



# Documenting occupational sorting by gender in the UK across three cohorts: does a grand convergence rely on societal movements?

Warn N. Lekfuangfu<sup>1</sup> · Grace Lordan<sup>2</sup>

Received: 17 October 2021 / Accepted: 4 September 2022  
© The Author(s) 2022

## Abstract

We consider the extent to which temporal shifts have been responsible for an increased tendency for females to sort into traditionally male roles over time, versus childhood factors. Drawing on three cohort studies, which follow individuals born in the UK in 1958, 1970 and 2000, we compare the shift in the tendency of females in these cohorts to sort into traditionally male roles compared to males, to the combined effect of a large set of childhood variables. For all three cohorts, we find strong evidence of sorting along gendered lines, which has decreased over time, yet there is no erosion of the gender gap in the tendency to sort into occupations with the highest share of males. Within the cohort, we find little evidence that childhood variables change the tendency for females of either the average or highest ability to sort substantively differently. Our work is highly suggestive that temporal shifts are what matter in determining the differential gendered sorting patterns we have seen over the last number of decades, and also those that remain today. These temporal changes include attitudinal changes, technology advances, policy changes and economic shifts.

---

We are thankful to the Centre for Longitudinal Studies (CLS), UCL Institute of Education, for the use of these data and to the UK Data Service for making them available. However, neither CLS nor the UK Data Service bear any responsibility for the analysis or interpretation of these data. Lekfuangfu acknowledges financial support from Spanish Ministry of Science and Innovation (Project Reference: AEI/10.13039/501100011033), MICIN/AEI/10.13039/501100011033, Comunidad de Madrid (EPUC3M11 (V PRICIT) and H2019/HUM-589).

---

✉ Grace Lordan  
g.lordan@lse.ac.uk

Warn N. Lekfuangfu  
nlekfuan@eco.uc3m.es

<sup>1</sup> Universidad Carlos III de Madrid, Madrid, Spain

<sup>2</sup> Department of Psychological and Behavioural Science, The London School of Economics and Political Science, London, UK

**Keywords** Occupational choice · Gender · Temporal change · Childhood influences · J16 · J4

**JEL Classification** J16 · J4

## 1 Background

For decades, economists have contributed to the literature that seeks to explain the gender wage gap.<sup>1</sup> A well-accepted conclusion is that the lack of women in high-paying, male-dominated professions is one major cause of this gap (Bayard et al. 2003; Goldin 2014; Blau and Kahn 2016). This has led to a search for the underlying causes of gender-based sorting. Explanations include differential human capital investments, (Altonji and Blank 1999), discrimination (Becker 1957), a lack of flexibility to combine a career and family in male-dominated jobs (Goldin 2014; Bertrand 2018) and differences in tastes and preferences (Lordan and Pischke 2022; Cortés and Pan 2017). In this work, we build on this literature and consider the extent to which temporal shifts have been responsible for an increased tendency for females to sort into traditionally male roles over time, versus childhood factors that have already been shown by economists to shape future successes of children in other life domains.<sup>2</sup> Considering that previous works have already demonstrated that gender gaps in childhood skills are key determinants for different occupational choices as well as college major choices between men and women,<sup>3</sup> our goal here is to examine the extent that gender predicts sorting patterns across time and ask whether this is moderated by childhood factors, such as early health, socio-economic status, parental investments and aspirations, characteristics of schooling and the child's own ability.<sup>4</sup>

The importance of large temporal movements for gendered sorting is intuitive. A convincing narrative points to the fact that until the 1970s, there were many more male-dominated roles, than today. Between 1970 and 2018, some stylised facts emerged. First, females sorted into many occupations that were traditionally male-dominated. Examples include law, accountancy and pharmacy. Second, females failed to converge into other occupations such as those requiring science, technology and engineering where the share of males in the UK, US and across the EU still exceed 80%.<sup>5</sup> Third, the revolution has been asymmetric—with males failing to sort into traditionally female occupations, such as social work, nursing and primary school teaching. These stylised facts are also visible in Fig. 1, which plots the occupations of three cohorts born in

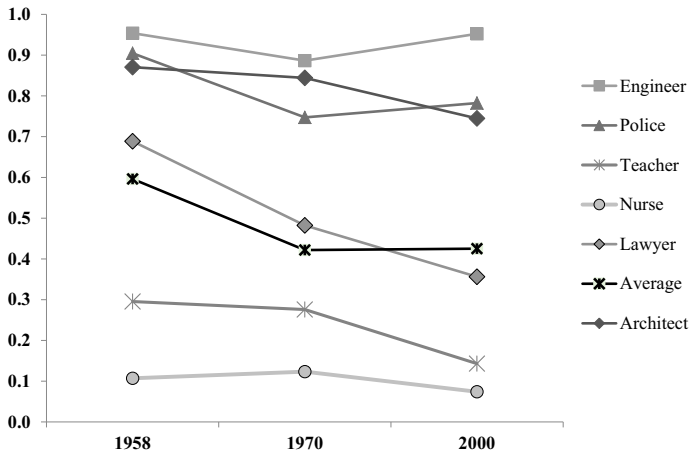
<sup>1</sup> See Daymont and Andrisani (1984), Blau and Kahn (1997, 2016), and Fortin (2008).

<sup>2</sup> For example, Akee et al. (2013), Angrist and Lavy (1999), Black et al. (2007), Becker and Tomes (1986), Blau (1999), Datcher (1982), Heckman et al. (2013), Luo and Waite (2005), and Van Den Berg et al. (2006).

<sup>3</sup> See, for instance, Speer (2017) for the importance of pre-market skills in gender gaps in STEM occupations; Cobb-Clark and Tan (2011) and Todd and Zhang (2020) for the role of childhood non-cognitive skills in gender sorting in occupational choices.

<sup>4</sup> Full details of the variables are provided in Appendix A.

<sup>5</sup> Based on calculations from the quarterly Labour Force Survey in the UK (years 2015–2017 combined), the Current Population Survey in the US (years 2015–2017 combined) and the EU Labor Force Survey (2014–2016 combined).



**Fig. 1** Share of men in an occupation in the NCDS, BCS and MCS. *Notes:* For the NCDS (1958) and the BCS (1970), share of men in an occupation is calculated from the occupation cohort member held at age 33 and 34 years old, respectively. For the MCS (2000), the occupation is the occupation cohort members (at age 11) aspired to be when they turn 30 years old

1958, 1970 and 2000 in the UK, respectively.<sup>6</sup> There is a downward trend in the share of males in law over time, but no real change in engineering. The proportion of males in nursing is flat over the three periods, with the share of males planning to go into teaching decreasing to even lower levels for the most recent cohort. This highlights that the asymmetric gender revolution remains in place.

Gendered sorting is intuitively influenced by temporal movements over time. For example, human capital investments, both in type and quantity, are affected by social norms. Systemic changes in attitudes over time can cause females to invest in different career paths if preferences are shaped by these norms. For example, the tolerance of discrimination has changed radically over the last 5 decades.<sup>7</sup> A systemic movement which causes males and females to share family responsibilities more equally removes constraints for females allowing them to have wider career choices. The influence of tastes and preferences on gendered sorting is also now being explored (Filer 1986; Lordan and Pischke 2022; Cortés and Pan 2017).

In addition, across their lifecycle, males and females interact with these social norms, updating their preferences, beliefs and decisions based on social norms. For example, one cause that has been identified as women having lower pay and access to senior leadership roles is their tendency to negotiate less as compared to men (Biasi and Sarsons 2021; Dreber et al. 2022). However, there is also evidence that women may choose not to negotiate as they are aware of the social expectations towards them

<sup>6</sup> These cohort data are used in this paper and are subsequently described. For the 1958 and 1970 cohorts, occupations are measured based on data collected when they were in their early 30s. For the cohort born in 2000, occupations are measured based on aspirations reported at age 12.

<sup>7</sup> Becker (1985), Katz and Murphy (1992) and Goldin (2006) have all suggested that the effects of gender discrimination are now much less relevant than other factors when it comes to explaining occupational segregation as compared to previous time periods.

and fear societal backlash (Kray and Gelfand 2009); as well as differential treatment when they do (Bowles et al. 2007)—causing them to negotiate less in the future. It is easy to imagine these dynamics playing out across a range of behaviours of both males and females, conditioning or resigning individuals to behave more in line with social norms in their future interactions.

While these papers suggest that tastes and preferences have a role to play in occupational sorting, they may be socially constructed. This fits with the idea that individual decisions are influenced by the opinions of others, which ultimately shape identity (Akerlof and Kranton 2000). Temporal changes in sorting patterns can also arise because the tasks within an occupation have changed over time, making them less physically demanding, and deemed more suitable for women (Black and Spitz-Oener 2010; Yamaguchi 2018, Alesina et al. 2013). In addition, temporal changes in sorting patterns can arise because of technology shifts. For example, Goldin and Katz (2002) demonstrate the role of the contraceptive pill, by allowing delayed fertility, on women’s career choices. Major policy changes have also been shown to affect sorting by gender, such as the introduction of credentials through the use of degrees and licenses and licences in the US in the 1970s and 1980s, which gave objective signals for women to demonstrate their suitability for jobs that previously had low shares of males.

At a more micro-level, gendered sorting has the potential to be influenced by childhood variables given that experiences in that period vary by gender. So, why do experiences vary? First, some people may have preferences for a particular gender or indeed have a preference for engaging children in different activities depending on whether they are a boy or a girl.<sup>8</sup> Second, people may hold certain *beliefs* that boys and girls have different production functions and as a result, decide to engage children in different gendered activities.<sup>9</sup> This concept also corresponds to a theory of biased choices (see, for instance, Berger et al. 1966; Oxoby 2002) whereby parents or children themselves make decisions that are conformed and biased by their hierarchically social characteristics. Third, there may be differential monetary and/or opportunity costs of engaging with boys over girls in specific activities. These explanations are not mutually exclusive, but together, can largely capture the underlying causes of why boys and girls are exposed to different experiences that ultimately may shape their futures. Examples of differential treatment by gender abound in the literature.<sup>10</sup> Differential treatment has the potential to impact cognitive development, including soft skills and motor skills.

We are interested in the extent to which these experiences that vary *within cohorts*, change gendered sorting patterns as compared to systemic changes, which occur *across cohorts*. Drawing on three British cohort studies, which follow children born in Britain

<sup>8</sup> See Dahl and Moretti (2008) and Kohler et al. (2005) for suggestive evidence that fathers strictly prefer sons can be found in the US and in Denmark, respectively.

<sup>9</sup> We intentionally write “belief” because a recent meta-analysis highlights that, among both children and adults, females perform equally to males on math assessments, and the gender difference in verbal skills is small and varies depending on the type of skill assessed (Hyde 2014).

<sup>10</sup> See Lundberg (2005), Lundberg and Pabilonia (2007) and Yeung et al. (2004) work on parental time allocation; Aznar and Tenenbaum (2015) research on parental communications; Mondschein et al. (2000) work on parental assessments of ability and Dee’s (2007) work on teach assignment.

in 1958, 1970 and 2000, we compare the generational shift in the tendency of females to sort into traditionally male roles, to the combined influence of a large set of childhood variables. We consider an exhaustive enough set of childhood variables that can be reasonably expected to be correlated with both gender and sorting patterns. These childhood variables capture cognitive, soft and motor skills, alongside socio-economic variables, health status, parental influences and peer influences. We consider several proxies that capture the differential aspects of traditionally male jobs. For those individuals born in 1958 and 1970, these proxies are based on their occupations in their early 30s. For the individuals born in 2000, the proxies are based on their aspirations for the future (when they turn 30 years old). These proxies range from the share of males in an occupation to variables which capture an occupation's content. We acknowledge that we do not capture a universe of childhood variables such that we can rule out every childhood factor that can possibly determine occupational sorting. However, our analysis does capture a large array of childhood variables—allowing us to consider many more measures of early life skills and preferences as inputs into the gender occupational process commonly considered in this literature.

Several stylised facts emerge from our analyses. First, for all three cohorts, we find strong evidence of sorting along gendered lines, regardless of the childhood variables that we include in our regressions. Second, the tendency to sort along gendered lines has decreased substantively over time. That is, the gender gap has narrowed *across* birth cohorts. Conversely, we find little evidence that childhood variables can considerably change the tendency for an average female to sort substantially differently *within* a cohort. Third, the same conclusions emerge even if we focus only on individuals with the highest childhood cognitive ability. These are the individuals who we expect to be able to move against gender stereotypes and subsequently sort into the top jobs. We view our work as underlining the importance of temporal shifts, over and above the role of *within-cohort* childhood variables that are observable in our samples, in determining the changes in gendered sorting patterns over the last number of decades, and also those that remain today. However, our work cannot pinpoint what exact temporal shift did the heavy lifting.

In all likelihood, it is a variety of structural shifts at the level of society that contributed to the trends that we observe today. This includes attitudinal, economic, policy and technological changes; as well as how quickly individuals interact with these shifts so that tipping points are reached. Other potential forces at play are the changing nature of tasks in many occupations, lower tolerance levels towards discrimination in the workplace, differential growth patterns across industries and the ability for women to delay their fertility.

## 2 Analytical framework and empirical design

We are interested in the extent to which a female's tendency to pursue traditionally male roles is influenced by childhood factors versus unobservable underlying factors that reflect societal shifts. The outcome variables considered are proxies for aspects of work where we expect the sexes will bifurcate in sorting tendencies. These proxies cover income, work hours, flexibility, job content and job competitiveness alongside

the share of males in an occupation. Drawing on individual-level data for cohorts of individuals born in the UK in 1958, 1970 and 2000, we relate each proxy in turn to a female dummy variable, and sequentially add groups of childhood variables, which we may expect to be correlated with both gender and the proxy. We take a holistic approach to specify these childhood factors and examine demographic and socio-economic variables from early childhood, alongside measures of cognitive and non-cognitive ability, childhood health, parental inputs and external influences.

We are interested in the extent to which the coefficient of the female dummy is attenuated with the addition of these childhood factors. If the coefficient is attenuated substantively, we argue that it reveals that malleable factors at the individual level, of which many can be readily influenced by parents, schools and policymakers, play a large role in determining gendered sorting across generations for an average female. However, if the female coefficient remains relatively stable *within* each birth cohort, it is highly suggestive that such childhood factors may matter little in explaining gendered sorting.<sup>11</sup> Rather, a relatively stable female coefficient *within* a given birth cohort, but a declining female coefficient *across* cohorts would therefore point in the direction of the importance of temporal movements in changing gendered sorting patterns.

Specifically, we estimate:

$$Y_{ij}^{\text{adult}} = \alpha_i + F_i \delta + X_i^{\text{child}} \beta + \varepsilon_{ij}^{\text{adult}} \quad (1)$$

where  $Y_{ij}^{\text{adult}}$  is a proxy for a component of the job,  $j$ , for individual,  $i$ , in adulthood.  $F_i$  is equal to 1 if the individual is female and 0 otherwise.  $X_i^{\text{child}}$  is a vector of childhood control variables, which will be subsequently discussed. We run Eq. (1) separately for each cohort and obtain the estimate of  $\delta$ , which indicates the extent of being a female influences occupational sorting across cohorts. We note that we care most about how  $\delta$  changes *within* as well as *across* cohorts when we sequentially add childhood variables. In this exercise, we do not seek to put structural interpretations on the coefficients of control variables,  $\beta$ . However, if  $F_i$  is attenuated by a particular set of  $X_i^{\text{child}}$ , we view this as suggestive evidence that something that is correlated with this same  $X_i^{\text{child}}$  is driving gender sorting *within* the cohort. Hence, childhood variables, in the general sense, matter in determining sorting.

In addition, we will check whether the same general patterns identified by Eq. (1) hold for the most intelligent children in the three cohorts. After all, these are the individuals we would expect to be most likely to reach the most prestigious jobs in society, at which we may care more about having a better representation of women. It also

<sup>11</sup> The major threat to this argument is that all of our childhood variables are measured with error. This seems highly unlikely, given the quantity of variables that we do consider, but later we return to this point. We do acknowledge, however, that the variables we have available in our data cannot represent every childhood variable that influence sorting. However, we can capture a large number of childhood factors, including those are most commonly considered in the related literature.

helps abate concerns that any effects found are owed to unobserved individual differences.<sup>12</sup> To consider this, we follow the psychometric literature and use exploratory factor analysis to reduce the dimensionality of our proxies for childhood intelligence in each cohort study into one variable (Gorsuch 1983; Thompson 2004).<sup>13</sup> From this factor, we repeat the analysis documented above on individuals who are in the top quantile of this distribution only. We note that the female share in this decile is 51.9%, 49.4% and 49.12% for children born in 1958, 1970 and 2000, respectively.

Note, however, that, an issue with estimating Eq. (1) is that there is a risk of overfitting, given that our estimation includes a large number of childhood variables. To mitigate this, we also estimate Eq. (1) by applying LASSO regression analysis.<sup>14</sup> For our purposes, LASSO is useful as a check as to whether the coefficient,  $\delta$ , on the female dummy remains nonzero after the shrinkage process, emphasising that of all the variables included in the regression it is one of the most strongly associated with the outcome. We report only  $\delta$  from the LASSO regressions in the main text and document the full results in Appendix D.

The first outcome we consider is the share of males in an individual's chosen occupation. This allows us to ask directly whether occupational segregation by gender has changed substantively for the three cohorts, and how these changes relate to the childhood variables we usually think of as determining a person's future that are commonly collected in surveys. We complement this with regressions that model the probability that a job with a share of males 80% or more is chosen, to allow us to quantify how this has changed over the three generations. We also consider an outcome that is equal to 1 if a female has opted out of the labour force and zero otherwise. This is our only outcome that is generated at the individual level, rather than the occupation level.

The next proxy we consider is average occupational income. For many years, economists have contributed to the literature that seeks to explain just why there is a gender wage gap. It is now clear that the lack of women in high-paying, male-dominated professions contributes significantly to this gap.<sup>15</sup> Therefore, considering the average income of an individual's chosen occupation as an outcome in Eq. (1) allows us to examine directly whether females have been choosing jobs with a significantly higher average income over time and how this is mediated by childhood variables.

We also consider the average hours in an individual's occupation as an outcome. Women who find it hard to juggle family and children—or indeed hard to imagine

<sup>12</sup> This concern may arise if females working in jobs with the highest shares of males have above average unobserved individual ability (see Ludsteck (2014) for evidence in this regard for Germany).

<sup>13</sup> A clear structure of one latent factor emerges in the first rotation (see Appendix A.2).

<sup>14</sup> LASSO is a shrinkage and variable selection method for linear regression models whose goal is to obtain the subset of predictors that minimizes prediction error for a quantitative response variable. The LASSO does this by imposing a constraint on the model parameters that causes regression coefficients for some variables to shrink toward zero. Variables with nonzero regression coefficients variables are most strongly associated with the response variable. LASSO is useful given that many of the variables we sequentially add are highly correlated, so disentangling their true coefficient size given potential issues with multicollinearity is difficult.

<sup>15</sup> See, for example, Blau (1977), Bielby and Baron (1984), Macpherson and Hirsch (1995), Carrington and Troske (1998), Bayard et al. (2003), and Blau and Kahn (2016).

juggling in the future—may ‘opt out’ of occupations that make this more difficult. This suggests a constrained choice. Therefore, we re-estimate Eq. (1) with average hours as the dependent variable.  $\delta$  is then indicative of how important it is for females to be in occupations with lower average hours as compared to males. This fits with work that suggests that females ‘opt elsewhere,’ choosing occupations that allow them to accommodate family responsibilities (Polachek 1981; Belkin 2003; Stone 2007) or choose to work fewer hours to balance family responsibilities (Antecol 2010).<sup>16</sup>

We complement the average hours proxy with another variable which captures nonlinear returns to hours worked. This follows Goldin (2014) who presents evidence for full-time college graduate workers in 95 high-paying occupations. Goldin’s metric for the flexibility of an occupation is the elasticity of individual earnings with respect to hours worked: high elasticities imply a penalty for workers seeking short hours and indicate a lack of flexibility. Goldin (2014) demonstrates that less flexible occupations have a larger pay gap, and it is also argued that less flexible work pushes females with children out of the labour market (Leber Herr and Wolfram 2012).

Our remaining proxies capture occupational content. This complements a recent emergence of explanations for occupational segregation which suggest that males and females have different tastes when it comes to the content of the work that they do. The psychologist Susan Pinker (2008) has pushed the idea that differences in the preferences of women and men are the main driver of gendered labour market choices. Pinker (2008) highlights that women may not like the nature of male-dominated jobs, preferring ‘people’ content over making ‘things’ (what we refer to as ‘brawn’ hereafter). Lordan and Pischke (2022) provide quantitative evidence from three countries and a discrete choice experiment which backs up this claim. Overall, their work suggests that females are more extrinsically motivated opting for jobs that are high in ‘people’ and ‘brains’ content, like medicine and law, over jobs that are relatively high in ‘brawn’ content, like engineering. By contrast, males care less about the job content. Cortés and Pan (2017) investigate the predictive power of a variety of occupational indices in regressions that model the rate of females in an individual’s occupation. They show that social contribution and physical skill dominate. Broadly, these findings lend support to the ‘people’ versus ‘brawn’ divide.

Together, these results raise the question of whether on average women prefer jobs with a societal contribution. These differences in tastes by gender can be innate, evolutionary or socialised.<sup>17</sup> Evidence that females differentially select into work of different content, but with a pattern changing over time, point to a temporal role in the formation of preferences. That is, what is going on in society influences sorting

<sup>16</sup> We note that a number of other studies have also considered whether ‘opt out’ of the labour market occurs conditional on having children and do not find any differences by education level (Boushey 2005; Goldin 2006; Vere 2007; Cohany and Sok 2007; Fortin 2008; Percheski 2008). For us, females do not necessarily opt out. Rather they may ‘opt elsewhere’ to allow them to better manage their current or expected family commitments, into occupations with lower average hours and enhanced flexibility. This is supported by Xie and Kimberlee (2004) who highlight that marriage and children move women from the male dominated fields of science and engineering towards other types of work.

<sup>17</sup> Evidence is also provided by Grove et al. (2011) who examine the pay gap of MBA graduates and find that female MBAs have a wage penalty owed to choosing occupations that contribute to society and have high ethical standards. Su et al. (2009) also emphasis sex differences in occupations preferences in an overview of the psychology literature on this topic.



patterns. To check this, we also estimate the model of Eq. (1). We create three proxies for job context, based on an approach introduced by Lordan and Pischke (2022) and drawn on ONET activities and job content data. Overall, these proxies represent ‘people,’ ‘brains’ and ‘brawn’ content.<sup>18</sup>

Competitiveness as another measure of job characteristic is also important. Experimental evidence has highlighted that females are more averse to competition as compared to males (Croson and Gneezy 2009).<sup>19</sup> Therefore, we also aim to check the extent to which gender explains an individual’s decision to enter and stay in an occupation, which has a highly competitive environment where the stakes are high. Therefore, our last proxy is a measure of competitiveness at the occupational level from the ONET database. This proxy is built on Cortés and Pan (2016, 2017) who use the same measure of competitiveness as we do here.

### 3 Data

We draw on the National Child Development Study (NCDS), a continuing study that follows the lives of 17,000 people living in Great Britain who were born in the week of March 3, 1958. We use data from the survey at birth, and ages 7, 11 and 16. We measure the NCDS child’s occupation variables at age 33. We also draw on the 1970 British Cohort Study (BCS70). The BCS70 began by including more than 17,000 children born between April 5 and 11 in 1970. We use the data from the survey at birth, and ages 5, 10 and 16. We measure the BCS child’s occupation variables at age 34.<sup>20</sup> To examine gender-based occupational sorting for a cohort entering the workforce soon, we draw on the Millennium Cohort Study (MCS). This group is now 17/18 years old and is about to decide what to do after their schooling finishes. The MCS follows the lives of around 19,000 children born in the UK in 2000 and 2001. We use the data from the survey at 9 months old (2001) and at ages 3 (2003/4), 5 (2005/6), 8 (2008/9) and 12 (2012/13). We utilise the *aspired* occupation, reported by the cohort member in the 2012/13 sweep.

#### 3.1 Outcome variables

For the NCDS and the BCS, the information on occupation is measured by four-digit socio-Economic Classification 2000 codes (SOC2000) at ages 33 and 34, respectively.

For the MCS, the cohort members were asked in the 2012/13 sweep “*by the time you are 30, which of the following would you be most likely to achieve?*” followed

<sup>18</sup> Occupations with relatively high ‘people’ content involve engaging with customers, clients or co-workers routinely (for example, nurses, physicians, social workers and teachers). Occupations with relatively high brains content are economists, financial managers, aerospace engineers and CEOs. Finally, occupations that are relatively high on ‘brawn’ include explosives workers, mechanical engineers and surveyors.

<sup>19</sup> Some of these works include Dohmen and Falk (2011), Nielderle and Vesterlund (2008), Gneezy and Rustichini (2004) and Gneezy et al. (2003).

<sup>20</sup> There are two waves of the BCS that surveyed cohort members when they were in their 30s. We use the 2004 wave (aged 34) as the main source of occupation. We supplement missing occupation information in the 2004 wave with occupation reported in the 2000 wave (when cohort members were 30 years old).

by a list of choices which follow the 4-digit UK Socio-Economic Classification 2010 (SOC2010). We convert SOC2010 to SOC2000 occupation coding using a crosswalk provided by Lordan (2019). We acknowledge that aspirations reflect an attitude or a perception of occupation opportunity, and not labour market outcome as used in the NCDS and the BCS data. However, we view this cohort as particularly special as they are just about to enter the labour market at the time of writing, and so are the most interesting with respect to learning if trends continue. We also have some faith that this proxy is meaningful. For instance, if we perfectly match aspirations to occupations in the NCDS data we observe a correlation of 0.30, with the correlation with the share of males in an occupation being approximately 0.60. Additionally, it has also been demonstrated elsewhere that aspirations measures are strongly indicative of actual labour market outcomes (Genicot and Ray 2017; La Ferrara 2019; Lekfuangfu and Odermatt 2022).<sup>21</sup>

Our occupation averages are calculated based on the 1993–2012 Quarterly Labour Force Survey data (QLFS) where we exploit the four-digit SOC00 occupational codes.<sup>22</sup> Therefore, the averages associated with each occupation are the same for the three cohorts, ensuring that  $\delta$  captures the change in sorting towards or away from particular occupation types rather than composition effects. We calculate averages at the 4-digit occupation-level (SOC00) of the log of gross income, average hours and share of males and, subsequently, match these directly to the NCDS, BCS and MCS's SOC00 codes.

We also create a variable to proxy the wage-hours elasticity used by Goldin (2014). We create this variable by running a regression of the log of wages on log hours, occupation fixed effects, the interaction between log hours and the occupation fixed effects and several other controls using the 1993–2012 QLFS and consistent British SOC00 codes. The proxy is then the coefficients on the interaction between occupation and log hours.

In addition, we consider a variable that is assigned equal to 1 if an individual is dis-employed and zero otherwise. For the MCS, this is based on the child's response (asked at age 11/12) to a question on whether they expect to have a 'good' job by age 30. We interpret this as a proxy which is likely to be highly and positively correlated with future labour market attachment. We assign a variable equal to 1 if a good job is expected and zero otherwise.

Our analysis also utilises three variables which capture what a job is about. These variables are created following the approach described by Lordan and Pischke (2022). Specifically, we retrieve the items from ONET (version 5) that are related to the activities and content of an individual's work. These 79 items report the level at which an occupation has a particular characteristic from 1 to 7. Next, three latent factors

<sup>21</sup> Occupation aspirations are also available for the BCS, but only in broad occupation groups (12 categories). Therefore, we report only the statistics deriving in the NCDS.

<sup>22</sup> The QLFS is the main survey of individual economic activity in Britain and provides the official measure of the national unemployment rate. The QLFS uses British SOC90 codes from 1993 through 2000 and SOC00 from 2001. We first assign to each SOC90 code a SOC00 value based on a crosswalk created from the British Household Panel Survey (BHPS). This is possible because in the BHPS after the year 2000 every individual is assigned a SOC90 and SOC00 code simultaneously. This information allows for a consistent coding system in the QLFS based on SOC00.

‘people,’ ‘brains,’ and ‘brawn’ (PBB) are calculated using this data. We then match the latent factors (in the US Standard Occupation Codes (SOC) 2000) to the QLFS using the British SOC00 codes and, finally, match these three factors for each occupation to the NCDS, BCS and MCS data.

Next, we also use the ONET database for our measure of occupation competitiveness. Specifically, incumbents are asked: “*To what extent does this job require the worker to compete or to be aware of competitive pressures?*” with response options of ‘not at all competitive’ ‘slightly competitive’ ‘moderately competitive’ ‘highly competitive’ and ‘extremely competitive’. We standardise this variable to have a mean of zero and a standard deviation of 1 and match to the NCDS, BCS, and MCS in the same manner described for the PBB factors. We note that summary statistics for all outcome variables are provided in Table 1.

### 3.2 NCDS, BCS and MCS control variables

We aim to consider a holistic set of controls that capture as many as possible the childhood variables that are simultaneously correlated with gender and the outcomes we consider. Across all three surveys, efforts are made for these controls to be measured at similar ages and have relatively consistent definitions. Fuller details of all controls, along with the relevant means and standard deviations, can be found in Appendix A. All covariates described below are standardised to have a mean of 0 and a standard deviation of 1. We run all our specifications using robust standard errors to allow for heteroscedasticity of an unknown form. The regression samples contain the same individuals across all specifications. Our estimation approach to Eq. (1) is sequential. First, we estimate Eq. (1) with the female dummy variable and no other controls. Subsequently, we add the following variables:

#### 3.2.1 Demographics and socio-economic variables ( $fam_{i,child}$ )

Childhood demographics and the socio-economic status variables are elicited as close to birth as possible (for the exact timing of the variables, see Appendix A). These variables are mother’s age, father’s age, social status, a set of marital status dummies indicating whether the child’s parents were together or not in each wave, a dummy variable indicating whether the child’s mother stayed in school beyond the minimum required age, household income, household tenure, a dummy indicating whether or not the child’s mother worked, a dummy variable indicating low birthweight, a dummy variable indicating whether the cohort child was the first-born, a dummy indicating whether the cohort child was breastfed, region of residence and a dummy variable indicating if the cohort child is White and zero otherwise.

#### 3.2.2 Cognitive ability scores (verbal and math/science) ( $cog_{i,child}$ )

We draw on all available measures of the child’s cognitive ability in the three surveys.<sup>23</sup> In the NCDS, at age 7 the child completed the Southgate Reading Test, a 35 Item

<sup>23</sup> See Shepherd (2012) for NCDS, Parsons (2014) for BCS and Johnson et al. (2015) for the MCS.

**Table 1** Summary statistics of outcome variables

Variable	Description	NCDS 1958		BCS 1970		MCS 2000	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
Female	A dummy variable equals to 1 if cohort member (CM) is female	0.5	(0.5)	0.48	(0.5)	0.53	(0.5)
Average share of males	Share of males in an occupation (QLFS 1993–2012)	0.53	(0.33)	0.55	(0.32)	0.56	(0.3)
Log of average gross income	Log of average hourly wage in an occupation (QLFS 1993–2012)	2.17	(0.38)	2.25	(0.4)	2.6	(0.44)
People	ONET Job content (QLFS 1993–2012) (standardised)	0	(0.99)	0	(0.99)	0	(0.99)
Brains	ONET Job content (QLFS 1993–2012) (standardised)	0	(0.99)	0	(1.02)	0	(1.01)
Brawn	ONET Job content (QLFS 1993–2012) (standardised)	0	(1.01)	0	(1.01)	0	(0.99)
Competitiveness	ONET measure of job competitiveness (7 levels) (standardised)	− 0.01	(1.00)	0	(1.06)	0.99	(1.01)
Flexibility	The coefficients on the interaction between occupation and log hours (QLFS)	− 0.12	(0.26)	− 0.13	(0.26)	− 0.22	(0.27)
Average hours	Average work hours per week in an occupation (QLFS 1993–2012)	41.85	(7.01)	42.54	(6.71)	44.86	(4.69)

**Table 1** (continued)

Variable	Description	NCDS 1958		BCS 1970		MCS 2000	
		Mean	(SD)	Mean	(SD)	Mean	(SD)
In a job with high share of men	A dummy variable equals to 1 if the occupation has more than 80% males	0.3	(0.46)	0.3	(0.46)	0.29	(0.45)
Dis-employment	A dummy variable equals to 1 if CM is currently out of job	0.16	(0.37)	0.07	(0.25)	NA	
Expect to be in a good job by age 30	A dummy variable equals to 1 if CM expects to have a good job by age 30	NA		NA		0.78	(0.41)

The table shows mean and standard deviation of variables for the balanced sample in each cohort survey. The table shows the statistics for non-imputed outcome variables. All outcomes variables are taken from NCDS age 33, BCS age 34 and MCS age 11/12

Reading Comprehension Test and the General Ability Test. At ages 11 and 16, they completed the 35 Item Reading Comprehension Test. These five proxies of verbal ability are included in our regressions.

The composite verbal ability score in the BCS is derived from tests conducted at ages 5 and 10. Specifically, at age 5, the 50-item Reading Test score, which is modified from the Schonell Reading Test, was administered. At age 10, the 21-item Word Similarity subscale of the British Ability Scale was administered. We also draw on the 67-item Shortened Edinburgh reading test, which assesses vocabulary, syntax, sequencing, comprehension and retention at age 10.

For the MCS, the composite verbal ability score is derived from instruments administered at ages 3, 5, 7 and 11. The British Ability Scale (BAS) was administered at ages 3 and 5, the Pattern Similarity and the Pattern Construction subscale of the BAS at ages 5 and 7, the 900-item Word Reading of the BAS at age 7 and the Verbal Similarity subscale of the BAS at age 11.

The Math-and-Science proxies for the NCDS include three tests measured at ages 7, 11 and 16, as well as teacher ratings for math and science. Specifically, these tests are the Problem Arithmetic Test score at age 7, a 40-item Arithmetic/Mathematics Test at age 11 and a 31-item Mathematics Test at age 16. We also include the teacher's ratings which reflect their perception of the child's ability in maths and science. These

rating scores equal 3, 2, 1 and 0 if the teacher thinks the child's ability is equivalent to A-level, high-graded GCSE, low-graded GCSE and below GCSE, respectively.<sup>24</sup>

For the BCS, we draw on tests administered at ages 10 and 16.<sup>25</sup> These are the Friendly Maths Test and the Recall of Digit subscale of the BAS administered at age 10, and the actual raw GCSE scores for Maths and four science subjects (science, chemistry, physics and biology) reported at age 16. Finally, for the MCS we draw on tests administered at ages 7 and 11. The NFER Number Skills test was taken at age 7. We also include the teacher's evaluation of the MCS member's ability at age 11 in maths, science and technology. These scores equal 5, 4, 3, 2 and 1, respectively, when the teacher evaluated the MCS member to be: well-below average, below average, average, above average and well-above average. Summary statistics can be found in Tables A.2 through A.3 in Appendix A.

### 3.2.3 Motor skills (gross motor and fine motor) ( $gross_{i,child}$ , $fine_{i,child}$ )

We consider instruments related to both gross motor and fine motor skills, captured across childhood as our proxies for motor skills. Specifically, gross motor skills are those which require whole body movement and involve the core stabilising muscles of the body to perform everyday functions, such as standing, walking, running, and sitting upright. It also includes eye-hand coordination skills such as ball skills. Fine motor skills are smaller movements. They include clothing fastenings, cleaning teeth, using cutlery drawing, writing and colouring, as well as cutting and pasting.

For the NCDS, we draw on teacher-assessed measures of gross motor skills at ages 7 and 11 that follow definitions by Sigurdsson et al. (2002).<sup>26</sup> This definition is consistent in the BCS with measurements taken at age 10. In the MCS, gross motor skills are estimated using a subset of the Denver Developmental screening test (Frankenburg and Dodds 1967) and are assessed at age 1 by the child's parent.

We have two measures of fine motor skills in the NCDS. These are based on the teacher assessment of the Human Figure Drawing test (age 7) and the Copying-Design test (ages 7 and 11). The same measures of fine motor ability are available in the BCS (age 10). For the MCS, we again draw on relevant sub-components of the Denver Developmental Screening test, assessed at age 1 by the primary carer. Summary statistics can be found in Tables A.2 and A.3 in Appendix A.

<sup>24</sup> Students completing A-levels stay in school until roughly 18 years and generally aim for third level education. A certificate of secondary education (CSE), O-levels or a general certificate of secondary education (GCSE) represent a low-level secondary school qualification that is usually achieved when the student is aged 15.

<sup>25</sup> These are the Friendly Maths Test and the Recall of Digit subscale of the BAS administered at age 10, and the actual raw GCSE scores for Maths and four science subjects (science, chemistry, physics and biology) reported at age 16.

<sup>26</sup> Because of the absence of direct and positive measures of gross motor skills in the NCDS and BCS, Sigurdsson et al. (2002) exploit the five measures of gross motor impairment (rated by class teacher) and calculate for the average score of gross motor deficiency. Smaller scores indicate more positive gross motor development. In the MCS, we can use Denver Developmental Scale to directly measure gross motor skills at early ages.

### 3.2.4 Non-cognitive skills (externalising and internalising behaviour) ( $\text{behav}_{i,\text{child}}$ )

We construct two separate measures of non-cognitive skills in childhood which proxy externalising behaviour and internalising behaviour, respectively. For all three surveys, we choose assessments provided by teachers over parents. In the NCDS, the behaviour scores are calculated from relevant items taken from the Bristol Social Adjustment Guide (BSAG) at ages 7 and 11. Essentially, we separate the questions into proxies that represent internalising and externalising behaviour. For the BCS, we utilize the Rutter's Behavioural Scale at age 10 (Rutter 1967) in the same way. Finally, measures of internalising and externalising behaviour in the MCS are constructed from the Strength and Difficulty Questionnaire (SDQ) at age 7. Summary statistics can be found in Tables A.2 and A.3 in Appendix A.

### 3.2.5 Health conditions (childhood physical health issues, psychological health issues) ( $\text{phy}_{i,\text{child}}$ , $\text{psy}_{i,\text{child}}$ )

We follow Goodman et al. (2011) when constructing the measures of psychological and physical health in childhood and classify physical health issues into (a) major physical health and (b) minor physical health issues. For the NCDS, medical assessments were conducted at ages 7 and 16. At age 11, the information came from the parent's report. In the BCS, the medical assessment was administered at ages 5, 10 and 16. For the MCS, we use parental reports on the child's medical conditions at ages 3, 5, 7 and 11.

The psychological health measures in the NCDS are calculated from medical examinations capturing emotional maladjustment at ages 7 and 16, as well as parental reports of the child's mental health support visits for ages 11 and 16. In the BCS, we follow similar classifications to the NCDS, drawing on data from parent reports of mental health support visits at ages 10 and 16. At age 16, the child also went through several medical assessments, capturing aspects of emotional maladjustment. Finally, for the MCS, psychological maladjustment is captured by teachers' and parents' reports of mental illness at ages 11/12, and reports of adolescent mental health services utilisation at school. Summary statistics can be found in Tables A.2 and A.3 in Appendix A.

### 3.2.6 Parental investments ( $\text{invest}_{i,\text{child}}$ )

We add measures of parental inputs, which mainly capture time inputs. For the NCDS, we draw on parent-reported frequencies at age 7 of how often the mother and father reading to their child; teacher-rated levels of parental interest in their child's education at age 7, and parental reports of engaging in various activities with their child at ages 7 and 11 (see Appendix A). For the BCS, we use the information on the parental-assessed frequency of the mother and father read to their child measured at age 5, teacher-rated levels of mother and father interest in their child's education at age 10 years and parental reports of engaging in various activities with their child at ages 5 and 10. For the MCS, we draw on the frequency of reading, and parent-assessed frequency reports at ages 3 and 5 of whether the parents: (i) read to the child, (ii) tell stories to the child, (iii) paint with the child, and (iv) play music with the child. We also draw on teacher-rated levels of parental interest in the child's education when aged 11 years,

and a set of variables on the frequency of parents visiting various places with the child at ages 3 and 5. Summary statistics can be found in Tables A.2 and A.3 in Appendix A.

### 3.2.7 Parental aspirations for the child ( $aspire_{i,child}$ )

We also consider measures of parental aspirations. For the NCDS, we draw on parent reports at age 16 that are equal to 1 if a parent wishes their child to leave school at age 15 years, and 0 otherwise. We consider additional measures of aspirations, measured at age 16, which correspond to how long the parent wishes their NCDS child to stay in education (16 years, 18 years, beyond 18 years or uncertain).

In the BCS, we draw on parent reports measured at age 16. We create a dummy variable that is equal to 1 if a parent advises their child to leave full-time education immediately after age 16, and 0 otherwise. We also consider additional measures of aspirations, measured at age 16, which correspond to how long the parent wishes the BCS child to stay in education (leaves at 16, finishes A-levels, goes to university, uncertain).

In the MCS, we draw on parent reports when the child is aged 11 on their perception of whether their child will attend university. Specifically, parents are asked to estimate the likelihood they think the child will attend university: very likely, fairly likely, not very likely and not at all. The responses are then added to the regression as a set of dummy variables.

### 3.2.8 External influences ( $external_{i,child}$ )

Finally, we try and capture aspects of the child's external environment. Given the data available, the variables we consider mainly capture the school environment. For the NCDS, we include a set of dummy variables that capture whether the child's teacher perceives they would benefit from further education (measured when the NCDS child is 16). We also include a separate set of dummy variables which capture the teacher's expectation of the highest level of education the child is likely to attain (university, lower college, advanced course, certificate, other further education, part-time professional qualification, other part-time education, and no other qualification). Similarly, for the BCS we include a set of dummy variables which capture the BCS child's teacher report on whether further education will benefit the BCS child. A further set of indicator variables are constructed from the teacher's expectation of whether the child would attend further education after age 16. For the MCS, we also add similar measures. These measures indicate the teacher's perception of whether the MCS child would (a) stay in full-time education after age 16, and (b) attend university. These variables were measured at when the MCS child was 11 years.

For the NCDS and BCS, we also capture some characteristics of the child's classmates. For the NCDS and the BCS, these variables are measured at age 16 and relate to the school-wide: (i) share of fathers from non-manual occupations, (ii) share of students staying on at school last year, (iii) share of girls obtaining at least two pass grades of GCSE or equivalent, (iv) share of 15-year old girls studying GCSE or equivalent only, (v) share of boys obtaining 2 passes of GCSE or equivalent, and (vi) share



of 15-year old boys studying GCSE or equivalent. Summary statistics can be found in Tables A.2 and A.3 in Appendix A.

### 3.3 Inclusion criteria

We begin with the total observations from the first sweep of each cohort survey (NCDS = 18,558, BCS = 18,752, MCS = 19,518). We drop observations with missing values on gender (NCDS = 4, BCS = 326, MCS = 700). We drop observations with missing values on realised occupation around age 33/34 years old (the NCDS and the BCS).<sup>27</sup> For the MCS, we drop observations with missing values on aspired occupation in the fifth sweep (age 11). The exclusions up to this stage reduce the NCDS, the BCS and the MCS samples to 11,469, 10,234 and 11,200 observations, respectively. We then match the occupation around age 33/34 to the associated occupation averages, generated from the QLFS and ONET. We are left with the samples of 9722, 8973 and 11,200 observations for NCDS, BCS and MCS, respectively. (See Appendix Table B.1).

An issue when working with cohort data, with this many variables and waves, is that there are many missing values. To address this problem, we apply mean imputation with missing indicator variables (the so-called Missing Indicator method) to the control variables described in (i) through (vii) above. When a variable is missing, we replace it with the average value from the non-missing sample. In the finalised sample, the share of females is 49.5%, 48.4% and 51.3% in the NCDS, the BCS and the MCS, respectively. (See Table B.1 in the Appendix.)

## 4 Results

### 4.1 Gender sorting in labour markets across birth cohorts

Table 2 details the coefficient on the female dummy with its associated standard error, alongside the adjusted  $R^2$  for each model. As we move through the rows from (i) to (ix) in each panel, we are estimating richer variants of Eq. (1). It is striking that while there are significant differences in the female coefficient across all three cohorts, the coefficient is not attenuated when childhood variables are added. For example, from panel A (i), a female born in 1958 chooses an occupation where the share of males is 45% lower, on average, as compared to their male peers. This compares to 41% for females born in 1970, and 34% for females born in 2000. This suggests that over time UK females have been more often choosing occupations with a higher share of males; however, the gap between the aspirations of females and males in the most recent cohort is still marked.

<sup>27</sup> There are two survey waves on life outcomes around age 30s for the BCS cohort (ages 30 and 34). In this analysis, we prioritise the responses given at the 34-year survey (1984). To maintain the size of observations of the BCS sample, we supplement any missing values of our variables at age 34 with the information given at age 30 to the exact questions. This strategy increases the final BCS sample from 6870 to 8973 observations. Running the models only with the 6,870 observations does not change the results.

Table 2 Estimates for share of males regressions

	NCDS (1958) 1991		BCS (1970) 2004		MCS (2000) 2012	
	Beta	[SE]	Beta	[SE]	Beta	[SE]
<i>Panel A: Share of men in an occupation</i>						
(i) + Female	- 0.446	[0.005]	- 0.411	[0.005]	- 0.335	[0.005]
(ii) + Baseline	- 0.446	[0.005]	- 0.41	[0.005]	- 0.335	[0.005]
(iii) + Cognitive skills	- 0.442	[0.005]	- 0.41	[0.005]	- 0.334	[0.005]
(iv) + Motor skills	- 0.441	[0.005]	- 0.409	[0.005]	- 0.332	[0.005]
(v) + Non-cognitive skills	- 0.439	[0.005]	- 0.405	[0.005]	- 0.328	[0.005]
(vi) + Health attainment	- 0.438	[0.005]	- 0.405	[0.005]	- 0.328	[0.005]
(vii) + Parent's investments	- 0.438	[0.005]	- 0.399	[0.006]	- 0.328	[0.005]
(viii) + Parent's aspirations	- 0.437	[0.006]	- 0.398	[0.006]	- 0.327	[0.005]
(ix) + External influences	- 0.452	[0.006]	- 0.4	[0.006]	- 0.325	[0.005]
<i>Panel B: Share of men is 80% or higher in an occupation</i>						
(i) + Female	- 0.486	[0.008]	- 0.486	[0.008]	- 0.459	[0.008]
(ii) + Baseline	- 0.487	[0.008]	- 0.484	[0.008]	- 0.459	[0.008]
(iii) + Cognitive skills	- 0.488	[0.008]	- 0.481	[0.008]	- 0.455	[0.008]
(iv) + Motor skills	- 0.486	[0.008]	- 0.48	[0.008]	- 0.452	[0.008]
(v) + Non-cognitive skills	- 0.482	[0.009]	- 0.473	[0.008]	- 0.447	[0.008]
(vi) + Health attainment	- 0.481	[0.009]	- 0.473	[0.008]	- 0.448	[0.008]
(vii) + Parent's investments	- 0.48	[0.009]	- 0.462	[0.009]	- 0.447	[0.009]
(viii) + Parent's aspirations	- 0.479	[0.009]	- 0.46	[0.009]	- 0.443	[0.009]
(ix) + External influences	- 0.506	[0.009]	- 0.464	[0.009]	- 0.438	[0.009]
	Adj R2		Adj R2		Adj R2	
	0.458		0.412		0.306	
	0.463		0.412		0.313	
	0.469		0.413		0.314	
	0.47		0.414		0.314	
	0.47		0.416		0.316	
	0.47		0.418		0.316	
	0.47		0.42		0.318	
	0.471		0.42		0.319	
	0.482		0.42		0.319	
	0.281		0.28		0.256	
	0.301		0.291		0.26	
	0.324		0.3		0.263	
	0.324		0.301		0.264	
	0.325		0.303		0.268	
	0.325		0.305		0.268	
	0.326		0.306		0.269	
	0.328		0.308		0.272	
	0.342		0.31		0.275	

**Table 2** (continued)

	NCDS (1958) 1991		BCS (1970) 2004		MCS (2000) 2012	
	Beta	[SE]	Beta	[SE]	Beta	[SE]
N of panels A and B		9722		8973		11,200
<i>Panel C: Probability of being dis-employed</i>						
(i) + Female	0.211	[0.008]	0.082	[0.005]	- 0.066	[0.008]
(ii) + Baseline	0.211	[0.008]	0.085	[0.005]	- 0.066	[0.008]
(iii) + Cognitive skills	0.204	[0.008]	0.092	[0.005]	- 0.066	[0.008]
(iv) + Motor skills	0.208	[0.008]	0.091	[0.005]	- 0.065	[0.008]
(v) + Non-cognitive skills	0.211	[0.008]	0.093	[0.005]	- 0.061	[0.008]
(vi) + Health attainment	0.212	[0.008]	0.096	[0.006]	- 0.062	[0.008]
(vii) + Parent's investments	0.211	[0.008]	0.097	[0.006]	- 0.063	[0.009]
(viii) + Parent's aspirations	0.212	[0.008]	0.095	[0.006]	- 0.062	[0.009]
(ix) + External influences	0.216	[0.009]	0.097	[0.006]	- 0.062	[0.009]
N of panel C		13,610		11,695		12,271

The tables show the estimated coefficient (beta), the standard error associated with the female dummy in each specification, and the adjusted R-square for each specification. (I) is the regression with only female dummy. (II) is (I) with the family variables. (III) is (II) with childhood cognitive skills. (IV) is (III) with childhood motor skills. (V) is (IV) with childhood non-cognitive skills. (VI) is (V) with childhood physical and psychological health conditions. (VII) is (VI) with parental investment variables. (VIII) is (VII) with variables indicating parental aspirations for children. Finally, (IX) is (VIII) with external influences from school peer and class teacher around age 16

Turning to Table 2 panel B, there is no difference in the probability that females born in 1958 or 1970 will choose a job that has a share of males of 80% or higher as compared to their male peers. The coefficient on female for those born in 2000 is slightly reduced, implying that females in this cohort are 46% less likely to choose occupations with the highest share of males as compared to comparable males. This emphasises that females still shy away from occupations with the highest shares of males.<sup>28</sup> Across all three cohorts, the female coefficient is not markedly attenuated with the addition of childhood variables, and the adjusted  $R^2$  remains flat. The result found here is consistent with the literature that shows that females are less likely to finish STEM majors (for instance, see Arcidiacono et al. 2016; Ost 2010). To explicitly check this result, we run an additional set of estimations with now the dependent variable is an indicator variable equal to 1 if the occupation is a STEM occupation and zero otherwise. Similarly, the female coefficients do not show much within-cohort variations after the inclusion of childhood variables. Instead, the female coefficient reduces its (negative) magnitude from the 1958 NCDS to the 1970 BCS, and eventually, becomes positive for the 2000 MCS (see Appendix Table D.1 for details).

Next, Table 2 panel C focuses on the probability a female is dis-employed (or for MCS children aspires to have a job that is not 'good'). For the NCDS, being female is associated with a 21% increase in the probability of being dis-employed as compared to male peers. This compares to 9% for the BCS. The female coefficient for the MCS is not significantly different from zero, emphasising that for the most recent cohort there is no difference in aspirations between males and females in their tendency to expect a 'good' job by age 30. The addition of the childhood variables does not either attenuate the female coefficient significantly for any of the cohorts, or change the  $R^2$  notably.

Table 3 panel A highlights that the BCS cohort have occupations with a smaller wage gap, as compared to females in the NCDS. However, the aspirations of the MCS cohort, if fulfilled, would cause a greater gender pay gap than those born in 1970 and 1958. Specifically, MCS girls are aspiring to do jobs that are paid 31% lower than males. The childhood variables do not attenuate the female coefficient significantly for any of the cohorts, and the  $R^2$  is flat. For the BCS cohorts, adding the childhood variables actually increases the female coefficient, with the addition of cognitive skills having the greatest impacts.

Table 3 panel B highlights that the propensity for females to sort into jobs with lower average hours, as compared to male peers, has declined significantly and substantively over the three cohorts. However, the difference between female aspirations and their male peers is still substantive for those born in 2000. The addition of the childhood variables gives modest attenuation across the three cohorts, but these additions do increase the adjusted  $R^2$  for the 1958 cohort (from 11% in the most basic model to 14% for the fullest model). For the other two cohorts, the  $R^2$  is relatively flat.

Table 3 panel C documents the results for the flexibility regressions. Overtime, the gap between males and females in terms of sorting into flexible jobs has narrowed. Notably, the coefficient for the aspirations of the 2000 cohort is roughly one third of

<sup>28</sup> In addition, the literature highlights that females may underestimate the gains from 'male-dominated' fields (see, for instance, Zafar 2011, Zafar 2013, Arcidiacono et al. 2012).

**Table 3** Estimates for income, average hours, flexibility and probability of being employed regressions

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel A: Log of average gross income</i>									
(i) + Female	- 0.238	[0.007]	0.1	- 0.145	[0.008]	0.033	- 0.314	[0.008]	0.126
(ii) + Baseline	- 0.237	[0.007]	0.212	- 0.149	[0.008]	0.158	- 0.313	[0.008]	0.142
(iii) + Cognitive Skills	- 0.23	[0.007]	0.327	- 0.158	[0.007]	0.231	- 0.311	[0.008]	0.148
(iv) + Motor Skills	- 0.234	[0.007]	0.329	- 0.163	[0.008]	0.238	- 0.31	[0.008]	0.148
(v) + Non-Cognitive Skills	- 0.238	[0.007]	0.33	- 0.167	[0.008]	0.24	- 0.309	[0.008]	0.154
(vi) + Health attainment	- 0.238	[0.007]	0.331	- 0.168	[0.008]	0.241	- 0.309	[0.008]	0.154
(vii) + Parent's investments	- 0.237	[0.007]	0.332	- 0.175	[0.008]	0.244	- 0.308	[0.009]	0.154
(viii) + Parent's aspirations	- 0.238	[0.007]	0.336	- 0.174	[0.008]	0.259	- 0.313	[0.009]	0.158
(ix) + External influences	- 0.244	[0.008]	0.342	- 0.173	[0.008]	0.261	- 0.312	[0.009]	0.158
<i>Panel B: Average hours</i>									
(i) + Female	- 8.213	[0.116]	0.343	- 6.917	[0.123]	0.265	- 2.71	[0.088]	0.083
(ii) + Baseline	- 8.194	[0.116]	0.348	- 6.926	[0.122]	0.277	- 2.703	[0.087]	0.111
(iii) + Cognitive skills	- 8.125	[0.125]	0.352	- 6.954	[0.123]	0.281	- 2.71	[0.089]	0.112
(iv) + Motor skills	- 8.127	[0.126]	0.354	- 6.969	[0.124]	0.284	- 2.695	[0.090]	0.113
(v) + Non-cognitive skills	- 8.107	[0.128]	0.354	- 6.881	[0.125]	0.286	- 2.628	[0.092]	0.114
(vi) + Health attainment	- 8.101	[0.128]	0.354	- 6.898	[0.126]	0.287	- 2.631	[0.092]	0.113
(vii) + Parent's investments	- 8.082	[0.130]	0.354	- 6.823	[0.131]	0.288	- 2.633	[0.094]	0.113
(viii) + Parent's aspirations	- 8.077	[0.130]	0.355	- 6.794	[0.133]	0.289	- 2.621	[0.094]	0.115
(ix) + External influences	- 8.447	[0.143]	0.363	- 6.844	[0.135]	0.289	- 2.584	[0.095]	0.116

Table 3 (continued)

	NCDS (1958) 1991		BCS (1970) 2004		MCS (2000) 2012	
	Beta	[SE]	Beta	[SE]	Beta	[SE]
<i>Panel C: Flexibility of an occupation</i>						
(i) + Female	0.184	[0.005]	0.164	[0.005]	0.063	[0.005]
(ii) + Baseline	0.183	[0.005]	0.164	[0.005]	0.063	[0.005]
(iii) + Cognitive skills	0.174	[0.005]	0.165	[0.005]	0.062	[0.005]
(iv) + Motor skills	0.175	[0.005]	0.165	[0.005]	0.062	[0.005]
(v) + Non-cognitive skills	0.174	[0.005]	0.164	[0.005]	0.059	[0.006]
(vi) + Health attainment	0.174	[0.005]	0.165	[0.005]	0.059	[0.006]
(vii) + Parent's investments	0.175	[0.005]	0.16	[0.005]	0.057	[0.006]
(viii) + Parent's aspirations	0.174	[0.005]	0.161	[0.006]	0.059	[0.006]
(ix) + External influences	0.181	[0.006]	0.161	[0.006]	0.058	[0.006]
Observations		9722		8973		11,200

See notes to Table 2. For the MCS, probability of being employed equals 1 if the child expects to have their own children and also does expects to have a good job by age 30, and 0 otherwise

that for NCDS children. For all three cohorts, the addition of the childhood variables does little to attenuate the female coefficient and the  $R^2$  is flat. This suggests that changing patterns over time determined the movement away from flexibility, over and above childhood factors.

Next, we turn to examine gender sorting in terms of job content. Table 4 documents the estimates from models which a series of job content ('people', 'brains', 'brawn') is used as the dependent variable. A few stylised facts emerge. First, over time females have moved more towards jobs that are high in brains as compared to their male peers, with the MCS girls being substantively more likely than the MCS boys to choose jobs with high brain content. Females across all three cohorts choose jobs with higher people content and lower brawn content, as compared to male peers. However, for the MCS cohorts the coefficients are about the half the size. This suggests that females in this cohort are still choosing job content along gender lines, but it is not as marked as it was for older cohorts. Finally, across all three cohorts, there is a gender gap in the propensity to choose jobs that are highly competitive, with males choosing work with higher competitive content. Markedly, there is no attenuation in this trend over time, with the MCS girls being substantively *less* likely to aspire to work in jobs that are competitive as compared to MCS males. Second, while there a couple of exceptions (that is, 'brains' for NCDS and 'brawn' for BCS), in general, the addition of the childhood variables does not attenuate the female coefficient in the job content regressions. Third, across the three cohorts, there are substantive changes to the  $R^2$  when we add the childhood variables for the 'brains' and 'brawn' regressions (in both cases the addition of the cognitive proxies is the most important). By contrast, the  $R^2$  is static despite these additions for the people and competitiveness content regressions.

## 4.2 Gender sorting among high cognitive ability individuals

Tables 5, 6 and 7 document estimates from Eq. (1) for children who are in the top 25% of the cognitive distribution. There has always been less of a difference between high skilled females and their male peers in their tendency to sort into jobs with high shares of males. While the gradient of the coefficient decreases across the three cohorts, it is flatter as compared to the average female regressions. Notably, females with high cognitive ability born in 2000 are still aspiring to enter occupations with 27% lower male shares, as compared to males. Panel B highlights that high ability females have become *less* likely than their male peers to sort/aspire into jobs with the highest shares of males. Given that these regressions pertain to the highest skilled females only, the trends of females sorting into science, technology and engineering are not improving across the three cohorts.

Consistent with the results for the average female, the gender gap in occupational hours (Table 6, panel B), flexibility (Table 6, panel C) and the potential to be disemployed (Table 5, panel C) has narrowed over time. High ability females have a lower occupation-level average pay gap, if we compare the 1970 cohort to the 1958 cohort. However, females born in 2000 are choosing jobs that have significantly *lower* average pay, as compared to their male peers. Together, these results suggest that females born in 2000 plan to have higher levels of labour force attachment as compared to females

**Table 4** Estimates for people, brains, brown and competitiveness regressions

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel A: People</i>									
(i) + Female	0.421	[0.019]	0.05	0.439	[0.019]	0.057	0.272	[0.019]	0.02
(ii) + Baseline	0.42	[0.019]	0.054	0.439	[0.019]	0.06	0.271	[0.019]	0.026
(iii) + Cognitive skills	0.418	[0.020]	0.056	0.442	[0.019]	0.062	0.277	[0.019]	0.035
(iv) + Motor skills	0.416	[0.020]	0.057	0.442	[0.019]	0.063	0.275	[0.019]	0.035
(v) + Non-cognitive skills	0.415	[0.020]	0.057	0.452	[0.020]	0.064	0.277	[0.020]	0.038
(vi) + Health attainment	0.413	[0.020]	0.056	0.452	[0.020]	0.065	0.276	[0.020]	0.038
(vii) + Parent's investments	0.41	[0.021]	0.058	0.461	[0.021]	0.065	0.277	[0.020]	0.04
(viii) + Parent's aspirations	0.411	[0.021]	0.059	0.462	[0.021]	0.065	0.274	[0.020]	0.04
(ix) + External influences	0.424	[0.023]	0.061	0.464	[0.021]	0.065	0.28	[0.020]	0.04
<i>Panel B: Brains</i>									
(i) + Female	-0.091	[0.019]	0.002	0.06	[0.021]	0.001	0.532	[0.020]	0.066
(ii) + Baseline	-0.089	[0.018]	0.091	0.05	[0.020]	0.097	0.532	[0.020]	0.072
(iii) + Cognitive skills	-0.058	[0.019]	0.189	0.032	[0.020]	0.147	0.532	[0.020]	0.079
(iv) + Motor skills	-0.066	[0.019]	0.19	0.024	[0.020]	0.15	0.528	[0.021]	0.079
(v) + Non-cognitive skills	-0.079	[0.019]	0.192	0.012	[0.020]	0.151	0.522	[0.021]	0.081
(vi) + Health attainment	-0.08	[0.019]	0.192	0.008	[0.021]	0.152	0.525	[0.021]	0.081
(vii) + Parent's investments	-0.08	[0.019]	0.193	-0.006	[0.022]	0.153	0.53	[0.021]	0.081
(viii) + Parent's aspirations	-0.083	[0.019]	0.198	-0.005	[0.022]	0.163	0.516	[0.021]	0.09
(ix) + External influences	-0.056	[0.021]	0.204	-0.001	[0.022]	0.165	0.509	[0.022]	0.09



Table 4 (continued)

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel C: Brown</i>									
(i) + Female	- 0.81	[0.020]	0.149	- 0.823	[0.019]	0.167	- 0.523	[0.015]	0.104
(ii) + Baseline	- 0.812	[0.019]	0.194	- 0.816	[0.019]	0.209	- 0.523	[0.015]	0.126
(iii) + Cognitive skills	- 0.79	[0.020]	0.258	- 0.803	[0.019]	0.238	- 0.518	[0.016]	0.138
(iv) + Motor skills	- 0.783	[0.020]	0.259	- 0.797	[0.019]	0.24	- 0.515	[0.016]	0.139
(v) + Non-cognitive skills	- 0.767	[0.020]	0.26	- 0.774	[0.019]	0.244	- 0.502	[0.016]	0.143
(vi) + Health attainment	- 0.769	[0.020]	0.261	- 0.772	[0.019]	0.245	- 0.503	[0.016]	0.143
(vii) + Parent's investments	- 0.761	[0.020]	0.265	- 0.742	[0.020]	0.247	- 0.502	[0.016]	0.143
(viii) + Parent's aspirations	- 0.755	[0.020]	0.267	- 0.739	[0.020]	0.251	- 0.491	[0.016]	0.151
(ix) + External influences	- 0.791	[0.022]	0.277	- 0.748	[0.020]	0.254	- 0.48	[0.016]	0.153
<i>Panel D: Competitiveness</i>									
(i) + Female	- 0.407	[0.020]	0.041	- 0.383	[0.022]	0.033	- 0.841	[0.028]	0.084
(ii) + Baseline	- 0.406	[0.020]	0.051	- 0.386	[0.022]	0.057	- 0.842	[0.028]	0.094
(iii) + Cognitive skills	- 0.4	[0.021]	0.058	- 0.393	[0.022]	0.063	- 0.844	[0.028]	0.094
(iv) + Motor skills	- 0.405	[0.021]	0.058	- 0.402	[0.022]	0.066	- 0.84	[0.028]	0.095
(v) + Non-cognitive skills	- 0.407	[0.022]	0.06	- 0.396	[0.022]	0.066	- 0.83	[0.029]	0.099
(vi) + Health attainment	- 0.406	[0.022]	0.06	- 0.4	[0.023]	0.066	- 0.835	[0.029]	0.099
(vii) + Parent's investments	- 0.408	[0.022]	0.059	- 0.397	[0.024]	0.065	- 0.843	[0.029]	0.101
(viii) + Parent's aspirations	- 0.41	[0.022]	0.059	- 0.392	[0.024]	0.067	- 0.836	[0.029]	0.102
(ix) + External influences	- 0.388	[0.024]	0.06	- 0.395	[0.024]	0.067	- 0.84	[0.030]	0.102
Observations		9722			8973			11,200	

See Table 2

**Table 5** Estimates for share of males and probability of dis-employment regressions for children with the highest cognitive ability

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel A: Share of men in an occupation</i>									
(i) + Female	-0.318	[0.009]	0.297	-0.317	[0.009]	0.3	-0.27	[0.008]	0.225
(ii) + Baseline	-0.318	[0.009]	0.297	-0.317	[0.009]	0.299	-0.271	[0.009]	0.228
(iii) + Cognitive skills	-0.304	[0.010]	0.301	-0.315	[0.009]	0.3	-0.271	[0.009]	0.229
(iv) + Motor skills	-0.305	[0.010]	0.302	-0.315	[0.009]	0.301	-0.27	[0.009]	0.228
(v) + Non-cognitive skills	-0.302	[0.010]	0.302	-0.311	[0.009]	0.304	-0.265	[0.009]	0.229
(vi) + Health attainment	-0.3	[0.010]	0.302	-0.311	[0.009]	0.306	-0.265	[0.009]	0.229
(vii) + Parent's investments	-0.297	[0.010]	0.302	-0.307	[0.010]	0.305	-0.262	[0.009]	0.228
(viii) + Parent's aspirations	-0.297	[0.010]	0.306	-0.306	[0.010]	0.305	-0.26	[0.009]	0.232
(ix) + External influences	-0.327	[0.013]	0.316	-0.307	[0.010]	0.306	-0.258	[0.009]	0.234
<i>Panel B: Share of men is 80% or higher in an occupation</i>									
(i) + Female	-0.279	[0.013]	0.124	-0.359	[0.014]	0.181	-0.380	[0.014]	0.184
(ii) + Baseline	-0.277	[0.013]	0.134	-0.356	[0.014]	0.191	-0.381	[0.014]	0.188
(iii) + Cognitive skills	-0.265	[0.015]	0.148	-0.357	[0.014]	0.199	-0.377	[0.014]	0.191
(iv) + Motor skills	-0.265	[0.015]	0.147	-0.357	[0.014]	0.198	-0.376	[0.014]	0.191
(v) + Non-cognitive skills	-0.262	[0.015]	0.147	-0.349	[0.015]	0.2	-0.374	[0.015]	0.192
(vi) + Health attainment	-0.26	[0.015]	0.146	-0.348	[0.015]	0.2	-0.374	[0.015]	0.191
(vii) + Parent's investments	-0.256	[0.015]	0.146	-0.338	[0.016]	0.202	-0.369	[0.015]	0.191

**Table 5** (continued)

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
(viii) + Parent's aspirations	- 0.256	[0.015]	0.152	- 0.336	[0.016]	0.207	- 0.366	[0.015]	0.195
(ix) + External influences	- 0.281	[0.019]	0.167	- 0.337	[0.016]	0.209	- 0.361	[0.015]	0.199
N of panels A and B		2952			3030			3877	
<i>Panel C: Probability of Being Dis-Employed</i>									
(i) + Female	0.209	[0.012]	0.098	0.098	[0.009]	0.037	- 0.075	[0.015]	0.007
(ii) + Baseline	0.208	[0.013]	0.096	0.098	[0.009]	0.04	- 0.074	[0.015]	0.014
(iii) + Cognitive skills	0.199	[0.014]	0.095	0.099	[0.009]	0.039	- 0.081	[0.015]	0.018
(iv) + Motor skills	0.2	[0.014]	0.094	0.099	[0.009]	0.038	- 0.081	[0.015]	0.019
(v) + Non-cognitive skills	0.203	[0.014]	0.096	0.099	[0.010]	0.037	- 0.076	[0.015]	0.018
(vi) + Health attainment	0.204	[0.014]	0.096	0.099	[0.010]	0.035	- 0.078	[0.015]	0.019
(vii) + Parent's investments	0.201	[0.014]	0.096	0.102	[0.010]	0.037	- 0.075	[0.016]	0.021
(viii) + Parent's aspirations	0.202	[0.014]	0.094	0.101	[0.010]	0.037	- 0.074	[0.016]	0.021
(ix) + External influences	0.216	[0.017]	0.094	0.104	[0.011]	0.032	- 0.072	[0.016]	0.022
N of Panel C		N = 4008			N = 3847			N = 3877	

See notes to Table 2 and also Appendix A for full details of how the children with highest cognitive ability are determined

**Table 6** Estimates for log of average income, average hours and flexibility regressions for children with the highest cognitive ability

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel A: Log of average gross income</i>									
(i) + Female	- 0.207	[0.013]	0.074	- 0.165	[0.014]	0.042	- 0.238	[0.013]	0.093
(ii) + Baseline	- 0.213	[0.013]	0.125	- 0.163	[0.014]	0.138	- 0.239	[0.013]	0.105
(iii) + Cognitive Skills	- 0.203	[0.014]	0.171	- 0.155	[0.014]	0.184	- 0.236	[0.013]	0.109
(iv) + Motor Skills	- 0.207	[0.014]	0.172	- 0.16	[0.014]	0.189	- 0.237	[0.013]	0.109
(v) + Non-Cognitive Skills	- 0.212	[0.014]	0.172	- 0.169	[0.014]	0.193	- 0.235	[0.013]	0.116
(vi) + Health attainment	- 0.211	[0.014]	0.173	- 0.17	[0.014]	0.195	- 0.236	[0.013]	0.116
(vii) + Parent's investments	- 0.214	[0.015]	0.176	- 0.178	[0.014]	0.198	- 0.233	[0.014]	0.116
(viii) + Parent's aspirations	- 0.214	[0.015]	0.179	- 0.179	[0.014]	0.22	- 0.236	[0.014]	0.121
(ix) + External influences	- 0.229	[0.018]	0.183	- 0.179	[0.015]	0.22	- 0.235	[0.014]	0.122
<i>Panel B: Average hours</i>									
(i) + Female	- 5.202	[0.198]	0.194	- 4.987	[0.190]	0.182	- 1.665	[0.147]	0.035
(ii) + Baseline	- 5.233	[0.198]	0.204	- 4.973	[0.191]	0.197	- 1.701	[0.146]	0.058
(iii) + Cognitive skills	- 4.983	[0.220]	0.213	- 4.872	[0.192]	0.201	- 1.732	[0.151]	0.057
(iv) + Motor skills	- 5.017	[0.219]	0.212	- 4.912	[0.195]	0.206	- 1.72	[0.152]	0.057
(v) + Non-cognitive skills	- 4.973	[0.221]	0.212	- 4.878	[0.197]	0.207	- 1.66	[0.156]	0.057
(vi) + Health attainment	- 4.917	[0.221]	0.215	- 4.874	[0.198]	0.206	- 1.651	[0.157]	0.056
(vii) + Parent's investments	- 4.877	[0.224]	0.215	- 4.84	[0.207]	0.206	- 1.605	[0.161]	0.055
(viii) + Parent's aspirations	- 4.875	[0.223]	0.214	- 4.816	[0.208]	0.206	- 1.575	[0.160]	0.062
(ix) + External influences	- 5.349	[0.271]	0.218	- 4.873	[0.216]	0.208	- 1.545	[0.162]	0.062

Table 6 (continued)

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel C: Flexibility of an occupation</i>									
(i) + Female	0.17	[0.009]	0.099	0.145	[0.009]	0.079	0.055	[0.009]	0.01
(ii) + Baseline	0.169	[0.009]	0.1	0.141	[0.009]	0.089	0.055	[0.009]	0.009
(iii) + Cognitive skills	0.151	[0.010]	0.108	0.139	[0.009]	0.089	0.054	[0.009]	0.008
(iv) + Motor skills	0.154	[0.010]	0.107	0.139	[0.009]	0.089	0.053	[0.009]	0.008
(v) + Non-cognitive skills	0.153	[0.011]	0.107	0.136	[0.009]	0.09	0.051	[0.010]	0.008
(vi) + Health attainment	0.151	[0.011]	0.107	0.137	[0.009]	0.089	0.051	[0.010]	0.009
(vii) + Parent's investments	0.151	[0.011]	0.105	0.138	[0.010]	0.086	0.05	[0.010]	0.011
(viii) + Parent's aspirations	0.151	[0.011]	0.107	0.139	[0.010]	0.09	0.051	[0.010]	0.013
(ix) + External influences	0.176	[0.013]	0.11	0.142	[0.010]	0.09	0.051	[0.010]	0.011
N		2952			3030			3877	

See notes to Table 1 and also Appendix A for full details of how the children with highest cognitive ability are determined

**Table 7** Estimates for people, brains, brawn and competitiveness regressions for children with the highest cognitive ability

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel A: People</i>									
(i) + Female	0.28	[0.034]	0.022	0.383	[0.033]	0.042	0.309	[0.034]	0.022
(ii) + Baseline	0.277	[0.034]	0.028	0.376	[0.033]	0.044	0.308	[0.034]	0.028
(iii) + Cognitive skills	0.272	[0.037]	0.032	0.376	[0.034]	0.047	0.315	[0.035]	0.043
(iv) + Motor skills	0.259	[0.037]	0.03	0.373	[0.034]	0.048	0.308	[0.035]	0.044
(v) + Non-cognitive skills	0.257	[0.038]	0.031	0.375	[0.034]	0.05	0.302	[0.036]	0.045
(vi) + Health attainment	0.268	[0.038]	0.034	0.376	[0.035]	0.054	0.297	[0.036]	0.044
(vii) + Parent's investments	0.264	[0.039]	0.037	0.381	[0.036]	0.052	0.296	[0.037]	0.047
(viii) + Parent's aspirations	0.264	[0.039]	0.045	0.38	[0.036]	0.053	0.29	[0.037]	0.048
(ix) + External influences	0.259	[0.046]	0.054	0.385	[0.037]	0.054	0.298	[0.037]	0.048
<i>Panel B: Brains</i>									
(i) + Female	-0.32	[0.035]	0.027	-0.163	[0.037]	0.006	0.335	[0.036]	0.024
(ii) + Baseline	-0.33	[0.035]	0.057	-0.154	[0.036]	0.078	0.33	[0.036]	0.027
(iii) + Cognitive skills	-0.283	[0.038]	0.095	-0.13	[0.036]	0.108	0.336	[0.037]	0.034
(iv) + Motor skills	-0.286	[0.038]	0.098	-0.139	[0.036]	0.11	0.335	[0.037]	0.034
(v) + Non-cognitive skills	-0.291	[0.038]	0.098	-0.154	[0.037]	0.11	0.325	[0.038]	0.035
(vi) + Health attainment	-0.295	[0.039]	0.095	-0.16	[0.037]	0.111	0.325	[0.038]	0.037
(vii) + Parent's investments	-0.293	[0.039]	0.095	-0.175	[0.039]	0.114	0.323	[0.039]	0.037
(viii) + Parent's aspirations	-0.29	[0.039]	0.101	-0.179	[0.039]	0.131	0.311	[0.039]	0.047
(ix) + External influences	-0.295	[0.048]	0.104	-0.181	[0.040]	0.129	0.303	[0.039]	0.049

Table 7 (continued)

	NCDS (1958) 1991			BCS (1970) 2004			MCS (2000) 2012		
	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2	Beta	[SE]	Adj R2
<i>Panel C: Brown</i>									
(i) + Female	- 0.488	[0.033]	0.067	- 0.526	[0.032]	0.084	- 0.343	[0.028]	0.043
(ii) + Baseline	- 0.481	[0.033]	0.08	- 0.525	[0.032]	0.108	- 0.347	[0.028]	0.056
(iii) + Cognitive skills	- 0.459	[0.036]	0.101	- 0.529	[0.032]	0.123	- 0.346	[0.028]	0.063
(iv) + Motor Skills	- 0.464	[0.037]	0.101	- 0.531	[0.032]	0.122	- 0.346	[0.028]	0.064
(v) + Non-cognitive skills	- 0.449	[0.037]	0.103	- 0.504	[0.032]	0.128	- 0.343	[0.029]	0.066
(vi) + Health attainment	- 0.44	[0.037]	0.102	- 0.498	[0.033]	0.13	- 0.347	[0.029]	0.066
(vii) + Parent's investments	- 0.417	[0.038]	0.108	- 0.473	[0.034]	0.13	- 0.347	[0.029]	0.066
(viii) + Parent's aspirations	- 0.418	[0.038]	0.111	- 0.468	[0.035]	0.139	- 0.338	[0.029]	0.076
(ix) + External influences	- 0.477	[0.046]	0.118	- 0.476	[0.036]	0.143	- 0.326	[0.029]	0.081
<i>Panel D: Competitiveness</i>									
(i) + Female	- 0.508	[0.038]	0.056	- 0.427	[0.039]	0.037	- 0.702	[0.046]	0.062
(ii) + Baseline	- 0.521	[0.038]	0.061	- 0.431	[0.039]	0.057	- 0.7	[0.046]	0.07
(iii) + Cognitive skills	- 0.472	[0.042]	0.065	- 0.419	[0.040]	0.059	- 0.71	[0.047]	0.072
(iv) + Motor skills	- 0.473	[0.042]	0.064	- 0.432	[0.040]	0.063	- 0.712	[0.047]	0.072
(v) + Non-cognitive skills	- 0.47	[0.043]	0.067	- 0.437	[0.041]	0.063	- 0.687	[0.048]	0.074
(vi) + Health attainment	- 0.47	[0.043]	0.069	- 0.438	[0.041]	0.064	- 0.693	[0.048]	0.075
(vii) + Parent's investments	- 0.486	[0.044]	0.068	- 0.437	[0.043]	0.062	- 0.687	[0.049]	0.077
(viii) + Parent's aspirations	- 0.487	[0.044]	0.067	- 0.437	[0.043]	0.061	- 0.678	[0.049]	0.079
(ix) + External influences	- 0.458	[0.053]	0.071	- 0.454	[0.045]	0.062	- 0.679	[0.049]	0.08
N		2952		3030		3877			

See notes to Table 2 and also Appendix A for full details of how the children with highest cognitive ability are determined

in earlier cohorts, but are less extrinsically motivated as compared to their current male peers.

Turning to Table 7, stylised facts consistent with Table 4 emerge. First, the gap between males and females in terms of job brains content has grown over time, with females choosing jobs with higher brain content as compared to their peers. Second, females across all cohorts choose jobs with higher people content, but lower brawn and competitiveness content. Notably, for the MCS cohort the difference between males and females in brawn content is lower than for the NCDS and BCS females, but the competitiveness gap is the larger.

Across Tables 5, 6 and 7, childhood variables do little to attenuate the coefficient of the female dummy. However, the addition of childhood variables does independently explain significant proportions of the variation for a number of outcomes (the log of average occupational income, brains content, and brawn content for all three cohorts).

### 4.3 LASSO estimations and gender effect

Table 8 documents LASSO estimates from the fullest model (see Appendix Tables D.2-D.4 for full LASSO results). First note that in no case is the coefficient on female shrunk to zero leading to the conclusion that gender has been and is now, a key factor in determining how individual's sort. The narrative from the OLS models remains. Overtime, females have sorted more regularly into traditionally male-dominated jobs, have decreased their propensity of being dis-employed as compared to male peers and the gender gap has narrowed in terms of flexibility and hours.

However, for both average and high ability children born in the year 2000 there is a larger greater gender gap in the propensity to pursue jobs that are with over 80% share of males, in average occupational income and occupational competitiveness. Over time, all females have preferred jobs with higher brains and brawn content, as compared to their male peers. The average female also has been choosing jobs with less people content over time, as compared to their male peer. However, for females with high ability the preference of the 2000 cohort is for jobs that are *higher* in people content as compared to those born in 1970.

## 5 Robustness checks

### 5.1 Measurement errors in childhood variables

The childhood variables we consider do not explain a significant proportion of the variation in any of our outcomes over and above what is explained the gender dummy. So, we conclude that the variables we usually think about as being important during childhood do not explain differences in gendered sorting, but gender itself is still a notable and independent determining factor. An obvious conclusion is perhaps the childhood variables we consider are simply measured with error. An easy way to explore this is to look at other outcomes that we think as important in adulthood, and



**Table 8** Lasso regression results

	Share of males	Share of males + 80%	Probability of being dis-employed	Log weekly income	Hours	Flexibility	People	Brains	Brawn	Competitiveness
<i>NCDS (1958) all children</i>										
Female	-0.448	-0.485	0.207	-0.235	-8.346	0.179	0.408	-0.060	-0.797	-0.394
R <sup>2</sup>	0.485	0.339	0.100	0.341	0.369	0.135	0.069	0.201	0.279	0.067
N	9722	9722	13,610	9722	9722	9722	9722	9722	9722	9722
<i>BCS (1970) all children</i>										
Female	-0.303	-0.329	0.077	-0.167	-6.761	0.158	0.425	0.000	-0.740	-0.371
R <sup>2</sup>	0.313	0.215	0.037	0.264	0.292	0.114	0.065	0.170	0.258	0.071
N	8973	8973	11,695	8973	8973	8973	8973	8973	8973	8973
<i>MCS (2000) all children</i>										
Female	-0.325	-0.442	-0.048	-0.307	-2.527	0.055	0.259	0.484	-0.481	-0.808
R <sup>2</sup>	0.321	0.276	0.020	0.160	0.116	0.020	0.044	0.090	0.154	0.104
N	11,200	11,200	11,200	11,200	11,200	11,200	11,200	11,200	11,200	11,200
<i>NCDS (1958) children in top 25% cognitive distribution</i>										
Female	-0.317	-0.248	0.193	-0.218	-5.034	0.148	0.200	-0.223	-0.445	-0.405
R <sup>2</sup>	0.330	0.170	0.104	0.204	0.223	0.104	0.070	0.113	0.131	0.065
N	2952	2952	4008	2952	2952	2952	2952	2952	2952	2952
<i>BCS (1970) children in top 25% cognitive distribution</i>										
Female	-0.300	-0.331	0.083	-0.157	-4.721	0.131	0.339	-0.116	-0.455	-0.386
R <sup>2</sup>	0.307	0.218	0.041	0.231	0.206	0.100	0.059	0.138	0.156	0.071
N	3030	3030	3847	3030	3030	3030	3030	3030	3030	3030

Table 8 (continued)

	Share of males	Share of males + 80%	Probability of being dis-employed	Log weekly income	Hours	Flexibility	People	Brains	Brawn	Competitiveness
<i>MCS (2000) children in top 25% cognitive distribution</i>										
Female	-0.256	-0.355	-0.055	-0.227	-1.431	0.029	0.273	0.281	-0.297	-0.627
$R^2$	0.240	0.198	0.031	0.130	0.069	0.008	0.063	0.061	0.086	0.086
N	3877	3877	3877	3877	3877	3877	3877	3877	3877	3877

look to see how our childhood variables relate to them. We are specifically interested in gauging their impact on the  $R^2$ .

Appendix Table D.5 documents the estimates from this exercise. We document for a number of adult outcomes in the NCDS and BCS (we do not yet observe adult outcomes for MCS children) the coefficient on the female dummy and its associated standard error, along with the  $R^2$  when we estimate regressions with the female dummy only and our fullest specification. We note that for most of these outcomes adding the full set of the childhood variables explains a significant amount of the variation in our outcome. Additionally, for many of our adult outcomes, the female coefficient does change substantively with the addition of the childhood variables. See, for example, the regressions that relate to general health status, attitudes towards racial issues, smoking behaviour in adulthood and the probability of attending university. Overall, we are therefore confident that the variables do measure something meaningful about childhood, but that these variables are not important determinants of occupational sorting nor correlated with gender. We acknowledge that there may be other childhood variables not included in our dataset that may determine occupational sorting.

## 5.2 Gender differences in childhood variables across the birth cohorts

Another potential explanation why we do not observe substantial roles of the childhood variables included in our estimation is that there may be large gender gaps in these measures, and that the gaps may also evolve over time. For instance, Baker and Milligan (2016) show that boys in the MCS receive fewer parental inputs and have lower cognitive test scores. To explicitly check this, we run supplementary linear regressions with each childhood variable as the dependent variable and the female dummy is the explanatory variable with no other control variables. Appendix Table D.6 shows the estimation results from the regression described above. Each cell in the Table reports the estimated coefficient of the female dummy for each individual regression) for each cohort (NCDS, BCS and MCS, respectively).

In panel A “family environment at childhood”, the majority of childhood variables show significant gender gaps. In panels B to G, we do observe that gender gaps exist across many of our childhood variables. However, it is worth noting that some gaps are in favour of males (negative values), while some gaps are in favour of females (positive values) with no particular patterns. For instance, some measures of cognitive skills show positive female coefficients, but others are negative. This is not different from previous literature, which agrees that boys generally do better at numeracy tasks while girls excel in literacy skills (see Cobb-Clark and Moschion (2017) for an extensive discussion). For parental beliefs about the child’s future, we observe that for the NCDS, parents are biased towards boys. This is also the case for teacher’s belief.<sup>29</sup> Nonetheless, this pattern of favouring boys then reverses when we check with the later cohorts (BCS and MCS). Instead, both parents and teachers become more optimistic towards girls’ academic and career future. Over the cohorts, we also cannot pinpoint

<sup>29</sup> See also Lavy and Sand (2015) and Gibbons and Chevalier (2008) for a review of the literature on the role of teacher’s bias.

clear trends in the evolution of the female coefficient. Based on this finding, we would argue that the issue arising from existing gender gaps in childhood measures may not be as concerning.

Since gender gaps in teacher's beliefs do exist, one may want to know how influential a teacher's belief on child's perception of their future, and whether such teacher's influence may evolve over the decades. Ideally, we would like to run a simple regression with actual occupational achievement as the dependent variable and teacher's occupational expectation as the explanatory variable. This set of variable exists only for the NCDS. Therefore, to be able to repeat the analysis across three birth cohorts, we have to use the educational variables as the alternative. In details, we run a simple regression with the educational attainment (an indicator whether attaining college or above) on the teacher's educational expectation, and including the female dummy and its interaction with the teacher's expectation. For the MCS, since we do not yet observe the actual educational outcome yet, the alternative dependent variable is the child's educational aspirations (whether they aspire to attain a college education or above). The result in Appendix Table D.7 shows that teacher's educational expectation can explain a large variation of the child's educational goal. But the size of the correlation reduces over the decades (0.693, 0.391, 0.203 for NCDS, BCS and MCS, respectively). For the earlier cohorts (NCDS and BCS), there is, in fact, no gender difference in the correlation of teacher's expectation and child's academic goal (that is, the interacted terms are insignificant). We observe around 8% gender difference only in the MCS.

### 5.3 Differential growth patterns by industry

In the UK, the services sector experienced significant growth between 1958 and 1970, and the landscape of the goods producing sector was shaped by both offshoring and automation, which contracted and changed the nature of work in terms physicality (Autor and Dorn 2013; Autor et al. 2015; Lordan and Neumark 2018). We are interested in learning whether our findings are driven solely by changes in industries that have experienced high growth levels, or whether they hold consistently across diverse industries. We follow Olivetti and Petrongolo (2016) and run separate sub-sample analysis for the NCDS and BCS cohorts separately by goods producing sector and services sector.<sup>30</sup> The estimates are documented in Appendix Table D.8. Within sector, in general the childhood variables do not attenuate the female coefficient significantly. The few exceptions are for share of males and brains BCS good's sector regressions and the NCDS's competitiveness regression for the services sector. We note that while there are differences in the female coefficient across the two sectors within the cohort data, both sectors mainly display the same gendered sorting patterns. However, over time, the change in the female coefficient for the goods sector is far greater—suggesting that

<sup>30</sup> Due to sample size limitation, we cannot produce estimates by one-digit industry codes. We are unable to include the MCS cohort as the MCS child's aspirations do not include industry. The sectors in the NCDS and BCS data are consistently coded according to the UK SIC-1992. Based on Olivetti and Petrongolo (2016), agriculture, hunting, forestry and fishing; mining, quarrying and utilities; manufacturing and construction are grouped as the goods producing sector. Wholesale and retail trade; hotels and restaurants; transport and storage; financial intermediation; business activities and real estate; public administration and defence; education; health and social work; and other services are classified together as the services sector.

trends in this industry away from physical strength may have greatly reduced aspects of gendered sorting.

#### 5.4 Shifting occupational preferences

It is also interesting to consider whether it changes in the preferences of males versus females that drive the conclusions. Appendix Table D.9 documents the average and standard deviation of our outcomes for each cohort, alongside the overall change occurred between 1958 and 2000. Table D.9 highlights that for the most recent cohort, it is males' increased propensity of pursuing high income and competitive work that is causing us to conclude that the sorting trends of the most recent cohort in these domains is less comparable to male peers as compared to the previous two cohorts. Across the British birth cohorts, females have sorted into occupations that have higher income and are more competitive, but not in the same extent as males. Moreover, males have also more regularly pursued jobs with high people content over time. Meanwhile, for all three cohorts, females more regularly choose jobs that are high in people content.

By contrast, females have continued to pursue jobs with high brawn content over time. Yet, males also regularly choose jobs that are high in brawn. Across cohorts, males have sorted into occupations with similar average weekly work hours (approximately 46 h), whereas there is an increase in females in occupations with higher average hours over time. British women in the recent cohorts have chosen occupations with *less* job flexibility, whilst *across cohorts*, men's jobs have continued to be in highly inflexible jobs. Overall, the comparison between the changes across cohorts and within cohorts suggests that the shifts in gender sorting patterns across cohorts are more marked for females as compared to males, with the exception of changes in aspirations for high income and competitive work by males in the most recent cohort.

## 6 Conclusions

Childhood factors, which encompass not only early skills (cognitive, non-cognitive, health), but also family environment, are known in the literature for their important role in shaping later achievements in life, including occupational choices. Moreover, the literature also documents gender gaps across several of these childhood factors (e.g. Buchmann and DiPrete 2008; Bertrand and Pan 2013; Cobb-Clark and Moschion 2017; Schleicher 2019). In this study, we begin by asking to what extent these important determinants can alter the role of gender in male and female sorting into different jobs *within cohorts*. Overall, we find that childhood factors, as measured by the variables that we capture in our data, do not noticeably change gendered sorting patterns, which is proxied by the coefficient of the female dummy, as we had anticipated. In contrast, by comparing the size of the coefficient in the same specification across birth cohorts, we observe that the female coefficient declines over the decades. We view this as highly suggestive evidence that it is temporal changes that matter the most in shaping these patterns.

We are aware that our childhood variables do not adequately capture every childhood factor that could potentially mediate gendered sorting. However, we would expect that some of the relevant omitted factors would be correlated with the variables that we do include so their signal should be picked up in our regressions. We acknowledge, however, that we do not capture a universe of childhood variables to allow us rule out every childhood factor that can determine occupational sorting. We also note that we cannot pinpoint what exact temporal shift shaped sorting and would speculate that it is a multitude of structural shifts, such as, the changing nature of tasks in many occupations, lower tolerance levels towards discrimination in the workplace, policy change, differential growth patterns across industries and the ability for women to delay their fertility.

Our analysis has also revealed several interesting stylised facts regarding gendered sorting over time. First, for all three cohorts, we find strong evidence of sorting along gendered lines, but this tendency has decreased substantively over time. That is, the gender gap has narrowed. Our analysis also reveals persistent gender gaps in the tendency to sort into occupations with the highest shares of males that have not changed over time. These jobs are often the golden pathway to C Suite positions and positions of power, and encapsulate science, technology and engineering posts as well as front office trading roles and politics. It may be tempting to conclude that the flatness in the gender gap in the tendency to sort into occupations with the highest share of males, particularly for children with the highest academic ability, reflects innate preferences. However, we note that over time both genders have significantly changed their tendency to sort into occupations that are high on people, brains and competitiveness content.<sup>31</sup> Some of this will be determined by labour markets (*i.e.* it is unsurprising that both genders sort towards jobs that are high in people, given the growth in services and jobs that require interpersonal skills). Yet, we also view these changes as highly suggestive that preferences are socialised, rather than representing innate differences by gender.

While all eyes are normally on the tendency for females to change preferences, our analysis reveals that the preferences of males are contributing to stubborn gender gaps in traditionally male-dominated positions. Noteworthy, is that males in the most recent cohort are aspiring to work in occupations with significantly higher levels of competitiveness and larger incomes as compared to previous cohorts and their current female peers. Therefore, even though the females born in 2000 have nearly closed the gender gap in terms of hours and flexibility they are demanding, the type of work they are aspiring to sort into suggests that the gender pay gap may prevail unless the rewards given to different occupations change, or indeed preferences change for even younger cohorts.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00181-022-02314-5>.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line

<sup>31</sup> Females have also changed their tendency to sort into jobs with high brawn content.

to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akee R, Simeonova E, Copeland W, Angold A, Costello J (2013) Young adult obesity and household income: effects of unconditional cash transfers. *Am Econ J Appl Econ* 5(2):1–28
- Akerlof G, Kranton RE (2000) Economics and identity. *Q J Econ* 115(3):715–753
- Alesina A, Giuliano P, Nunn N (2013) On the origins of gender roles: women and the plough. *Q J Econ* 128(2):469–530
- Altonji JG, Blank RM (1999) Race and gender in the labor market. *Essay. Handb Labor Econ* 3C:3143–3259
- Angrist J, Lavy V (1999) Using maimonides rule to estimate the effect of class size on student achievement. *Q J Econ* 114(2):533–575
- Antecol H (2010) The opt-out revolution: a descriptive analysis. IZA Discussion Paper, No 5089
- Arcidiacono P, Hotz JV, Kang S (2012) Modeling college major choices using elicited measures of expectations and counterfactuals. *J Econom* 166(1):3–16
- Arcidiacono P, Aucejo EM, Hotz JV (2016) University differences in the graduation of minorities in STEM fields: evidence from California. *Am Econ Rev* 106(3):525–562
- Autor DH, Dorn D (2013) The growth of low-skill service jobs and the polarization of the US labor market. *Am Econ Rev* 103(5):1553–1597
- Autor D, Dorn D, Hanson G (2015) Untangling trade and technology: evidence from local labor markets. *Econ J* 125(584):621–646
- Aznar A, Tenenbaum HR (2015) Gender and age differences in parent–child emotion talk. *Br J Dev Psychol* 33:148–155
- Baker M, Milligan K (2016) Boy-girl differences in parental time investments: evidence from three countries. *J Hum Cap* 10(4):399–441
- Bayard K, Hellerstein J, Neumark D, Troske K (2003) New evidence on sex segregation and sex differences in wages from matched employee–employer data. *J Law Econ* 21(4):887–922
- Becker GS (1957) *The economics of discrimination*, 2nd edn. University of Chicago Press, Chicago
- Becker GS (1985) Human capital, effort, and the sexual division of labor. *J Law Econ* 3(1):S33–S80
- Becker GS, Tomes N (1986) Human capital and the rise and fall of families. *J Law Econ* 4(3):1–39
- Belkin L (2003) The opt-out revolution. *The New York Times Magazine*, pp 43–86
- Berger J, Zelditch M, Anderson B (1966) *Sociological theories in progress*. Houghton Mifflin, Boston, MA
- Bertrand M (2018) Coase lecture—the glass ceiling. *Economica* 85(338):205–231
- Bertrand M, Pan J (2013) The trouble with boys: social influences and the gender gap in disruptive behavior. *Am Econ J Appl Econ* 5(1):32–64
- Biasi B, Sarsons H (2021) Information, confidence, and the gender gap in bargaining. *AEA Pap Proc* 111:174–178
- Bielby WT, Baron JN (1984) A woman's place is with other women: sex segregation within organizations. *Sex Seg Workpl Trends Expl Rem* 145
- Black SE, Spitz-Oener A (2010) Explaining women's success: technological change and the skill content of women's work. *Rev Econ Stat* 92(1):187–194
- Black S, Devereux PJ, Salvanes KG (2007) From the cradle to the labor market? The effect of birth weight on adult outcomes. *Q J Econ* 122(1):409–439
- Blau FD (1977) Sex, discrimination, and the division of labor. *Indust Labor Relat Rev* 30(2):254
- Blau D (1999) The effect of income on child development. *Rev Econ Stat* 81(2):261–276
- Blau F, Kahn LM (1997) Swimming upstream: trends in the gender wage differential in the 1980s. *J Law Econ* 15:1–42
- Blau F, Kahn LM (2016) The gender wage gap: extent, trends, and explanations. NBER Working Paper, No 21913
- Boushey H (2005) Are women opting out? Debunking the myth. Center for Economic and Policy Research, Briefing Paper

- Bowles HR, Babcock L, Lai L (2007) Social Incentives for gender differences in the propensity to initiate negotiations: sometimes it does hurt to ask. *Organ Behav Hum Decis Process* 103(1):84–103
- Buchmann C, DiPrete TA (2008) Gender inequalities in education. *Ann Rev Sociol* 34:319–337
- Carrington WJ, Troske KR (1998) Sex segregation in US manufacturing *ILR Rev* 51(3):445–464
- Cobb-Clark D, Moschion J (2017) Gender gaps in early educational achievement. *J Popul Econ* 30(4):1093–1134
- Cobb-Clark DA, Tan M (2011) Noncognitive skills, occupational attainment, and relative wages. *Labour Econ* 18(1):1–13
- Cohany SR, Sok E (2007) Trends in labor force participation of married mothers of infants. *Monthly Lab Rev* 130:9
- Cortés P, Pan J (2016) Prevalence of Long hours and skilled women's occupational choices. IZA Discussion Paper, No 10225
- Cortés P, Pan J (2017) Occupation and gender. IZA Discussion Paper, No 10672
- Crosnon R, Gneezy U (2009) Gender differences in preferences. *J Econ Lit* 47(2):448–474
- Dahl G, Moretti E (2008) The demand for sons. *Rev Econ Stud* 75(4):1085–1120
- Datcher L (1982) Effects of community and family background on achievement. *Rev Econ Stat* 64:32–41
- Daymont T, Andrisani PJ (1984) Job preferences, college major, and the gender gap in earnings. *J Hum Resour* 19:408–428
- Dee TS (2007) Teachers and the gender gaps in student achievement. *J Hum Resour* 42(3):528–554
- Dohmen T, Falk A (2011) Performance pay and multidimensional sorting: productivity, preferences, and gender. *Am Econ Rev* 101(2):556–590
- Dreber A, Heikensten E, Säve-Söderbergh J (2022) Why do women ask for less? *Labour Econ* 78:102204
- Filer RK (1986) The role of personality and tastes in determining occupational structure. *ILR Rev* 39(3):412–424
- Fortin NM (2008) The gender wage gap among young adults in the united states the importance of money versus people. *J Human Res* 43(4):884–918
- Frankenburg WK, Dodds JB (1967) The Denver developmental screening test. *J Pediatr* 71(2):181–191
- Genicot G, Ray D (2017) Aspirations and inequality. *Econometrica* 85(2):489–519
- Gibbons S, Chevalier A (2008) Assessment and age 16+ education participation. *Res Pap Educ* 23(2):113–123
- Gneezy U, Rustichini A (2004) Gender and competition at a young age. *Am Econ Rev* 94(2):377–381
- Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: gender differences. *Q J Econ* 118(3):1049–1074
- Goldin C (2006) The quiet revolution that transformed women's employment, education, and family. *Am Econ Rev* 96(2):1–21
- Goldin C (2014) A grand gender convergence: its last chapter. *Am Econ Rev* 104(4):1091–1119
- Goldin C, Katz LF (2002) The power of the pill: oral contraceptives and women's career and marriage decisions. *J Polit Econ* 110(4):730–770
- Goodman SH, Rouse MH, Connell AM, Broth MR, Hall CM, Heyward D (2011) Maternal depression and child psychopathology: A meta-analytic review. *Clin Child Family Psychol Rev* 14(1):1–27
- Gorsuch RL (1983) Factor analysis, 2nd edn. Lawrence Erlbaum Associates, Hillsdale, NJ
- Grove WA, Hussey A, Jetter M (2011) The gender pay gap beyond human capital heterogeneity in noncognitive skills and in labor market tastes. *J Hum Resour* 46(4):827–874
- Heckman J, Pinto R, Savelyev P (2013) Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *Am Econ Rev* 103(6):2052–2086
- Herr L, Wolfram CD (2012) Work environment and OPT-out rates at motherhood across high-education career paths. *ILR Rev* 65(4):928–950
- Hyde J (2014) Gender similarities and differences. *Annu Rev Psychol* 65(3):373–398
- Johnson J, Atkinson M, Rosenberg R (2015) Millennium Cohort Study: psychological, developmental and health inventories. Centre for Longitudinal Studies, London
- Katz L, Murphy KM (1992) Changes in relative wages, 1963–1987: supply and demand factors. *Q J Econ* 107(1):35–78
- Kohler H-P, Behrman JR, Skytthe A (2005) Partner + children = happiness? An assessment of the effect of fertility and partnerships on subjective well-being in Danish twins. *Popul Dev Rev* 31(3):407–445
- Kray LJ, Gelfand MJ (2009) Relief versus regret: the effect of gender and negotiating norm ambiguity on reactions to having one's first offer accepted. *Soc Cogn* 27(3):418–436
- La Ferrara E (2019) Aspirations, social norms, and development. *J Eur Econ Assoc* 17(6):1687–1722



- Lavy V, Sand E (2015) On the origins of gender human capital gaps: short and long term consequences of teachers' stereotypical biases. NBER Working Paper, No w20909
- Lekfuangfu W, Odermatt R (2022) All I have to do is dream? The role of aspirations in intergenerational mobility and well-being. *Eur Econ Rev* 148:104193
- Lordan G (2019) People versus machines in the UK: minimum wages, labor reallocation and automatable jobs. *PLoS ONE* 14(12):e0224789
- Lordan G, Neumark D (2018) People versus machines: the impact of minimum wages on automatable jobs. *Labor Econ* 52:40–53
- Lordan G, Pischke J-S (2022) Does Rosie like riveting? Male and female occupational choices. *Economica* 89(353):110–130
- Ludsteck J (2014) The impact of segregation and sorting on the gender wage gap: Evidence from German linked longitudinal employer-employee data. *ILR Rev* 67(2):362–394
- Lundberg S, Pablonia S (2007) Time allocation of parents and investments in sons and daughters. Department of Economics, University of Washington Working Paper
- Lundberg S (2005) The division of labor by new parents: does child gender matter? IZA Discussion Paper, No 1787
- Luo Ye, Waite LJ (2005) The impact of childhood and adult SES on physical, mental, and cognitive well-being in later life. *J Gerontol B Psychol Sci Soc Sci* 60(2):93–101
- Macpherson DA, Hirsch BT (1995) Wages and gender composition: why do women's jobs pay less? *J Labor Econ* 13(3):426–471
- Mondschein E, Adolph KE, Tamis-LeMonda CS (2000) Gender bias in mothers' expectations about infant crawling. *J Exp Child Psychol* 77:304–316
- Nielderle M, Vesterlund L (2008) Gender differences in competition. *Negot J* 24(4):447–463
- Olivetti C, Petrongolo B (2016) The Evolution of gender gaps in industrialized countries. *Ann Rev Econ* 8(1):405–434
- Ost B (2010) The role of peers and grades in determining major persistence in the sciences. *Econ Educ Rev* 29(6):923–934
- Oxoby RJ (2002) Status characteristics, cognitive bias, and incentives in teams. *J Socio-Econ* 31(3):301–316
- Parsons S (2014) Childhood cognition in the 1970 British cohort study. CLS Data Note. University of London, London
- Percheski C (2008) Opting out? Cohort differences in professional women's employment rates from 1960 to 2005. *Amer Soc Rev* 73(3):497–517
- Pinker S (2008) *The Sexual Paradox: Men, Women and the Real Gender Gap*. Simon and Schuster
- Polachek S (1981) Occupational self-selection: a human capital approach to sex differences in occupational structure. *Rev Econ Stat* 63(1):60–69
- Rutter M (1967) A children's behaviour questionnaire for completion by teachers: preliminary findings. *J Child Psychol Psychiatry* 8(1):1–11. <https://doi.org/10.1111/j.1469-7610.1967.tb02175.x>
- Schleicher A (2019) PISA 2018: insights and interpretations. OECD Publishing, Berlin
- Shepherd P (2012) Measures of ability at ages 7 to 16. National Child Development Study User Guide. University of London, London
- Sigurdsson E, Jim Van O, Eric F (2002) Are impaired childhood motor skills a risk factor for adolescent anxiety? Results from the 1958 UK birth cohort and the National Child Development Study. *Amer J Psych* 159(6):1044–1046
- Speer JD (2017) Pre-market skills, occupational choice, and career progression. *J Hum Resour* 52(1):187–246
- Stone P (2007) *Opting out? Why women really quit careers and head home*, 1st edn. University of California Press, Berkeley, CA
- Su R, Rounds J, Armstrong PI (2009) Men and things, women and people: a meta-analysis of sex differences in interests. *Psychol Bull* 135(6):859–884
- Thompson B (2004) *Exploratory and confirmatory factor analysis: understanding concepts and applications*. American Psychological Association, Washington, DC
- Todd PE, Zhang W (2020) A dynamic model of personality, schooling, and occupational choice. *Quant Econ* 11(1):231–275
- Van Den Berg G, Lindeboom M, Portrait F (2006) Economic conditions early in life and individual mortality. *Am Econ Rev* 96:290–302
- Vere, JP (2007) "Having it all" no longer: Fertility, female labor supply, and the new life choices of generation X. *Demography* 44(4):821–828

- Xie Y, Kimberlee AS (2004) Women in science: career processes and outcomes. *Soc For* 82(4):1669–1671
- Yamaguchi S (2018) Changes in returns to task-specific skills and gender wage gap. *J Hum Resour* 53(1):32–70
- Yeung J, Sandberg JF, Davis-Kean PE, Hofferth SL (2004) Children's time with fathers in intact families. *J Marriage Fam* 63:136–154
- Zafar B (2011) How do college students form expectations? *J Law Econ* 29(2):301–348
- Zafar B (2013) College major choice and the gender gap. *J Hum Resour* 48(3):545–595

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.