I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13380

Things versus People: Gender Differences in Vocational Interests and in Occupational Preferences

Andreas Kuhn
Stefan C. Wolter

JUNE 2020

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 13380

Things versus People: Gender Differences in Vocational Interests and in Occupational Preferences

Andreas Kuhn
SFIVET, University of Bern and IZA

Stefan C. Wolter
University of Bern, SKBF-CSRE, CESifo and IZA

JUNE 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Things versus People: Gender Differences in Vocational Interests and in Occupational Preferences*

Occupational choices remain strongly segregated by gender, for reasons not yet fully understood. In this paper, we use detailed information on the cognitive requirements in 130 distinct learnable occupations in the Swiss apprenticeship system to describe the broad job content in these occupations along the things-versus-people dimension. We first show that our occupational classification along this dimension closely aligns with actual job tasks, taken from an independent data source on employers' job advertisements. We then document that female apprentices tend to choose occupations that are oriented towards working with people, while male apprentices tend to favor occupations that involve working with things. In fact, our analysis suggests that this variable is by any statistical measure among the most important proximate predictors of occupational gender segregation. In a further step, we replicate this finding using individual-level data on both occupational aspirations and actual occupational choices for a sample of adolescents at the start of 8th grade and the end of 9th grade, respectively. Using these additional data, we finally also show that the gender difference in occupational preferences is largely independent of individual, parental, and regional controls.

JEL Classification: J16, J24, D91

Keywords: occupational choice, occupational segregation, things versus people, preferences, gender differences, job content

Corresponding author:
Andreas Kuhn
Swiss Federal Institute for Vocational Education and Training
Kirchlindachstrasse 79
3052 Zollikofen
Switzerland

E-mail: andreas.kuhn@ehb.swiss

* We thank Manuel Aepli, Anne Brenøe and Jürg Schweri for comments on an early draft of the paper. We also thank Walter Götze for providing us with the raw data describing the cognitive requirements of the various occupations, Christoph Düby from the cantonal department of education of the canton of Bern for providing us with the data on the population of apprenticeship contracts in the canton of Bern, Miriam Grønning for helping us with the occupational crosswalks used to merge the different data sets, and Sally Gschwend-Fisher for proofreading the manuscript.

1 Introduction

Many countries have seen profound and often surprisingly rapid changes in various measures of women's participation and performance in both the educational system and in the labor market, paralleled by corresponding changes in individuals' attitudes towards the equality between women and men. Nonetheless, there remains a persistently high degree of occupational gender segregation, even in the most progressive countries in this regard (see, for example, Charles and Grusky, 2004) – and thus often despite explicit and considerable public effort to decrease the extent of occupational segregation.[1] These remaining differences in occupational choice show up in women's underrepresentation in STEM jobs (e.g. Kahn and Ginther, 2017), and they are related to the remaining gender gap in wages (e.g. Blau and Kahn, 2017; Olivetti and Petrongolo, 2016). Therefore, understanding the underlying causes of the persistence in occupational gender segregation is of obvious academic and public interest (Bertrand, 2011; Cortes and Pan, 2018).

One rather obvious potential explanation starts from the observation that men and women differ in preferences that may influence their occupational choices. Indeed, a voluminous and growing number of empirical studies documents gender differences in various economic preferences (Bertrand, 2011; Croson and Gneezy, 2009) and across many countries (Falk *et al.*, 2018). More specifically, empirical research has documented gender differences in time preferences (e.g. Dittrich and Leipold, 2014), risk preferences (e.g. Charness and Gneezy, 2012), social preferences (e.g. Kamas and Preston, 2015), and competitiveness (e.g. Niederle and Vesterlund, 2011), to name but the most prominent examples. Moreover, a few studies have explicitly studied whether these gender differences in preferences are related to differences in occupational choices. For example, Bonin *et al.* (2007) showed that more risk-tolerant individuals select into occupations with a higher earnings risk. In a closely related study, Fouarge *et al.* (2014) showed that both risk preferences and time preferences are related to occupational choices, with more patient individuals choosing occupations with a steeper earnings profile later on. In

[1] In fact, several cross-country studies have consistently documented that gender differences in economic preferences, as well as in personality traits, tend to increase, rather than decrease, in countries with more progressive views towards equality between men and women (Falk and Hermle, 2018; Schmitt *et al.*, 2008). The same holds true for occupational gender segregation (e.g. Charles, 2017; Charles and Bradley, 2009). In contrast, gender differences in academic achievement do not appear to be consistently related to measures of gender equality (e.g. Stoet and Geary, 2015).

the context of occupational choice, however, competitiveness has presumably received the most attention from economists. In one prominent study on the subject, Buser *et al.* (2014) found that the gender difference in competitiveness can explain some of the gender gap in choice of study subject in the Netherlands. Other studies have found similar results, such as Kleinjans (2009) for Denmark or Buser *et al.* (2017a), who estimate the effect of competitiveness on study choices among Swiss college students (see also Buser *et al.*, 2017b, who extend the analysis to non-college students). Moreover, empirical studies also found that both gender differences in economic preferences (see, for example, Gneezy and Rustichini, 2004, on competitiveness) as well as in occupational preferences (Kooreman, 2009) appear early in life, further suggesting some connection between gender differences in economic preferences and in occupational preferences and choices, respectively.

One closely related, and in some way even more obvious hypothesis is that men and women may differ in their occupational choices simply because they have different vocational interests, i.e. that they differ in their preferences over the general task content within a given occupation. Indeed, there is a considerable amount of empirical evidence, mainly from psychologists, consistently documenting large and persistent gender differences in vocational interests, especially along the things-versus-people dimension, an idea that goes back at least to the work of Holland (1959).[2] This concept classifies the task content of occupations according to the extent to which people working in a given occupation deal with inanimate things or with other people, respectively. Empirically, it appears that this dimension can best discriminate between tasks men and women prefer, respectively (e.g. Lippa, 1998). Interestingly, this specific dimension of vocational interests may also explain a substantive part of the gender segregation within relatively narrow groups of occupations, such as within STEM occupations (e.g. Cheryan *et al.*, 2017; Su and Rounds, 2015). Somewhat surprisingly, however, economists have not yet given this hypothesis much, if any, attention.

A few recent papers by economists have focused on closely related concepts, however. For example, using a hypothetical choice experiment, Wiswall and Zafar (2018) show that women

[2]It turns out that observed gender differences in vocational interests along the things-people dimension are among the largest differences between men and women measured by any psychometric instrument. For example, in their meta-study on the subject, Su *et al.* (2009) document an average effect size of 0.93. Similarly, Lippa (2010) reports an average absolute effect size of 1.18. As we will show in section 4.3 below, we find a gender difference in vocational interests along the things-versus-people dimension of a comparable size.

have stronger preferences for workplace hours flexibility and job security than men, and that men hold stronger preferences for higher earnings growth than women. Similarly, Lordan and Pischke (2016) take up the argument in their study that men and women have different tastes regarding the task content of work by analyzing data on job satisfaction in different countries. They find that female workers' job satisfaction is lower when the share of male workers is higher in an occupation, and that part of this effect is picked up by variables describing the broad task content of jobs. In another study closely related to ours, Baker and Cornelson (2018) focus on gender differences in motor skills and visuo-spatial aptitudes and how they relate to the proportion of males and females in a given occupation, respectively. They find that, overall, the shares of males and females, respectively, in an occupation strongly aligns with gender differences in these skills, and that gender differences in these skills can account for a large share of the observed occupational segregation. Another closely related study is by Gelblum (2020), who estimates individuals' willingness to pay for different job tasks also using a hypothetical choice experiment. She finds significant gender differences in the willingness to pay for some, but not all, gender-typical tasks (e.g. women have a higher willingness to pay for "helping and caring for others" than men). Moreover, she also finds that these differences can account for a sizeable proportion of the observed occupational segregation. Finally, another closely related study shows that men and women differ in their preferences for meaning at work, especially meaning derived from job mission (Burbano *et al.*, 2020).

In this paper, we add further empirical evidence to this important and fascinating debate, using a unique combination of different data sources to describe gender differences in occupational preferences among Swiss adolescents. In a first step, however, we show that occupational aspirations are indeed highly segregated, consistent with analogous findings for the country as a whole (e.g. Aepli *et al.*, 2019). We then show that the proportion of men and women, respectively, in an occupation is strongly correlated with its position on the things-versus-people dimension – in fact, we find that the broad task content of an occupation is an extremely powerful predictor of whether males or females predominantly choose the occupation. In the second part of our empirical analysis, we replicate this finding using individual-level data for a sample of adolescents from the canton of Bern. The majority of these individuals were surveyed twice, at the start of 8th grade as well as at the end of 9th grade. As we will explain below, this implies that the survey contains information on participants' occupational aspirations at the time they started the vocational selection process as well as about their actual occupational choices later on. This allows us not only to see whether we can replicate the analysis based on the occupational level data, but also to show that the association between gender and the task content of an occupation shows up already in adolescents' occupational aspirations. Moreover, using these additional data, we are also able to see whether the gender difference in occupational preferences is robust to the inclusion of additional controls at the individual, parental and regional level.

The remainder of this paper proceeds as follows. In the following section, we shortly describe some of the key features of the Swiss educational system, mainly focusing on the apprenticeship system. Section 3 discusses the different data sources used in the empirical analysis. Our empirical analysis is presented and discussed in several consecutive steps in section 4. In the first part of the analysis (sections 4.1 and 4.2), we analyze data at the occupational level. In the second part of the analysis (section 4.3), we use individual-level data to replicate and expand on the main finding from the analysis at the occupational level. Section 5 concludes.

2 The Swiss educational system

Compulsory schooling in Switzerland lasts eleven years, of which two years are spent in kindergarten, six in primary school, and three in secondary school (see SCCRE, 2014, for a much more detailed description of the Swiss educational system). Children usually enter primary school in the year they turn seven years old, and thus most of them are round 15 years old when they finish the mandatory part of education. Consequently, they usually enter post-mandatory, upper-secondary schooling/training in the year they reach the age of 16.

At the upper-secondary level, there are two main options (see appendix figure A.1 for a schematic illustration). First, there is the possibility to enter further general education via a baccalaureate school ("Gymnasium" in German, about equivalent to high school), which prepares for and gives access to further studies at the tertiary (usually university) level. The other, far more popular route at this stage is to enter the apprenticeship system, which also gives opportunities to enter further education and training at the tertiary level later on.

2.1 The apprenticeship system

At the national level, a large majority of about 62% of adolescents eventually enters the apprenticeship system, typically through a dual-track apprenticeship which combines practical training within a firm with vocational school (e.g. SERI, 2017).[3] In the canton of Bern, from where our sample of adolescents is drawn (see section 3 below for details), the proportion of adolescents entering the apprenticeship system is actually higher than the national average, with almost 70% of recent cohorts entering the apprenticeship system.

Most apprenticeships last three or four years, and one day per week (in some cases two days per week) are spent in vocational school (SERI, 2017). There are also two-year apprenticeships for adolescents with with lower academic standards. Overall, there are about 240 different learnable occupations within the Swiss apprenticeship system. It is also worth noting that the apprenticeship system is regulated at the national level, in contrast to the rest of the educational system (where both the cantons and the municipalities play the lead role; that is also one of the main reasons why there are large regional differences in fraction of adolescents entering vocational training versus further general education at the upper-secondary level).

Moreover, there is basically a market for apprenticeship positions, with adolescents looking for apprenticeship positions and with employers simultaneously advertising vacant positions (furthermore, apprentice wages may differ across employers for the same occupation). In case of a mutual match, the host company and the apprentice both sign an apprenticeship contract, which lasts until the completion of the apprenticeship. This implies that an apprenticeship does not automatically lead to an employment contract with the training firm, even though many training firms retain their apprentices after they have successfully finished their apprenticeships. During their training period, apprentices are paid an apprentice wage, which is substantively lower than that of a fully trained worker in the same occupation.

3 Data

We next describe the different data sources used in our empirical analysis.

[3] See Wettstein *et al.* (2017) for a detailed discussion of the Swiss apprenticeship system.

3.1 Describing the task content of occupation

We use two different and independent data sources to describe the task content of occupations.

Cognitive requirements by occupation

First, we use data on the cognitive requirements in the different learnable occupations within the Swiss apprenticeship system. These data come from a project initiated, and partially funded, by the Swiss State Secretariat for Education, Research and Innovation (SERI), and with the primary aim of providing adolescents and their parents, as well as people working within the VET system (such as teachers in vocational schools or vocational advisers) with comparable information on the cognitive requirements in the different learnable occupations (labelled *"Anforderungsprofile"* in German).[4] For this purpose, the data contain quite detailed requirements in native and foreign languages, mathematics, and natural sciences for the various learnable occupations within the Swiss apprenticeship system. Within each of the four subjects, there is a more detailed breakdown by subtopics.[5] However, we will almost exclusively focus on the aggregated scores by main topic in our own analysis (we do use some of the subcategories for a validation exercise; see section 4.1 below). This is our main data source for describing the task content of occupations, as explained in more detail in section 4.1 below.

Task descriptions from actual job postings

We complement this information with data from actual job advertisements, which are taken from an additional and independent source of data collected for the main purpose of monitoring the demand side of the Swiss job market over time (the Swiss Job Market Monitor, *"Stellenmarktmonitor"*).[6] These data contain samples of actual job advertisements by both private and public employers from the years 1950 until 2018 (currently, the data collection is still going) and sampled from newspapers, company websites, as well as online job portals.

[4]More information about these data is available online (however, only in German, French, and Italian) at: www.anforderungsprofile.ch. Appendix figure A.2 shows an example comparison between two occupations (healthcare assistant versus mechanical engineer) as available directly from the website.

[5]For example, in mathematics, there are the following five subtopics: algebra ("Zahl und Variable"), geometry ("Form und Raum"), units of measurement ("Grössen und Masse"), calculus ("Funktionale Zusammenhänge") and statistics ("Daten und Zufall'). See again appendix figure A.2 for a concrete example.

[6]Additional details are available online from the project website: www.stellenmarktmonitor.uzh.ch.

For our purpose, we focus on the more recent advertisements from the years 2010 to 2015, with a total of 24,368 individual job postings from all over Switzerland. Among other things, the data record the main activity of each job posting.[7] We use this variable to validate our approach of categorizing the different occupations along the interest in things versus people dimension based on the data describing the cognitive requirements (see section 4.1 below for details).

3.2 Describing the extent of occupational gender segregation

Moreover, we have also access to a dataset that contains the full population of individual-level apprenticeship contracts involving either apprentices and/or employers from the canton of Bern as of August 2015, with more than 45,000 apprenticeship contracts covered. We use these data to compute the proportion of females by occupation, which in turn allows us to describe the extent of occupational gender segregation. We can do this with reasonable precision for most occupations because the data cover so many individual-level contracts. Nonetheless, some occupations are still only rarely chosen, and we exclude these occupations from most of the analysis, as discussed in more detail below.[8]

3.3 Individual-level survey among adolescents

Our final data source is a computer-assisted individual-level survey among 1,514 adolescents at the start of 8th grade (with an average age of about 14 years at the time the main survey was administered) from the German language area of the canton of Bern (see Buser *et al.*, 2017a, for additional details). The adolescents are from 28 different schools spread across 24 different municipalities in the German language area of the canton of Bern.[9] The main

[7]The variable containing the main activity has 21 different values designating broad groups of tasks, such as "planning, engineering, designing/drawing" or "educating/teaching, advising", and it is available from 1995 onwards only (see table 2 below for the full list of activities coded). Moreover, the data also contain several occupational codes, allowing us to merge the two datasets.

[8]It is worth noting that the distribution of individuals across occupations is highly skewed, with a large proportion of all apprenticeship contracts concentrated in relatively few occupations only. This is clearly evident from appendix table A.1, which shows the ten most popular occupations among girls and boys, respectively. Among boys (girls), the ten most popular occupations account for 42.3% (65.3%) of all apprenticeship contracts.

[9]These data have been used before to study whether competitiveness has an influence on adolescents' occupational aspirations (Buser *et al.*, 2017a,b). Jaik and Wolter (2019) look at the correlation between occupational aspirations and occupational choices. Finally, Kuhn and Wolter (2019) use the same data to study the impact of societal gender norms on gender-stereotypical occupational aspirations.

survey was administered in the summer of 2015, and a majority of adolescents (about 96%) was successfully contacted a second time in the spring of 2017, at the end of 9th grade.

Occupational aspirations and occupational choices among adolescents

In the first and main round of the survey, adolescents were asked about their occupational aspirations, i.e. adolescents were simply asked in which occupation they would like to work ("What apprenticeship would you most like to complete?"). This information is in raw-text form, but we assigned an occupational code to each occupation that is learnable through an apprenticeship occupation. The adolescents were further asked about their actual occupational choices in a second, later round of the survey. At this stage, most adolescents tend to have chosen their occupation and already have signed an apprenticeship contract with their prospective employer and instructor (or have decided instead that they want to go on with further general education, in which case there is no occupational choice to be made yet).

In the empirical analysis reported below, we will report results for both actual choices and aspirations. The data on individual-level occupational choices obviously allow us to replicate the analysis based on the occupational-level data. At the same time, occupational aspirations are also of interest because they are arguably not (or less) influenced by external restrictions on occupational choice, such as by the availability of apprenticeship positions in the desired occupation, or by their prospective employers and/or their parents. On the other hand, however, it is often argued that people are more prone to be influenced by societal norms when they are younger.[10] This allows us to show that the results using actual choices, both at the individual and occupational levels, are not simply driven by such external forces.

4 Empirical analysis

The primary aim of our analysis is to see whether there are gender differences in occupational preferences that align with differences in the task content of the occupations. Obviously, how-

[10]See also Jaik and Wolter (2019), who study how occupational aspirations in 8th grade deviate from actual occupational choices later on using the same data. Kuhn and Wolter (2019) find a null effect of societal gender norms on both occupational aspirations and choices. Together, the two studies suggest that adolescents' aspirations may often differ from their choices, without influencing the gender-stereotypicity of their preferences, however. Our results are also in line with this conclusion.

ever, we need to classify the different occupations according to their task content, along the things-versus-people dimension, in a first step.

4.1 Describing the task content of occupations

We therefore first describe how the learnable occupations differ in their task content.

The task content of occupations: things versus people

We first run a principal-components analysis to describe the task content of occupations (e.g. Gorsuch, 2014; James *et al.*, 2013). More specifically, we use the four variables describing the cognitive requirements in mathematics, natural sciences, as well as native and foreign languages as input variables into the analysis. The results show that the first three principal components can reproduce almost 98% of the overall variation in the four original variables (cf. appendix table A.2). Also note that, by construction, the resulting principal components are uncorrelated with each other.

Not surprisingly, the first principal component (PC) loads on all four input variables, which suggests that this PC can be interpreted as the overall level of cognitive demands in a given occupation, and we will thus use the shorthand C_j^{demand} for this variable subsequently. Indeed, comparing the occupations with the highest and lowest values on this PC supports this interpretation (cf. appendix table A.3). The ten occupations with the lowest overall cognitive requirements, for example, are exclusively apprenticeships that only last for two years (i.e. these are occupations which were intentionally set up for academically underachieving youths), such as timer worker or tire work assistant. At the upper end of the distribution of this variable, on the other hand, we find occupations such as optometrist or geomatics expert.

In what follows, however, we will mainly focus on the second and the third PCs, which together describe the task content of the occupations along the interest in things versus interest in people dimension, as we will now argue. The second PC loads positively on mathematics and natural sciences as well as negatively on both native and foreign languages, while the third PC loads positively on mathematics and foreign languages, but negatively on natural sciences and native languages. Thus the second PC classifies occupations according to whether they are oriented towards the importance of mathematics and sciences rather than towards the usage of languages, and we use the shorthand C_j^{math} for this PC in what follows. The third PC is more ambiguous than the previous two, but we think that one may interpret this PC as indicating whether an occupation deals with things or people in an either more abstract or a more direct and practical way, and we thus denote this PC as C_j^{abstract} subsequently.[11] Occupations oriented towards things are presumably tilted towards both mathematics/sciences and abstract content, and we thus combine the second and the third PCs into a common variable, simply by summing them up.[12] In what follows, we use the shorthand C_j^{things} (i.e. $C_j^{\text{things}} = C_j^{\text{math}} + C_j^{\text{abstract}}$) for this variable, with larger values indicating that the broad task content in an occupation is tilted towards working with inanimate things, rather than with people.

Table 1

A look at some specific occupations immediately shows that this classification is both plausible and meaningful. Table 1 therefore lists the ten occupations with the largest and with the smallest values on C_j^{things}, respectively. Not surprisingly, technical occupations such as mechanical engineer or carpenter are found on the things-intense tail of the distribution of this variable, while occupations such as healthcare assistant or hairdresser are found on the opposite end of the distribution.[13]

Validating our classification of the task content of occupations

In a next step, we use the data on employers' actual job postings to validate our classification of the occupations along the main dimension of interest (i.e. things-versus-people) because one might raise the objection that the data on the cognitive requirements may not have much to do with what people working in these occupations actually do. Using the additional data

[11]The fourth principal component has no obvious interpretation and is not used in the empirical analysis at all (also note that this PC explains only a negligible fraction of the variation in the input variables; cf. appendix table A.2).

[12]Remember that the two PCs are independent of each other at the occupational level by construction. This implies that C_j^{things} still has a mean of zero, and a variance that equals the sum of the variances of the two variables that are added up.

[13]See also appendix table A.4, which shows the breakdown between C_j^{math} and C_j^{abstract} for the occupations from table 1. The table suggests that, while all of the most math-intensive occupations are also high on abstraction, there is notable variation on C_j^{abstract} among the most language-intensive occupations. That is, among the language-intensive occupations, there are occupations that score either low (e.g. healthcare assistant) or high (e.g. customer dialogue specialist) in terms of abstract job content. Quite obviously, this mainly reflects the fact that these occupations differ in the extent to which they have direct (e.g. physical contact with patients) or only indirect (e.g. contact with customers by phone or email) contact with other people.

from employers' job postings allows us to check empirically whether our classification of the occupations based on the cognitive requirements alone roughly corresponds to what employers really expect people actually working in these occupations to do.

Table 2

Table 2 describes the full correlational pattern by regressing our measure of the task-intensity along the things-people dimension C_j^{things} on a set of 21 variables, each essentially measuring the relative frequency with which a given main activity was explicitly mentioned in a real job posting (i.e. each of these variables may take on values within the unit interval). The first column of table 2 shows unweighted estimates, while the second column weights the occupations by the number of apprenticeship contracts. Both columns show that the value on the things-versus-people variable of an occupation closely aligns with the actual activity pattern in an occupation as made salient through job postings by employers looking for individuals to work in these occupations. Correspondingly, the resulting R-squared is very high, 0.673 in the unweighted and 0.877 in the weighted case. Given that we use two fully independent sources of data, we view these results as a strong confirmation that our occupational classification based on the cognitive requirements is meaningful.[14]

4.2 Analysis of the occupational-level data

In the second part of our analysis, we show that men and women chose occupations that differ strongly in their broad task content along the things-versus-people dimension.

Describing occupational gender segregation

In a preliminary step, however, we first show that there indeed is strong occupational gender segregation within the Swiss apprenticeship system (cf. Aepli *et al.*, 2019; Kuhn and Wolter, 2019). As we mentioned above, access to data on the population of apprenticeship contracts in

[14] An additional validation exercise is shown in appendix table A.5. The table shows that the five subcategories for the cognitive requirements in native language, such as "reading" or "writing", are all negatively associated with C_j^{things}, but to very different degrees. For example, "reading" is not significantly associated with C_j^{things}, while both "participation in discussions" and "coherent speech" are strongly associated with C_j^{things}. Again, this aligns well with the claim that things-intense occupations are characterized by a low level of interaction with other people. Because only the aggregate-level information on native language is used in constructing C_j^{things}, this further corroborates our occupational classification.

the canton of Bern allows us to calculate the proportion of female apprentices within each of the 214 different occupations covered in the data with reasonable accuracy. We do, however, exclude occupations with very low overall frequencies, i.e. occupations with less than ten contracts as of August 2015.[15] This restriction leads to the exclusion of 44 (of a total of 214) occupations. However, because these occupations are rarely chosen, they represent only 0.32% of the total of all apprenticeship contracts. Another restriction is due to the fact that we do not have information on the cognitive requirements for all of the occupations; due to this restriction, we lose another 40 occupations. We thus end up with a total of 130 different occupations for which we have at least ten apprenticeship contracts as well as information on the cognitive requirements in that occupation (unless stated otherwise, all results in this section are based on this set of occupations). These 130 occupations cover about 90% of the total of apprenticeship contracts in the canton of Bern as of August 2015.

Figure 1

Figure 1 shows a histogram of the proportion of females by occupation. The most obvious feature is the bimodality of the distribution, with many occupations clustering at the two extreme values (specifically, there are many occupations that are primarily chosen by men; obviously, the least frequent occupations are those with a more or less balanced sex ratio).

Figure 2

This is more explicit evident from figure 2, which plots the number of individual apprenticeship contracts by males (females) in an occupation characterized by a given overall proportion of females choosing an occupation. Obviously, and as expected, both male and female adolescents tend to cluster in occupations mainly chosen by individuals of the same gender (cf. Kuhn and Wolter, 2019). On average, the typical occupation chosen by female adolescents is 71.4% female, while the typical occupation of male adolescents is 77.5% male. Thus, without any doubt, the data show that there is strong occupational gender segregation in the Swiss apprenticeship system (Aepli *et al.*, 2019, show similar results for the country as a whole).

[15] Architectural model builder ("Architekturmodellbauer") or glass painter ("Glasmaler") are just two examples of rarely chosen occupations. The occupational-level results are robust to the inclusion of the smaller occupations, however, as shown in columns 3 and 4 of appendix table A.6.

Men and women choose occupations with different task content

In a next step, we merge the variables describing the task content of occupations with the information on the proportion of females within each of these occupations. This allows us to see whether occupations females predominantly choose differ in their task content from those occupations males primarily select. Based on the existing empirical literature on the subject, we expect to find a negative association between the degree to which an occupation is heavy on things-related content and the proportion of women in that occupation.

Figure 3

Figure 3 thus shows a scatterplot of the fraction of females on the x-axis against C_j^{things}, the principal component describing the task content of an occupation j along the people-versus-things dimension on the y-axis (the dotted horizontal line corresponds to the overall proportion of women in the population of all apprenticeship contracts). The relation between the two variables could hardly be more evident. As expected, occupations mainly chosen by males are oriented towards things, while occupations predominantly chosen by females are tilted towards working with people.[16]

In a next step, we run several linear regression models of the following form:

$$F_j = \pi_0 + \pi_1 C_j^{\text{things}} + \pi_2 C_j^{\text{demand}} + \epsilon_j, \tag{1}$$

where the dependent variable, F_j, denotes the proportion of women choosing occupation j, and thus $F_j \in [0, 1]$. The regressor of main interest is C_j^{things}, which describes whether occupation j is oriented towards working with things (as discussed in section 4.1 above). We include the overall demand level C_j^{demand} as an additional regressor in some of the specifications, and we show both unweighted estimates as well as estimates which are weighted by the total number of apprenticeship contracts within a given occupation. We report heteroscedasticity-robust standard errors throughout.

Table 3

[16] A further notable feature of figure 3 is the fact that two of the most popular occupations among both males and females are located somewhere in the middle along the things-versus-people axis: "commercial employee" and "retail professional" (cf. appendix table A.1).

Table 3 shows the resulting parameter estimates. In the first column, we include C_j^{things} as the only regressor. This specification yields an estimate of $\widehat{\pi}_1 = -0.202$ with a robust standard error of about 0.014. The point estimate of π_1 implies that the predicted proportion of women in an occupation essentially shifts from zero to one as we move from the lowest to the highest values on C_j^{things}, consistent with the pattern shown in figure 3. Also note the large R-squared of 0.477 associated with this simple regression – also suggesting that the variable describing the task content along the things-versus-people dimension is likely one of the more important predictors of whether an occupation is predominantly chosen by men or women. In the second column, we add the overall demand level C_j^{demand} as regressor. By construction, this does not change the point estimate of π_1 – remember that the two variables are statistically independent from each other at the occupational level by construction – but it shows that the predictive value of the overall level of cognitive requirements is far lower than that of things-versus-people dimension. In substantive terms, it shows that women tend to favor somewhat more demanding occupations than men.[17] In the next two columns, we replicate the specifications from columns 1 and 2, but we weight the observations by the number of apprenticeship contracts. This yields similar point estimates for the two regressors, an even higher R-squared in each of the two specifications (0.771 and 0.799, respectively).[18]

Overall, the estimates reported in table 3 are not only in line with our expectations (and also with common sense, we would add), they also turn out to be surprisingly large, by any statistical measure. This in turn suggests that vocational interests along the things-people dimension presumably ranks amongst the most important proximate predictors of occupational gender segregation. In the final part of our empirical analysis, we will replicate this result in our individual-level data and then show that the gender difference in occupational preferences is robust to the inclusion of a variety of control variables.

[17]This is well in line with existing empirical evidence on gender differences in academic achievement, showing that girls tend to (slightly) outperform boys academically (e.g. Voyer and Voyer, 2014).

[18]Appendix table A.6 shows that we get very similar estimates when using both C_j^{math} and C_j^{abstract} as regressors (instead of C_j^{things} only). This is also the case if we include occupations with less than ten apprenticeship contracts or allow for a nonlinear relation between C_j^{things} and the proportion of women in an occupation.

4.3 Analysis of individual-level data

For the final part of the analysis, we therefore combine the information describing occupations' task content with the individual-level survey among the adolescents from the German language area of the canton of Bern. This will allow us to go beyond the analysis at the occupational level, checking whether the finding from section 4.2 is robust to the inclusion of additional predictors of occupational aspirations or choices, such as school grades or parental characteristics.

Gender differences in occupational aspirations

We therefore next estimate linear regressions that take the following form:

$$c_{j[i]}^{\text{things}} = \alpha + \beta f_i + \gamma_x x_i + \gamma_p p_i + \delta_p (f_i \times p_i) + \gamma_r r_{[i]} + \delta_r (f_i \times r_{r[i]}) + \epsilon_i, \tag{2}$$

where the dependent variable c_i^{things} characterizes the job content of occupation j for which individual i aspires while in 8th grade (i.e. c_i^o corresponds to the value of C_j^{things} associated with the occupation that individual i aspires for).[19] The main regressor in this case is a dummy variable indicating whether individual i is a female, in which case $f_i = 1$. Step by step, we will then also control for various individual-level controls x_i, parental-level controls p_i, as well as regional-level controls $\psi_{r[i]}$. Note that controls at both the parental and the regional level are expected to be uncorrelated with an adolescent's gender, and simply including these variables as additional controls will therefore not have any noticeable impact on the estimate of β. Instead, we thus also include the interaction terms between the female dummy and these controls and then check whether or not we can reject the null hypothesis that the parameters associated with the interaction terms are simultaneously equal to zero (we will show robust F-statistics and associated p-values). Throughout, we are primarily interested in the estimated size of gender differences in occupational preferences, which is given by parameter β in equation (2), as well

[19]More precisely, in the survey, individuals stated their occupational aspirations (with zero, one, or more than one possible occupations given). In a first step, and where possible, we assigned the same occupational coding to the raw text as those used for the occupational-level data. We then merged the corresponding value on C_j^{things} for each occupation with a valid code. In cases where the same individual named more than one occupation, we simply use the mean value of C_j^{things} across all occupations that individual in question named. We lose some observations because of missing answers ("don't know") or because we are not able to assign a learnable occupation to the answer (e.g. in cases where the adolescent stated an occupation which requires studying at the university, such as a medical doctor for example).

as in the coefficients associated with the interaction terms, i.e. δ_p and δ_r. We report robust standard errors for the full set of estimates.

At the individual level, we include an adolescent's exact age, an indicator for being a single child, the number of siblings, school grades (in mathematics as well as in German, French, and English), an indicator of the educational track, as well as survey measures of competitiveness, risk preferences, and locus of control as additional regressors (see Buser *et al.*, 2017a; Jaik and Wolter, 2019, for additional details). Parental-level controls include the highest educational attainment of both mother and father, as well as their occupations (major, one-digit ISCO group). Finally, in the full-blown specification, we include a full set of local labor markets dummies to further control for regional differences in the availability of apprenticeship positions in different occupations.

Table 4

Table 4 presents the resulting OLS estimates. In the first column, we simply regress the things-orientation of individual's occupational aspirations on the female dummy. This yields an estimate of $\widehat{\beta} = -1.775$, which implies that male and female adolescents in the sample differ widely in terms of the task content of their occupational aspirations. More specifically, this estimate implies that the mean difference in occupational preferences along the things-versus-people dimension between male and female adolescents equals more than one standard deviation, a result which is consistent with previous empirical evidence (e.g. Lippa, 2010; Su *et al.*, 2009).[20] With a robust standard error of about 0.062, the estimate is also statistically highly significant. Moreover, and similar to the occupational-level regressions, the resulting R-squared of 0.409 is unusually large. Overall, this first specification based on individual-level data is thus consistent with our finding based on the occupational-level data from section 4.2 above.

We next check whether this difference still holds when we add various control variables. We first add individual-level controls in the second column, which results in an estimate of $\widehat{\beta} = -1.721$ (with a robust standard error of 0.065, the estimate also remains highly statis-

[20] In the occupational level data, the standard deviation of C_j^{things} equals about 1.234. In the individual-level survey data, the standard deviation of $c_{j[i]}^{\text{things}}$ amounts to about 1.387. Both Su *et al.* (2009) and Lippa (2010) report similar-sized gender differences along the things-versus-people dimension of job content.

tically significant). Obviously, adding these controls has hardly any impact on the size of $\widehat{\beta}$, either because there are no or only small gender differences in these variables and/or because these variables do not predict occupational preferences along the things-versus-people dimension. In the next column, we further add parental-level controls as well as the interaction terms between these and the female dummy. Here we are mainly interested in the coefficients associated with the interaction terms.[21] The F-test associated with the null hypothesis that the gender difference in occupational preferences does not vary with parental education is small and statistically insignificant ($F = 1.236$, with a p-value of 0.253); similarly for parental occupation ($F = 1.350$, $p = 0.184$). In column 4, we further add dummies representing local labor markets and their interactions with the female dummy. This also yields a small and insignificant test statistic ($F = 0.900$, $p = 0.546$).

Taken together, and consistent with the results from our occupational-level analysis above, the estimates from table 4 show that there is a large and statistically significant gender difference over the task content along the things-versus-people dimension of individual occupational aspirations. Moreover, we also find that this difference is not driven by any of the individual- or parental-level control variables considered in the regression analysis.

Gender differences in actual occupational choices

In the final step of our analysis, we rerun the analysis above, but focus on the job content along the things-versus-people dimension of actual occupational choices instead of occupational aspirations.[22] By and large, these additional estimates mirror those associated with individuals' occupational aspirations. There is, however, a small yet statistically insignificant difference in the average value in the "things"-intensity of occupational aspirations ($\bar{c} = -0.433$) and actual occupational choices ($\bar{c} = -0.361$)

Table 5

[21]As expected, however, and consistent with the results of Kuhn and Wolter (2019), the F-tests associated with the main effects of the parental controls tend to be statistically significant. Thus, children from different parental backgrounds do differ in their occupational preferences along the things-people dimension, but there appears to be no differential pattern between girls and boys in this regard.

[22]Focusing on actual choices implies that we must focus on a smaller, and potentially somewhat selective, sample. Specifically, both individuals who decided to go on with general education (i.e. baccalaureate school) and individuals who have not yet found an apprenticeship position, or who opted for an interim solution, are not included in this subsample.

Again, the first specification of table 5 simply regresses the things-orientation of an adolescent's actual occupational choice on the female dummy. This yields an estimate of $\widehat{\beta} = -1.691$ with an standard error of 0.068; somewhat smaller, though not significantly so, than the corresponding estimate from table 4. Thus, in the unconditional case we find a similar sized gender difference along the things-versus-people dimension for occupational aspirations as well as occupational choices. As evident from column 2, adding individual-level controls has only a small impact on the estimate of β, which becomes somewhat smaller than in the unconditional case ($\widehat{\beta} = -1.664$, robust standard error of 0.070). Again, this parallels the corresponding result related to occupational aspirations from table 4.

Results differ somewhat between occupational choices and aspirations when we further add parental-level controls and the corresponding interaction terms with the female dummy, as done in column 3 of table 5. In the case of actual occupational choices, both the F-test associated with the interaction terms between the female dummy and parents' education ($F = 2.437, p = 0.024$) and with the interaction terms between the female dummy and parental occupation ($F = 2.921, p < 0.001$) turns out to be statistically significant, suggesting that parental background does have some differential impact on the things-intensity of actual occupational choices between girls and boys.

In general, however, the impact of parental background remains very limited, as the gender difference in occupational choices along the things-people dimension shows up across most attributes characterizing parental background. Thus one main conclusion from the analysis using individual-level data is that the gender difference in vocational interests is not driven by any of the controls at the individual or at the parental level, and that it already shows up in adolescents' aspirations.

5 Conclusions

Occupational gender segregation remains at persistently high levels, for reasons not yet fully understood. In this paper, we add to this important discussion and argue that gender differences in vocational interests are among the most important proximate determinants of occupational gender segregation. More precisely, motivated by previous empirical evidence, mainly by psychologists, we use a unique combination of different data sources to show that male adolescents clearly tend to favor occupations which require creating and/or manipulating objects (i.e. "things"), while female adolescents prefer to work in occupations in which interacting with customers or patients is important (i.e. "people").

For our own empirical analysis, we use data on the cognitive requirements in 130 different learnable occupations in the Swiss apprenticeship system to classify them according to their broad task content along the things-versus-people dimension, using a data-driven process. Moreover, we validate our classification results using an alternative, fully independent source of data that describes the actual task-content of different occupations. We then show at the occupational level that this simple, unidimensional classification of the occupations explains a surprisingly large part of the observed variation in the proportion of females in an occupation (with an unusually large R-squared of 0.477, and an even larger R-squared of 0.771 in the case of weighting the regression by the number of apprenticeship contracts in a given occupation). In other words, knowing whether an occupation is tilted towards working with things rather than people allows one to formulate a reasonable guess as to whether men rather than women predominantly choose it.

In the second part of our empirical analysis, we replicate this finding using individual-level data for a sample of adolescents from the German language area of the canton of Bern. Taken together, we find that variation in the task content of occupations along the things-versus-people is a very powerful predictor of whether males or females predominantly choose an occupation; actually, it appears that this variable is likely one of the most important proximate predictors of occupational gender segregation – by any statistical measure applied. For example, a simple, univariate regression of the things-intensity of individuals' occupational aspirations on a female dummy yields an estimate of -1.775, which gives an effect size of about 1.28. At this point, it is perhaps worth pointing out just how large the effect size associated with this estimate actually is: it is, for example, larger than the effect size associated with the gender difference in body height.[23] As we mentioned above, however, this finding is well in line with many previous empirical studies, mostly by psychologists. For example, both Su *et al.* (2009) and Lippa (2010)

[23]For example, data from the Swiss Health Survey ("Schweizer Gesundheitsbefragung") for the year 2017 show a mean difference of about 12.7 cm in height between men and women, with a standard deviation of about 12.89 cm. Thus, the associated effect size equals 0.98.

report gender differences in vocational interests along the things-versus-people dimension that are as large as those reported in this paper.

Our findings also overlap with and complement a number of recent empirical studies which show that various characteristics of (the task content of) an occupation are important in explaining the segregation of men and women into different occupations. Most obviously, perhaps, the results of our analysis align closely with the findings from Baker and Cornelson (2018), who document gender differences in visuo-spatial and motor aptitudes (see also Halpern, 2013, on a more general discussion of gender differences in cognitive aptitudes). Overall, and based on our own results as well as on the results of related studies, it appears obvious to us that gender differences in these traits are absolutely key for understanding why men and women tend to prefer to work in distinct occupations.

References

Aepli, M., Kuhn, A., and Schweri, J. (2019). Frauen und Männer lernen andere Berufe. EHB, skilled 2/19, 4–5.

Baker, M. and Cornelson, K. (2018). Gender-based occupational segregation and sex differences in sensory, motor, and spatial aptitudes. *Demography*, **55**(5), 1749–1775.

Bertrand, M. (2011). New perspectives on gender. *Handbook of Labor Economics*, **4b**, 1543–1590.

Blau, F. D. and Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, **55**(3), 789–865.

Bonin, H., Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, **14**(6), 926–937.

Burbano, V., Padilla, N., and Meier, S. (2020). Gender differences in preferences for meaning at work. IZA Discussion Paper No. 13053.

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

Buser, T., Peter, N., and Wolter, S. C. (2017a). Gender, competitiveness, and study choices in high school: Evidence from Switzerland. *American Economic Review*, **107**(5), 125–30.

Buser, T., Peter, N., and Wolter, S. C. (2017b). Gender, willingness to compete and career choices along the whole ability distribution. IZA Discussion Paper No. 10976.

Charles, M. (2017). Venus, Mars, and math: Gender, societal affluence, and eighth graders' aspirations for STEM. *Socius*, **3**, 1–16.

Charles, M. and Bradley, K. (2009). Indulging our gendered selves? Sex segregation by field of study in 44 countries. *American Journal of Sociology*, **114**(4), 924–976.

Charles, M. and Grusky, D. B. (2004). *Occupational ghettos: The worldwide segregation of women and men*. Stanford, CA : Stanford University Press.

Charness, G. and Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, **83**(1), 50–58.

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

Cortes, P. and Pan, J. (2018). Occupation and Gender. In S. L. Averett, L. M. Argys, and S. D. Hoffman, editors, *The Oxford Handbook of Women and the Economy*, chapter 18, pages 425–452. Oxford, UK: Oxford University Press.

Croson, R. and Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, **47**(2), 448–74.

Dittrich, M. and Leipold, K. (2014). Gender differences in time preferences. *Economics Letters*, **122**(3), 413–415.

Falk, A. and Hermle, J. (2018). Relationship of gender differences in preferences to economic development and gender equality. *Science*, **362**(6412), eaas9899.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global evidence on economic preferences. *Quarterly Journal of Economics*, **133**(4), 1645–1692.

Fouarge, D., Kriechel, B., and Dohmen, T. (2014). Occupational sorting of school graduates: The role of economic preferences. *Journal of Economic Behavior & Organization*, **106**, 335–351.

Gelblum, M. (2020). Preferences for job tasks and gender gaps in the labor market. Unpublished manuscript, Harvard University.

Gneezy, U. and Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, **94**(2), 377–381.

Gorsuch, R. L. (2014). *Factor analysis*. Taylor and Francis, 2nd edition.

Halpern, D. F. (2013). *Sex differences in cognitive abilities*. Psychology Press, 4th edition.

Holland, J. L. (1959). A theory of vocational choice. *Journal of Counseling Psychology*, **6**(1), 35–45.

Jaik, K. and Wolter, S. C. (2019). From dreams to reality: market forces and changes from occupational intention to occupational choice. *Journal of Education and Work*, **32**(4), 320–334.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

Kahn, S. and Ginther, D. (2017). Women and STEM. NBER Working Paper 23525.

Kamas, L. and Preston, A. (2015). Can social preferences explain gender differences in economic behavior? *Journal of Economic Behavior & Organization*, **116**, 525–539.

Kleinjans, K. J. (2009). Do gender differences in preferences for competition matter for occupational expectations? *Journal of Economic Psychology*, **30**(5), 701–710.

Kooreman, P. (2009). The early inception of labor market gender differences. *Labour Economics*, **16**(2), 135–139.

Kuhn, A. and Wolter, S. C. (2019). The Strength of Gender Norms and Gender-Stereotypical Occupational Aspirations Among Adolescents. IZA Discussion Paper No. 12861.

Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people–things dimension. *Journal of Personality and Social Psychology*, **74**(4), 996–1009.

Lippa, R. A. (2010). Gender differences in personality and interests: When, where, and why? *Social and Personality Psychology Compass*, **4**(11), 1098–1110.

Lordan, G. and Pischke, J.-S. (2016). Does Rosie like riveting? Male and female occupational choices. NBER Working Paper Nr. 22495.

Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, **3**(1), 601–630.

Olivetti, C. and Petrongolo, B. (2016). The evolution of gender gaps in industrialized countries. *Annual Review of Economics*, **8**, 405–434.

SCCRE (2014). *Swiss Education Report 2014*. Aarau: Swiss Coordination Centre for Research in Education (SCCRE).

Schmitt, D. P., Realo, A., Voracek, M., and Allik, J. (2008). Why can't a man be more like a woman? sex differences in big five personality traits across 55 cultures. *Journal of Personality and Social Psychology*, **94**(1), 168–182.

SERI (2017). Vocational and Professional Education and Training in Switzerland. Facts and Figures 2017. Bern: State Secretariat for Education, Research and Innovation (SERI).

Stoet, G. and Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence*, **48**, 137–151.

Su, R. and Rounds, J. (2015). All STEM fields are not created equal: People and things interests explain gender disparities across STEM fields. *Frontiers in Psychology*, **6**, 189.

Su, R., Rounds, J., and Armstrong, P. I. (2009). Men and things, women and people: a meta-analysis of sex differences in interests. *Psychological Bulletin*, **135**(6), 859–884.

Voyer, D. and Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, **140**(4), 1174–1204.

Wettstein, E., Schmid, E., and Gonon, P. (2017). *Swiss Vocational and Professional Education and Training (VPET)*. Bern: hep Verlag.

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

Table 1: The occupations ranked highest/lowest along the things–people dimension

Rank	Occupation	C_j^{things}
(a) The ten most things-oriented occupations		
1.	Plant and apparatus engineer ("Anlagen- und Apparatebauer")	2.260
2.	Design engineer ("Konstrukteur")	2.084
3.	Mechanical engineer ("Polymechaniker")	2.052
4.	Automatic technician ("Automatikmonteur")	2.027
5.	Mechanical technician ("Produktionsmechaniker")	2.027
6.	Gunsmith ("Büchsenmacher")	1.974
7.	Insulation contractor ("Isolierspengler")	1.933
8.	Automation engineer (Automatiker")	1.884
9.	Carpenter ("Schreiner")	1.538
10.	Industrial ceramist ("Industriekeramiker")	1.517
(b) The ten most people-oriented occupations		
1.	Podiatrist ("Podologin")	−3.049
2.	Healthcare assistant ("Fachfrau Gesundheit")	−2.715
3.	Information and documentation specialist ("Fachfrau Information und Dokumentation")	−2.634
4.	Certified social care worker ("Fachfrau Betreuung")	−2.590
5.	Druggist ("Drogistin")	−2.538
6.	Pharmacy assistant ("Pharma-Assistentin")	−2.234
7.	Hairdresser ("Coiffeuse")	−2.233
8.	Bookseller ("Buchhändlerin")	−2.218
9.	Hairdresser ("Coiffeuse")	−2.136
10.	Customer dialogue specialist ("Fachmann Kundendialog")	−2.072

Notes: The table shows the ten occupations with the highest (lowest) scores on C_j^{things}. The official German description is given in parentheses, along with the English translation suggested by the State Secretariat for Education, Research and Innovation (SERI), where available (otherwise we chose our own translation).

Table 2: Validating our classification of occupations using information from actual job postings

	C_j^{things}	
Agricultural tasks	$-0.460^{\star}$	-0.220
Manufacturing	0.125	0.421
Installation, assembly, construction	$0.875^{\star\star}$	$1.309^{\star\star\star}$
Set up, operate	0.669	2.555
Repair, restore	$0.823^{\star\star\star}$	$1.363^{\star\star\star}$
Store and transport	$0.352^{\star\star\star}$	$0.320^{\star\star\star}$
Purchasing/sales, cashier, customer service	$-2.101^{\star\star\star}$	$-1.183^{\star\star\star}$
Writing, correspondence, administration	$-4.958^{\star\star\star}$	$-3.965^{\star}$
Accounting and finance	7.292	1.300
IT, programming	-0.742	0.098
Hospitality services	$-2.977^{\star\star\star}$	$-2.255^{\star\star}$
Ironing, cleaning, waste management	-0.116	-0.462
Guarding	-4.339	-5.727
Analysis/research, controller	-0.481	-0.027
Planning, engineering, designing/drawing	$1.918^{\star\star\star}$	$2.018^{\star\star\star}$
Supervising, hiring	9.892	-17.663
Disposing, organizing, leading	3.282	1.964
Educating/teaching, advising	$-4.056^{\star}$	-3.968
Administration of justice	-1.569	110.618
Medical and cosmetical care	$-1.944^{\star\star\star}$	$-2.490^{\star\star\star}$
Publishing, creative work	-0.881	-3.225
Weighted regression	No	Yes
Number of observations	130	130
R-Squared	0.673	0.877
Adjusted R-Squared	0.610	0.853

Notes: $\star$, $\star\star$, $\star\star\star$ denotes statistical significance on the 10%, 5%, and 1% level, respectively. The table shows point estimates from a regression suppressing the constant term (robust standard errors are not shown to keep the table compact). The labels for the activities are carried over from the "Stellenmarktmonitor" data. Weights are equal to the number of apprenticeship contracts in the canton of Bern in August 2015 in a given occupation.

Table 3: Occupational-level regressions

	F_j			
C_j^{things}	$-0.202^{\star\star\star}$	$-0.202^{\star\star\star}$	$-0.228^{\star\star\star}$	$-0.220^{\star\star\star}$
	(0.014)	(0.015)	(0.013)	(0.016)
C_j^{demand}		$0.050^{\star\star\star}$		$0.045^{\star\star}$
		(0.014)		(0.019)
Weighted regression	No	No	Yes	Yes
Number of observations	130	130	130	130
R-Squared	0.477	0.524	0.771	0.799
Adjusted R-Squared	0.473	0.517	0.769	0.795

Notes: *, **, *** denotes statistical significance on the 10%, 5%, and 1% level, respectively. Robust stanard errors are given in parentheses. Weights are equal to the number of apprenticeship contracts in the canton of Bern as of August 2015 in a given occupation.

Table 4: Individual-level regressions, occupational aspirations at the start of 8th grade

	$c^{\text{things}}_{j[i]}$			
Female$_i$ (yes = 1)	-1.775^{***}	-1.721^{***}	-1.556^{***}	-1.792^{***}
	(0.062)	(0.065)	(0.242)	(0.267)
Individual-level controls:				
Demographics	No	Yes	Yes	Yes
School grades	No	Yes	Yes	Yes
Preferences	No	Yes	Yes	Yes
Parental-level controls:				
Education	No	No	Yes	Yes
Occupation	No	No	Yes	Yes
Regional-level controls:				
Local labor market	No	No	No	Yes
Main effects:				
Parental education			0.477	0.492
			(0.929)	(0.921)
Parental occupation			2.737	2.280
			(0.000)	(0.003)
Local labor market				1.325
				(0.198)
Interaction effects:				
Female × parental education			1.236	1.350
			(0.253)	(0.184)
Female × parental occupation			1.305	1.310
			(0.186)	(0.183)
Female × local labor market				0.900
				(0.546)
Number of observations	1,191	1,191	1,191	1,191
R-Squared	0.409	0.423	0.455	0.473
Adjusted R-Squared	0.409	0.416	0.421	0.428

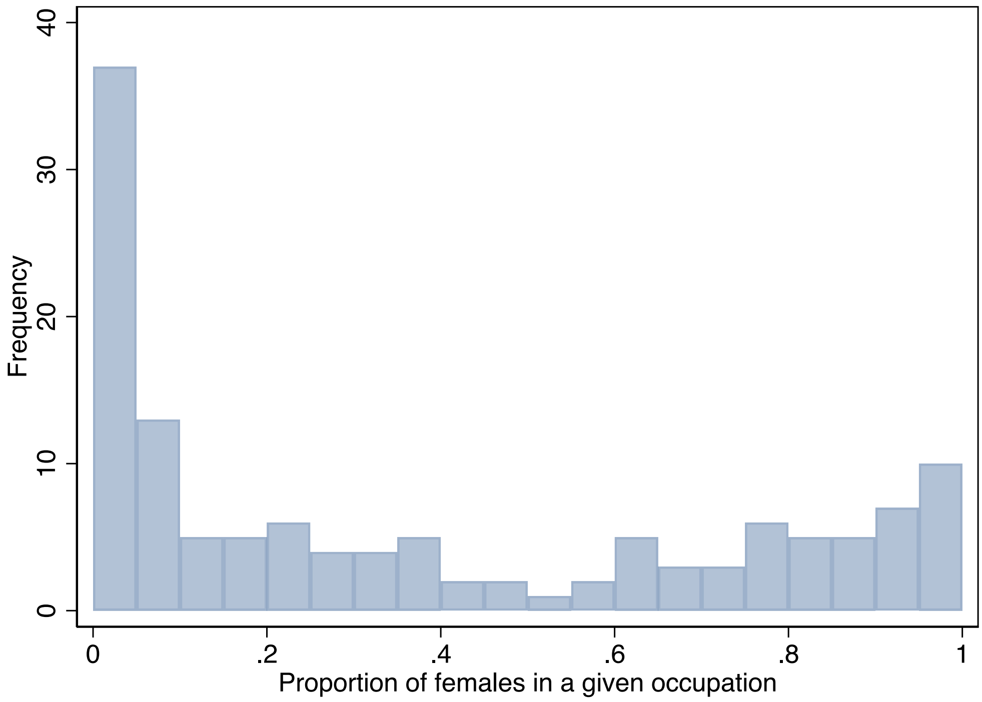
Notes: *, **, *** denotes statistical significance on the 10%, 5%, and 1% level, respectively. Robust standard errors are given in parentheses. The table also reports robust F-statistics (with associated p-values in parentheses below) from testing the null hypothesis that the corresponding interaction terms are equal to zero. Full regression results are available upon request.
The full list of controls is as follows: "demographics" includes the age at the time of the survey, the number of siblings, and a dummy for being a single child; "school grades" includes grades in mathematics, German, French and English; "preferences" includes survey measures of competitiveness, risk preferences, and locus of control; parental education includes a full set of dummies for the highest attained education of both mother and father; parental occupation includes a full set of dummies for the occupation of both mother and father (major ISCO group); local labor market includes a full set of local-labor markets dummies.

Table 5: Individual-level regressions, occupational choices at the end of 9th grade

	$c_{j[i]}^{\text{things}}$			
Female$_i$ (yes = 1)	-1.691^{***}	-1.664^{***}	-2.182^{***}	-2.581^{***}
	(0.068)	(0.070)	(0.480)	(0.624)
Individual-level controls:				
Demographics	No	Yes	Yes	Yes
School grades	No	Yes	Yes	Yes
Preferences	No	Yes	Yes	Yes
Parental-level controls:				
Education	No	No	Yes	Yes
Occupation	No	No	Yes	Yes
Regional-level controls:				
Local labor market	No	No	No	Yes
Main effects:				
Parental education			1.974	2.102
			(0.024)	(0.015)
Parental occupation			2.921	3.366
			(0.000)	(0.000)
Local labor market				1.681
				(0.066)
Interaction effects:				
Female × parental education			2.437	2.494
			(0.004)	(0.003)
Female × parental occupation			4.693	4.232
			(0.000)	(0.000)
Female × local labor market				0.789
				(0.662)
Number of observations	953	953	953	953
R-Squared	0.397	0.411	0.465	0.479
Adjusted R-Squared	0.396	0.403	0.423	0.423

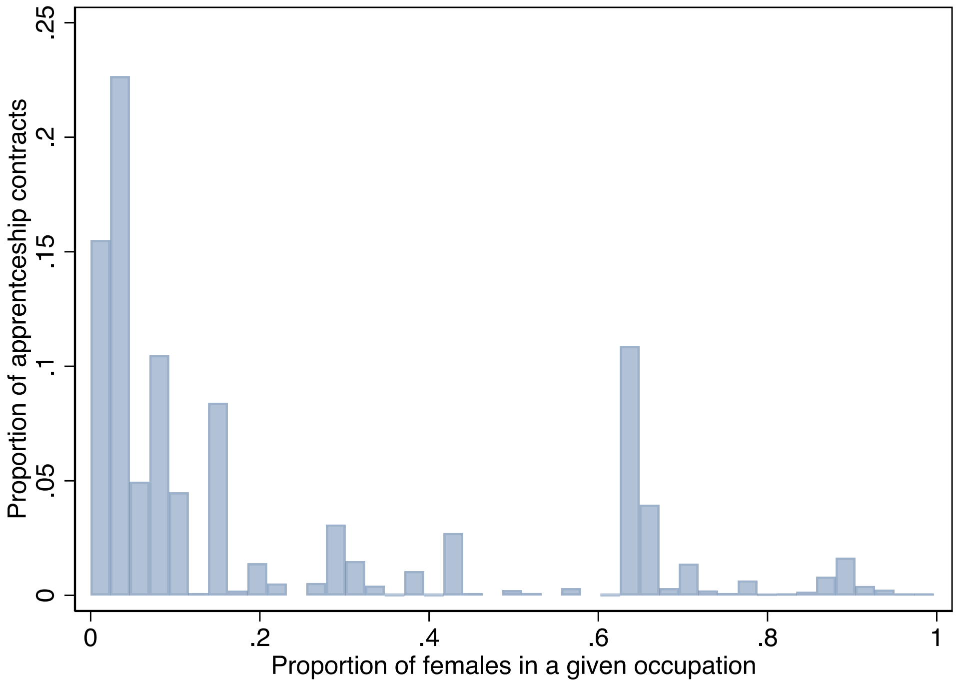
Notes: *, **, *** denotes statistical significance on the 10%, 5%, and 1% level, respectively. Robust standard errors are given in parentheses. The table also reports robust F-statistics (with associated p-values in parentheses below) from testing the null hypothesis that the corresponding interaction terms are equal to zero. Full regression results are available upon request.
The full list of controls is as follows: "demographics" includes the age at the time of the survey, the number of siblings, and a dummy for being a single child; "school grades" includes grades in mathematics, German, French and English; "preferences" includes survey measures of competitiveness, risk preferences, and locus of control; parental education includes a full set of dummies for the highest attained education of both mother and father; parental occupation includes a full set of dummies for the occupation of both mother and father (major ISCO group); local labor market includes a full set of local-labor markets dummies.

Figure 1: Proportion of females, by occupation



Notes: The figure shows the absolute number of learnable occupations with a given proportion of females choosing that occupation. Only occupations with ten or more apprenticeship contracts are included in the figure ($J = 130$; see sections 3 and 4.2 in the main text for details). There are 12 (3) occupations with a female share of exactly 0 (1).

Figure 2: Occupational gender segregation

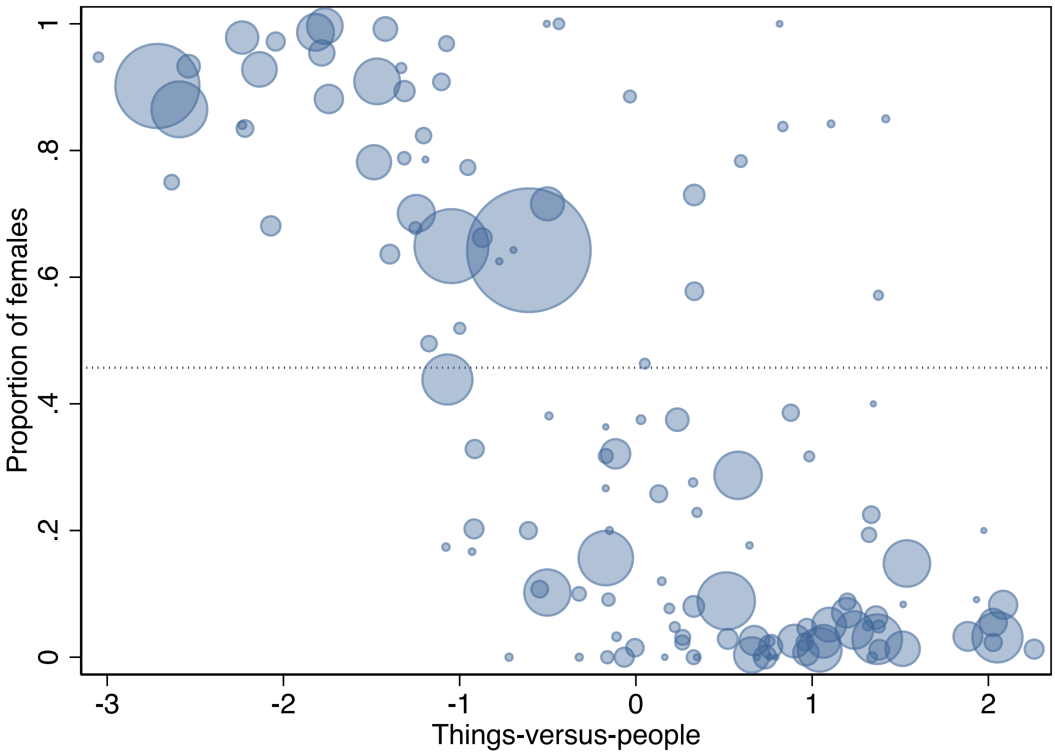


(a) Male apprentices

(b) Female apprentices

Notes: The figure shows the distribution of individuals across occupations characterized by varying proportions of females (shown on the x-axis of each panel). Panel (a) shows the distribution of male apprentices; panel (b) the distribution of female apprentices. The two figures are based on the population of apprenticehip contracts in the canton of Bern as of 2015. See also sections 3 and 4.2 in the main text for details.

Figure 3: Proportion of females versus broad task-content, by occupation



Notes: The figure plots, by occupation, the proportion of females against the task content along the people-versus-things dimension (positive values indicate that an occupation is more oriented towards things, negative values indicate than an occupation is oriented towards people). Only occupations with at least ten apprenticeship contracts are considered ($J = 130$). The size of the markers is proportional to the number of apprenticeship contracts in an occupation in the canton of Bern as of August 2015. The dotted horizontal line denotes the mean fraction of females in the population of apprenticeship contracts in the canton of Bern. See main text for additional details.

A Additional tables and figures

Table A.1: The ten most popular apprenticeship occupations among male and female apprentices in the canton of Bern

Rank	Occupation	f_j	F_j
(a) Male apprentices			
1.	Commercial employee (“Kaufmann”)	0.093	0.643
2.	Information technologist (“Informatiker”)	0.051	0.089
3.	Farmer (“Landwirt”)	0.042	0.156
4.	Mechanical engineer (“Polymechaniker”)	0.041	0.031
5.	Licensed electrician (“Elektroinstallateur”)	0.040	0.028
6.	Retail professional (“Detailhandelsfachmann”)	0.033	0.649
7.	Logistics expert (“Logistiker”)	0.033	0.102
8.	Carpenter (“Zimmermann”)	0.032	0.012
9.	Carpenter (“Schreiner”)	0.031	0.148
10.	Gardener (“Gärtner”)	0.027	0.287
(b) Female apprentices			
1.	Commercial employee (“Kauffrau”)	0.213	0.643
2.	Healthcare assistant (“Fachfrau Gesundheit”)	0.138	0.902
3.	Retail professional (“Detailhandelsfachfrau”)	0.077	0.649
4.	Certified social care worker (“Fachfrau Betreuung”)	0.058	0.865
5.	Specialist in hotel housekeeping (“Hotelfachfrau”)	0.041	0.909
6.	Dental assistant (“ Dentalassistentin”)	0.029	0.987
7.	Medical practice assistant (“Medizinische Praxisassistentin”)	0.027	0.996
8.	Hairdresser (“Coiffeuse”)	0.024	0.928
9.	Chef (“Köchin”)	0.024	0.438
10.	Pharmacy assistant (“Pharma-Assistentin”)	0.022	0.978

Notes: The table shows the ten most popular occupations among male and female apprentices in the canton of Bern as of August 2015 (the official German description of the occupation is given in parentheses, along with the English translation proposed by the State Secretariat for Education, Research and Innovation (SERI), where available). f_j denotes the proportion of male (female) apprentices choosing a specific occupation among all male (female) apprentices, while F_j denotes the proportion of female apprentices in the corresponding occupation.

Table A.2: Principal-components analysis of the occupational-level data describing the cognitive requirements in four different subjects

Variable	PC-1	PC-2	PC-3	PC-4
Mathematics	0.3617	0.7195	0.4990	0.3202
Native language	0.5694	−0.3266	−0.3653	0.6601
Natural sciences	0.5661	0.3110	−0.4754	−0.5974
Foreign language	0.4739	−0.5281	0.6258	−0.3238
Proportion	0.616	0.287	0.075	0.023

Notes: The table shows the results from the principal-components analysis of the four main subjects covered by the data on cognitive requirements in learnable occupations.The upper part of the table shows the factor loadings for the four PCs, the lower part shows the proportion of variance explained by any of the four PCs.

Table A.3: Task content of occupations ranked highest/lowest along the overall level of cognitive requirements

Rank	Occupation	C_j^{demand}
(a) Ranked highest on overall cognitive demands		
1.	Mediamatician ("Mediamatiker")	3.600
2.	Optometrist ("Augenoptikerin")	3.445
3.	Interactive media designer ("Interactive Media Designer"	3.133
4.	Geomatics expert ("Geomatiker")	3.028
5.	Information and documentation specialist ("Fachfrau Information und Dokumentation")	2.977
6.	Druggist ("Drogistin")	2.848
7.	Laboratory technician ("Laborantin")	2.656
8.	Automation engineer ("Automatiker")	2.476
9.	Bookseller ("Buchhändlerin")	2.467
10.	Information technologist ("Informatiker")	2.461
(b) Ranked lowest on overall cognitive demands		
1.	Mechanical assistant ("Mechanikpraktiker")	−2.972
2.	Automobile assistant ("Automobil-Assistent")	−2.793
3.	Timber worker ("Holzbearbeiter")	−2.792
4.	Logistician ("Logistikerin")	−2.722
5.	Dairy industry assistant ("Milchpraktiker")	−2.573
6.	Tire work assistant ("Reifenpraktiker")	−2.569
7.	Food production assistant ("Lebensmittelpraktiker")	−2.564
8.	Horse keeper ("Pferdewartin")	−2.548
9.	Metal construction practitioner ("Metallbaupraktiker")	−2.542
10.	Building services assistant ("Haustechnikpraktiker")	−2.463

Notes: The table shows the ten occupations ranked highest (lowest) in terms of their overall cognitive requirement level, C_j^{demand}.

Table A.4: Disaggregated task content of occupations ranked highest/lowest along the things-versus-people dimension

Rank	Occupation	C_j^{math}	C_j^{abstract}	C_j^{things}
(a) The most things-oriented occupations				
1.	Plant and apparatus engineer (“Anlagen- und Apparatebauer”)	1.713	0.547	2.260
2.	Design engineer (“Konstrukteur”)	1.434	0.650	2.084
3.	Mechanical engineer (“Polymechaniker”)	1.389	0.662	2.052
4.	Automatic technician (“Automatikmonteur”)	1.178	0.850	2.027
5.	Mechanical technician (“Produktionsmechaniker”)	1.178	0.850	2.027
6.	Gunsmith (“Büchsenmacher”)	1.380	0.594	1.974
7.	Insulation contractor (“Isolierspengler”)	1.698	0.235	1.933
8.	Automation engineer (Automatiker”)	1.056	0.828	1.884
9.	Carpenter (“Schreiner”)	1.515	0.023	1.538
10.	Industrial ceramist (“Industriekeramiker”)	0.599	0.918	1.517
(b) The most people-oriented occupations				
1.	Podiatrist (“Podologin”)	−1.293	−1.757	−3.049
2.	Healthcare assistant (“Fachfrau Gesundheit”)	−1.242	−1.474	−2.715
3.	Information and documentation specialist (“Fachfrau Information und Dokumentation”)	−2.784	0.150	−2.634
4.	Certified social care worker (“Fachfrau Betreuung”)	−1.060	−1.530	−2.590
5.	Druggist (“Drogistin”)	−1.582	−0.956	−2.538
6.	Pharmacy assistant (“Pharma-Assistentin”)	−1.504	−0.730	−2.234
7.	Hairdresser (“Coiffeuse”)	−1.660	−0.573	−2.233
8.	Bookseller (“Buchhändler”)	−2.403	0.185	−2.218
9.	Hairdresser (“Coiffeuse”)	−1.207	−0.929	−2.136
10.	Customer dialogue specialist (“Fachmann Kundendialog”)	−3.048	0.976	−2.072

Notes: The variable C_j^{things} is equal to the sum of C_j^{math} and C_j^{abstract}. Therefore, column 3 is identical to table 1 in the main text.

Table A.5: Validating our occupational classification using subcategories in cognitive requirements (native language only)

	Cognitive requirements in native language					
	Mean	Subcat-1 Reading	Subcat-2 Listening	Subcat-3 Writing	Subcat-4 Conversation	Subcat-5 Speech
C_j^{things}	-4.758^{***}	-0.567	-6.928^{***}	-3.406^{***}	-5.798^{***}	-7.092^{***}
	(0.187)	(0.637)	(0.378)	(0.408)	(0.399)	(0.497)
C_j^{demand}	8.071^{***}	8.361^{***}	6.590^{***}	8.560^{***}	8.079^{***}	8.764^{***}
	(0.175)	(0.406)	(0.392)	(0.326)	(0.339)	(0.315)
Number of observations	130	130	130	130	130	130
R-Squared	0.962	0.738	0.814	0.861	0.861	0.856
Adjusted R-Squared	0.962	0.734	0.811	0.859	0.859	0.854

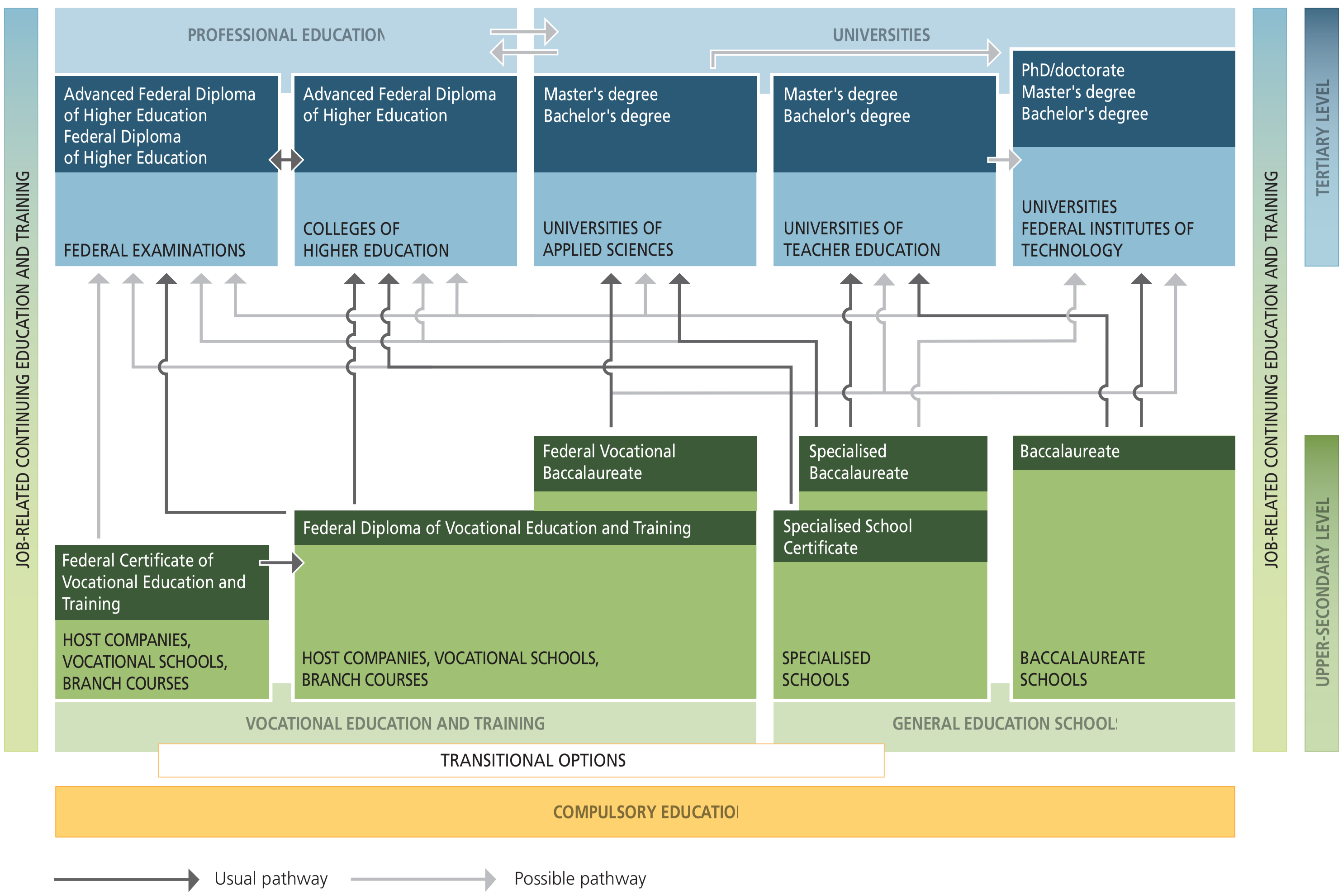
Notes: *, **, *** denotes statistical significance on the 10%, 5%, and 1% level, respectively. Robust standard errors are given in parentheses. Weighted regressions are not shown, but yield very similar estimates.

Table A.6: Occupational-level regressions, robustness

	F_j					
C_j^{math}	-0.206^{***}	-0.239^{***}				
	(0.018)	(0.014)				
C_j^{abstract}	-0.186^{***}	-0.178^{***}				
	(0.028)	(0.015)				
C_j^{demand}	0.050^{***}	0.030^{*}	0.048^{***}	0.045^{**}	0.048^{***}	0.049^{***}
	(0.014)	(0.018)	(0.014)	(0.019)	(0.015)	(0.015)
C_j^{things}			-0.188^{***}	-0.220^{***}	-0.266^{***}	-0.334^{***}
			(0.019)	(0.016)	(0.031)	(0.024)
$(C_j^{\text{things}})^2$					0.036^{***}	0.041^{***}
					(0.012)	(0.012)
$(C_j^{\text{things}})^3$					0.022^{***}	0.034^{***}
					(0.007)	(0.006)
Weighted regression	No	Yes	No	Yes	No	Yes
Number of observations	130	130	153	153	130	130
R-Squared	0.525	0.817	0.430	0.797	0.553	0.852
Adjusted R-Squared	0.514	0.813	0.423	0.795	0.539	0.848

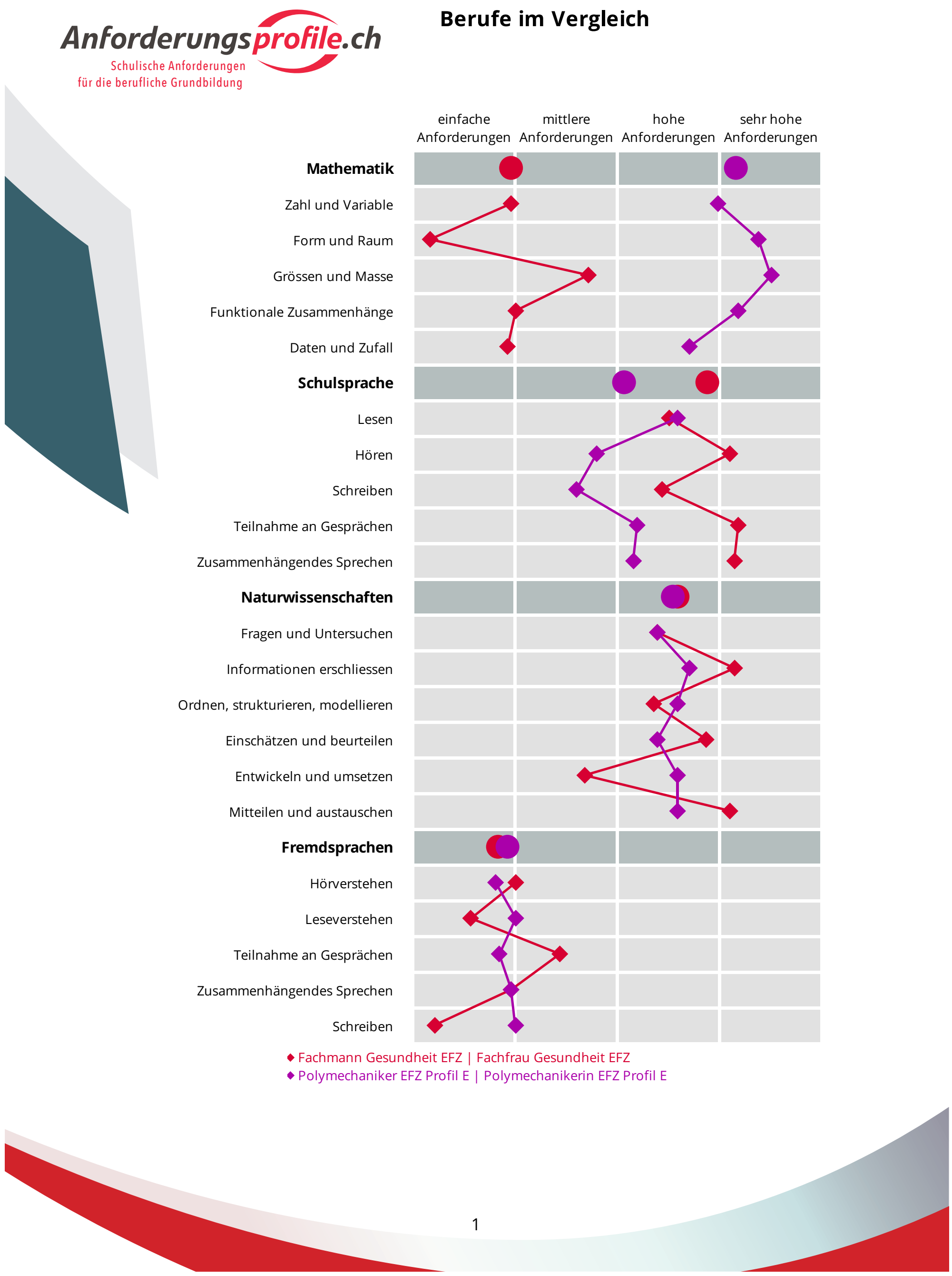
Notes: *, **, *** denotes statistical significance on the 10%, 5%, and 1% level, respectively. Robust standard errors are given in parentheses. Weights are equal to the number of apprenticeship contracts in the canton of Bern as of August 2015 in a given occupation.

Figure A.1: The Swiss educational system



Source: State Secretariat for Education, Research and Innovation (SERI).

Figure A.2: Two example profiles from the "Anforderungsprofile" website



Notes: The figure shows an example comparison in the cognitive requirements for two different occupations, healthcare assistant ("Fachmann/-frau Gesundheit") and mechanical engineer ("Polymechaniker/-in"), as available from the website (www.anforderungsprofile.ch). See also table 1 in the main text.