447. 2, 3.2.1

Bussey, Kay and Albert Bandura (1999) "Social cognitive theory of gender development and differentiation," *Psychological Review*, 106 (4), 676–713, `https://doi.org/10.1037/0033-295X.106.4.676`. 10

Carneiro, Pedro, Karsten T Hansen, and James Heckman (2003) "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," *International Economic Review*, 44 (2), 361–422. F.1, F.2, 26, 30, 31

Carrell, Scott E., Marianne E. Page, and James E. West (2010) "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *The Quarterly Journal of Economics*, 125 (3), 1101–1144, 10.1162/qjec.2010.125.3.1101. 4.1

Carver, Priscilla R, Jennifer L Yunger, and David G Perry (2003) "Gender identity and adjustment in middle childhood," *Sex roles*, 49 (3), 95–109. 4.3

Cattell, Raymond B. (1971) *Abilities: Their Structure, Growth, and Action*: New York: Houghton Mifflin. 4, C

Ceci, Stephen J., Donna K. Ginther, Shulamit Kahn, and Wendy M. Williams (2014) "Women in Academic Science: A Changing Landscape," *Psychological Science in the Public Interest*, 15 (3), 75–141, 10.1177/1529100614541236, PMID: 26172066. 1

Chetty, Raj, Matthew O. Jackson, Theresa Kuchler, and et al. (2022) "Social Capital I: Measurement and Associations with Economic Mobility," *Nature*, 608, 108–121, https://doi.org/10.1038/s41586-022-04996-4. 6, C, 25

Cleveland, H. Harrington, J. Richard Udry, and Kim Chantala (2001) "Environmental and Genetic Influences on Sex-Typed Behaviors and Attitudes of Male and Female Adolescents," *Personality and Social Psychology Bulletin*, 27 (12), 1587–1598, 10.1177/01461672012712003. 3.2

Coleman, James (1961) *The Adolescent Society*, New York: Free Press of Glencoe. 6

Coleman, James S. (1961) "Athletics in High School," *The ANNALS of the American Academy of Political and Social Science*, 338 (1), 33–43, 10.1177/000271626133800105. 1, 3.2.1

Cortes, Patricia, Jessica Pan, Laura Pilossoph, and Basit Zafar (2021) "Gender differences in job search and the earnings gap: Evidence from business majors,"Technical report, National Bureau of Economic Research. 2

Cribb, Jonathan (2024) "Labour market and income inequalities in the United Kingdom, 1968–2021," *Fiscal Studies*, 45 (2), 131–142, https://doi.org/10.1111/1475-5890.12373. 4.2

Croson, Rachel and Uri Gneezy (2009) "Gender Differences in Preferences," *Journal of Economic Literature*, 47 (2), 448–74, 10.1257/jel.47.2.448. 2

Cullen, Zoe B. and Ricardo Perez-Truglia (2021) "The Old Boys' Club: Schmoozing and the Gender Gap," NBER Working Paper Series, 26530. 2

Davis, Jac T. M. and Melissa Hines (2020) "How Large Are Gender Differences in Toy Preferences? A Systematic Review and Meta-Analysis of Toy Preference Research," *Archives of Sexual Behavior*, 49 (2), 373–394, 10.1007/s10508-019-01624-7. 1

Davis, Shannon N. and Barbara J. Risman (2015) "Feminists wrestle with testosterone: Hormones, socialization and cultural interactionism as predictors of women's gendered selves," *Social Science Research*, 49, 110–125, https://doi.org/10.1016/j.ssresearch.2014.07.012. 10

De Haas, Ralph, Victoria Baranov, Ieada Matavelli, and Pauline Grosjean (2024) "HeforShe: Bargaining Power, Parental Beliefs, and Investments in Children," CEPR Discussion Paper, Paris & London, Paris & London, `https://cepr.org/publications/dp19493`. 1, 2, 2

Diamond, Lisa M. (2020) "Gender Fluidity and Nonbinary Gender Identities Among Children and Adolescents," *Child Development Perspectives*, 14 (2), 110–115, https://doi.org/10.1111/cdep.12366. 1

DiFulvio, Gloria T (2011) "Sexual minority youth, social connection and resilience: From personal struggle to collective identity," *Social Science & Medicine*, 72 (10), 1611–1617. 4.3

DiPrete, Thomas A. and Claudia Buchmann (2013) *Rise of Women, The: The Growing Gender Gap in Education and What it Means for American Schools*: Russell Sage Foundation, `http://www.jstor.org/stable/10.7758/9781610448000`. 1

Doepke, Matthias and Fabian Kindermann (2019) "Bargaining over babies: Theory, evidence, and policy implications," *The American economic review*, 109 (9), 3264–3306. 4.3

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2010) "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, 100 (3), 1238–60, 10.1257/aer.100.3.1238. 2

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011) "INDIVIDUAL RISK ATTITUDES: MEASUREMENT, DETERMINANTS, AND BEHAVIORAL CONSEQUENCES," *Journal of the European Economic Association*, 9 (3), 522–550, https://doi.org/10.1111/j.1542-4774.2011.01015.x. 2

Dollmann, Jörg (2021) "Ethnic inequality in choice- and performance-driven education systems: A longitudinal study of educational choices in England, Germany, the Netherlands, and Sweden," *The British Journal of Sociology*, 72 (4), 974–991, https://doi.org/10.1111/1468-4446.12854. 3.2.3

Eagly, A. H. and V. J. Steffen (1984) "Gender stereotypes stem from the distribution of women and men into social roles," *Journal of Personality and Social Psychology*, 46 (4), 735–7545, `https://doi.org/10.1037/0022-3514.46.4.735`. 1, 1

Eagly, Alice Hendrickson. (1987) *Sex differences in social behavior : a social-role interpretation / Alice H. Eagly.*, John M. MacEachran memorial lecture series ; 1985, Hillsdale, N.J: L. Erlbaum Associates. 2, 10, 3.2.3

Eckel, Catherine C. and Philip J. Grossman (2008) "Sex and risk: experimental evidence," in Plott, Charles and Vernon Smith eds. *Handbook of Experimental Economics Results*, 1: Elsevier. 2

Eriksen, Tine Louise Mundbjerg, Helena Skyt Nielsen, and Marianne Simonsen (2014) "Bullying in Elementary School," *Journal of Human Resources*, 49 (4), 839–871, 10.3368/jhr.49.4.839. 6

Exley, Christine L and Judd B Kessler (2022) "The Gender Gap in Self-Promotion," *The Quarterly Journal of Economics*, 137 (3), 1345–1381, 10.1093/qje/qjac003. 2

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti (2004) "Mothers and Sons: Preference Formation and Female Labor Force Dynamics," *The Quarterly Journal of Economics*, 119 (4), 1249–1299, 10.1162/0033553042476224. 1

Fleming, Paul J., Kathleen Mullan Harris, and Carolyn Tucker Halpern (2017) "Description and Evaluation of a Measurement Technique for Assessment of Performing Gender," *Sex Roles*, 76, 731–746, https://doi.org/10.1007/s11199-016-0657-3. 1, 3.2

Fogli, Alessandra and Laura Veldkamp (2011) "Nature or Nurture? Learning and the Geography of Female Labor Force Participation," *Econometrica*, 79 (4), 1103–1138, https://doi.org/10.3982/ECTA7767. 1

Folke, Olle and Johanna Rickne (2020) "All the Single Ladies: Job Promotions and the Durability of Marriage," *American Economic Journal: Applied Economics*, 12 (1), 260–87. 4.3

Folkierska-Zukowska, Monica, Qazi Rahman, Artur Marchewka, Marek Wypych, Dawid Drozdziel, Andrzej Sokolowski, and Wojciech L. Dragan (2020) "Male sexual orientation, gender nonconformity, and neural activity during mental rotations: an fMRI study," *Scientific Reports*, 10 (18709), `https://doi.org/10.1038/s41598-020-74886-0`. 10

Fortin, Nicole M (2005) "Gender Role Attitudes and the Labour-market Outcomes of Women across OECD Countries," *Oxford Review of Economic Policy*, 21 (3), 416–438, 10.1093/oxrep/gri024. 2, 2

Fryer Jr, Roland G and Steven D Levitt (2010) "An empirical analysis of the gender gap in mathematics," *American Economic Journal: Applied Economics*, 2 (2), 210–40. 3.2.1

Genicot, Garance (2022) "Tolerance and Compromise in Social Networks," *Journal of Political Economy*, 130 (1), 94–120, 10.1086/717041. 8

Gibbons, Judith L, Maria Lynn, and Deborah A Stiles (1997) "Cross-National Gender Differences in Adolescents' Preferences for Free-Time Activities," *Cross-cultural research*, 31 (1), 55–69. 3.2.1

Giuliano, Paola (2018) "Gender: A Historical Perspective," in *The Oxford Handbook of Women and the Economy*: Oxford University Press, 10.1093/oxfordhb/9780190628963.013.29. 1

Goldbach, Jeremy T., Emily E. Tanner-Smith, Meredith Bagwell, and Shannon Dunlap (2014) "Minority Stress and Substance Use in Sexual Minority Adolescents: A Meta-analysis," *Prevention Science*, 15, 350–363, https://doi.org/10.1007/s11121-013-0393-7. 6

Goldin, Claudia (2004) "The Long Road to the Fast Track: Career and Family," *The Annals of the American Academy of Political and Social Science*, 596, 20–35, http://www.jstor.org/stable/4127648. 4.3

——— (2014) "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104 (4), 1091–1119, 10.1257/aer.104.4.1091. 1, 20

——— (2021) *Career and Family: Women's Century-Long Journey Toward Equity*: Princeton University Press. 4.3, 5

Goldin, Claudia and Lawrence F. Katz (2002) "The Power of the Pill: Oral Contraceptives and Women's Career and Marriage Decisions," *Journal of Political Economy*, 110 (4), 730–770, 10.1086/340778. 1

Goldin, Claudia, Lawrence F Katz, and Ilyana Kuziemko (2006) "The homecoming of American college women: The reversal of the college gender gap," *Journal of Economic perspectives*, 20 (4), 133–156. 6

Halpern, Diane F. (2011) *Sex differences in cognitive abilities (4th ed.)*: New York: Psychology Press. 10, 4

Hansen, Karsten T, James J Heckman, and Kathleen J Mullen (2004) "The effect of schooling and ability on achievement test scores," *Journal of Econometrics*, 121 (1-2), 39–98. F.1, F.2

Heckman, James J. and Bo E. Honoré (1990) "The Empirical Content of the Roy Model," *Econometrica*, 58 (5), 1121–1149, http://www.jstor.org/stable/2938303. 5

Heckman, James J., John Eric Humphries, and Gregory Veramendi (2018) "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 126 (S1), S197–S246, 10.1086/698760. F.1, F.2

Heckman, James J. and Salvador Navarro (2007) "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136 (2), 341–396. F.1

Heckman, James J. and Edward J. Vytlacil (2007) "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation," in Heckman, James J. and Edward E. Leamer eds. *Handbook of Econometrics*, 6, 4779–4874: Elsevier, https://doi.org/10.1016/S1573-4412(07)06070-9. F.1

Heckman, James, Jora Stixrud, and Sergio Urzua (2006) "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24 (3), 411–482. 5, F.2, F.3

Hernandez, Stephanie M., Carolyn T. Halpern, and Kerith J. Conron (2024) "Sexual orientation, gender expression and socioeconomic status in the National Longitudinal Study of Adolescent to Adult Health," *Journal of Epidemiology & Community Health*, 78 (2), 121–128, 10.1136/jech-2022-220164. 1, 2, 6, 24

Hines, Melissa (2004) *Brain Gender*: Oxford University Press. 10, 4

Hoxby, Caroline (2000) "Peer Effects in the Classroom: Learning from Gender and Race Variation," Working Paper 7867, National Bureau of Economic Research, 10.3386/w7867. E

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow (2019) "The allocation of talent and us economic growth," *Econometrica*, 87 (5), 1439–1474. 2

Hsu, Jaime (2024) "Penalizing Nonconformity: Gender Nonconformity at the Intersection of Gender and Sexual Identity in the Labor Market," *Social Problems*, spae050, 10.1093/socpro/spae050. 1, 2, 6

Humphries, Kayleigh D., Li Li, Gary A. Smith, Jeffrey A. Bridge, and Motao Zhu (2021) "Suicide Attempts in Association With Traditional and Electronic Bullying

Among Heterosexual and Sexual Minority U.S. High School Students," *Journal of Adolescent Health*, 68 (6), 1211 – 1214, https://www.jahonline.org/article/S1054-139X(20)30863-6/fulltext. 6

Husen, Torsten (1961) "School Reform in Sweden: A Liberal Democracy Adopts the Comprehensive School System," *The Phi Delta Kappan*, 43 (2), 86–91, http://www.jstor.org/stable/20342681. E

Hyde, Janet Shibley (2014) "Gender Similarities and Differences," *Annual Review of Psychology*, 65 (1), 373–398, 10.1146/annurev-psych-010213-115057, PMID: 23808917. 1, 1

Hyde, Janet Shibley, Rebecca S. Bigler, Daphna Joel, Charlotte Chucky Tate, and Sari M. van Anders (2019) "The Future of Sex and Gender in Psychology: Five Challenges to the Gender Binary," *American Psychologist*, 74, 171–193. 1, 8, 7

Jiang, Xuan (2021) "Women in STEM: Ability, preference, and value," *Labour Economics*, 70, 101991, https://doi.org/10.1016/j.labeco.2021.101991. 2

Joensen, Juanna Schrøter and Helena Skyt Nielsen (2016) "Mathematics and gender: Heterogeneity in causes and consequences," *The Economic Journal*, 126 (593), 1129–1163. 3.2.1

Judd, Kenneth L (1998) *Numerical Methods in Economics*, Cambridge, Massachusetts: The MIT Press. F.2

Justman, Moshe and Susan J Méndez (2016) "Gendered selection of STEM subjects for matriculation," Melbourne institute working paper. 3.2.1

Kahn, Shulamit and Donna Ginther (2018) "Women and Science, Technology, Engineering, and Mathematics (STEM): Are Differences in Education and Careers Due to Stereotypes, Interests, or Family?" in Averett, Susan L., Laura M. Argys, and Saul D. Hoffman eds. *The Oxford Handbook of Women and the Economy*: Oxford University Press, https://doi.org/10.1093/oxfordhb/9780190628963.013.13. 4.1

Kalter, Frank, Anthony F. Heath, Miles Hewstone, Jonsson Jan O., Irena Kogan, and Frank van Tubergen (2016) "Children of Immigrants Longitudinal Survey in Four European Countries (CILS4EU) - Reduced version. Reduced data file for download and off-site use," GESIS Data Archive, Cologne. ZA5656 Data file Version 1.2.0, https://doi.org/10.4232/cils4eu-de.6656.6.0.0, 10.4232/cils4eu.5656.1.2.0. 3.2.3, 3, B.4, B.4, B.5, B.6, B.7, B.5

Karimi, Arizo, Charlotte Lucke, and Mårten Palme (2024) "Components of the evolution of income inequality in Sweden, 1990–2021," *Fiscal Studies*, 45 (2), 187–204, https://doi.org/10.1111/1475-5890.12367. 4.2

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019) "Children and Gender Inequality: Evidence from Denmark," *American Economic Journal: Applied Economics*, 11 (4), 181–209, 10.1257/app.20180010. 1, 4.3

Kotlarski, I (1967) "On characterizing the gamma and the normal distribution.," *Pacific Journal of Mathematics*, 20 (1), 69–76. F.3

Landolt, Monica A, Kim Bartholomew, Colleen Saffrey, Doug Oram, and Daniel Perlman (2004) "Gender Nonconformity, Childhood Rejection, and Adult Attachment: A Study of Gay Men," *Archives of sexual behavior*, 33 (2), 117–128. 4.3

Lindqvist, Erik and Roine Vestman (2011) "The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment," *American Economic Journal: Applied Economics*, 3 (1), 101–28, 10.1257/app.3.1.101. C

Lippa, R. (2005) "How Do Lay People Weight Information About Instrumentality, Expressiveness, and Gender-Typed Hobbies When Judging Masculinity–Femininity in Themselves, Best Friends, and Strangers?" *Sex Roles*, 53, 43–55, https://doi.org/10.1007/s11199-005-4277-6. 3.2.1

Lippa, R. and S. Connelly (1990) "Gender diagnosticity: A new Bayesian approach to gender-related individual differences," *Journal of Personality and Social Psychology*, 59 (5), 1051–1065, https://doi.org/10.1037/0022-3514.59.5.1051. 3.2, 3.2.1, 3.2.3

Lippa, Richard A (2008a) "The relation between childhood gender nonconformity and adult masculinity–femininity and anxiety in heterosexual and homosexual men and women," *Sex Roles*, 59 (9), 684–693. 4.3

Lippa, Richard A. (2008b) "The Relation Between Childhood Gender Nonconformity and Adult Masculinity–Femininity and Anxiety in Heterosexual and Homosexual Men and Women," *Sex Roles*, 59, 684–693, https://doi.org/10.1007/s11199-008-9476-5. 6

Lippa, Richard A (2010) "Gender Differences in Personality and Interests: When, Where, and Why?: Gender Differences in Personality and Interests," *Social and personality psychology compass*, 4 (11), 1098–1110. 3.2.1

List, John A., Fatemeh Momeni, and Yves Zenou (2020) "The Social Side of Early Human Capital Formation: Using a Field Experiment to Estimate the Causal Impact of Neighborhoods," Working Paper 187, University of Chicago, Becker Friedman Institute for Economics, 10.2139/ssrn.3760076. 6

Lundberg, Shelly (2020) "Educational gender gaps," *Southern Economic Journal*, 87 (2), 416–439, https://doi.org/10.1002/soej.12460. 1

——— (2023) "Gender Economics: Dead-Ends and New Opportunities," in Polachek, Solomon W. and Konstantinos Tatsiramos eds. *50th Celebratory Volume (Research in Labor Economics)*, 50, 151–189: Emerald Publishing Limited, Leeds, `https://doi.org/10.1108/S0147-912120230000050006`. 1, 7

Maccoby, Eleanor E. (1998) *The two sexes: Growing up apart, coming together*, Cambridge, MA: Belknap Press/Harvard University Press. 3.2.1

MacMullin, Laura N, Lisa M Bokeloh, A Natisha Nabbijohn, Alanna Santarossa, Anna I R van der Miesen, Diana E Peragine, and Doug P VanderLaan (2021) "Examining the Relation Between Gender Nonconformity and Psychological Well-Being in Children: The Roles of Peers and Parents," *Archives of sexual behavior*, 50 (3), 823–841. 4.3

Magliozzi, Devon, Aliya Saperstein, and Laurel Westbrook (2016) "Scaling Up: Representing Gender Diversity in Survey Research," *Socius*, 2, 2378023116664352, 10.1177/2378023116664352. 3.2

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001) "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27 (1), 415–444, 10.1146/annurev.soc.27.1.415. 3.2.1

Miho, Antonela, Alexandra Jarotschkin, and Ekaterina Zhuravskaya (2023) "Diffusion of Gender Norms: Evidence from Stalin's Ethnic Deportations," *Journal of the European Economic Association*, 22 (2), 475–527, 10.1093/jeea/jvad040. 1

Mittleman, Joel (2022) "Intersecting the Academic Gender Gap: The Education of Lesbian, Gay, and Bisexual America," *American Sociological Review*, 87 (2), 303–335, 10.1177/00031224221075776. 1, 3.2, 6, B.7

Moss-Racusin, Corinne, Julie Phelan, and Laurie Rudman (2010) "When Men Break the Gender Rules: Status Incongruity and Backlash Against Modest Men," *Psychology of Men & Masculinity*, 11, 140–151, 10.1037/a0018093. 2, 2, 4.3

NAS (2016) "Preventing Bullying Through Science, Policy, and Practice," National Academies of Sciences, Engineering, and Medicine. Washington, DC, https://books.google.com/books/about/Preventing_Bullying_Through_Science_Poli.html?id=zDwUDQAAQBAJ. 6

Nollenberger, Natalia, Núria Rodríguez-Planas, and Almudena Sevilla (2016) "The Math Gender Gap: The Role of Culture," *American Economic Review*, 106 (5), 257–61, 10.1257/aer.p20161121. 1

OECD (2014) *Skills for Social Progress*: OECD. 4

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou (2018) "Mothers, Peers, and Gender-Role Identity," *Journal of the European Economic Association*, 18 (1), 266–301, 10.1093/jeea/jvy050. 4

Olivetti, Claudia and Barbara Petrongolo (2016) "The Evolution of Gender Gaps in Industrialized Countries," *Annual Review of Economics*, 8 (1), 405–434, 10.1146/annurev-economics-080614-115329. 1

Ors, Evren, Frédéric Palomino, and Eloïc Peyrache (2013) "Performance Gender Gap: Does Competition Matter?" *Journal of Labor Economics*, 31 (3), 443–499, http://www.jstor.org/stable/10.1086/669331. 2

Pan, Jessica (2015) "Gender Segregation in Occupations: The Role of Tipping and Social Interactions," *Journal of labor economics*, 33 (2), 365–408. 4.2

Paulston, Rolland G. (1966) "The Swedish Comprehensive School Reform: A Selected Annotated Bibliography," *Comparative Education Review*, 10 (1), 87–94, http://www.jstor.org/stable/1186319. E

Pew Research Center (2020) "On the Cusp of Adulthood and Facing an Uncertain Future: What We Know About Gen Z So Far,"Technical report, Pew Research Center, Washington, D.C., https://www.pewresearch.org/social-trends/2020/05/14/on-the-cusp-of-adulthood-and-facing-an-uncertain-future-what-we-know-about-gen-z-s 1

Prada, María F. and Sergio Urzúa (2017) "One Size Does Not Fit All: Multiple Dimensions of Ability, College Attendance, and Earnings," *Journal of Labor Economics*, 35 (4), 953–991, 10.1086/692477. F.2

Rudman, Laurie A (1998) "Self-Promotion as a Risk Factor for Women: The Costs and Benefits of Counterstereotypical Impression Management," *Journal of personality and social psychology*, 74 (3), 629–645. 6

Rudman, Laurie A. and Julie E. Phelan (2008) "Backlash effects for disconfirming gender stereotypes in organizations," *Research in Organizational Behavior*, 28, 61–79, https://doi.org/10.1016/j.riob.2008.04.003. 6

Sacerdote, Bruce (2011) "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," Elsevier, 10.1016/b978-0-444-53429-3.00004-1. 6

Santavirta, T and M Sarzosa (2024) "Effects of disruptive peers in endogeneous social networks." E

Sarzosa, Miguel (2021) "Victimization and Skill Accumulation: The Case of School Bullying," *Journal of Human Resources*, 10.3368/jhr.0819-10371R2. 6

Sarzosa, Miguel and Sergio Urzua (2016) "Implementing Factor Models for Unobserved Heterogeneity in Stata," *The Stata Journal*, 16 (1), 197–228. 26, 29

Sarzosa, Miguel and Sergio Urzúa (2021) "Bullying among adolescents: The role of skills," *Quantitative economics*, 12 (3), 945–980. 4.3, 6

Shurchkov, Olga and Catherine C Eckel (2018) *Gender differences in behavioral traits and labor market outcomes*: Oxford, UK: Oxford University Press. 2

SOU1961:30 (1961) *Grundskolan. Betänkande avgivet av 1957 års skolberedning*, 1961:30 of Statens offentliga utredningar 1961:30: Ivar Häggströms Boktryckeri AB, Stockholm, `https://lagen.nu/sou/1961:30`. E

Spence, J. T., R. Helmreich, and J. Stapp (1975) "Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity," *Journal of Personality and Social Psychology*, 32 (1), 29–39, `https://doi.org/10.1037/h0076857`. 3.2

Stehlé, Juliette, François Charbonnier, Tristan Picard, Ciro Cattuto, and Alain Barrat (2013) "Gender homophily from spatial behavior in a primary school: A sociometric study," *Social Networks*, 35, 604–613, 10.1016/j.socnet.2013.08.003. 3.2.1

Stenberg, Sten-Åke and Denny Vågerö (2006) "Cohort Profile: The Stockholm birth cohort of 1953," *International journal of epidemiology*, 35 (3), 546–548. 11

Su, Rong, James Rounds, and Patrick Ian Armstrong (2009) "Men and things, women and people: a meta-analysis of sex differences in interests," *Psychological bulletin*, 135 (6), 859—884, 10.1037/a0017364. 1, 1

Svensson, Allan (1964) *Sociala och regionala faktorers samband med över- och underrepresentation i skolarbetet*, Rapporter från Pedagogiska institutionen, Göteborgs universitet: (Mimeographed). C

——— (1971) *Relative Achievement. School performance in relation to intelligence, sex and home environment.* in , Göteborg Studies in Educational Sciences (6): Almqvist & Wiksell, Stockholm. C

Toomey, Russell B. and Stephen T. Russell (2016) "The Role of Sexual Orientation in School-Based Victimization: A Meta-Analysis," *Youth & Society*, 48 (2), 176–201, 10.1177/0044118X13483778, PMID: 26997680. 6

Urzua, Sergio (2008) "Racial Labor Market Gaps," *Journal of Human Resources*, 43 (4), 919. F.1, F.3

Ushchev, Philip and Yves Zenou (2020) "Social norms in networks," *Journal of Economic Theory*, 185, 104969, https://doi.org/10.1016/j.jet.2019.104969. 3

Walker, Sarah ⓡ Pauline Grosjean ⓡ Alejandrina Cristia ⓡ Adeline Delavande (2024) "HeforShe: Bargaining Power, Parental Beliefs, and Parental Investments," CEPR Discussion Paper, Paris & London, Paris & London, `https://cepr.org/publications/dp18775`. 4

West, Candace and Don H. Zimmerman (1987) "Doing Gender," *Gender & Society*, 1 (2), 125–151, 10.1177/0891243287001002002. 1, 10, 3.2.4

Willis, Robert J. and Sherwin Rosen (1979) "Education and Self-Selection," *Journal of Political Economy*, 87 (5), S7–S36, `http://www.jstor.org/stable/1829907`. F.1

Wiswall, Matthew and Basit Zafar (2014) "Determinants of College Major Choice: Identification using an Information Experiment," *The Review of Economic Studies*, 82 (2), 791–824, 10.1093/restud/rdu044. 1

Wood, W. and A.H. Eagly (2015) "Two Traditions of Research on Gender Identity," *Sex Roles*, 73, 461–473, https://doi.org/10.1007/s11199-015-0480-2. 3.2.1

Yavorsky, Jill E. and Claudia Buchmann (2019) "Gender Typicality and Academic Achievement among American High School Students," *Sociological Science*, 6 (25), 661–683, 10.15195/v6.a25. 1, 3.2, 6

Yunger, Jennifer L, Priscilla R Carver, and David G Perry (2004) "Does gender identity influence children's psychological well-being?" *Developmental psychology*, 40 (4), 572. 4.3

Zafar, Basit (2013) "College Major Choice and the Gender Gap," *Journal of Human Resources*, 48 (3), 545–595, 10.3368/jhr.48.3.545. 1

Appendix

A Descriptive statistics

Table A.1: Descriptive Statistics by Sex and Gender Conformity

	Male				Female			
	CGN	GC	Diff	S.E.	CGN	GC	Diff	S.E.
Life Outcomes								
GPA in grade 9	316.28	320.93	-4.65	2.83	345.61	318.35	27.26***	2.61
Upper secondary dropout	0.44	0.41	0.02	0.02	0.33	0.45	-0.12***	0.02
Any post secondary	0.41	0.41	0.01	0.02	0.54	0.43	0.11***	0.02
STEM secondary track	0.44	0.63	-0.19***	0.02	0.32	0.14	0.18***	0.02
Any college	0.27	0.24	0.03*	0.02	0.33	0.22	0.11***	0.02
Log earnings age 37	1.78	1.86	-0.08***	0.02	1.44	1.41	0.03	0.02
Log average earnings age 37-47	1.86	1.94	-0.08***	0.02	1.63	1.59	0.04**	0.02
Full time in 1980	0.77	0.80	-0.03**	0.01	0.55	0.51	0.03*	0.02
Not employed in 1980	0.14	0.11	0.02*	0.01	0.21	0.21	-0.00	0.01
Unemployed in 2000	0.07	0.05	0.02	0.01	0.07	0.06	0.01	0.01
Professional	0.14	0.15	-0.01	0.01	0.12	0.07	0.05***	0.01
Legal or business	0.19	0.17	0.02	0.01	0.18	0.19	-0.00	0.01
STEM	0.09	0.16	-0.08***	0.01	0.08	0.05	0.03***	0.01
Blue collar	0.37	0.40	-0.02	0.02	0.09	0.10	-0.01	0.01
Clerical support	0.07	0.04	0.03***	0.01	0.17	0.20	-0.03*	0.01
Teacher-other health	0.08	0.07	0.01	0.01	0.25	0.22	0.04**	0.01
Service and sales	0.11	0.09	0.03***	0.01	0.12	0.15	-0.03***	0.01
Married by 1980	0.24	0.26	-0.02	0.02	0.41	0.44	-0.03	0.02
Married by 1990	0.58	0.63	-0.05**	0.02	0.71	0.73	-0.01	0.02
Married by 2000	0.67	0.70	-0.03	0.02	0.76	0.77	-0.01	0.01
Divorced by 1980 | married	0.09	0.07	0.02	0.02	0.15	0.10	0.04**	0.02
Divorced by 1990 | married	0.14	0.13	0.01	0.02	0.20	0.17	0.02	0.02
Divorced by 2000 | married	0.29	0.25	0.03	0.02	0.34	0.31	0.03	0.02
Teenage childbearing					0.01	0.02	-0.01*	0.01
Age at first birth					24.25	23.79	0.46***	0.13
Childlessness in 1980					0.53	0.45	0.08***	0.02
Childlessness in 1990					0.21	0.16	0.05***	0.01
Total fertility in 1980					0.70	0.84	-0.14***	0.03
Total fertility in 1990					1.69	1.76	-0.07	0.04
Mental health disorders	0.10	0.08	0.02**	0.01	0.07	0.07	-0.00	0.01
Substance abuse	0.07	0.05	0.02***	0.01	0.03	0.02	0.01	0.01
Leadership ability	-0.12	0.04	-0.16***	0.04				
Ability to function under stress	-0.09	0.07	-0.17***	0.04				
Student and family attitudes toward school								
Feeling safe at school	6.66	6.83	-0.17*	0.08	6.41	5.95	0.46***	0.08
Interest in school work	4.29	5.00	-0.71***	0.09	5.68	4.94	0.73***	0.09
Family's attitude toward ed	6.20	6.32	-0.13	0.08	6.58	5.83	0.75***	0.08
Adolescent risky behavior (age 14-18)								
Stealing	0.14	0.12	0.02	0.01	0.03	0.03	-0.00	0.01
Crimes of violence	0.06	0.05	0.01	0.01	0.00	0.01	-0.00	0.00
Abuse of alcohol & narcotics	0.04	0.03	0.02***	0.01	0.01	0.02	-0.00	0.00
Drunkenness & abuse of solvents	0.07	0.05	0.02**	0.01	0.01	0.02	-0.01*	0.00
Acting out	0.04	0.03	0.01**	0.01	0.03	0.03	0.00	0.01
Observations	1,001	3,982			1,042	4,129		

Note: The table provides statistics based on individual sex and gender conformity. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.2: Descriptive Statistics: Sociodemographic Background & IQ Components

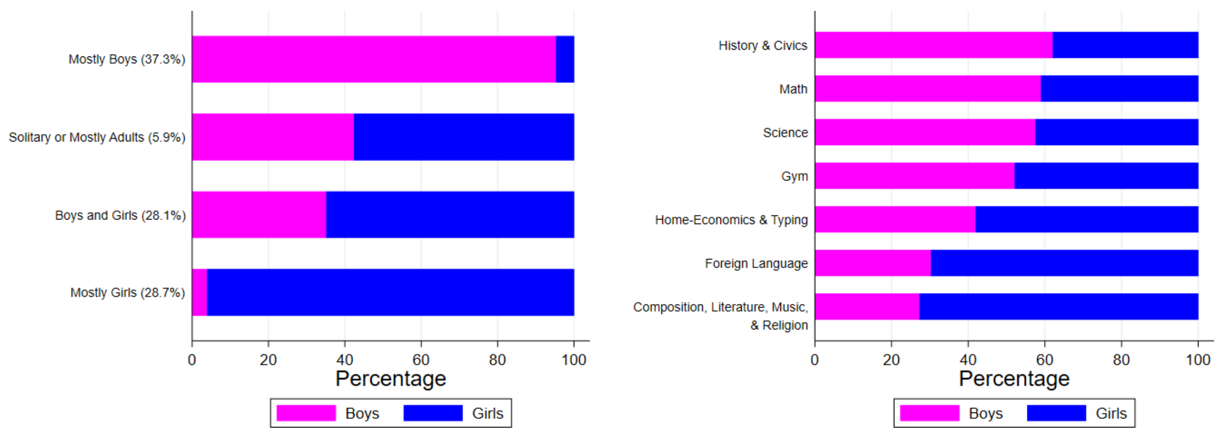
	Male				Female			
	CGN	GC	Diff	S.E.	CGN	GC	Diff	S.E.
Household characteristics								
Older brother	0.39	0.33	0.07***	0.02	0.33	0.37	-0.04**	0.02
Older sister	0.33	0.32	0.01	0.02	0.35	0.33	0.02	0.02
Younger brother	0.34	0.33	0.01	0.02	0.34	0.34	0.00	0.02
Younger sister	0.31	0.31	-0.00	0.02	0.33	0.33	0.00	0.02
Professional mother	0.05	0.04	0.01	0.01	0.05	0.03	0.02***	0.01
Working mother	0.19	0.19	0.01	0.01	0.18	0.19	-0.01	0.01
Female head of house	0.07	0.07	0.00	0.01	0.08	0.08	-0.00	0.01
Mother less than HS	0.93	0.93	-0.00	0.01	0.91	0.94	-0.03***	0.01
Mother any college	0.02	0.02	0.00	0.00	0.02	0.01	0.01**	0.00
Father less than HS	0.71	0.74	-0.02	0.02	0.68	0.76	-0.07***	0.02
Father any college	0.11	0.09	0.02**	0.01	0.12	0.08	0.04***	0.01
Single father	0.01	0.01	0.00	0.00	0.02	0.01	0.00	0.00
Single mother	0.06	0.05	0.00	0.01	0.05	0.06	-0.00	0.01
Home-ownership	0.20	0.18	0.01	0.01	0.21	0.17	0.04***	0.01
Cognitive test at age 13								
Numeric score	22.99	22.50	0.49	0.27	22.60	20.46	2.14***	0.25
Verbal score	25.81	25.82	-0.02	0.21	27.10	25.08	2.03***	0.22
Spatial score	22.28	24.63	-2.35***	0.24	24.08	22.11	1.97***	0.23
Observations	1,001	3,982			1,042	4,129		

Note: The table provides statistics by sex and gender conformity. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Construction of the GCI

Figure B.1: Underlying Gendered Variation

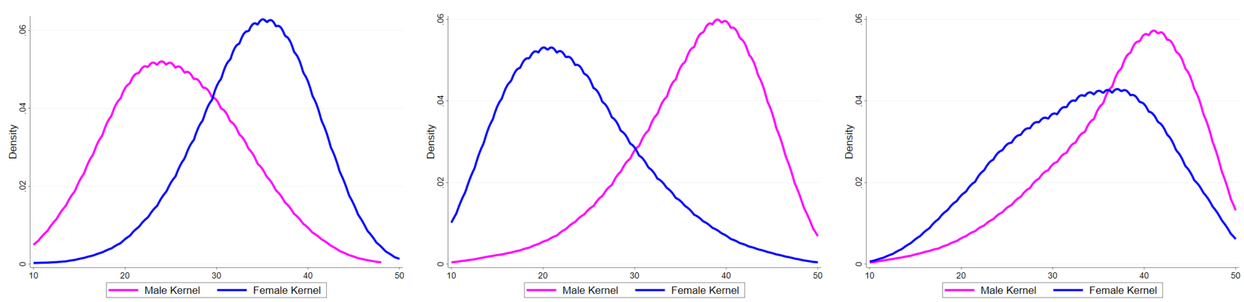
(a) Gender Homophily (b) Preferred School Subject



Note: (a) Students report with whom they spend their time. We present their responses in this figure by the respondents own gender. (b) Students report their favorite subject of study in school. We present their responses in this figure by sex of the individual in the study. The favorite school subject input variable is categorical. To use it in the PCA, we coded it to reflect whether it is female dominated (-1), male dominated (1), or neither (0). Data from Stockholm Birth Cohort.

Figure B.2: Interests in Leisure Activities by Gender

(a) Domestic interests (b) Mechanical interests (c) Sports



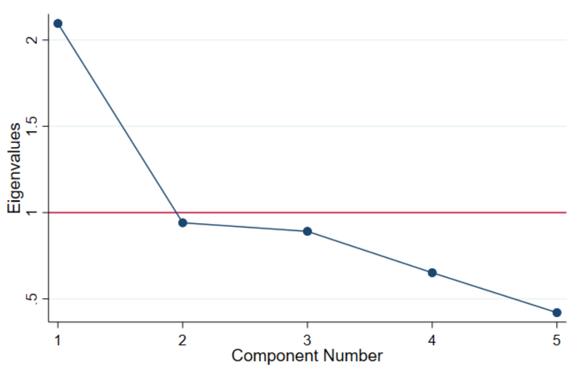
Note: The given three variables in the data each provide a numerical score from 10 to 50, where 10 indicates that the study person finds the area of activities “very boring” and 50 indicates that the study person finds the area of activities “very much fun”. Each figure was constructed using an in-built kernel density command while also splitting the data by gender. Data from Stockholm Birth Cohort.

Table B.1: Preferences for Leisure Time Interests

Domestic interests	**Mechanical interests**	**Sports**
Making clothes	Playing with model railways	Practicing gymnastics
Interior decoration	Visiting a museum of technology	Bike racing
Visiting an furniture exhibition	Repairing bikes	Practicing high jumping
Baking bread	Figuring out how a washing machine works	Practicing winter sports
Using a washing machine	Building a radio set	Coaching athletes
Using kitchen appliances	Mending mechanical toys	Playing basketball for a club
Working as a chef in a hotel	Assisting with the construction of a television	Cross-country running
Prepare a sausage dish for guests	Reading about space ships	Sailing
Cooking foreign dishes	Building models	Participating in an athletics event
Cooking a school meal	Constructing high-jump hurdles	Attending an athletics event

Note: The school survey questionnaire presented a list of items on preferences for leisure time interests. For each interest the student rated their preferences using a 4-step scale that ranged from "would be very much fun" to "would be very boring" to practice it. All items were scrambled in the questionnaire in order to elicit the true preference for each particular interest.

Figure B.3: Eigenvalues after PCA



Note: To create this figure, an in-built scree plot command was used after running principal component analysis. Data from Stockholm Birth Cohort.

Table B.2: PCA Input Variable Loadings

	Loadings
Domestic interests	0.4276
Mechanical interests	0.5454
Sports	0.3669
Who do you spend time with?	0.5538
Favorite school subject	0.2801

Note: The table provides loadings for each of the input variables of the principal component analysis.

B.1 Auxiliary data: Youth in Europe Study, 2010-2012

This longitudinal in-class school survey was conducted in three waves beginning with the cohort of 14-year-old students in 2010. The sampling of schools was stratified at three levels (school, classroom and student), to guarantee a sufficient representation of immigrants. Two complete classrooms were then drawn at random in each school within relevant grades. The school interviews consisted of a 45 minute self-completion questionnaire (including questions on friendship and classmate networks (6 nominations each) as well as a 30 minute written test in basic cognitive and language achievement. Table B.3 below describes the sampling frame and the number of YES! survey participants in waves 1 to 3. Figures B.4a and B.4b, and Table B.4 describe the inputs of the principal component analysis that was conducted as described in Section 3.2.3.

Table B.3: Youth in Europe Study: Study samples of the three selected countries

	Schools	Classrooms	Students		
			Wave 1	Wave 2	Wave 3
England	107	214	4,315	3,389	2,284
Germany	144	271	5,013	4,256	3,427
Sweden	129	251	5,025	4,531	2,768
Total	380	736	14,353	12,176	8,479

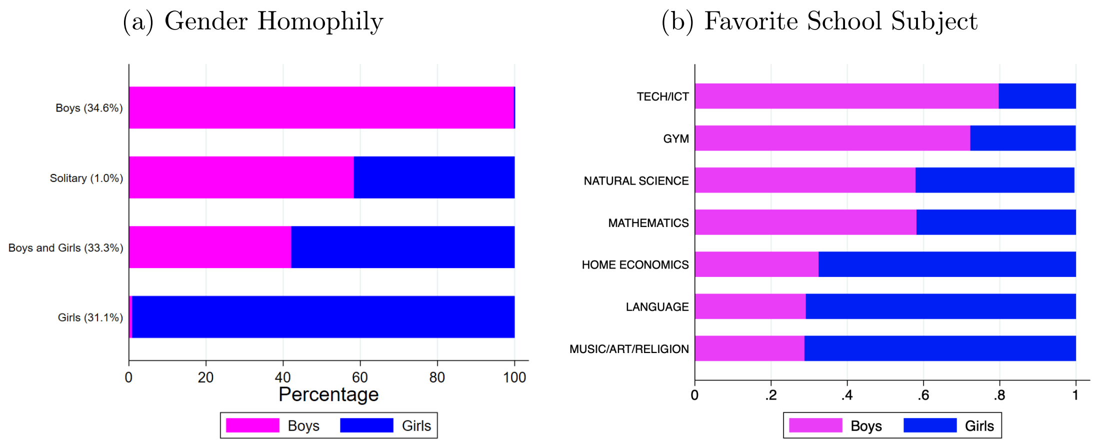
Note: Our study samples exclude all immigrants based on country of birth. The surveys of wave one and two were conducted in-class whereas for the third wave survey, the students were contacted through postal and web questionnaires or phone interviews when necessary. In the third wave the students were age 16 and some had already left school while others had transitioned to upper secondary school.

Table B.4: Youth in Europe Study: Preferences for Leisure Time Activities and Attitudes Towards Domestic Activities

Leisure time activities				Domestic activities	
Feminine	t-value	Masculine	t-value	Feminine	t-value
Reading	14.4	Video games, alone	36.75	Child care	10.9
Chatting	4.74	Video games,		Cooking	10.5
Homework	6.33	with others	38.1	Cleaning	12.68

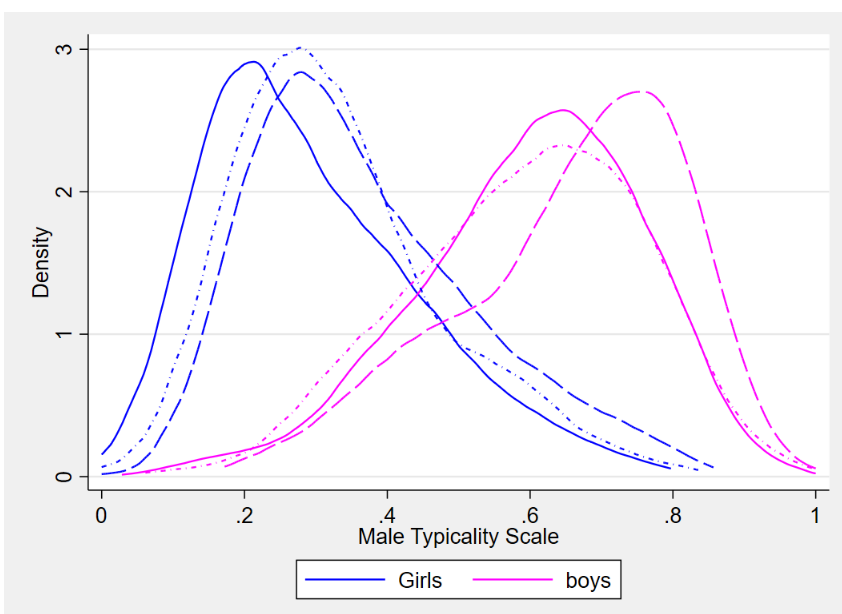
Note: Data from YES! (Kalter et al., 2016). Here, to describe the data, we show results for the Swedish survey (results for England and Germany are available from the authors). Leisure time activities inquire on a 5-step scale about the time spent on practicing the particular activity ranging from "No time at all" to "More than two hours a day". Domestic activities inquire about the attitudes towards whether the man or the woman in a household should do the domestic activities, the 3-step scale ranging from "Mostly the man" to "Mostly the woman". The t-value reports the statistic of the t-test of whether the activities are female-typed or male-typed among both genders pooled.

Figure B.4: Youth in Europe Study: Distribution of Gendered Variation



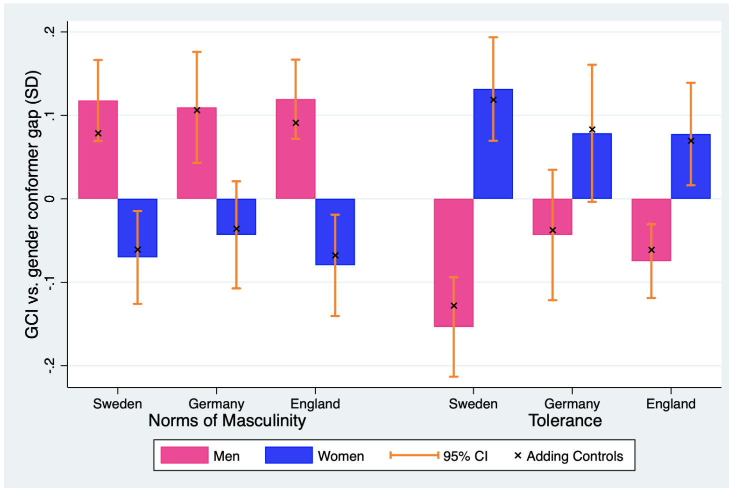
Note: Data from YES! (Kalter et al., 2016). Here, to describe the data, we show results for the Swedish survey (results for England and Germany are available from the authors). (a) Students nominate their five best friends with whom they spend their time within or outside of school and the gender of each, respectively. We coded groups of five best friends including both boys and girls as mixed gender reference groups. We required all five friends to be of opposite gender for opposite-gender reference groups. "Solitary" was categorized based on answer (yes/no) to the statement "Í do not have any friends". (b) Students report their favorite subject of study in school. We present their responses in this figure by gender. The favorite school subject input variable is categorical. To use it in the PCA, we coded it to reflect whether it is female dominated (-1), male dominated (1), or neither (0).

Figure B.5: Gender Typicality by Sex in England, Germany and Sweden in 2010



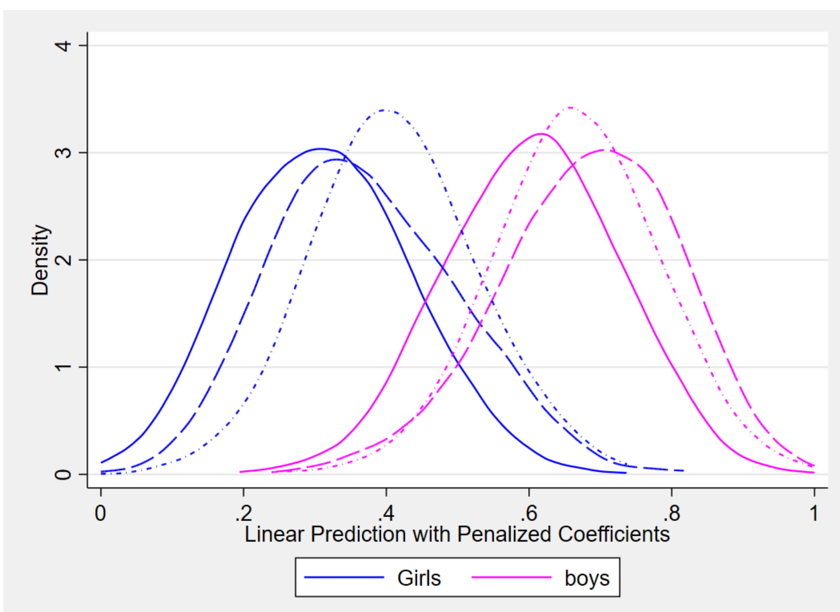
Note: Data from YES! (Kalter et al., 2016). The gender typicality distributions for England (solid), Germany (dashed) and Sweden (dashed and dotted) are based on the normalized first factor of the pca equivalent to the one presented for our primary sample in Section 3.2.2.

Figure B.6: GCI and Gender Norms in Contemporary School Surveys



Note: Data from YES! (Kalter et al., 2016). Samples exclude migrant children. Vertical axis measures the gap in the average norms due to a change of a standard deviation in GCI across genders and countries. Masculine norms collects (by Crombachs alfa) information on how much the respondent agrees with men having to use violence in different contexts (y1_masc1-y1_masc3). Tolerance collects (by Crombachs alfa) information on how much the respondent is at ease with a couple cohabiting without being married, divorce, homosexuality and abortion (y1_tol1-y1_tol4). We include school fixed-effects. Black x symbols indicate the value of the coefficient when the list of controls is expanded to include cognitive achievement.

Figure B.7: Predicted Gender by Assigned Sex in England, Germany and Sweden in 2010



Note: Data from YES! (Kalter et al., 2016). The predicted probability of being boy distributions for England (solid), Germany (dashed) and Sweden (dashed and dotted) are based on the linear prediction with penalized coefficients following (Mittleman, 2022). See Section 3.2.4 for the details of the construction of this measure.

Table B.5: Comparing the GCI to Predicted Gender in Contemporary School Surveys

	Sweden		Germany		England	
	Men	Women	Men	Women	Men	Women
	Dependent variable: Norms of Masculinity					
Gender conformity index (GCI)	0.766***	-0.413**	0.688***	-0.271	0.752***	-0.502**
	(0.171)	(0.203)	(0.253)	(0.244)	(0.181)	(0.232)
Predicted gender	1.069***	-0.327	1.023***	-0.066	1.234***	-0.448
	(0.266)	(0.249)	(0.384)	(0.243)	(0.209)	(0.390)
Observations	1,290	1,345	1,467	1,464	1,312	1,354
	Dependent variable: Tolerance					
Gender conformity index (GCI)	-0.945***	0.831***	-0.272	0.492	-0.471***	0.488**
	(0.224)	(0.232)	(0.297)	(0.312)	(0.168)	(0.234)
Predicted gender	-0.753***	0.787**	-0.254	0.288	-0.658**	1.003**
	(0.272)	(0.315)	(0.397)	(0.407)	(0.428)	(0.297)
Observations	1,284	1,350	1,460	1,467	1,300	1,349

Note: Data from YES! (Kalter et al., 2016). *Predicted gender* refers to the predicted probability of being assigned the sex boy based on survey answers. See Section 3.2.4 for the details of the construction of this measure. Samples exclude migrant children and schools with less than 15 students in 8th grade (i.e, below 4% of schools in all three auxiliary samples). All entries in the table come from separate multivariate regressions using as the main explanatory variable either Gender conformity index or Predicted gender. All regressions include school fixed effects. Standard errors are clustered at the school level.

C Construction of Variables

Socio-emotional variables. The psychologists assign each conscript's military aptitude (ability to cope with stress) a Stanine score from 1 to 9, a high score meaning high aptitude. This score is in turn based on four different subscores which range from 1 to 5. The subscores work only as a guide to the psychologists—two conscripts with the same sequence of subscores could still get different final scores. In addition to ability to cope with stress, leadership skills are evaluated for those that high enough cognitive scores to become considered for leadership training. The scale of leadership capacity follows the same Stanine distribution, ranging from 1 to 9, as the more general "ability to function under stress" variable. See Lindqvist and Vestman (2011) for a detailed account of the military enlistment's psychological test procedure. We normalize both variables to have mean zero and variance one.

Cognitive tests. We use the three components of ability collected by the school survey in 1966 (at age 13): numeric, verbal, and spatial ability. The tests were constructed at the Swedish Institute for Educational Research in the early 1960s and have served to this date as the default cognitive tests in elementary school (Svensson, 1964).

The verbal and numeric tests are weighted more toward crystallized intelligence (Cattell, 1971). Scores on crystallized intelligence tests are in part determined by innate ability but also by acquired skills and knowledge and are thus depending on educational opportunity and motivation (Borghans et al., 2008, 2016). Some researchers suggest that the numeric and verbal ability tests might therefore more appropriately be called achievement tests than intelligence test (Almlund et al., 2011). The spatial ability test is weighted more towards fluid intelligence (Cattell, 1971), which is often considered the more innate of the two measures of intelligence (Svensson, 1971).

- Numeric ability - The test of numeric ability posed 40 numerical sequences of six numbers, each of which follows a logical pattern based on elementary arithmetic concepts. The students were asked to predict the next two numbers following the same pattern in the sequence. Our measure reports the test score.
- Verbal ability - The verbal ability test presented the student with 40 words, for which the student had to find antonyms among four options.
- Spatial ability - The spatial ability test consisted of 40 unfolded figures that needed to be folded mentally.

Network graphs. The classroom survey conducted for sixth graders (age 13) asked students were asked to nominate their three best classroom friends (the nominations only concerned friends within the same classroom). Of all students in our data who participated in the school survey and completed the friendship nomination component (n=11,854) , 7,497 nominated three friends (63.2%), 3,211 nominated two friends (27.1%), 766 nominated only one friend (6.5%), and 380 did not nominate any friends (3.2%). In our analytic sample (n=10,154), 6,586 students nominated three friends (65.4%), 2,610 students nominated two friends (25.9%), 558 students nominated only one friend (5.5%), and 324 students did not nominate any friends (3.2%).

- Nominating only one friend - This binary variable states whether the student only nominated one friend in their classroom given they had the option to nominate more.
- At least one friend is CGN - This binary variable states whether the student nominated one or more friends who were gender-nonconforming.
- Average verbal/numeric/spatial score of friends - This continuous variable gives the average verbal/numeric/spatial score of the student's nominated friends.
- Clustering - The degree to which students in the network tend to group in cliques. See Section Measuring social capital in Chetty et al. (2022) for an illustrative description.
- Diameter - The shortest distance in terms of number of edges between the two most distant nodes in the network.

Attitude toward school. The school survey in 1966 (at age 13) was a classroom survey that collected information on students' feelings toward school. These variables are the sum of ten questions. Each of the ten questions was a "yes" or "no" question. Answers were assigned 0 or 1 points and tallied to create a final variable.

- Student's feeling safe at school - This variable states how safe/secure the student felt at school. The variable is from 0 (very unsafe/insecure) to 10 (very safe/secure).
- Student's interest in school work - This variable states how interested the student was in school. The variable is from 0 (very uninterested) to 10 (very interested).

D Robustness Checks

Table D.1: Gender Conformity Index Score and Life Outcomes

	Explanatory variable: *Gender Conformity Index*					
	Sample: Men			Sample: Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Educational outcomes						
GPA in grade 9	19.378***	(6.456)	4,746	-56.031***	(5.534)	4,955
Upper secondary dropout	-0.072*	(0.041)	4,588	0.195***	(0.036)	4,808
Any post secondary	0.022	(0.041)	4,588	-0.201***	(0.038)	4,808
STEM secondary track	0.651***	(0.050)	3,374	-0.413***	(0.037)	3,469
Any college	-0.042	(0.036)	4,588	-0.169***	(0.033)	4,808
Log earnings outcomes						
Log earnings age 37	0.263***	(0.056)	4,716	-0.055	(0.050)	4,880
Log average earnings age 37-47	0.297***	(0.064)	4,735	-0.105**	(0.044)	4,903
Labor market outcomes						
Full time in 1980	0.096***	(0.037)	4,803	-0.064	(0.039)	4,960
Not employed in 1980	-0.055*	(0.030)	4,803	0.029	(0.032)	4,960
Unemployed in 2000	-0.062***	(0.022)	4,585	0.003	(0.020)	4,788
Professional	0.019	(0.032)	4,090	-0.040	(0.026)	3,699
Occupational outcomes†						
Legal or business	-0.088***	(0.029)	4,947	-0.060**	(0.027)	5,138
STEM	0.243***	(0.030)	4,947	-0.077***	(0.016)	5,138
Blue collar	0.056	(0.039)	4,947	-0.007	(0.023)	5,138
Clerical support	-0.060***	(0.016)	4,947	0.092***	(0.029)	5,138
Teacher-other health	-0.030	(0.020)	4,947	-0.042	(0.030)	5,138
Service and sales	-0.064***	(0.022)	4,947	0.080***	(0.027)	5,138
Did not work	-0.058***	(0.021)	4,947	0.014	(0.022)	5,138
Marriage & fertility outcomes						
Married by 1980	0.077**	(0.036)	4,825	0.015	(0.038)	4,991
Married by 1990	0.139***	(0.042)	4,825	-0.003	(0.035)	4,991
Divorced by 1980 | married	-0.050	(0.052)	1,217	-0.114***	(0.041)	2,159
Divorced by 1990 | married	-0.041	(0.037)	2,974	-0.095***	(0.036)	3,614
Teenage childbearing				0.030***	(0.010)	5,171
Age at first birth by 1980				-0.638**	(0.287)	2,749
Childlessness in 1980				-0.115***	(0.038)	5,171
Childlessness in 1990				-0.066**	(0.030)	4,914
Total fertility by 1980				0.163**	(0.065)	5,171
Total fertility by 1990				0.034	(0.095)	4,880
Social emotional outcomes						
Mental health disorders	-0.093*	(0.026)	4,826	0.025	(0.020)	4,991
Substance abuse	-0.084***	(0.022)	4,826	-0.007	(0.012)	4,991
Leadership ability	0.556***	(0.103)	3,612			
Ability to function under stress	0.516***	(0.089)	4,492			

Note: This table provides the results of regressions of the variables in the rows on the continuous gender conformity index as shown in Figure 1(*b*). See the top panel of Table A.2 for the complete set of sociodemographic covariates included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.2: Gender Conformity and Life Outcomes adjusting for Cognitive Ability

	Explanatory variable: *Gender nonconformity*					
	Sample: Men			Sample: Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Educational outcomes						
GPA in grade 9	1.518	(2.598)	4,746	15.803***	(2.380)	4,955
Upper secondary dropout	0.004	(0.017)	4,588	-0.055***	(0.016)	4,808
Any post secondary	0.021	(0.017)	4,588	0.055***	(0.017)	4,808
STEM secondary track	-0.168***	(0.021)	3,374	0.150***	(0.017)	3,469
Any college	0.032**	(0.015)	4,588	0.075***	(0.016)	4,808
Log earnings outcomes						
Log earnings age 37	-0.074**	(0.023)	4,716	0.022	(0.021)	4,880
Log average earnings age 37-47	-0.069***	(0.024)	4,735	0.015	(0.019)	4,903
Labor market outcomes						
Full time in 1980	-0.036**	(0.015)	4,803	0.027	(0.017)	4,960
Not employed in 1980	0.019	(0.012)	4,803	0.004	(0.015)	4,960
Unemployed in 2000	0.013	(0.009)	4,585	0.013	(0.009)	4,788
Professional	>0.001	(0.013)	4,090	0.027**	(0.012)	3,699
Occupational outcomes†						
Legal or business	0.020	(0.013)	4,947	-0.005	(0.013)	5,138
STEM	-0.073***	(0.014)	4,947	0.018**	(0.007)	5,138
Blue collar	-0.030*	(0.016)	4,947	-0.002	(0.011)	5,138
Clerical support	0.025***	(0.007)	4,947	-0.010	(0.014)	5,138
Teacher-other health	0.015*	(0.009)	4,947	0.025*	(0.014)	5,138
Service and sales	0.024***	(0.009)	4,947	-0.020	(0.013)	5,138
Did not work	0.018**	(0.009)	4,947	-0.005	(0.010)	5,138
Marriage & fertility outcomes						
Married by 1980	-0.021	(0.016)	4,825	-0.024	(0.018)	4,991
Married by 1990	-0.043**	(0.018)	4,825	-0.019	(0.016)	4,991
Divorced by 1980 | married	0.018	(0.023)	1,217	0.059***	(0.021)	1,930
Divorced by 1990 | married	0.007	(0.016)	2,974	0.038**	(0.018)	3,246
Teenage childbearing				-0.008**	(0.004)	5,171
Age at first birth by 1980				0.256*	(0.132)	2,749
Childlessness in 1980				0.057***	(0.017)	5,171
Childlessness in 1990				0.049***	(0.014)	4,914
Total fertility by 1980				-0.097***	(0.030)	5,171
Total fertility by 1990				-0.083*	(0.043)	4,880
Social emotional outcomes						
Mental health disorders	0.015	(0.011)	4,826	0.008	(0.009)	4,991
Substance abuse	0.017*	(0.009)	4,826	0.013**	(0.006)	4,991
Leadership ability	-0.152***	(0.042)	3,612			
Ability to function under stress	-0.149***	(0.039)	4,492			

Note: This table provides the results of the regressions of equation (2). All models adjust for spatial ability test scores in grade 6. See top panel of Table A.2 for the complete set of sociodemographic covariates included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

† Results from multinomial logit (excluding school fixed effects). Marginal effect reported.

Table D.3: Gender Conformity and Life Outcomes with Restricted Sample

	Explanatory variable: *Gender nonconformity*					
	Sample: Men			Sample: Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Educational outcomes						
GPA in grade 9	-5.378*	(2.762)	4,730	22.264***	(2.731)	4,454
Upper secondary dropout	0.030*	(0.017)	4,573	-0.081***	(0.017)	4,323
Any post secondary	-0.005	(0.017)	4,573	0.082***	(0.018)	4,323
STEM secondary track	-0.204***	(0.022)	3,364	0.170***	(0.019)	3,139
Any college	0.014	(0.015)	4,573	0.084***	(0.017)	4,323
Log earnings outcomes						
Log earnings age 37	-0.076***	(0.024)	4,701	0.023	(0.023)	4,386
Log average earnings age 37-47	-0.086***	(0.025)	4,720	0.023	(0.021)	4,407
Labor market outcomes						
Full time in 1980	-0.033**	(0.015)	4,787	0.024	(0.018)	4,455
Not employed in 1980	0.019	(0.012)	4,787	-0.006	(0.015)	4,455
Unemployed in 2000	0.013	(0.009)	4,570	0.015	(0.010)	4,305
Professional	-0.015	(0.014)	4,080	0.028**	(0.013)	3,320
Occupational outcomes†						
Legal or business	0.014	(0.013)	4,930	0.009	(0.014)	4,615
STEM	-0.088***	(0.015)	4,930	0.027***	(0.007)	4,615
Blue collar	-0.010	(0.017)	4,930	0.002	(0.011)	4,615
Clerical support	0.025***	(0.007)	4,930	-0.022	(0.015)	4,615
Teacher-other health	0.012	(0.009)	4,930	0.024	(0.015)	4,615
Service and sales	0.029***	(0.010)	4,930	-0.037***	(0.014)	4,615
Did not work	0.017**	(0.009)	4,930	-0.003	(0.011)	4,615
Marriage & fertility outcomes						
Married by 1980	-0.021	(0.016)	4,809	-0.027	(0.019)	4,486
Married by 1990	-0.049***	(0.018)	4,809	-0.015	(0.017)	4,486
Divorced by 1980 | married	0.016	(0.022)	1,212	0.055***	(0.019)	2,159
Divorced by 1990 | married	0.009	(0.017)	2,968	0.037**	(0.016)	3,614
Teenage childbearing				-0.010**	(0.004)	4,646
Age at first birth by 1980				0.285**	(0.134)	2,473
Childlessness in 1980				0.067***	(0.018)	4,646
Childlessness in 1990				0.059***	(0.015)	4,419
Total fertility by 1980				-0.111***	(0.031)	4,646
Total fertility by 1990				-0.098**	(0.044)	4,386
Social emotional outcomes						
Mental health disorders	0.018*	(0.010)	4,810	0.003	(0.009)	4,486
Substance abuse	0.019**	(0.008)	4,810	0.012*	(0.006)	4,486
Leadership ability	-0.187***	(0.041)	3,602			
Ability to function under stress	-0.150***	(0.037)	4,479			

Note: For these results, we drop individuals in the bottom fifth percentile of reported interest in all three categories of leisure interests (domestic interests, mechanical interests, and sports) taken together. See the top panel of Table A.2 for the complete set of sociodemographic covariates included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

† Results from multinomial logit (excluding school fixed effects). Marginal effect reported.

Table D.4: Gender Conformity (PCA without Favorite School Subject) and Life Outcomes

	Explanatory variable: *Gender nonconformity*					
	Sample: Men			Sample: Women		
Outcome:	Coeff.	S.E.	Obs.	Coeff.	S.E.	Obs.
Educational outcomes						
GPA in grade 9	-2.965	(2.717)	4,746	17.528***	(2.558)	4,955
Upper secondary dropout	-0.000	(0.018)	4,588	-0.077***	(0.017)	4,808
Any post secondary	0.022	(0.018)	4,588	0.085***	(0.017)	4,808
STEM secondary track	-0.168***	(0.021)	3,374	0.151***	(0.016)	3,469
Any college	0.021	(0.015)	4,588	0.058***	(0.015)	4,808
Log earnings outcomes						
Log income age 37	-0.072***	(0.024)	4,716	0.020	(0.022)	4,880
Log average income age 37-47	-0.081***	(0.024)	4,735	0.041**	(0.019)	4,903
Labor market outcomes						
Full time in 1980	-0.065***	(0.015)	4,803	0.027	(0.018)	4,960
Not employed in 1980	0.027**	(0.012)	4,803	-0.015	(0.015)	4,960
Unemployed in 2000	0.013	(0.009)	4,585	0.005	(0.009)	4,788
Professional	-0.019	(0.014)	4,090	0.014	(0.011)	3,699
Occupational outcomes†						
Legal or business	0.027**	(0.013)	4,947	0.005	(0.014)	5,138
STEM	-0.044***	(0.013)	4,947	0.024***	(0.007)	5,138
Blue collar	-0.045***	(0.017)	4,947	0.010	(0.010)	5,138
Clerical support	0.016**	(0.007)	4,947	-0.051***	(0.015)	5,138
Teacher-other health	0.010	(0.009)	4,947	0.040***	(0.014)	5,138
Service and sales	0.007	(0.009)	4,947	-0.034***	(0.013)	5,138
Did not work	0.028***	(0.008)	4,947	0.006	(0.010)	5,138
Marriage & fertility outcomes						
Married by 1980	0.006	(0.016)	4,825	-0.020	(0.018)	4991
Married by 1990	0.006	(0.017)	4,825	-0.041**	(0.017)	4991
Divorced by 1980 | married	0.027	(0.023)	1,217	0.015	(0.018)	2,159
Divorced by 1990 | married	0.019	(0.015)	2,974	0.016	(0.016)	3,614
Teenage childbearing				0.002	(0.005)	5,171
Age at first birth by 1980				-0.118	(0.118)	2,749
Childlessness in 1980				0.012	(0.018)	5,171
Childlessness in 1990				0.037**	(0.015)	4,914
Total fertility by 1980				-0.022	(0.031)	5,171
Total fertility by 1990				-0.077*	(0.041)	4,880
Social emotional outcomes						
Mental health disorders	0.028***	(0.010)	4,826	0.006	(0.009)	4,991
Substance abuse	0.024***	(0.008)	4,826	0.011*	(0.006)	4,991
Leadership ability	-0.257***	(0.041)	3,612			
Ability to function under stress	-0.305***	(0.037)	4,492			

Note: For these results, we drop the favorite school subject variable from the principle component analysis. This table provides the results of the regressions of equation (2). See Table 2 for the details of the sociodemographic covariates included. All regressions include school fixed effects. The sample has 189 schools and 543 classrooms. Standard errors (in parentheses) are bootstrapped (1,000 draws) using random sampling stratified on school. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

† Marginal effects from a multinomial logit (excluding school fixed effects).

E CGN Students and the Classroom Network

In this appendix, we explore how the availability of CGN students affect their classmates' social capital as measured by social network cohesion. To do so, we use the sociometric matrices based on the friendship links (three best friends) ascertained through the in-class school survey in 1966. Given that we observe an entire cohort, we are able to characterize complete classroom networks in grade six. We ask specifically whether having more CGN students in the classroom changes its social cohesion. Thus, we regress standard network-cohesion statistics (i.e., clustering and diameter) of classroom c in school s, NC_{cs}, with the fraction of CGN students in the classroom $\text{C}\bar{\text{G}}\text{N}_{cs}$.[25]

$$\text{NC}_{cs} = \alpha_0 + \text{C}\bar{\text{G}}\text{N}_{cs}\beta + \omega_s + \epsilon_{cs} \tag{3}$$

We adopt an identification strategy similar to Hoxby (2000), that exploits the within-school across-classroom variation in the share of CGN students in the classroom (all students in our data belong to the same cohort). The school fixed effects, ω_s, in equation (3) controls for sorting of students across schools. Causal identification of β requires random allocation of students into classrooms. Here, we take advantage of the institutional framework that administered student assignments to Swedish primary schools (grades 1 to 6) in the 1960s. In particular, students attended the nearest school in the neighborhood, and tracking based on ability or background was not allowed in the first six grades (Husen, 1961; Paulston, 1966; SOU1961:30, 1961). Further, a homogenous curriculum and a fixed number of weekly hours of instruction resulted in what explicitly was called a primary school system absent of any "organizational differentiation" of students with respect to ability or social background (SOU1961:30, 1961). The only homogenous groupings allowed during the first six grades of Swedish comprehensive school were special education classes for students with special needs (Husen, 1961). Santavirta and Sarzosa (2024) show reassuring evidence of non-sorting of abused and neglected students in the same context and cohort.

[25]See C for variable definitions of both network-cohesion statistics. Also, Chetty et al. (2022) provide an illustrative description of the measurement of clusterning.

Table E.1: Effect Share of Gender Nonconformers on Social Cohesion

Explanatory variable	Outcomes: Clustering		Diameter	
Class CGN fraction male	-0.027	-0.020	-0.132	-0.148
	(0.040)	(0.040)	(0.137)	(0.139)
Class CGN fraction female	-0.121***	-0.109***	-0.372***	-0.373***
	(0.040)	(0.040)	(0.137)	(0.136)
Constant	0.494***	0.490***	0.789***	0.792***
	(0.013)	(0.013)	(0.044)	(0.044)
Observations	536	536	536	536
PCA includes homophily?	Yes	No	Yes	No

Note: This table provides the results of the regressions of equation where the fraction of gender-nonconforming men and women are included as explanatory variables (3). These regressions control for school fixed effects. The sample has 185 schools and 536 classrooms. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results in Table E.1 shows that as the fraction of gender-nonconforming students in the classroom increases, the classroom networks are significantly less clustered and the relative diameter of the classroom is significantly less. That means that the presence of female CGN students makes classrooms' social networks more cohesive and less fractured. That is, CGN women and not men serve as bridge-builders between social cliques–usually ones defined by gender in this age group. CGN girls help connect students who otherwise would belong to disjoint social networks, despite being in the same classroom. In addition, our results do not change when we exclude homophily from the principle component analysis, which indicates that the results are not mechanically driven by the homophily input variable.

F Extended Roy Model with Factor Structure

F.1 Model of occupational choice and earnings

We consider occupational choices stemming from the perceived monetary rewards and the nonpecuniary benefits (e.g., job flexibility) or costs (e.g., social rejection) that can affect the individual's fit for the occupation. We assume that individuals make their occupation decisions based on a comparison of the expected benefits and costs associated with each alternative. Specifically, if V_o denotes the expected benefit associated with occupation o, then

$$V_o = E\left[\sum_{t=1}^{T} \rho^{t-1} u(Y_o(t), C_o(t)) | \mathcal{I}_0, \mathcal{P}_0\right]$$

where $u(\cdot)$ represents the per period utility function, $Y_o(t)$ represents the total earnings received in period t given occupation o, $C_o(t)$ is a psychic benefit (or cost) associated with occupation o, ρ is the discount factor, $\mathcal{I}_o$ represents the information set available to the agent at $t = 0$, and $\mathcal{P}_o$ represents the gender norms affecting each occupation. We can write earnings and psychic benefits in occupation o as

$$Y_o(t) = \mu_{ot}^{Y}(X_Y) + \eta_{Y_o t}, \qquad C_o(t) = \mu_{ot}^{C}(X_C) + \eta_{C_o t}$$

where X_Y and X_C are observable variables, $\eta_{Y_o t}$ and $\eta_{C_o t}$ are unobservables, and $(X_Y, X_C) \perp (\eta_{Y_o t}, \eta_{C_o t})$ (Heckman and Navarro, 2007; Heckman and Vytlacil, 2007). We assume that $\eta_{Y_o t}$ and $\eta_{C_o t}$ follow a factor structure that separates the individual's endowments $\boldsymbol{\Theta}$ from uncorrelated variation (Carneiro et al., 2003; Hansen et al., 2004; Heckman et al., 2018). As in our conceptual framework in Section 2, these endowments include gender nonconformity and skills. Thus, $\boldsymbol{\Theta} = [\theta^{CGN} \quad \theta^{H}]$, where θ^{CGN} and θ^{H} represent the unobserved gender nonconformity and skills, respectively.

$$\eta_{Y_o t} = \alpha_t^{Y_o,G} \theta^{CGN} + \alpha_t^{Y_o,H} \theta^{H} + e_t^{Y_o}, \qquad \eta_{C_o t} = \alpha_t^{C_o,G} \theta^{CGN} + \alpha_t^{C_o,H} \theta^{H} + e_t^{C_o}$$

where $\boldsymbol{\Theta} \perp (e_t^{Y_o}, e_t^{C_o})$ and e's are mutually independent. We assume that, although the endowments are unobserved to the econometrician, they are known to the agent and are constant over time (Urzua, 2008). Thus, $\theta^H \in \mathcal{I}_0$ and $\theta^{CGN} \in \mathcal{P}_0$ as agents use them to make decisions just like they do with observable characteristics X_Y and X_C. The individual selects her occupation o^* by comparing the expected utility levels V_o across the different alternatives in the set $\mathcal{O}$. If we assume, for simplicity a linear utility function, and given the elements in sets $\mathcal{I}_0$ and $\mathcal{P}_0$, the occupation choice is given by

$$o^* = \arg\max_{o \in \mathcal{O}} \sum_{t=1}^{T} \rho^{t+1} \left(\mu_{ot}^Y(X_Y) + \alpha_t^{Y_o,G}\theta^{CGN} + \alpha_t^{Y_o,H}\theta^H + \mu_{ot}^C(X_C) + \alpha_t^{C_o,G}\theta^{CGN} + \alpha_t^{C_o,H}\theta^H \right)$$

Suppose there are two possible occupation types: STEM s and Non-STEM s' such that $\mathcal{O} = \{s, s'\}$. A person will choose a STEM occupation if

$$\begin{aligned} \mu_s^Y(X_Y) - \mu_{s'}^Y(X_Y) + \mu_s^C(X_C) - \mu_{s'}^C(X_C) + (\alpha^{Y_{s,G}} - \alpha^{Y_{s',G}} + \alpha^{C_{s,G}} - \alpha^{C_{s',G}})\theta^{CGN} \\ (\alpha^{Y_{s,H}} - \alpha^{Y_{s',H}} + \alpha^{C_{s,H}} - \alpha^{C_{s',H}})\theta^H > 0 \end{aligned}$$

If $X_Y \subseteq X_C$, and given the independence of e's, we can describe the occupation choice as a function of $(X_C, \theta^{CGN}, \theta^H)$ such as STEM $= \mathbb{1}\left[\mu_t^C(X_C) + \alpha^{C_G}\theta^{CGN} + \alpha^{C_H}\theta^H + e^C > 0\right]$, where $\mathbb{1}$ is an indicator function that takes the value of 1 if the condition holds and 0 otherwise (Willis and Rosen, 1979). Depending on the occupational choice, we observe Y_o. If the individual's optimal response is $o^* = STEM$ then we observe Y_1, otherwise we observe Y_0. Thus, empirically, the model can be described by

$$\begin{aligned} \text{STEM} &= \mathbb{1}[\mathbf{x}_C\beta^C + \alpha^{C_G}\theta^{CGN} + \alpha^{C_H}\theta^H + e^C > 0] \\ y_0 &= \mathbf{x}_Y\beta^{Y_0} + \alpha^{Y_{0,G}}\theta^{CGN} + \alpha^{Y_{0,H}}\theta^H + e^{Y_0} \quad \text{if } D = 0 \\ y_1 &= \mathbf{x}_Y\beta^{Y_1} + \alpha^{Y_{1,G}}\theta^{CGN} + \alpha^{Y_{1,H}}\theta^H + e^{Y_1} \quad \text{if } D = 1 \end{aligned} \tag{4}$$

F.2 Test scores as measurement system of latent factors

In order to estimate the Roy model of potential outcomes in (4), we rely on the factor structures governing η's that result in the independence of e's, once we identify $\boldsymbol{\Theta}$ (Heckman et al., 2006; Prada and Urzúa, 2017; Heckman et al., 2018).[26] For this purpose, we consider that ability and gender nonconformity are latent factors. In this sense, we acknowledge that we do not directly observe skills or the degree to which an individual conforms to the prevailing gender norms and prescriptions. Rather, we indirectly infer skills and conformity from the variation we observe in manifest scores (in the case of skills) or in responses to seemingly irrelevant questions about preferences and behaviors (in the case of gender nonconformity). In a latent factor framework, researchers recover unobserved variation $\mathbf{\Theta}$ using manifest information $\mathbf{T}$ that is known to be affected by said variation, some contexts $\mathbf{X}_T$ and some error $\mathbf{e}^{\mathbf{T}}$ (Heckman et al., 2006; Bartholomew et al., 2011). That is, we consider a linear relation between the unobserved factors θ^{CGN} and θ^H, gender-nonconformity and skills, and the (vectors of) manifest variables $\mathbf{T}^{CGN}$ and $\mathbf{T}^H$. Namely,

$$\mathbf{T}^{CGN} = \mathbf{X}_T\beta^{T_G} + \alpha^{T_G}\theta^{CGN} + \mathbf{e}^{\mathbf{T_G}} \tag{5}$$

$$\mathbf{T}^{H} = \mathbf{X}_T\beta^{T_H} + \alpha^{T_H}\theta^{H} + \mathbf{e}^{\mathbf{T_H}} \tag{6}$$

where $\mathbf{X}_T$ is a matrix with all observable controls for each measurement and α^{T^G} and α^{T^H} are vectors containing the factor loadings in each measurement T. In Sub-section F.3, we show that based on the system of equations (5) and (6), we can identify the distribution of the unobserved heterogeneity θ, clean from the influence of $\mathbf{X}_T$ or the error term $\mathbf{e}^{\mathbf{T}}$ (Carneiro et al., 2003; Hansen et al., 2004). There are multiple advantages of pursuing this approach. First, we flexibly estimate the unobserved factors' distributions $f_{\theta^{CGN}}(\cdot)$ and $f_{\theta^H}(\cdot)$ using mixtures of normals. That help us fit the factor's true underlying distribution without strong functional form assumptions (Judd, 1998). Second, we identify the factor while controlling for predetermined characteristics $\mathbf{X}_T$ that have an influence on reporting of the measures we use. For instance, our results partial-out the correlation that SES might have on the *reporting* of the manifest variables and the outcomes. Third, the factor is continuous. That way, we observe the effects of different degrees of gender nonconformity

[26]Identification of the parameters of the Roy model require the assumption of $(e^C \perp e^{Y_0} \perp e^{Y_1})$. We assume that this assumption holds conditioned on observable $\mathbf{x}$ and unobservable characteristics θ^H and θ^{CGN}. Though $\mathbf{x}_C \neq \mathbf{x}_Y$, we recognize the lack of natural exclusion restrictions. Yet, they are not needed to formally secure the identification of the parameters of interest. Rather than natural exclusion restrictions, we rely on identification through functional form (Carneiro et al., 2003; Sarzosa and Urzua, 2016).

on adult outcomes.

As we indicate below, identification of the unobserved heterogeneity requires at least three measures per factor. We estimate $f_{\theta^H}(\cdot)$ using three components of IQ—numeric, verbal and spatial ability—ascertained when students are 13 years old. We identify a gender-nonconformity factor $f_{\theta^{CGN}}(\cdot)$ using as manifest scores the reported preferences for domestic interests, mechanical interests, and sports.[27] We estimate the joint models comprising measurement systems (5) and (6), and outcome equations (4) separately for men and women using maximum likelihood.[28] To ease interpretation, and given that gender-nonconformity and skills are unobserved, we rely on simulations of the expected outcome as a function of the unobserved heterogeneity. We randomly draw 20,000 θ^{CGN} and θ^H from the estimated distributions $\hat{f}_{\theta^{CGN}}(\cdot)$ and $\hat{f}_{\theta^H}(\cdot)$ and construct $E[Y|\theta^{CGN}, \theta^H]$.

F.3 Identification and Estimation of Models with Unobserved Heterogeneity

The type of models developed in this section can be described as a set of measurement systems that are linked by a factor structure.[29] We start with a measurement system which we use to identify the distributional parameters of q unobserved factors. The measurement system would have the following form:

$$\mathbf{T} = \mathbf{X}_T\beta^T + \alpha^{\mathbf{T},\mathbf{A}}\theta^A + \mathbf{e}^{\mathbf{T}} \tag{7}$$

[27]In fact, given how the manifest scores are set out, from measurement system (5), we identify the distribution of a *masculinity* factor separate for the samples of men and women. For both genders, higher scores of this measure implies behaviors or tastes that are more in line with the typical male. As in Section 3.2.2, we prefer men and women to be on the same scale going from gender conformity to nonconformity. Thus, we simply recode the identified factor for the male sample, so that for both genders high values in the factor mean greater gender-nonconformity.

[28]As controls for the measurement systems $\mathbf{X}_T$, we use a dummy of having a female head of household, father's education and home-ownership status. Table F.1 in the Appendix presents the estimates of measurement system (5) and (6).

[29]This Appendix follows closely the argument put forth in (Sarzosa and Urzua, 2016).

where $\mathbf{T}$ is a $L \times 1$ vector of measurements (e.g., test scores, measures of behaviors), $\mathbf{X}_T$ is a matrix with all observable controls for each measurement and $\alpha^{T,A}$ is a vector containing the loadings of unobserved factor A in each mean measurement T. We assume that $\left(\theta^A, \mathbf{X}_T\right) \perp \mathbf{e}^{\mathbf{T}}$, that all the elements of the $L \times 1$ vector $\mathbf{e}^{\mathbf{T}}$ are mutually independent and have associated distributions $f_{e^h}(\cdot)$. To explain how the parameters of the measurement system (7) are identified, let us focus on the matrix $COV\left(\mathbf{T}\left|\mathbf{X}_T\right.\right)$ whose elements in the diagonal are of the form $COV\left(T_i, T_i\left|\mathbf{X}_T\right.\right) = \left(\alpha^{T_i,A}\right)^2 \sigma^2_{\theta^A} + \sigma^2_{e^{T_i}}$, and the off-diagonal elements are of the form $COV\left(T_i, T_j\left|\mathbf{X}_T\right.\right) = \alpha^{T_i,A}\alpha^{T_j,A}\sigma^2_{\theta^A}$.

As it is, the model is underidentified. To see this, note that the is no way to use the off-diagonal elements of $COV\left(\mathbf{T}\left|\mathbf{X}_T\right.\right)$ to come up with unique values for the parameters we intend to estimate. More precisely, note that

$$\frac{COV\left(T_2, T_3\left|\mathbf{X}_T\right.\right)}{COV\left(T_1, T_2\left|\mathbf{X}_T\right.\right)} = \frac{\alpha^{T_3,A}}{\alpha^{T_1,A}}, \frac{COV\left(T_2, T_3\left|\mathbf{X}_T\right.\right)}{COV\left(T_1, T_3\left|\mathbf{X}_T\right.\right)} = \frac{\alpha^{T_2,A}}{\alpha^{T_1,A}}, \frac{COV\left(T_1, T_3\left|\mathbf{X}_T\right.\right)}{COV\left(T_1, T_2\left|\mathbf{X}_T\right.\right)} = \frac{\alpha^{T_3,A}}{\alpha^{T_2,A}}$$

Therefore, identification requires some assumptions. First, we acknowledge that latent factors have no metric or scale of their own (Bartholomew et al., 2011). Hence, we need to normalize to unity one loading, and the remaining loadings should be interpreted as relative to the one used as nummeraire. If, with out loss of generality, we normalize $\alpha^{T_3,A} = 1$, the remaining loadings are identified from the quotients of the off-diagonal elements of $COV\left(\mathbf{T}\left|\mathbf{X}_T\right.\right)$. This also shows that identification requires $L \geq 3$. That is, we need at least three scores per factor.[30] Having identified the loadings, we can further use the off-diagonal elements of $COV\left(\mathbf{T}\left|\mathbf{X}_T\right.\right)$ to identify the factor variance σ_{θ^A} and the diagonal elements of $COV\left(\mathbf{T}\left|\mathbf{X}_T\right.\right)$ to identify σ_{θ^A}. Equivalent steps across the q dimensions of unobserved heterogeneity yield the identification all the unobserved factors.

Now that we have identified all the loadings, factor variances and measurement residual variances, together with the fact that the means of θ^A, θ^B and $\mathbf{e}^{\mathbf{T}}$ are finite—in fact, equal to zero because we allow the measurement system (7) to have intercepts—we can invoke

[30]Analogously, we can also infer the minimum number of scores by acknowledging that we can use the $\frac{L(L-1)}{2}$ off-diagonal elements to identify $L-1$ loadings—taking into account we normalized one loading—and the factor variance (Carneiro et al., 2003) . We then need that $\frac{L(L-1)}{2} \geq (L-1)+1$. Thus $L \geq 3$.

the Kotlarski Theorem to use the manifest variables $\mathbf{T}$ to non-parametrically identify the distributions of $f_{\theta^A}(\cdot)$ (Kotlarski, 1967).[31]

With the distribution of the factors in hand, we can consider a model linking the outcome variables we observe with the factor structure. That is,

$$\mathbf{Y} = \mathbf{X}_Y \beta^Y + \mathbf{\Lambda}^Y \Theta + \mathbf{e}^Y \tag{8}$$

where $\mathbf{Y}$ is a $M \times 1$ vector of outcomes, $\mathbf{\Lambda}^Y$ is an $M \times q$ matrix containing the factor loadings for each of the M outcome equations and q dimensions of unobserved heterogeneity, $\mathbf{e}^Y$ is a vector of error terms with distributions $f_{e^y}(\cdot)$. We assume that $\mathbf{e}^Y \perp (\Theta, \mathbf{X_Y})$ and also that $e^{Y_m} \perp e^{Y_{m'}}$ for $m, m' = 1, \ldots, M$ and $m \neq m'$. This is the type of models considered, for instance, by Heckman et al. (2006) and Urzua (2008) often with $M = 1$ (e.g., employment or earnings). In the case of the Roy model $M = 3$, $\mathbf{Y} = [D \quad Y_0 \quad Y_1]'$, where D takes the value of one if the agent chooses sector 1 (e.g., a STEM occupation) or zero if the agent chooses sector 0 (e.g., a non-STEM occupation). Y_0 is the outcome observed for those who select into sector zero, and Y_1 is the observed outcome who sort into sector one. Note that the econometrician does not observe the actual value of Θ for each observation. Rather, in the first stage, s/he estimates the distributions they are drawn from and uses it to integrate it out in (8). We estimate the model using Maximum Likelihood. In the case of the Roy model, we considering the following likelihood function:

$$\mathcal{L} = \prod_{i=1}^{N} \int_q \begin{bmatrix} \left[\left(1 - f^D\left(\mathbf{X}_D, Y_D, \zeta\right)\right) f^{Y_0}\left(\mathbf{X}_Y, Y_0, \zeta\right)\right]^{1-D} \\ \times \left[f^D\left(\mathbf{X}_D, Y_D, \zeta\right) f^{Y_1}\left(\mathbf{X}_Y, Y_1, \zeta\right)\right]^D \\ \times f_{e^1}\left(\mathbf{X}_{T_1}, T_1, \zeta\right) \times \cdots \times f_{e^L}\left(\mathbf{X}_{T_L}, T_L, \zeta\right) \end{bmatrix} dF_\Theta(\zeta)$$

[31] The basic idea of the Kotlarski Theorem is that if there are three independent random variables e_{T_1}, e_{T_2} and θ and define $T_1 = \theta + e_{T_1}$ and $T_2 = \theta + e_{T_2}$, the joint distribution of (T_1, T_2) determines the distributions of e_{T_1}, e_{T_2} and θ, up to one normalization. Note that, given that we have already identified all the loadings, we can write (7) in terms of $T_\tau = \theta + e_{T_\tau}$ by dividing both sides by the loading. See more details in Carneiro et al. (2003).

F.4 Additional Results

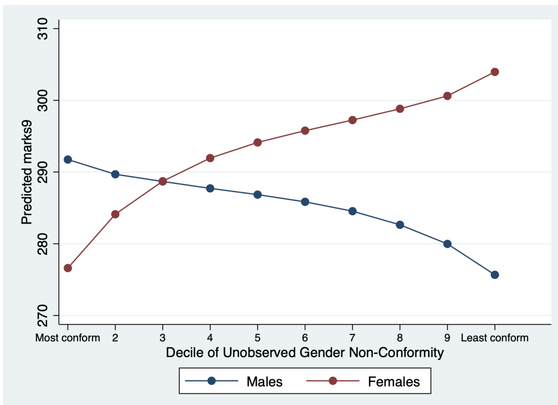
In this subsection, we present estimates of the model with unobserved heterogeneity (8) with $M = 1$. That is, we analyze the relationship between selected outcomes—not mediated by a choice—and gender conformity in the form of unobserved heterogeneity. Our results corroborate our regression estimates in Section 4 by showing clear CGN gradients. Figure F.1(a) shows that gender nonconformity is associated with lower 9^{th} grade GPA for boys, but higher scores for girls. The performance gaps between the extremes of the CGN distribution are sizable. The least gender-conforming boys (bottom 10%) score 0.16 points less than the most gender-conforming boys (top 10%). On the contrary, the least gender-conforming girls score 0.28 points *more* than the most gender-conforming girls. Figure F.1(b) indicates that pursuing a STEM track in upper secondary is mostly a masculine trait. While 83% of the gender-conforming boys choose a STEM track, only about half of the gender-nonconforming boys do. In the same vein, while only 43% of the most gender-conforming women choose a STEM track, 53% of the nonconforming ones do. Interestingly, the most gender-conforming men and the least gender-conforming women have roughly the same probability of selecting a STEM track in upper secondary.

Figure F.1(c) presents a wide gap between gender-typical and gender-nonconforming women in the likelihood of them reaching any kind of tertiary education. CGN women are 7 percentage points (20% in relative terms) more likely to take on post secondary education than gender-typical women. This gap contrast with the almost flat relation between gender conformity a tertiary education attainment among men. Gender-typical women have about the same chances of going to tertiary education as men, while CGN women are about 16% more likely than men to do so. This is another example of how the education gaps in favor of women we have become accustomed to are driven by those who challenged gender norms during early adolescence.

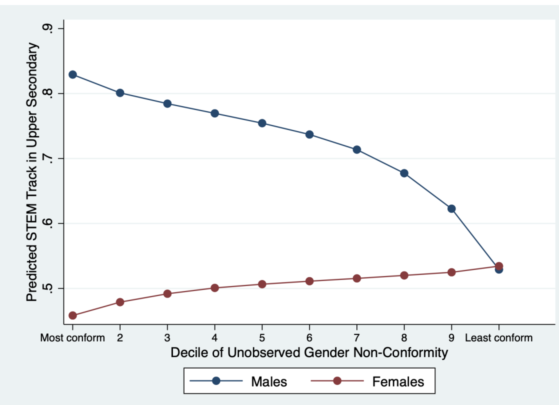
Figure F.1(d) explores occupation choices. We zoom in on two types of occupations: STEM and Services & Sales. We choose them because they are opposites in terms of women's representation. Men dominate STEM occupations, while Services & Sales occupations are dominated by women. Figure F.1(d) shows that CGN men are twice more likely to sort into service and sales (12.9%) than into STEM (6.4%) occupations. On the contrary, gender-conforming men are about three times *more* likely to sort into STEM occupations (23%) than into service and sales (8%). In contrast, CGN women are 4 percentage points (18%) *less* likely to sort into a services or sales occupation than gender-conforming women.

Figure F.1: Gender Nonconformity and Education Outcomes

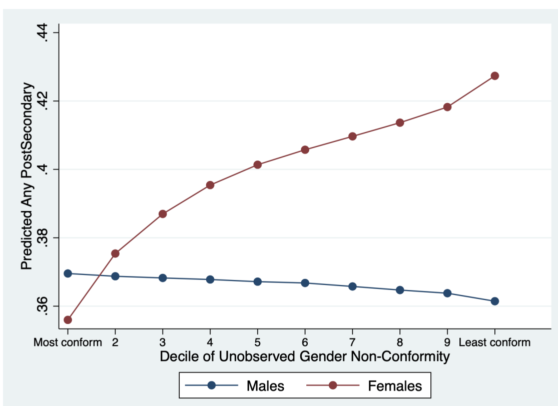
(a) GPA at 9^{th} Grade



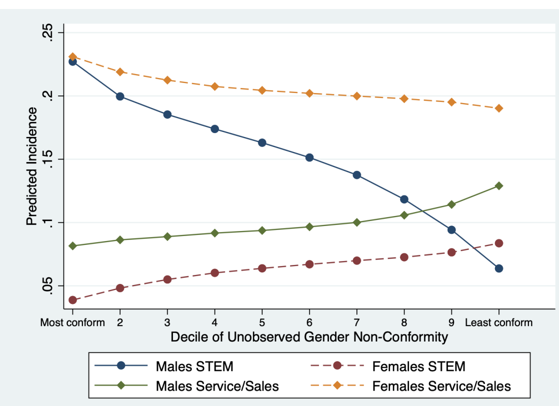
(b) STEM Track in Upper Secondary



(c) Any Post Secondary



(d) Occupation Sorting



Note: Each figure presents $E[Y|\theta^{CGN}]$ for GPA in $9th$ grade, the choice of a STEM track in upper secondary, the choice of going into any kind of tertiary education, the choice of STEM and Service & Sales occupations, and the age of first birth in the vertical axis. Each is the product of 20,000 simulations based on the findings model (8). The horizontal axes in all panels displays the deciles of the gender-nonconforming factor. Data from Stockholm Birth Cohort.

Table F.1: System of Manifest Variables and the Identification of Latent Cognitive Skills and Masculinity Factor

	Cognitive Scores			Male-biased Behaviors					
					(Males)			(Females)	
	Spatial	Verbal	Numeric	Sports	Domestic	Mechanical	Sports	Domestic	Mechanical
Cognitive	0.731***	0.711***	1						
	(0.020)	(0.019)							
Masculinity				5.435***	-0.147*	1	3.538***	-0.163***	1
				(0.463)	(0.084)		(0.175)	(0.040)	
Female	-1.532***	-0.160	-1.444***						
	(0.127)	(0.112)	(0.139)						
Female Head HHld	0.018	0.354	-0.294	0.298	-0.266	-0.361	-0.735*	-0.271	0.027
	(0.248)	(0.219)	(0.273)	(0.315)	(0.386)	(0.375)	(0.399)	(0.311)	(0.381)
FatherEduc: HS	1.638***	2.505***	2.437***	-0.504**	0.095	-0.071	-0.123	-0.686***	0.617**
	(0.175)	(0.155)	(0.193)	(0.226)	(0.264)	(0.257)	(0.272)	(0.224)	(0.273)
FatherEduc: College	2.131***	4.119***	4.384***	-0.649**	0.576*	-0.590*	0.144	-1.368***	1.635***
	(0.233)	(0.207)	(0.257)	(0.313)	(0.349)	(0.340)	(0.381)	(0.305)	(0.372)
Homeowner	0.562***	0.381**	0.273	-0.664***	-0.055	-0.013	-0.239	-0.750***	0.635**
	(0.169)	(0.150)	(0.186)	(0.222)	(0.256)	(0.249)	(0.278)	(0.218)	(0.267)
Constant	25.073***	24.936***	23.010***	37.901***	25.624***	36.968***	33.365***	34.591***	23.040***
	(0.207)	(0.184)	(0.228)	(0.122)	(0.123)	(0.120)	(0.136)	(0.103)	(0.127)
Observations			10,967			5,403			5,575

Note: This table provides the estimates of measurement system (5) and (6). Data from Stockholm Birth Cohort. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.