Does Rosie Like Riveting?
Male and Female Occupational Choices

Grace Lordan and Jörn-Steffen Pischke
LSE

June 2018

Abstract: Occupational segregation and pay gaps by gender remain large while many of the constraints traditionally believed to be responsible for these gaps seem to have weakened over time. We explore the possibility that women and men have different tastes for the content of the work they do. We run regressions of job satisfaction and job mobility on measures that proxy for the content of the work in an occupation, which we label 'people,' 'brains,' and 'brawn.' The results suggest that women value jobs that are relatively high on 'people' content and low on 'brawn.' Men care about job content in a similar fashion but seem to have much weaker preferences. These relationships hold up in a separate analysis that includes controls for firm fixed effects, suggesting that these findings do not just reflect differences in the work environment. To substantiate that our results indicate differences in preferences for job content rather than some other unobserved aspect of jobs we conducted a discrete choice experiment with high school students. The students' hypothetical choices roughly mimic the actual behavior of adults in the labor market. The majority of students report that they are choosing between jobs on the basis of preferences for the work itself rather than other factors.

JEL classifications: J16, J4
Keywords: Occupational choice, job content, gender, preferences

Email addresses: g.lordan@lse.ac.uk and s.pischke@lse.ac.uk. We thank Krittika Ray and Liaoliang Zhang for excellent research assistance and Alan Manning, Kelly Bedard, Dan Black, Alex Bryson, Deborah Cobb Clark, Stephen Machin, David Neumark, Ron Oaxaca, Barbara Petrongolo, and David Ribar for helpful discussions and participants of the CEP Labor Market Workshop at the LSE, 14th IZA SOLE transatlantic workshop, 18th IZA European Summer School in Labor Economics, 2016 Royal Economic Society meeting, and 2016 Society of Labor Economics Meeting for their comments. We are grateful to the ESRC for support through the Centre for Economic Performance at the LSE.

"And finally, in our time a beard is the one thing that a woman cannot do better than a man." - John Steinbeck, Travels with Charley: In Search of America.

Introduction

Women's progress in the labor market has been dramatic since the 1960s. The female employment rate has risen, the pay gap with men has declined, and occupational segregation has decreased. Despite all this progress, it is striking that female convergence has slowed and possibly stopped since about the turn of the millennium, while sizeable gaps remain in pay and hours. Figure 1 tracks the share of males in the occupations in which women work in the US over time. It shows that substantial differences remain between the jobs done by women and men. One particular concern is that females are still under-represented in many high paying professional and managerial occupations (see Figure 2 and Goldin, 2014). Although a lot of the gender wage gap is within occupations, the lack of women in these high-paying, male dominated professions contributes to the gap (Bayard et al., 2003; Blau and Kahn, 2016). For example, in 2014, the average hourly wage of individuals in the US who work in majority male occupations (proportion of males >=0.70) was $23.67, versus $19.30 for those in minority male occupations (proportion of males <=0.30).[1] And understanding occupational segregation may help us better understand the pay gap within occupations as well. It is also important to inform policies which are designed to help women in the labor market.

The traditional explanations for these wage gaps are discrimination, labor supply, and human capital investments (Altonji and Blank, 1999). The role of these factors has doubtlessly declined massively over the past half-century: equal pay legislation has eroded much overt discrimination, women's participation and experience in the labor market has much increased, and women are now more educated than men. More recently, the literature has turned towards the role of attitudes, personality traits, and gender identity as possible explanations for different labor market choices and outcomes of men and women (see Croson and Gneezy, 2009 and Bertrand, 2010). However, the role of many of the variables suggested as explanations for lower female earnings remains empirically elusive (Manning and Swaffield, 2008; Fortin, 2008).

1. Based on the 2014 Current Population Survey (CPS) merged outgoing rotation group data.

The predominant view among economists seems to be that the main remaining obstacle to more equal labor market outcomes between the sexes is a lack of flexibility to combine a career and family. Goldin (2014) argues this point most forcefully but it is also shared by Bertrand (2018).[2] Kleven, Landais, and Sogaard (2018) provide a powerful demonstration of the continuing sharp decline in wages and earnings once a woman has children in Denmark, a country with a long history of comparatively equal gender attitudes.

But there are other intriguing empirical regularities, which suggest that factors beyond the flexibility story matter as well. Some of these suggest that women and men may value the content of jobs differently. For example, women do not necessarily gravitate towards the most flexible occupations and sometimes seem to do the exact opposite. Goldin (2014) presents evidence for full-time college graduate workers in 95 high paying occupations. One of her metrics 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. She demonstrates that less flexible occupations have a larger pay gap. Goldin (2014) classifies these occupations into five groups: health, business, tech, science, and other. Business occupations are the least flexible group with an average elasticity of 0.93 but women's share in this group is about the same as their overall representation in all these occupations, around 40%. On the other hand, women make up only 20% of workers in the much more flexible tech group (with an elasticity of 0.47). Across all the 95 occupations, the share of men (SOM) in an occupation is basically uncorrelated with the earnings-hours elasticity.[3] Goldin (2014) shows that the lack of flexibility is related to the amount of contact with others and the importance of building relationships in a job: where workers have to communicate with co-workers or clients both parties have to be present at the same time, limiting flexibility. Our conjecture is that women may actually value jobs which incorporate some interpersonal elements over purely abstract tasks (and it seems Claudia Goldin would agree with this idea, see EPL Cornell, 2014, 1:21:53-1:23:35).

[2] In the social sciences more broadly, Hochschild (1989) is an early advocate of this view. See also Cortés and Pan (2016).

[3] These results are from our own calculations based on the data posted with Goldin's (2014) article using file AllOccsWageGaps.xlsx, sheet FullBA, EducTime plus Hours.

Another piece of puzzling evidence is the sorting of women into subfields within occupations. A familiar example is the female share in the research fields of academic economists. Dolado, Felgueroso, and Almunia (2012) present breakdowns into different sets of fields. Most striking is their Figure 2, which classifies economists in top 50 departments into 34 fields. The top five fields by female share are wages/income distribution, economics of education, health care/demographics/social security, labor markets/unemployment, and social choice/allocative efficiency/public goods. The bottom five are agricultural economics, business cycles, general equilibrium/cooperative games, alternative approaches/comparative systems, and corporate finance. Similarly, Chari and Goldsmith-Pinkham (2017) find that the representation of women on the program of the NBER Summer Institute has been highest in the children, health economics, health care, crime, and price dynamics workshops, and lowest in the impulse and propagation mechanisms, asset pricing, forecasting and empirical methods, dynamic equilibrium models, and monetary economics ones. It is difficult not to notice an applied/theory division and a pattern in how directly the work in these fields relates to the welfare of individuals. On the other hand, it is difficult to think of other differences in job attributes across sub-specialties in academia. In particular, there tends to be no variation in arrangements for part-time work or career interruptions.

In this paper, we explore whether preferences for the content and context of the work done in particular jobs might explain some of the occupational segregation we see in the labor market. We link job satisfaction to attributes of the work done in various occupations. We parsimoniously summarize occupational content into three latent factors, distilled from descriptions in the ONET database, which we label 'people,' 'brains,' and 'brawn.' These occupational content measures matter for both male and female job satisfaction. Both men and women value 'people' and 'brain' jobs but 'people' matter more to women than to men. Conversely, female job satisfaction is lower for 'brawn,' while males care less about 'brawn' content. The content measures are also important in explaining job choices, in line with related results by Cortés and Pan (2017). We document these relationships using panel data for the US, Britain, and Russia. We complement this analysis with cross-sectional data from the British Workplace Employment Relations Study (WERS), which lets us control for firm effects instead. We find similar patterns with respect to the occupation attributes as before. This suggests that our job attributes are not simply proxies for other factors, which might differ across firms, like work environments more or less conducive to one of the sexes.

While these results point towards a role for preferences, it is also possible that the job attributes simply correlate with the constraints individuals face. In order to probe more directly why individuals make the choices they do, we conducted a discrete choice experiment with high school students. We asked the students to choose between six pairs of occupations. We chose the occupations in each pair in order to provide variation particularly along the 'people' and 'brawn' dimensions. The choices also highlight a preference for 'people' among the students. In order to pinpoint what drives these differences we asked the respondents to explain why they made the particular choices they did. The vast majority of answers indicate that students prefer the activities in one of the jobs, or that their abilities are a better match. None of our respondents mentions work hours, flexibility, or the opportunity to combine a career and family as a factor.

Our findings are suggestive of a role for preferences but we cannot fully rule out alternative explanations. In particular, we do not profess to be able to separate preferences and skills, both of which the students mention. A related literature has focused on biological differences in skills between men and women. Baker and Cornelson (2016) link the share of men in an occupation to DOT codes that capture the sensory, motor, and spatial skills required in that occupation. They find that occupational segregation would have been about 25% lower if these skills did not vary by gender but that the skills did not play a role in the narrowing of the occupation gap during the past 40 years.[4] We suspect that their skills pick up some variation that is similar to our 'people,' 'brains,' and 'brawn' factors; however, Baker and Cornelson (2016) do not relate their skills to job satisfaction.[5]

A large literature in psychology has been classifying occupational attributes, the vocational intersts of individuals, and investigated differences between men and women. This work is largely based on the vocational interest model developed by John Holland (1959, 1992), who classified individual's interests into the six types: realistic, investigative, artistic, social, enterprising, and conventional. Prediger (1982) simplified this to the two polar dimensions things (realistic) versus people (social) and data (combining enterprising and conventional) versus ideas (investigative and artistic). While our classification of the ONET context variables mirrors Prediger's people—things dimension, we find only a single (important) factor related

[4] Beaudry and Lewis (2014) using a different approach, which does not involve occupations, find a large role for male-female skill differences in the narrowing of gender wage gap.
[5] Weinberg (2000) is an earlier analysis along these lines and Fortin (2008) also uses skill measures.

to analytical work (our ‘brain’ factor), rather than his data—ideas distinction. This literature has persistently pointed out important sex differences in the people versus things dimension. Su, Rounds, and Armstrong (2009) is a recent analysis and overview. However, the idea that men and women bifurcate on the people—things dimension predates the work of Holland and Prediger, see Thorndike (1911) and Strong (1943). This psychology literature has mostly documented the fact that the interests of men and women differ. Morris (2003) complements this work by showing that a greater congruence between individual interests and job content is correlated with job satisfaction and retention in the job, consistent with our argument.

The sociologist Catherine Hakim (2000) and the psychologist Susan Pinker (2008) have gone further and pushed the idea that these differences in preferences of women and men are a primary driver of the persistent differences in labor market choices. Hakim’s interest is in women’s attitudes towards a role as homemaker, a full-time labor market career, or a combination of family and work. While Hakim offers quantitative evidence using similar variables as we do, occupational choice plays a minor role in her account—it matters primarily to the degree that some occupations are more likely to offer part-time work or accommodate less committed careers. Pinker’s (2008) work is closer to our idea that women may not like the nature of male dominated jobs, and supports a division along the people—things dimension, but only contains a narrative analysis.

A closely related, concurrent analysis to ours is Cortés and Pan (2017). Their paper discusses a wider range of explanations for occupational segregation of men and women but they consider very similar ONET variables as we do. Fortin (2008) uses a narrower set of survey based variables related to skills and preferences in wage regressions. She shows that they do not explain any of the gender wage gap but does not analyze occupational choice. Also related is Usui (2008), who uses the National Longitudinal Survey of Youth 1979 (NLSY79) from 1979-1982 and shows that women are less satisfied in male dominated jobs**.** Hunt (2016) demonstrates that female college graduates in the US are more likely than males to leave engineering jobs but shows that this is mostly due to the fact that women are more likely to leave male dominated occupations in general.

Analytical Framework and Methods of Analysis

We are interested in an individual's preferences for the content of the work they do on their job, whether these preferences differ between men and women, and whether such differences explain differences in occupational choices. The economics literature on sex differences in preferences has typically focused on three main aspects: risk attitudes, attitudes towards competition and bargaining, and social preferences; the surveys of both Croson and Gneezy (2009) and Bertrand (2010) are organized around these themes. These are certainly all related to preferences but our interest is more squarely on preferences for the *activities, content, and context of the work* an individual does. Risk preferences and attitudes towards competition probably play some role in choices of a particular occupation and job but they are doubtlessly only a small slice of the relevant factors individuals consider when choosing an occupation or field. Social preferences, i.e. attitudes such as altruism, trust, or reciprocity, seem more directly connected to the content of work. Such preferences may lead individuals to choose jobs more oriented towards helping others, for example. The findings in the literature are not clear-cut. Croson and Gneezy (2009) conclude that there is little evidence that women are more socially oriented than men, while Bertrand (2010) is willing to interpret the existing literature slightly more favorable towards the view that women are more altruistic.

The preferences we are interested in are both broader and narrower than often defined in the literature. We would like to know why female academics are more likely to be found in the life sciences than the physical sciences, or why women are more likely to work as financial analyst than electrical drafters. Our hunch is that these disparate choices have something to do with the content and context of the work in these occupations, rather than with the more extraneous factors related to the organization of work. For example, Wiswall and Zafar (2016) also talk about preferences for job attributes but focus exclusively on aspects like hours, the availability of part-time work, layoff probabilities, the structure of earnings, and career progression. We think of these organizational factors as related to the constraints an individual faces (for example, needing flexibility to care for children) rather than their preferences for a certain type of work.

To set the stage for our investigation, suppose utility is given by U(C, JC, OTH) where C is consumption, JC is (for simplicity) a unidimensional aspect capturing the content of the work or "job content," and OTH stands for other aspects of the job. A job amenity like JC is typically valued by computing the marginal rate of substitution (dU/dJC)/(dU/dC); how many units of consumption is the individual willing to give up to gain one more unit of the job attribute JC? Our conjecture is that (dU/dJC)/(dU/dC) may differ for men and women, and these differences influence the choices of jobs by gender.

How would we assess this? The economics literature uses three main methods to study preferences: studying choices, asking individuals directly about their preferences, and estimating satisfaction equations. We will discuss these three methods in turn.

Studying choices: The most standard method to evaluate preference parameters is the study of choices. If women like the attribute JC more than men we should see more women in high JC jobs. We can evaluate this by regressing individual job choices or the share of men (SOM) in an occupation on attributes including JC. There are two obvious complications with this approach. The first is the fact that the list of relevant job attributes may be long, and many of these attributes might be unobservable. If any omitted attributes are correlated with JC, we might get the estimate wrong. The second complication is that choices are not determined solely by preferences but by the interaction between preferences and constraints. It may be the differences in constraints, which give rise to different choices of men and women, not differences in preferences.

One way to address these issues is not to rely on real choices but rather present individuals with hypothetical choices in a survey. The options given to individuals in such a setting (called vignettes) are often designed specifically to vary one particular aspect of a job in a set of choices (or one aspect in addition to the relevant price) in order to minimize the risk of omitted variable bias. This methodology (often called choice experiments, although typically there is no experimental manipulation as all subjects are given the same choices) has various advantages. Individuals can be confronted with choices from many sets, which produces individual level panel data and the attributes presented can be chosen so as to create a large amount of relevant variation. This circumvents many of the problems associated with actual choices.

Examples of choice experiments are Wiswall and Zafar (2016), who presented university students with hypothetical vignettes, and Mas and Pallais (2017), who varied job attributes in a field setting with actual online job applicants. The advantages of hypothetical choice experiments are balanced by fact that the cost of these choices do not have real consequences for individuals. Field settings avoid this but are often forced to focus on narrow groups of respondents. For example, Mas and Pallais (2017) study applicants for call center jobs. And just like real choices, it is unlikely that this methodology is able to distinguish cleanly between preferences and constraints either.

Asking individuals about their preferences: An alternative to studying choices is to simply ask people directly about their preferences. This methodology has been widely used in settings where valuations are not priced directly by markets, like environmental policy, so that looking at choices is simply not feasible. Such contingent valuation methods have been criticized because individuals tend to find it difficult to think about hypothetical choices in areas they are not typically faced with, and as a result give inconsistent responses (see e.g. Diamond and Hausman, 1994). This should be less of an issue in a job choice context. Working individuals have faced the relevant choices and are at least familiar with a subset of them. We will ask high school students about their preferences for different occupations. Although this group has no direct experience with these jobs yet, the students are beginning to think actively about their future in terms of subject choice which might lead them to future careers.

Estimating satisfaction equations: An alternative approach is to interpret survey measures of satisfaction (with the job or with life) as measures of U(.), estimate such a satisfaction equation, and treat the estimates as preference parameters. If one of the arguments in the satisfaction equation is income or consumption, the estimates can again be used to calculate a willingness to pay, (dU/dJC)/(dU/dC). Frijters and van Praag (1998) have applied this idea to valuing climate and van Praag and Baarsma (2005) to value airport noise. Finkelstein, Luttmer, and Notowidigdo (2013) use a similar idea to estimate marginal utilities like dU/dJC directly.

Estimating satisfaction equations suffers from the same problem as studying choices: the included job attributes might proxy for omitted ones. One advantage over studying choices is that variation in job attributes which comes about because different individuals face different constraints (or prices), should still lead to valid inferences. As long as variation in constraints move an individual along a single indifference curve, the individual should report the same satisfaction level.

However, reported job satisfaction may not be the same as choice utility, in which case estimating satisfaction equations will not give the same result as evaluating choices. Kimball and Willis (2006) and Benjamin et al. (2012) propose a utility function which, adapted to our setting, would have the form U(C, JC, S(JC)), where S(.) is the job satisfaction function. For simplicity, here we ignore aspects of the job other than JC. JC matters for job satisfaction, and job satisfaction matters for utility relevant for decision making. But JC also enters utility directly. For example, JC may affect the happiness of one's family if a person's feelings about their work spills over to the home environment. This is not part of job satisfaction but is something individuals will take into account in their choices. As a result

$$\frac{dU}{dJC} = \frac{\partial U}{\partial JC} + \frac{\partial U}{\partial S}\left(\frac{dS}{dJC}\right). \tag{1}$$

This framework highlights that preferences of men and women may differ because of differences in either dS/dJC, ∂U/∂S, or ∂U/∂JC. Estimating satisfaction equations at best yields information on the term dS/dJC. But differences in dS/dJC between men and women, which would be revealed by satisfaction equations, likely imply differences in dU/dJC (since it seems a knife edge case that sex differences in ∂U/∂S and ∂U/∂JC would exactly offset those in dS/dJC). Of course, the direction and magnitude of the sex differences in dS/dJC and dU/dJC may possibly differ.

Benjamin et al. (2012) compare vignette based choices from a variety of diverse scenarios with rankings based on subjective well-being (SWB) measures. Benjamin et al. (2014) make similar comparisons between real choices in the medical Resident Matching Program and SWB measures related to the options. In both studies, there is a fair alignment between choices and SWB ranking but there are also some systematic deviations. In Benjamin et al. (2012), the differences in rankings are related to other life domains, like control over one's life and a sense of purpose. Various choice scenarios in their paper are work related, and they find a large role for the term (∂U/∂S)(dS/dJC) in choices. Not everything in these papers is comparable to the setup of research based on satisfaction equations but the results suggest that satisfaction equations will contain some useful information content as well as highlighting limitations. Comforting for our purpose, they find no systematic differences in the way choices versus SWB rankings differ for men and women. Any differences we find should therefore reflect real differences in preferences rather than, for example, different uses of satisfaction scales across sexes.

The previous discussion highlights the drawbacks associated with each of these methods. As a result, neither method is likely to give a definite answer to the question whether preferences play a role in the diverging occupational choices of men and women. Here we combine elements of all of these approaches. We start with simple satisfaction equations, relating job satisfaction to a variety of occupation characteristics. In particular, we construct measures to reflect the activities of an occupation, which we obtain from the ONET database. We show that job satisfaction varies with job content, and job content matters more for women than for men. We also show that the job content measures matter in choice equations (either in cross-sectional sorting equations for the SOM in an occupation or in longitudinal equations for job mobility).

Preferences for the content of a job are one possible explanation for our results but we acknowledge that there could be others, for example flexibility. In order to probe the role of preferences in job choices further we conducted a choice experiment with high school students. We asked the students to make choices between six paired occupations, where the pairs are matched on average income and hours but distinct in terms of work content. We then asked the respondents *why* they made their choices, not a common methodology in economics. The results corroborate our view that preferences for the content of work play an important role in occupational choices.

There are some remaining caveats to our interpretation. Closely related to preferences are talents. In fact, the surveyed students tell us that being good at the skills required in an occupation is an important reason for their choices, in addition to liking or disliking the activities in the job. We make no claim that we can separate talents from preferences. In fact, it is not clear whether such a separation is possible or even desirable. There is doubtlessly a lot of feedback throughout childhood and early adulthood between talents and preferences: we gain practice by engaging in the activities we like, and we enjoy the activities at which we excel.

Similarly, preferences are difficult to distinguish from norms or gender identities, which might prescribe that women should take on certain roles. Our results would likely look very similar if women felt that they *should* be doing certain jobs, gravitated towards those jobs, and felt dissatisfied or switched jobs if they violate the prevailing norms. There is potentially a large role for socialization in childhood for the formation of both own preferences and for gender conformity. All of the approaches we have outlined would have a difficult time distinguishing between societal norms and personal preferences. One advantage of our interviews of high school students is that they have not started their working lives yet, and their views are at least unlikely to be tainted by ex-post rationalizations of choices, which might have been driven by personal constraints or societal pressures rather than their own preferences.

Measuring Job Content

To measure job content we use ONET version 5, which provides a diverse set of information on occupational attributes, requirements, and characteristics of the workers in an occupation; all in all, it offers about 249 distinct items. Out of these, we use the 79 items describing the work activities and context of a person's occupation. For each individual item, an incumbent in randomly selected firms reports a level from 1 to 7. For example, in activities, an item might describe to which degree an occupation involves 'assisting and caring for others,' 'analyzing data or information,' or the 'repairing and maintaining of mechanical equipment.' Examples for context are the level of 'contact with others,' 'the importance of being exact or accurate,' and 'being exposed to hazardous conditions' (see Appendix A Table A.1 for all attributes). We standardize each of these variables to have a mean of 0 and a standard deviation of 1. These variables are then matched to country specific survey data as described in the Data Section below.

Rather than add the 79 context and activities variables to our regressions directly and risk over-fitting, we follow the psychometric literature and use exploratory factor analysis to reduce the dimensionality first (Gorsuch, 1983; Thompson, 2004). A clear structure of three latent factors emerges in the first rotation. We then follow Heckman et al. (2012) and drop ONET items that are weakly associated with the factors or those that are associated with more than one factor.[6] We then repeat the factor analysis using the remaining ONET items and extract the final latent variables, which we allow to be freely correlated. We loosely label the three factors as 'people,' 'brains,' and 'brawn,' or PBB. These labels appear natural to us based on the items that load on each factor (see Appendix A Tables A.1 and A.2). An alternative would have been to call them factor '1,' '2,' and '3' rather than choosing labels with strong connotations but we feel it would be difficult for readers to recall which factors we were talking about. We note that we have explored variations of how to extract these factors. For all three countries, our analysis

[6] For the former, we remove items with a loading of 0.4 or less. For the latter we remove items that have a loading that is greater than 0.4 on more than one factor.

suggested three main factors and results are robust to exactly how we obtained these factors (for fuller details see Appendix A).

Our approach differs, for example, from that taken by Beaudry and Lewis (2014), who use the DOT (the predecessor to ONET) and manually pick groups of attributes they view as being associated with physical, cognitive, and people skills in an occupation. We rely on a more mechanical method to reduce the dimensionality of the data to avoid handpicking occupational attributes, which may or may not fit our prejudices. Nonetheless, we arrive at a roughly similar classification. Table 1 lists the top and bottom ten occupations for each of the three factors. In addition, Table 2 documents the scores for a number of occupations, which we find useful for thinking about occupational segregation, together with the share of men in 1930 and now. The factors have a mean 0 and standard deviation 1, so electricians, for example, score about one standard deviation below the mean on 'people,' a quarter of a standard deviation above the mean on 'brains,' and a two standard deviations above the mean on 'things.' We note that women have always been dominant in occupations which score high on 'people' (e.g. teaching, nursing, and social work), and underrepresented in occupations high on 'brawn.' This latter factor does not simply capture physical strength but occupational content related to making and manipulating things as well (see Table 2). The 'brawn' factor is strongly associated with traditional blue-collar occupations but also with engineering fields and plays an intermediate role in other occupations like nursing. As expected, professional and technical jobs tend to be associated with positive 'brains.' The most cerebral occupations are the hard sciences, engineering, and mathematics but also financial managers. This same group of occupations often tends to be characterized by having low 'people' content (see Tables 1 and 2).

Empirical Analysis

We wish to investigate how job content – measured by our three PPB factors – affects job choice by gender. Our starting point in this regard is a linear regression for job satisfaction or job mobility of the form

$$Y_{ijt} = JC_j\delta' + X_j\beta' + X_{ijt}\gamma' + \mu_t + \omega_a + \varepsilon_{ijt} \qquad (2)$$

where Y_{ijt} is either job satisfaction or a binary variable which indicates whether a person stayed in the same occupation in the next period for individual *i* in occupation *j* and year *t*. JC_j refers to the 'people,' 'brains,' or 'brawn' content of the occupation, X_j is a vector of other occupational averages, X_{ijt} is a vector of individual-level control variables, μ_t are wave effects, and ω_a are region effects.[7] In our baseline specifications, X_j contains average wages, hours, age, and the proportion college graduates, while X_{ijt} contains age and age squared.

To understand differences by gender, we present estimates separately for males and females. The coefficients of interest in equation (2) are δ. For example, in the stayer regressions, a positive 'people' coefficient implies that higher levels of 'people' content are associated with an increased tendency to stay in an occupation. For the job satisfaction regressions, a positive coefficient implies that a high 'people' content is associated with higher levels of job satisfaction. To make the interpretation of δs more intuitive in the job satisfaction regressions (given that the job satisfaction scales differ across country) we follow van Praag and Ferrer-i-Carbonell (2008) and normalize the job satisfaction variables by using the fitted values from an ordered probit on the raw sample fractions. Since we also standardize the job content variables, our estimates have the interpretation of effect sizes.

An important issue in interpreting the results from a regression like (2) is how workers sort into heterogeneous occupations. The standard compensating differentials framework suggests that workers pick among packages of wages and job attributes while employers offer such packages in order to attract workers. To the degree that workers differ, they will sort into the type of jobs they prefer in equilibrium. Wages adjust to eliminate any excess supplies and demands, so that occupation wage differentials reflect the compensating differentials required by marginal workers who are indifferent between two alternative jobs. This framework predicts that men and women may end up working in different jobs in equilibrium if they have different preferences for job attributes or if they face different constraints (say in terms of hours choices or flexible schedules an occupation offers). In this scenario, it is unlikely that job satisfaction will reflect preferences. One reason is that most of the variation in (2) is cross-sectional, and it is unclear whether the answers to job satisfaction questions are comparable across individuals. Another reason is that in the competitive compensating differentials model

[7] For the BHPS this amounts to the inclusion of 19 fixed effects representing the following regions: inner London, outer London, rest of the South East, South West, East Anglia, East Midlands, West Midlands Conurbation, Rest of the West Midlands, Greater Manchester, Merseyside, Rest of the North West, South Yorkshire, West Yorkshire, Rest of Yorks and Humberside, Tyne and Wear, Rest of the North, Wales, Scotland and Northern Ireland. For the United States, regions are the North East, North Central, South or West. For Russia we include eight individual residential site indicators.

everybody works in their most preferred occupation, given equilibrium wages, and hence should report their maximum job satisfaction attainable.

Therefore, we add individual fixed effects to equation (2) which amounts to identifying the effect of job attributes from occupation switchers, while controlling for time invariant individual differences.

$$Y_{ijt} = \alpha_i + JC_j\delta' + X_j\beta' + X_{ijt}\gamma' + \mu_t + \omega_a + \varepsilon_{ijt} \quad (3)$$

Our baseline results start with estimates of equation (3), however estimates for equation (2) are documented in Appendix B Tables B.1 and B.2. We calculate standard errors using two-way clustering by individual and occupation.[8]

The frictionless, full information framework underlying the standard compensating differentials model is unlikely to be a good representation of actual labor markets, where individuals often make choices subject to constraints, imperfect information regarding what an occupation's content is in practice, and other frictions. Occupations are also bundles of attributes but not all possible combinations may be on offer fitting all individual tastes. Modelling occupational choices and wage differentials in a framework with frictions can lead to very different equilibrium outcomes (see Hwang, Mortensen, and Reed, 1998; Manning, 2003; and Lang and Majumdar, 2004). One implication is that wages no longer reflect compensating differentials. Rather, employers with wage setting power will use wage-amenity packages to attract workers, and wages and amenities may be positively correlated in equilibrium.

Importantly, in a setting with frictions, workers may end up in jobs other than their preferred one, but they will switch jobs in future periods in search of better matches. This "frictional disequilibrium" constitutes a natural source for interpreting the results from a job satisfaction equations like (3). As there are good jobs and bad jobs, as well as high and low quality job matches for particular individuals in this framework, the coefficients on occupation characteristics have a more natural interpretation as individual preferences for these

[8] See Cameron, Gelbach, and Miller (2011). Practically this is implemented using ivreg2 and xtivreg2 as appropriate in Stata.

characteristics. Frictions also offer a natural point of departure for interpreting the stayer regressions, as there is no reason for systematic job changes in the frictionless model. This discussion reinforces that within person comparisons over time as in eq. (3) are more meaningful than cross-sectional ones in eq. (2).

Of course, even in the framework with frictions individuals are not randomly assigned to occupations. This gives rise to two complications. One is the possibility of reverse causality: the choices women and men make may influence the way the work in an occupation is structured. For example, Chang (2018) points out that the share of female computer programmers used to be higher in the 1970s than it is now. Programming also used to be organized in a more interactive fashion then. This could be due to the fact that there were enough women in the occupations so that they were able to structure their work environment to suit their own preferences. Once men dominated the profession, work organization changed to a more solitary model.

The second complication with the types of regression strategies we are employing relates to the problem that the ONET variables we are using may proxy for other relevant aspects of the occupations, as we discussed above. In order to get at the most important ones of these, we control for average wages, hours, age, and the proportion college graduates in an occupation, which are all important factors in the job satisfaction and stayer equations. But we note that the SOM in an occupation is likely to affect variables like wages and hours worked as well, so that these attbributes become endogenous. While the controls we use don't vary at the individual level (except for age), the variation in job content we are interested in is an occupation level variable, and we would expect that the bad controls issue to spill over to the occupation level when the SOM varies across occupations. Like everybody else in the literature on sex differences, we have no solution to offer to this problem.

Another issue in evaluating the valuation of job attributes is that individuals face both a set of jobs with different attributes but also an outside option of not working. For a non-employed individual we have no information on job satisfaction. We may not see an individual working if a particular job attribute is very important to them (for example, enough flexibility to be able to care for children or other household members) and employers may not provide certain amenities because there is no interior market equilibrium where such trade takes place. As a result, those individuals for whom we see job satisfaction may not value an underprovided amenity as much or at all. This selection problem, similar to the problem of estimating wage equations in the presence of employment participation, may distort estimates relating satisfaction to amenities in the sample of working individuals. While we do not address the selection into employment directly, we note that it will likely bias the coefficient estimates on the PBB factors towards zero if the non-employment option offers a better amenity package than the available jobs. The same selection issue also affects the study of observed choices (as we observe no occupation for individuals who do not work) but survey based methods like choice experiments allow us to elicit responses, which are not subject to this problem.

Many analysts favor explanations of the gender wage gap and occupational choice centered on fertility and children. Children may be an important constraint on female choices in the labor market, which swamps other considerations. One worry might be that our job content effects spuriously reflect these constraints. In this case, we should see females without children making less constrained choices. If it is children driving the choices, females without children should look more like males. If female preferences drive the choices, then females with and without children should look more alike as compared to males. We therefore show results for females without children as well.

It is typical in the evaluation of job attributes to measure marginal rates of substitution, i.e. (dU/dJC)/(dU/dC). Instead, we simply look at the coefficients of job attributes in the satisfaction equations directly, i.e. dU/dJC. There are a number of reasons for this. First of all, we estimate simple linear satisfaction equations. With a linear income term, the implied MRS is constant. Of course, we could add non-linear terms of income to the regressions or use a more structural utility framework but we are worried that there is not enough information in the job satisfaction measure, which is measured coarsely in the surveys we use (on a 4 to 7 point scale), and the same is true for our binary mobility equations. We don't believe that these data are particularly well suited to estimate the marginal utility of income well (but see Finkelstein, Luttmer, and Notowididgo, 2013, for an alternative view), and we worry that poor estimates of dU/dC might cloud our results. The costs of this are that our estimates do not have a simple numerical interpretation. We are willing to live with this drawback, as our main interest is the contrast between estimates for females and males.

Another reason why we are hesitant to rely on income estimates is the fact that we include various human capital variables like education and age among the occupational averages X_j.

These variables will capture a lot of the permanent income components, and the interpretation of the coefficients on average earnings in the occupation or own earnings of the respondent becomes much more dubious. Average age and education of an occupation are important correlates of job satisfaction, presumably because more educated and experienced workers get paid more but also because they often get to work in more interesting jobs. Finally, even leaving this last issue aside, Benjamin et al. (2012) find that income coefficients are typically underestimated in satisfaction equations compared to the role of income in choice.

Data

We analyze four public use datasets in addition to collecting our own survey data. Our main analysis uses the US National Longitudinal Survey of Youths 1979 (NLSY79), the British Household Panel Study (BHPS) and the Russian Longitudinal Monitoring Survey (RLMS). We supplement this with results from the British Workplace Employment Relations Study (WERS). We obtain information on the job content of different occupations from the US ONET database.

US NLSY79

We use the NLSY79, a panel of 12,686 individuals who were between 14 and 22 years old when first surveyed in 1979. These individuals were interviewed annually through 1994 and then on a biennial basis. The NLSY79 sample spans 1979 to 2014.

The question on job satisfaction was asked in every wave. Specifically, respondents were asked, “How do you feel about the job you have now?” and were given the following response option: ‘I like it very much,’ ‘I like it fairly well,’ ‘I dislike it somewhat,’ ‘I dislike it very much.’ We coded responses so that higher values represent higher satisfaction. Our analysis uses an unbalanced panel of employees who responded to this job satisfaction question. The NLSY79 uses the US Census Bureau occupation definitions. Specifically, the 1982-2000 and 2002-2014 waves use the 1980 and 2000 codes, respectively. Our analysis sample spans the years 1982 to 2014.

We create an additional dependent variables that captures movements in the labor market.[9] This variable is defined equal to 1 if a person has the same three digit occupation code in year t+2 compared to the occupation that they held in t. Conversely, the variable is defined equal to 0 if an individual has a different occupation code in t+2 or has left employment. We call this variable 'stayers.' The variable is defined on a biennial basis given the interview schedule of the NLSY79 post 1994.[10]

We pool the 1980, 1990, and 2000 Census IPUMS and 2001-2014 American Community Survey (ACS) samples to create averages of an hourly wage, weekly hours, the proportion college graduates, and age in each occupation. We also create a variable for the share of males (SOM) in an occupation by dividing the number of men working in the occupation by all employees. In order to link the Census/ACS and ONET variables to the NLSY we utilize a crosswalk provided by the Bureau of Labor Statistics to assign a Census 2000 occupation code to each occupation in the ONET file. We then rely on the crosswalks from Autor and Dorn (2013) and Dorn (2009) in order to create a consistent set of occupations across the 1980, 1990, 2000, 2002, and 2010 Census codes, which can be linked to the codes used in the NLSY. We calculate occupation averages for this consistent set of occupations.

The occupation distribution in the Census/ACS implicitly provides weights for the factor analysis where we derive the PBB factors. Similar to the Census/ACS averages, we match the PBB variables to the NLSY based on the consistent occupation codes from the crosswalks (for full details of the cross-walking and matching, see Appendix E). We use sampling weights in the analysis that reflect that the NLSY79 oversampled blacks, Hispanics, and the economically disadvantaged (see Appendix C for the unweighted results).

British Household Panel Survey (BHPS)

We use all 18 waves of the original sample of the British Household Panel Survey (BHPS), a longitudinal study of around 5,500 households and over 10,000 individuals in England, Wales and Scotland that began in 1991. This main sample was supplemented in later years with a Welsh extension from 1999 (about 1500 households), a Scottish extension from 1999 and a

[9] Give that this outcome relies on comparing occupation codes across periods, this analysis omits the year 2000 from the analysis given the change in occupation coding.

[10] We utilize the 1980 wave of the NLSY to create the stayers variable for 1982, so the stayers sample starts in 1982 comparable to the one for the job satisfaction regressions.

Northern Ireland extension from 2001 (about 1900 households). We present unweighted results from the unbalanced panel of all individuals including the extensions between 1991 and 2008.[11]

The BHPS contains a number of different job satisfaction questions, which are available for the full 18 waves. We use the two questions asking respondents how satisfied or dissatisfied they are with i) their current job overall and ii) the actual work itself; we present some additional results on satisfaction with other job domains in Appendix B Table B.3. Answers are on a 7-point scale. We again create an additional binary dependent variable that captures whether a person stayed in the same occupation. We measure mobility in the BHPS between two consecutive years.[12]

The BHPS uses occupation codes based on the Standard Occupational Classification 1990 (SOC90) up to 2001; in 2002 this was replaced with SOC 2000 (SOC00). We calculate occupation averages in a three-digit occupation using the 1993-2008 Quarterly Labor Force Survey (QLFS). The QLFS is the main survey of individual economic activity in the Britain, and provides the official measure of the national unemployment rate. It uses SOC90 codes from 1993 through 2000 and SOC00 from 2001. Thus, we first assign to each SOC90 code a SOC00 value based on a crosswalk created from the BHPS.[13] We next calculate the same occupation averages and variables as for the NLSY based on the SOC00 codes. We then match the occupation averages to the BHPS data.

To create the three PBB factors for the British analysis we match the occupation codes in the ONET data directly to the British SOC00 in the QLFS data using a crosswalk provided by Anna Salomons.[14] We then proceed to extract the underlying latent factors (see Appendix A). These differ only from the US analysis in the fact that the distribution of workers across occupations is slightly different in Britain. Unsurprisingly, we again obtain three latent factors corresponding to 'people,' 'brains,' and 'brawn' from the QLFS analysis, which we match to the BHPS.

[11] We have investigated the sensitivity of our results to i) unweighted regressions of the original BHPS sample only and ii) weighted regressions of the main BHPS sample. See Appendix C for these results.
[12] This outcome relies on comparing occupation codes across periods, therefore this analysis omits the year 2002 from the analysis given the change in the occupation codes.
[13] See Appendix E
[14] For the years in the QLFS where the UK SOC90 code is used, we use a translation to SOC00 that is implicitly provided by the BHPS. That is, SOC00 appears for the respondent's primary occupation post 2000 and SOC90 appears for all waves of the survey. So we have a translation between the two coding systems.

Russian Longitudinal Monitoring Survey (RLMS)

Our measure of job satisfaction for Russia comes from the Russian Longitudinal Monitoring Survey (RLMS). This is a nationally representative annual survey, with data available from 1994-2012. However, job satisfaction data is only available from 2002-2012. We restrict our sample to employees who answer the question: 'How satisfied or unsatisfied are you with your job in general?' Response options are absolutely satisfied, mostly satisfied, neutral, not very satisfied and absolutely unsatisfied. We code responses so that higher values represent being more satisfied. We create a binary dependent variable that captures whether a person stayed in the same occupation over two consecutive years.

We do not have a large labor force survey that allows us to calculate occupation averages for Russia, like the US CPS or British QLFS. Instead, we rely on pooling the RLMS from 1994-2012 with two other data sources, the International Social Survey Program (ISSP) 1995-2011 and the European Social Survey (ESS) from 2002-2012.[15] We calculate the occupation averages for age, hours, proportion of college graduates, and the SOM in the sample pooled across these three data sources. Only the RLMS reports individual earnings and, as a result, we calculate the average wage from this data source only. Our RLMS regressions use weights that allow for the complex design of the RLMS where many observations are derived from following the housing unit rather than the person, as well as having oversamples from the first wave to allow for forecasted attrition. However, the overall conclusions are not sensitive to weighting, and we show unweighted regressions in Appendix C.

In order to create the three PBB factors for the Russian analysis, we begin again by complementing the RLMS with ISSP and ESS data in order to obtain occupation cells with more observations. Next, we match the ISCO code to the ONET occupation codes using a crosswalk provided by the Bureau of Labor Statistics. The factor analysis again yields the three familiar PBB factors.

British Workplace Employment Relations Study (WERS)

The British Workplace Employment Relations Study (WERS) is a national survey of people at work in Britain, which collects data from employees, employee representatives, and employers

[15] ISSP: http://www.issp.org/page.php?pageId=4. ESS: http://www.europeansocialsurvey.org/

in about 2,500 firms. Multiple employees are interviewed in each firm. The WERS is conducted every six to eight years but is not a panel of firms or workers. We use the 2004 and 2011 surveys, which included an individual's three-digit occupation code using the British SOC00 codes (previous versions did not), along with a series of questions concerning various elements of job satisfaction. We utilize the employee responses to the question about satisfaction with the work itself as there is no overall job satisfaction question. Response options are on a 5-point scale. We calculate the occupation averages from the QLFS and the ONET variables in the same manner as described for the BHPS data and match this to the WERS based on the employee's three-digit occupation code.

Results

We start in Table 3 by presenting a simple linear regression of the SOM on the three latent factors along with the other occupational averages, time dummies, and area dummies. We run this at the individual level but note that this is essentially an occupation level regression and the individuals here only serve to give different weights to different occupations. These regressions use the Census/ACS, QLFS, and RLMS.

Table 3 highlights that there is substantial sorting in all three countries along the dimension of 'people,' 'brains,' and 'brawn.' Women are overrepresented in 'people' jobs, men in 'brawn' jobs, and they roughly share 'brain' jobs. The pattern is stronger in Russia than in the US and Britain but is important in all three countries. The 'brawn' component seems to be the more potent predictor of sorting by gender than the 'people' factor. We suspect that this is due to the role of blue-collar jobs in the occupation distribution at large, few of which appear in Table 2.

In Table 4 we turn to individual fixed effects regressions of job satisfaction and occupational mobility on PBB as in equation (3). In all three countries, women like 'people' and 'brain' jobs and dislike 'brawn' jobs, with the 'brain' coefficient for Russia being an exception. Coefficients for women are generally bigger in absolute value than those for men, suggesting that women may have stronger preferences for these job attributes. In the US, the signs for men and women are similar and only magnitudes differ, while in Britain and Russia, male patterns are sometimes reversed. The stayer regressions tend to match these patterns overall although there are discrepancies for a few coefficients. In general, these results closely mirror the ones we saw for sorting into occupations in Table 3.[16] The most notable pattern is that women seem to prefer 'people' jobs more than men.

In order to get a sense of the magnitudes of these effects, consider forming predicted values by multiplying the PBB coefficients from the NLSY job satisfaction equation with the values of the three factors (but ignoring other occupation averages). The female predicted value for heavily female dominated social work (SOM = 0.25) is 0.14, while for male dominated mechanical engineering (SOM = 0.94) it is 0.04. This reflects the fact that mechanical engineering scores much lower on 'people' and somewhat higher on 'brawn' than social work. Moving between these occupations changes job satisfaction by 0.10 of a standard deviation. For comparison, Stevenson and Wolfers (2008) find that a 30% difference in income is associated with about 0.10 of a standard deviation difference in life satisfaction in within country cross-sectional data.[17] This suggests a sizeable role for job content to us.

For men, the predicted values are 0.06 for social work and 0.04 for mechanical engineering, indicating that men are slightly more satisfied with the social worker bundle of job content as well (since most men dislike the solitary nature of engineering too). The occupations with the most negative predicted values for women are blue collar jobs with values ranging from 0.0 to -0.2. Men dislike these jobs as well but less so than women. The fact that men generally care less about the PBB factors is also reflected in the standard deviation of these predicted values across the entire 310 occupations, which is 0.03 for men and 0.09 for women. But for both sexes the influence of the PBB variables on job satisfaction is sizeable.[18]

Women's satisfaction translates into decisions whether to stay in a job or not as well but the magnitudes are relatively small. The same comparison of the values of PBB implies a 0.3 percentage points higher probability of a woman quitting her career in mechanical engineering as opposed to one in social work.

---

[16] In Appendix B Table B.8 we also show estimates for college educated females. While individual coefficients jump around the general pattern of results is very similar to those in Table 4.

[17] Using their central estimate of 0.3 (Stevenson and Wolfers, 2008, p. 31).

[18] We note that personal income is also more significant in explaining job satisfaction and the propensity to stay for males as compared to females (see Appendix B Tables B.4 through B.6). This suggests that males may be more extrinsically motivated than females. Together with the PBB results, this might explain why females sort more frequently into careers like social work, which are low paid but relatively high on 'people.'

Together Tables 3 and 4 suggest an important role for the PBB variables for satisfaction and job choice. These effects are more important for women than they are for men. Because women strongly shy away from 'brawn' jobs, these jobs are left to fill for men who are less averse to them.

Despite the fact that our regressions control for average hours, it might also be that the PBB variables, and in particular the 'people' factor, do not just proxy for the content of work but pick up constraints women face when juggling child care with working as well. To address this, we also present separate estimates for women with and without children in Table 5. If constraints are purely related to children, then women without children should look more like men. If the PBB variables pick up preferences for the content of work and women's preferences differ from men's then women with and without children should look more alike and differ from men.

For about half the coefficients in Table 5, job satisfaction and retention in the occupations high in the 'people' and 'brain' factors and low in 'brawn' tends to be as strong or stronger for women without children as it is for women with children. In most of the remaining cases, the results for women without children fall inbetween women with children and men. Only three of the coefficients in the table are virtually the same for women without children as they are for men. While the results are far from clear-cut, they are more aligned with the idea that women's preferences for the occupational content factors are stronger for women than men, regardless of whether women have children or not. There is certainly no evidence that women without children look just like men. Because the subsamples tend to get small, we pool both older women, who may have decided not to have children at all, and younger women, who might have children later, which might attenuate the difference with men. Overall, this suggests that the PBB variables indeed relate to the content of work rather than other aspects of the work environment.[19]

[19] In Appendix B Table B.3, we also show similar regressions for satisfaction with hours, pay, and job security. The PBB variables have consistently smaller effects in these regressions than in the one for satisfaction with work itself. We also show results including the wage-hours elasticity used by Goldin (2014) in Table B.7. The PBB coefficients are mostly only slightly attenuated compared to Table 4.

Work environment

The results we have presented so far are consistent with the idea that tastes for the content of work influence occupation choices of women and men. However, the PBB variables are crude measures of work content, and may pick up yet other attributes about the workplace. In particular, they may proxy for environmental or organizational factors, which affect men and women differently.

A lot of aspects related to the work environment might be specific to a workplace and shaped by managers and co-workers, i.e. a firm level characteristic rather than a characteristic of the occupation per se. None of the datasets we have analyzed allows us to incorporate this in our analysis. We therefore turn to the British Workplace Employment Relations Study (WERS). This data set samples workplaces, and within these workplaces surveys managers, worker representatives, and a subsample of employees. The WERS data also contains a measure of satisfaction with work itself, consistent with the BHPS. In our analysis of the WERS data we have to rely on cross-sectional specifications, as we do not observe individuals over time. However, we can include firm fixed effects to capture aspects of the environment that may affect females at work. Therefore, we identify the coefficients on PBB from variation caused by having individuals from multiple occupations working in the same firm.

The WERS estimates are documented in Table 6, where the baseline specification is a simple cross-sectional regression. The pattern of results is very similar to that in Table 4 although coefficients are slightly bigger and the female ‘brawn’ coefficient is small but positive. Including firm fixed effects mostly attenuates the ‘people’ and ‘brawn’ estimates but less so the ‘brains’ coefficient. The basic conclusion that females care more about ‘people’ and ‘brains’ compared to males remains intact.

Schools survey

In order to get at job preferences at an earlier stage in life, and in order to probe for the reasons of individuals’ choices, we conducted our own survey among students in Year 11 (about age 15 – 16) in two secondary schools in Greater London. Both are high performing/relatively advantaged schools (with students going on to university at a rate that puts them in the top third in the country). These students are at an age where they are actively thinking about subject and job choices for the future but will not have engaged in actual work experience. The students completed the surveys in an assembly hall on a day when one of us visited the school. All students who were present on the day participated with nobody choosing to opt out.[20] We received 311 responses and dropped four who provided no sex information. The resulting dataset contains 157 males and 150 females.

The survey presented students with a list of 12 occupations and gave each student six choices among pairs of occupations. We started by splitting occupations into three classes by earnings, and then each of these into occupations with high or low average hours. These matches, particularly on earnings, are relatively coarse in practice. We picked a pair of occupations for each of these groups. Starting with a list of occupations, in which both male and female graduates commonly work, we picked the pairs in order to obtain a large amount of variation in the 'people' and 'brawn' factors within the pair as these are the dimensions where we found the largest differences between females and males; see the Appendix D for more details.

Table 7 lists the six pairs of occupations, together with the average earnings and hours as well as the PBB scores for each occupation. Within each of the six questions, the occupations are listed in the table with the one with the highest SOM in the QLFS first (in the survey, we listed male and female dominated occupations first in three cases each). The last column in Table 7 shows the fraction of males among the students who chose this occupation. With the exception of question 3, the students' choices mimic the sex distribution among actual workers. Some of the sex differences of the choices are sizeable.

In order to relate the six occupational choices to the PBB factors, we treat the resulting data as a set of binary choices from a multinomial list of preferences over a large set of occupations. We show in Appendix D that a standard random utility model gives rise to a simple pooled logit regression for these data. Because the choice is one between a pair of occupations, it is only the relative characteristics of the two occupations that matter. Our covariates are therefore the differences in the occupation specific variables between the first and second occupation in the group, and the dependent variable is 1 if the first occupation is chosen.

Table 8 shows odds-ratios from these logit regressions of the occupational choices on the PBB factors. Both genders prefer 'people' oriented jobs and are relatively indifferent to the 'brain'

[20] Students were advised beforehand they could opt out or choose to passively not answer any or all questions. Ethics approval was received by the authors from their home institution.

and ‘brawn’ aspects of the jobs. Despite the qualitative similarities, females have a stronger people orientation as compared to males. Curiously, in terms of the point estimates, males dislike brawn jobs, while females are indifferent to brawn. However, the male effect is not significant.

We are worried that these choices might be spuriously driven by skills the students possess rather than their preferences for the job content because we only have a small set of occupations. In columns (3) to (6), we therefore control for whether the skills required in the occupation are a particularly good match for the specific talents of the students. We asked students in the survey which subjects they are taking, and which subject is their best one. We combined this information with the fields of study listed by respondents to the American Community Survey from 2009 to 2015 to create measures for the skill match between the best subject of the students and the fields highly represented in the occupation (see Appendix D for more details).

We define two measures of a skill match for a student-occupation pair, a continuous and a discrete one. Columns (3) and (4) in Table 8 show the results adding the continuous skill match measure, and columns (5) and (6) display estimates with the discrete measure. Skills are important in occupational choices for both females and males. Adding the skill match measures lowers the estimates on the ‘people’ factor a bit, raises estimates on the ‘brains’ factor, and further reduces the ‘brawn’ coefficient for males. In fact, in columns (4) and (6), males’ dislike of ‘brawn’ jobs is significant at conventional levels, and larger than their preference for people jobs. This is a consequence of our choice of the twelve occupations we analyze here. If we restrict our QLFS sample to these six occupations, and repeat the sorting regressions from table 3, we also find that men are less likely to choose occupations high on ‘brawn’ in this subsample.[21] So we conclude that the choices of the students along the PBB dimension actuallymatch those of adults fairly closely.

How do the students in our survey make choices? In order to gain more insight on this, we asked them directly: “For each of the six job choices you made, tell us in a few words why you picked the job you did?” The students gave answers in free form, without any prompts. There was a fair amount of coherency in the answers, and we coded the answers by hand into eight

[21] See Appendix D.

categories ourselves as shown in Table 9. In most cases, this was straightforward to do. When respondents indicated more than one reason for their choice we coded the one mentioned first.

Roughly two thirds of responses indicated that they found one of the activities more interesting, or that the job related to some desirable goals like helping people. About another 15% of responses indicated that they felt they better qualified for one of the jobs. Another 5% indicated some other clearly articulated reason, like higher pay or status, that one of the jobs is easier, and a hodgepodge of other things. In no case did any respondent indicate that the work hours or flexibility of the job played any role in their answer. There is little difference between males and females in why they made their choices, and there is little difference across the individual six questions. The answers indicate very clearly that interest in the activity by far dominates the thoughts of the students as to their job choices.

Overall, the results from the survey of school aged youths closely mimics the findings we obtain for actual adult job choices. Our survey of the students suggests that in the vast majority cases these choices are driven by preferences for the content of the jobs and not environmental factors like hours, flexibility, or other job amenities. This reinforces the idea that preferences play an important role in the differences in job choices of men and women.

Discussion

Stigler and Becker (1977) have famously cautioned economists against relying on variation in preferences to explain economic outcomes, suggesting that the most worthwhile focus is on the comparative statics induced by variation in constraints. The literature on differences in labor market outcomes and behaviors between men and women has indeed for a long time adopted this approach, and studied the impact of discrimination, human capital investments, and labor supply. Less than two decades ago, Altonji and Blank (1999) devoted two paragraphs of their handbook chapter on race and gender to differences in preferences before moving on to the traditional constraint based explanations.

But stubborn differences in male and female pay and occupational segregation persist while many of the constraints faced by women in the workplace seem to have diminished (which does not mean that these constraints are all gone). At the same time, economists have grown more relaxed about thinking about differences in tastes. The handbook chapter by Bertrand (2010), a mere ten years after Altonji and Blank, focuses almost entirely on explanations based on differences in psychological traits between men and women, as well as gender identity. A powerful form in which such psychological differences manifest themselves is in different tastes of men and women for the content of the work they do. We argue that economists should be open-minded that this may help explain occupational sorting, and subject this idea to scrutiny.

Here we have offered an initial attempt at this by analyzing the differences in job satisfaction of women in jobs which we characterize by their 'people,' 'brain,' and 'brawn' content. We find that women care more about these job characteristics than men. The same job content measures predict retention in the occupation more strongly for women than for men. To us, the results point towards a role for differences in preferences for the content of the work individuals do for their job choices and how they feel about their work. We realize that our results are suggestive at best and may not sway an avid sceptic. Our discrete choice experiments, administered to high school students, may be more compelling as many students indicate clearly that preferences for job content played the main role in their choices.

Overall, it seems that Rosie is able and willing to be a riveter if asked to do so but it is not her preferred line of work. Economists should explore the possibility that gender specific tastes matter for two reasons: We believe that preferences play a role for understanding the outcomes for women in the labor market, and taste based sorting into occupations may lead to policy prescriptions which differ from those traditionally advocated.

Many economists have so far dismissed explanations like ours for explaining the gender pay gap. For example, Fortin (2008) finds little role for attitude variables in wage regressions. Goldin (2014) dismisses explanations based on occupational sorting because most of the pay gap manifests itself within and not between occupations. We feel that these arguments fall short because occupational choice and gender differences in outcomes within occupations may interact.

To us, our results together with the existing literature point towards a story which runs along the following lines. This study suggests that women care about the content of the work they do more than men, and this influences their occupational choices. Most importantly, women stay away from traditional blue collar jobs, probably because of a combination of tastes and skill based comparative advantage (Weinberg, 2008, and Baker and Cornelsen, 2016). But even within white collar jobs, women sort systematically into occupations which are people rather than object oriented. In particular, this may explain why women choose occupations in business, law, and the health sector over technical and scientific jobs. Unfortunately, jobs with a lot of human contact are also typically jobs which require coordination and restrictions on work schedules and flexibility (Goldin, 2014). Advancement in these occupations often requires substantial dedication to the job, and career interruptions or part-time work are heavily penalized (see also Landers, Rebitzer, and Taylor, 1996).

Since the second half of the 20^{th} century, many (though not all) women strive to combine a meaningful career with a family (Hakim, 2000). But the occupational choices made by many of these often well educated women frequently expose them to a large pay penalty once they decide to have children. As a result, differences in labor market outcomes between young, childless women and men are slight but large pay gaps emerge once women have children (Kleven, Landais, and Sogaard, 2018). This story might be stylized, hides a lot of heterogeneity, and leaves out other factors which matter, but we believe it captures important elements.

The remaining differences in labor outcomes for men and women persist despite decades of policies, which have tried to address discrimination and make it easier for women to combine work and family. Strikingly, Kleven, Landais, and Sogaard (2018) find that the motherhood penalty has been extremely stable in Denmark for over two decades. This suggests that the existing policies may be reaching the limits of what they can achieve. Many policies try to change women’s occupational choices. Examples are quotas and efforts to shepherd girls into STEM fields. But these efforts may fail if women simply don’t want to work in the occupations the policy makers picked for them, or are more likely to leave these occupations after trying them initially. Instead, it may be more fruitful to think about policies which acknowledge women’s choices. For example, if many women want to be teachers, social workers, and nurses rather than programmers and engineers but these female professions are not paid as well then it may be necessary to think about policies which change pay levels more directly. An important first step is to increase our understanding about the role of women’s preferences in the labor market.

References

Altonji, Joseph G. and Rebecca M. Blank (1999). "Race and Gender in the Labor Market." In: Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics*, volume 3C, Amsterdam: Elsevier, 3143-3259.

Autor, David H. and David Dorn (2013). "The Growth of Low Skill Service Jobs and the Polarization of the U.S. Labor Market." *American Economic Review*, 103(5): 1553-1597.

Baker, Michael and Kirsten Corneslon (2016). "Gender Based Occupational Segregation and Sex Differences in Sensory, Motor and Spatial Aptitudes." *NBER Working Paper Series.* No. 22248.

Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske (2003). "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data." *Journal of Labor Economics,* 21(4): 887-922.

Beaudry, Paul, and Ethan Lewis (2014). "Do Male-Female Wage Differentials Reflect Differences in the Return to Skill? Cross-City Evidence from 1980-2000." *American Economic Journal: Applied Economics*, 6(2): 178-194.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones (2012). "What Do You Think Would Make You Happier? What Do You Think You Would Choose?" *American Economic Review,* 102(5): 2083-2110.

Benjamin, Daniel J., Ori Heffetz, Miles S. Kimball, and Alex Rees-Jones (2014). "Can Marginal Rates of Substitution Be Inferred from Happiness Data? Evidence from Residency Choices." *American Economic Review*, 104(11): 3498-3528.

Bertrand, Marianne. (2010). "New Perspectives on Gender." In: Orley Ashenfelter and David Card (eds), *Handbook of Labor Economics*, volume 4B, Amsterdam: Elsevier, 1545-1592.

Bertrand, Marianne (2018). "Coase Lecture – The Glass Ceiling." *Economica*, 85(338): 205–231.

Blau, Francine D. and Lawrence M. Kahn (2016). "The Gender Wage Gap: Extent, Trends, and Explanations." NBER Working Paper No. 21913.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). "Robust Inference with Multi-Way Clustering." *Journal of Business and Economic Statistics*, 29 (2): 238-249.

Chari, Anusha and Paul Goldsmith-Pinkham (2017). "Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute." NBER Working Paper No. 23953.

Chang, Emily (2018) *Brotopia: Breaking Up the Boys' Club of Silicon Valley*. New York: Portfolio Penguin.

Cortés, Patricia and Jessica Pan (2016) “Prevalence of Long Hours and Skilled Women’s Occupational Choices.” IZA Discussion Paper 10225.

Cortés, Patricia and Jessica Pan (2017) “Occupation and Gender.” IZA Discussion Paper 10672.

Croson, Rachel and Uri Gneezy (2009). “Gender Differences in Preferences”. *Journal of Economic Literature* Vol. 47, No. 2 (Jun., 2009), pp. 448-474.

Diamond, Peter A., and Jerry A. Hausman (1994). “Contingent Valuation: Is Some Number Better than No Number?” *Journal of Economic Perspectives*, 8(4): 45-64.

Dolado, Juan J., Dolado, Florentino Felgueroso, and Miguel Almunia (2012). “Are Men and Women-Economists Evenly Distributed across Research Rields? Some New Empirical Evidence.” *SERIEs*, 3(3): 367–393.

Dorn David (2009). “Essays on Inequality, Spatial Interaction, and the Demand for Skills.” Dissertation, University of St. Gallen no. 3613.

EPL Cornell (2014). *Claudia Goldin on Gender Equality in the Labor Market*. [online video] Available at: https://www.youtube.com/watch?v=9kgmmPHxe1E [Accessed 4 January 2018].

Finkelstein, Amy, Erzo F. P. Luttmer, and Matthew J. Notowidigdo (2013). “What Good is Wealth Without Health? The Effect of Health on the Marginal Utility of Consumption.” *Journal of the European Economic Association*, 11(S1), 221–258.

Fortin, Nicole M. (2008). “The Gender Wage Gap Among Young Adults in the United States – The Importance of Money versus People.” *The Journal of Human Resources* 43(4): 884-918.

Frijters, Paul and Bernard M.S. Van Praag (1998). “The Effects of Climate on Welfare and Well-Being in Russia.” *Climatic Change*, 39(1): 61–81.

Goldin, Claudia (2014). “A Grand Gender Convergence: Its Last Chapter.” *American Economic Review*, 104(4): 1091-1119.

Gorsuch, Richard L. (1983). *Factor Analysis*. Second edition, Hillsdale: Lawrence Erlbaum Associates.

Hakim, Catherine (2000). *Work-Lifestyle Choices in the 21st Century. Preference Theory*. Oxford: Oxford University Press.

Heckman, James, Rodrigo Pinto and Peter Savelyev (2013). “Understanding the Mechanisms through which an Influential Early Childhood Program Boosted Adult Outcomes.” *American Economic Review*, 103(6): 2052-86.

Hochschild, Arlie with Anne Machung (1989). *The Second Shift: Working Parents and the Revolution at Home*. New York: Viking.

Holland, John L. (1959). “A Theory of Vocational Choice.” *Journal of Counseling Psychology*, 6(1): 35-45.

Holland, John L. (1992). *Making Vocational Choices: A Theory of Vocational Personalities and Work Environments* (2nd ed.). Odessa, FL: Psychological Assessment Resources.

Hunt, Jenny (2016). “Why Do Women Leave Science and Engineering?” *Industrial and Labor Relations Review*, 69(1), 199-226.

Hwang, Hae-shin, Dale T. Mortensen, and W. Robert Reed (1998). “Hedonic Wages and Labor Market Search.” *Journal of Labor Economics*, 16(4): 815-847.

Kimball, Miles, and Robert Willis (2006). “Utility and Happiness.” Mimeographed, University of Michigan.

Kleven, Henrik, Camille Landais and Jakob Egholt Sogaard (2018). “Children and Gender Inequality: Evidence from Denmark.” NBER Working Paper No. 24237.

Landers, Renée M., James B. Rebitzer, and Lowell J. Taylor (1996). “Rat Race Redux: Adverse Selection in the Determination of Work Hours in Law Firms.” *American Economic Review*, 86(3): 329-348.

Lang, Kevin and Sumon Majumdar (2004). “The Pricing Of Job Characteristics When Markets Do Not Clear: Theory And Policy Implications.” *International Economic Review*, 45(4): 1111-1128.

Manning, Alan (2003). *Monopsony in Motion: Imperfect Competition in Labor Markets*. Princeton: Princeton University Press.

Manning, Alan and Joanna Swaffield (2008). “The Gender Gap in Early-Career Wage Growth.” *Economic Journal,* 118(530): 987-1024.

Mas, Alexandre and Amanda Pallais (2017). “Valuing Alternative Work Arrangements.” *American Economic Review*. 107(12): 3722-3759.

Morris, Mark A. (2003). “A Meta-analytic Investigation of Vocational Interest-based Job Fit, and Its Relationship to Job Satisfaction, Performance, and Turnover.” PhD Dissertation, University of Houston.

Prediger, Dale J. (1982). “Dimensions Underlying Holland’s Hexagon: Missing Link Between Interests and Occupations?” *Journal of Vocational Behavior*, 21(3): 259–287.

Pinker, Susan (2008). *The Sexual Paradox: Troubled Boys, Gifted Girls and the Real Difference Between the Sexes*. New York: Macmillan.

Stevenson, Betsey, and Justin Wolfers (2008). "Economic Growth and Happiness: Reassessing the Easterlin Paradox." *Brookings Papers on Economic Activity*, 2008(1): 1-87.

Stigler, George J. and Gary S. Becker (1977). "De Gustibus Non Est Disputandum." *American Economic Review*, 67(2): 76-90.

Strong, Edward K. (1943). *Vocational Interests of Men and Women*. Stanford: Stanford University Press.

Su, Rong, James Rounds, and Patrick Ian Armstrong (2009). "Men and Things, Women and People: A Meta-Analysis of Sex Differences in Interests." *Psychological Bulletin*, 135(6): 859-884.

Taylor, Marcia F. (ed.) with Brice John, Nick Buck, and Elaine Prentice-Lane (2010). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. Colchester: University of Essex.

Thorndike, Edward L. (1911). *Individuality*. Boston: Houghton Mifflin.

Thompson, Bruce. (2004). *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications*. Washington, DC: American Psychological Association.

Usui, Emiko (2008). "Job Satisfaction and the Gender Composition of Jobs." *Economics Letters*, 99(1): 23-26.

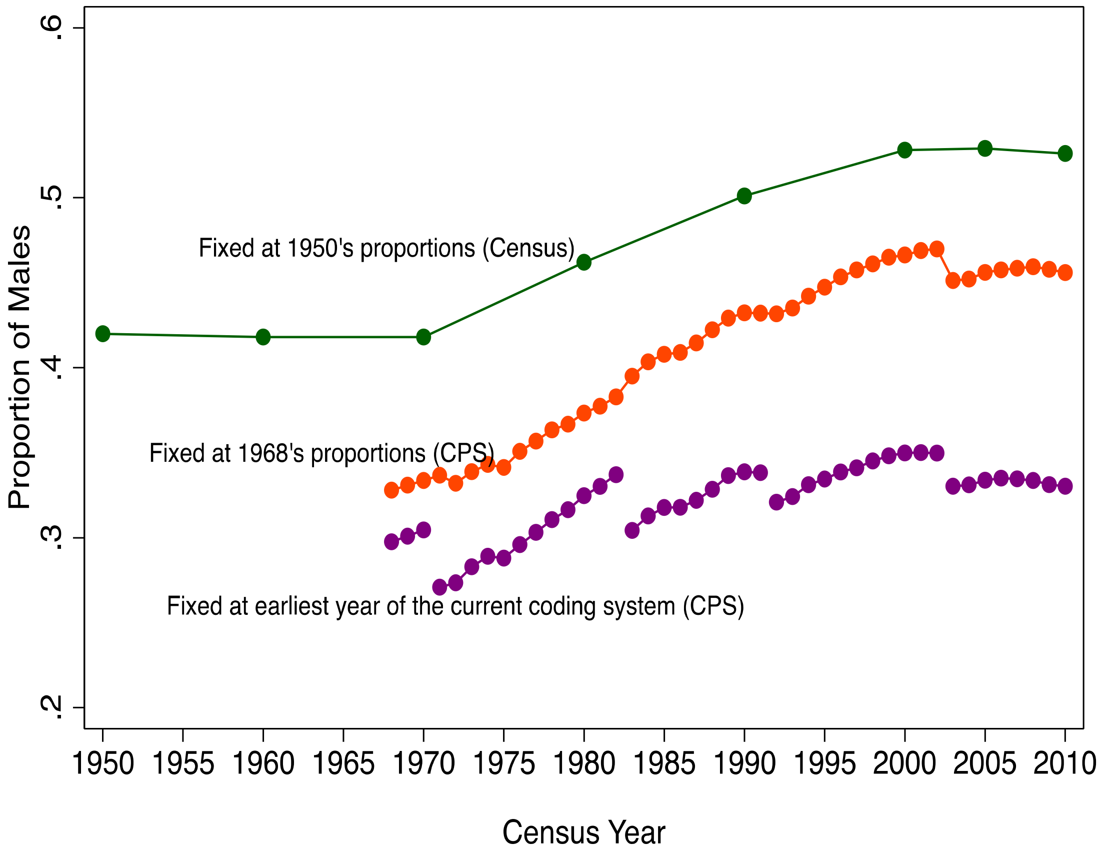
Van Praag, Bernard M. S. and Barbara E. Baarsma (2005). "Using Happiness Surveys to Value Intangibles: The Case of Airport Noise." *Economic Journal*, 115(500): 224-246.

Van Praag, Bernard M.S. and Ada Ferrer-i-Carbonell (2008). *Happiness Quantified: A Satisfaction Calculus Approach*. Revised edition, Oxford: Oxford University Press.

Weinberg, Bruce (2000). "Computer Use and the Demand for Female Workers." *Industrial and Labor Relations Review*, 53(2): 290-308.

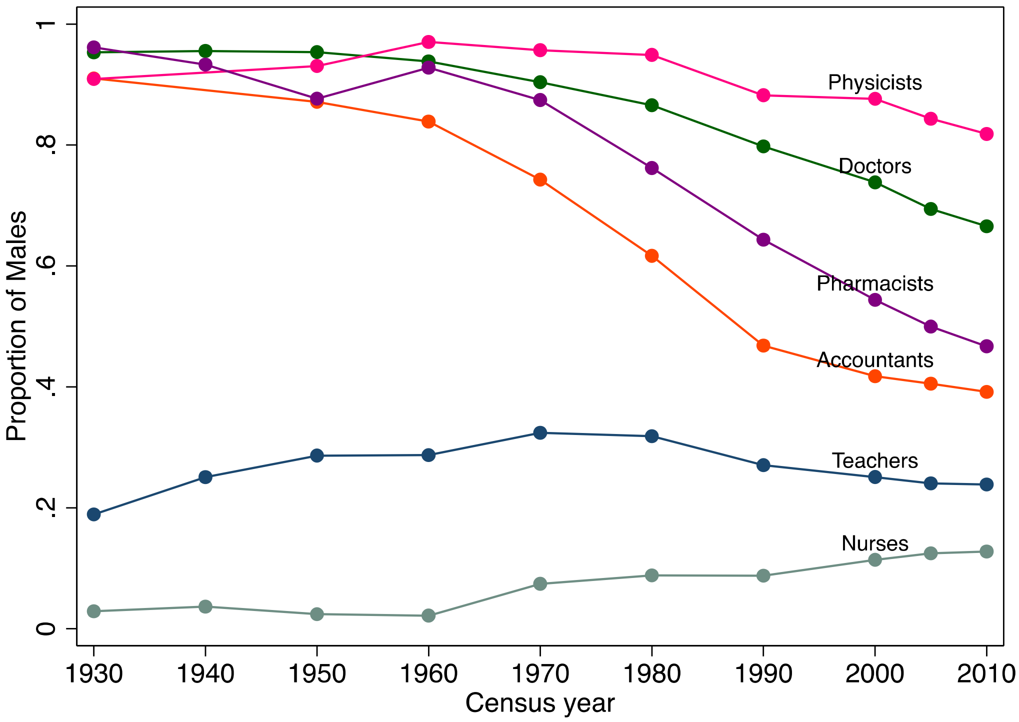
Wiswall, Matthew, and Basit Zafar (2018) "Preference for the Workplace, Human Capital, and Gender." *Quarterly Journal of Economics*, 133(1): 457-507.

Figure 1: The Share of Males in Female Jobs



Notes: The lines in this graph show the share of males (SOM) in the occupations in which females work in a particular year in the US. The top line uses Census data and is based on the SOM in each occupation in 1950 using the IPUMS 1950 consistent occupation code. The other lines use annual CPS data. In the second line, SOM in an occupation is calculated based on the 1968 data. The bottom line uses the current occupation codes and fixes the SOM in the year the current code was first introduced. The line is broken whenever a new set of occupation codes comes into use.

Figure 2: Trends in the Share of Males in Selected White Collar Jobs



Notes: This graph shows the share of males in selected white-collar occupations in the US Census.

Table 1: High and Low Ranked Occupations According to the Content of the Work (US)

Rank	People	Brains	Brawn
		10 Highest Ranked Occupations	
1	Physicians Assistants (106)	CEOs, public administrators and legislators (4)	Explosives workers (615)
2	Protective Services (427)	Chemical engineers (48)	Other mining occupations (617)
3	Licensed Practical Nurses (207)	Engineers and other professionals n.e.c. (59)	Roofers and slaters (595)
4	Occupational Therapists (99)	Financial Managers (7)	Miners (616)
5	Other health and therapy occupations (89)	Human resources and labor relations managers (8)	Heating, AC, and refrigeration mechanics (534)
6	Sheriffs and Bailiffs (423)	Petroleum mining and geological engineers (47)	Boilermakers (643)
7	Registered nurses (95)	Aerospace Engineers (44)	Water and sewage treatment plant operator (694)
8	Physicians (84)	Architects (43)	Carpenters (567)
9	Veterinarians (86)	Urban and regional planners (173)	Plasterers (584)
10	Dieticians and Nutritionists (97)	Civil engineers (43)	Helpers, surveyors (866)
		10 Lowest Ranked Occupations	
1	Mathematicians and Statisticians (68)	Garbage and recyclable material collectors (875)	Insurance underwriters (24)
2	Statistical Clerks (386)	Clothing pressing machine operator (747)	Statistical Clerks (386)
3	Chemical Technicians (224)	Machine operators (779)	Actuaries (66)
4	Physicists and Astronomists (69)	Housekeepers, maids, butlers and cleaners (405)	Economists, market and survey researchers (166)
5	Actuaries (66)	Proofreaders (384)	Lawyers and Judges (178)
6	Biological Scientists (78)	Parking lot attendants (813)	Interviewers, enumerators, and surveyors (316)
7	Chemical Engineers (48)	Personal services n.e.c (469)	Payroll and timekeeping clerks (338)
8	Machinists (637)	Mail carriers for post office (355)	Art/entertainment performers and related (194)
9	Engineering technicians (214)	Packers and packagers by hand (888)	Millwrights (544)
10	Programmers of Numerically Controlled Machines (233)	Helpers, construction (865)	Drillers of oil wells (614)

Table 2: Factor Scores for Selected Occupations (US)

Occupation	1930 SOM	SOM	Factor Scores People	Brains	Brawn
Electricians (575)	0.993	0.979	-0.977	0.227	1.907
Miners (686)	0.997	0.972	-0.099	-1.230	0.600
Mechanical Engineers (57)	0.993	0.937	-2.154	1.288	0.393
Architects (43)	0.979	0.784	0.207	1.951	0.075
Physicians (84)	0.944	0.722	0.384	1.587	0.081
Butchers and Meat Cutters (686)	0.992	0.746	-0.098	-1.230	0.600
Mathematicians and Statisticians (66)	N/A	0.640	-0.259	0.158	-1.632
Financial managers (7)	N/A	0.572	-0.587	2.078	-1.100
Economists, market and survey researchers (166)	0.810	0.510	0.297	2.066	-0.843
Bartenders (434)	0.960	0.434	0.717	-1.015	-0.177
Accountants and auditors (23)	0.912	0.423	-0.936	0.981	-1.482
Social Workers (174)	0.265	0.251	1.670	0.968	-0.982
Primary School Teachers (156)	0.188	0.165	1.345	0.567	-0.622
Nurses (207)	0.025	0.068	1.140	1.347	-0.110

Notes: The occupation codes in parentheses are occ1990dd codes from Dorn (2009). SOM is the share of males in an occupation based on the Census and ACS from 1980-2014. 1930 SOMs calculated using the 1930 Census.

Table 3: The Relationship Between the Share of Males and People, Brains, and Brawn

	Samples		
	US – Census	Britain – LFS	Russia – RLMS
People	-0.031	-0.057	-0.124
	(0.014)	(0.013)	(0.029)
Brains	-0.012	-0.029	-0.001
	(0.017)	(0.022)	(0.021)
Brawn	0.067	0.102	0.183
	(0.024)	(0.018)	(0.025)
Number of Observations	14464167	4266356	328371

Notes: All regressions also include the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, as well as time and area effects. Standard errors are clustered by occupation.

Table 4: Individual Fixed Effects Regressions

	Samples							
	US – NLSY		Britain – BHPS		Britain – BHPS		Russia – RLMS	
	Females	Males	Females	Males	Females	Males	Females	Males
Dependent Variable	Overall Job Satisfaction		Overall Job Satisfaction		Satisfaction with Work Itself		Overall Job Satisfaction	
People	0.021	0.011	0.028	0.022	0.063	0.036	0.022	-0.003
	(0.006)	(0.006)	(0.010)	(0.009)	(0.014)	(0.010)	(0.015)	(0.017)
Brains	0.072	0.046	0.029	-0.006	0.032	-0.012	-0.009	0.024
	(0.008)	(0.008)	(0.013)	(0.011)	(0.018)	(0.012)	(0.013)	(0.014)
Brawn	-0.031	-0.000	-0.046	-0.016	-0.053	-0.010	-0.060	-0.040
	(0.008)	(0.006)	(0.014)	(0.012)	(0.017)	(0.013)	(0.016)	(0.015)
Number of Observations	91234	97638	49606	46099	49606	46099	35443	27117
Dependent Variable	Stayers							
People	0.002	0.008	0.033	0.019			0.003	-0.026
	(0.003)	(0.003)	(0.010)	(0.009)			(0.015)	(0.015)
Brains	0.033	-0.001	0.022	-0.009			0.030	0.001
	(0.004)	(0.004)	(0.017)	(0.012)			(0.012)	(0.012)
Brawn	0.000	0.012	-0.044	0.012			-0.023	0.012
	(0.004)	(0.003)	(0.013)	(0.012)			(0.015)	(0.014)
Number of Observations	91234	97638	48116	44862			23449	16792

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2.

Table 5: Individual Fixed Effects Regressions Distinguishing Women with and without Children

	Samples											
	US – NLSY			Britain – BHPS			Britain – BHPS			Russia – RLMS		
	Females with Children	Females without Children	Males	Females with Children	Females without Children	Males	Females with Children	Females without Children	Males	Females with Children	Females without Children	Males
Dep. Variable	Overall Job Satisfaction			Overall Job Satisfaction			Satisfaction with Work Itself			Overall Job Satisfaction		
People	0.023	0.019	0.011	0.021	0.028	0.022	0.055	0.053	0.036	0.032	0.021	-0.003
	(0.008)	(0.009)	(0.006)	(0.016)	(0.013)	(0.009)	(0.020)	(0.015)	(0.010)	(0.019)	(0.017)	(0.017)
Brains	0.057	0.082	0.046	0.075	0.014	-0.006	0.071	0.024	-0.012	-0.047	0.024	0.024
	(0.012)	(0.012)	(0.008)	(0.024)	(0.018)	(0.011)	(0.023)	(0.018)	(0.012)	(0.017)	(0.023)	(0.014)
Brawn	-0.024	-0.038	-0.000	-0.073	-0.048	-0.016	-0.086	-0.039	-0.010	-0.054	-0.059	-0.040
	(0.011)	(0.012)	(0.006)	(0.023)	(0.020)	(0.012)	(0.025)	(0.020)	(0.013)	(0.021)	(0.026)	(0.015)
Number of Obs	53648	36857	97638	21660	27946	46099	21660	27946	46099	19926	15517	27117
Dep. Variable	Stayers											
People	0.005	0.002	0.008	0.022	0.035	0.019				0.042	0.012	-0.026
	(0.004)	(0.008)	(0.003)	(0.011)	(0.013)	(0.009)				(0.014)	(0.016)	(0.015)
Brains	0.025	0.040	-0.001	-0.014	0.015	-0.009				-0.007	0.002	0.001
	(0.006)	(0.006)	(0.004)	(0.016)	(0.020)	(0.012)				(0.021)	(0.019)	(0.012)
Brawn	-0.002	-0.010	0.012	-0.017	-0.037	0.012				-0.033	-0.030	0.012
	(0.005)	(0.005)	(0.003)	(0.015)	(0.016)	(0.012)				(0.019)	(0.020)	(0.014)
Number of Obs.	53648	36857	97638	21660	27946	44862				14140	9309	16792

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects interacted with the sub-periods with consistent occupation codes. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2.

Table 6: Satisfaction with Work Itself Regressions in the WERS

	Samples			
	Females	Males	Females	Males
	Baseline		Firm Fixed Effects	
People	0.106	0.067	0.038	0.006
	(0.010)	(0.009)	(0.011)	(0.012)
Brains	0.052	0.030	0.070	0.020
	(0.010)	(0.009)	(0.013)	(0.013)
Brawn	0.010	0.026	0.000	0.009
	(0.012)	(0.010)	(0.015)	(0.013)
Number of Observations	20964	17231	20964	17231

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, along with time effects. Standard errors are two-way clustered by firm and worker's occupation and shown in parentheses. Models are estimated using ivreg2.

Table 7: Fraction of Males among Those Choosing an Occupation in the Secondary School Survey

Question	Occupation	Weekly Gross Wage	Avg. Hours	Factor Scores: People	Brains	Brawn	Fraction of males choosing occ in survey
1	Public Relations Specialist	£424	40.4	0.612	0.671	-0.657	0.723
	Mental Health Social Worker	£441	39.5	1.301	0.817	-1.082	0.308
2	Architectural Drafter	£619	41.4	-1.306	0.370	-0.476	0.617
	Retail Buyer	£615	43.2	1.238	0.589	-0.578	0.417
3	Town Planner	£558	41.9	0.884	1.929	-0.171	0.481
	Insurance Underwriter	£532	40.1	-1.155	-0.029	-1.852	0.616
4	Physicist	£703	43.6	-1.471	1.706	1.037	0.763
	Art Director	£539	45.9	0.289	0.991	-0.761	0.355
5	Accountant	£584	43.7	-0.654	1.516	-1.206	0.709
	Health Services Manager	£676	42.8	0.542	2.056	-0.231	0.256
6	Architect	£620	44.7	-0.212	1.792	-0.255	0.561
	Human Resources Manager	£758	45.5	0.552	2.109	-1.349	0.433

Notes: Occupational averages are calculated based on the 2002-2012 QLFS SOC 2000 3 digit occupation codes.

Table 8: Logit Regressions of Occupational Choices on People, Brains, and Brawn

	(1) Females	(2) Males	(3) Females	(4) Males	(5) Females	(6) Males
People	1.63	1.23	1.46	1.19	1.56	1.25
	(0.13)	(0.09)	(0.13)	(0.09)	(0.13)	(0.10)
Brains	0.92	0.81	1.13	0.92	1.07	1.07
	(0.16)	(0.14)	(0.21)	(0.16)	(0.20)	(0.20)
Brawn	1.02	0.82	0.97	0.76	0.94	0.65
	(0.12)	(0.09)	(0.11)	(0.09)	(0.11)	(0.08)
Skill match (continuous)			1.31	1.33		
			(0.07)	(0.07)		
Skill match (discrete)					1.68	2.28
					(0.23)	(0.29)
Equality of male and female PBB coefficients (p-value)	0.000		0.000		0.000	

Notes: Coefficients shown are odds ratios. Regressions have 886 observations on 150 females and 936 observations on 157 males. Robust standard errors in parentheses.

Table 9: Justification Given for Occupation Choice

Reason	(1) Females	(2) Males
Like the activity/job interesting	0.471	0.434
Other job is unappealing/lack skills	0.211	0.207
Like the environment of the job	0.093	0.065
Good at the skills required	0.096	0.117
Indifferent between the choices	0.007	0.013
Other	0.032	0.045
Uninformative/illegible	0.027	0.030
No answer	0.063	0.088

Notes: Based on the question "For each of the six job choices you made, tell us in few words why you picked the job you did?" Answers are in free form, without any prompts, and responses are coded into the eight categories above.

Does Rosie Like Riveting?
Male and Female Occupational Choices

Appendix
Not for Publication

Grace Lordan and Jörn-Steffen Pischke
LSE

Appendix A: Construction of Latent Factors from ONET

In order to create measures of job content we use the ONET database version 5. ONET provides information on occupational attributes and requirements, as well as the characteristics of the workers in an occupation in the US. We focus on the 79 items describing work activities and context. For each individual item, a level from 1 to 7 is reported by an incumbent. We standardize each of these variables to have a mean of 0 and a standard deviation of 1.

We follow the psychometric literature (Gorsuch, 1983, 2003; Thomson, 2004) and use exploratory factor analysis to reduce the dimensionality of the ONET data. To extract the underlying latent factors, we first determine the number of factors to retain. Looking at a scree plot from an orthogonal exploratory analysis and the eigenvalue of each individual factor as depicted for the US in Figure A.1, the slope of the curve levels off after the fourth factor (in Figure A.1 the eigenvalues are on the y-axis and the number of factors on the x-axis). Only two variables load onto the fourth factor, which look like 'routine task intensity.' Specifically, these are degree of automation and importance of repeating same tasks. We decided to drop this fourth factor and only retain the first three, which load on many more variables. The results are very similar for Britain and Russia. For all three countries, the first three factors explain between 65% and 70% of the variability in the data.

Using orthogonal rotation, we next perform Confirmatory Factor Analysis (CFA) to extract three latent variables. Table A.1. documents how the items load onto each factor. The table omits weak loadings which are below 0.4 (in absolute value). Using these three factors with loadings on all items directly in our job satisfaction and stayer regressions does not change the conclusions drawn in the main text (see Tables A.3 and A.4).

The results in the main text follow an approach recommended by Heckman et al. (2012). Specifically, once the first confirmatory analysis is performed, to identify three latent uncorrelated factors, we review how every item loads on each factor and drop items that are weakly associated with all three factors or those that are associated with two or more factors. Specifically, we remove items with a loading of 0.4 or less on all factors (weak loaders) and items that have a loading that is greater than 0.4 on more than one factor (cross-loaders). We then repeat the factor analysis using the remaining ONET items and extract the final latent variables which have no items that are weakly loaded or cross loaded and are freely correlated.

These latent factors are used in the main analysis. Table A.2 documents how each item loads on these final factors.

Figure A.1 Scree Plot of the US Exploratory Analysis

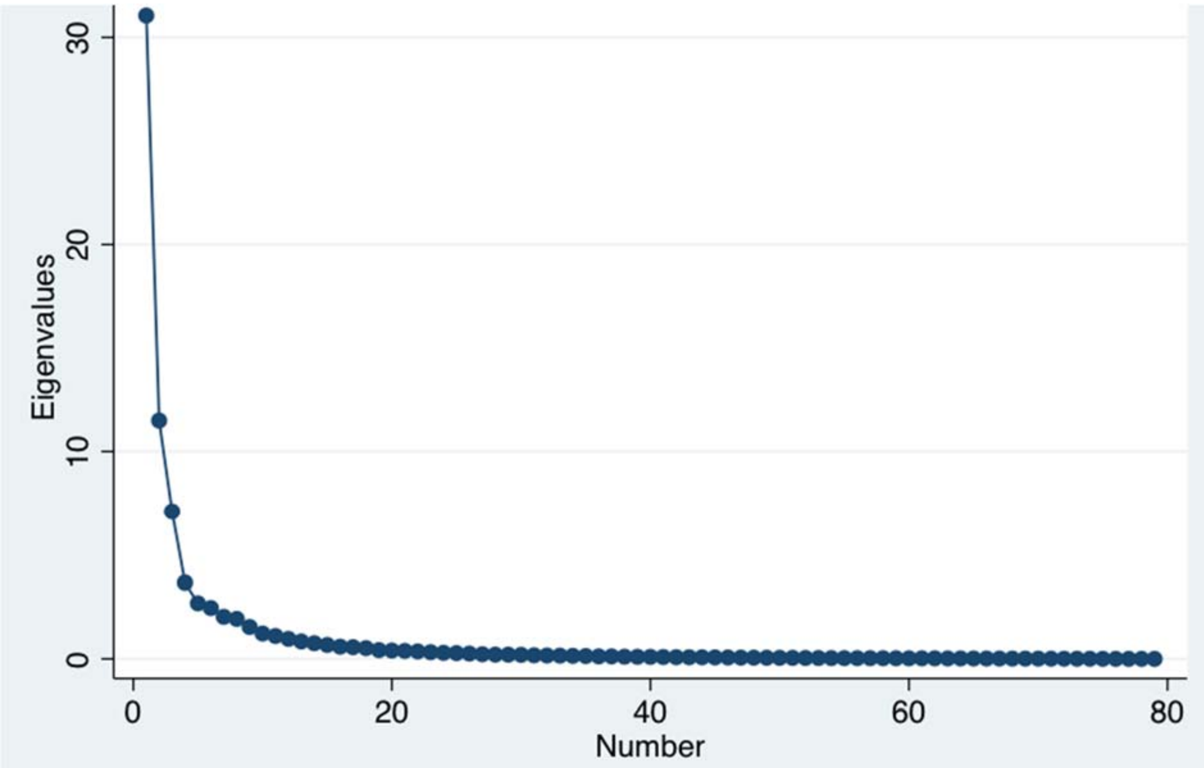


Table A.1 Rotated Factor Loadings of First Rotation (<0.40 is blank)

ONET Item	People	Brains	Brawn
Monitoring and Controlling Resources (A)		0.717	
Staffing Organizational Units (A)		0.644	
Performing Administrative Activities		0.695	-0.448
Provide Consultation and Advice to Others (A)		0.855	
Coaching and Developing Others (A)	0.737	0.512	
Getting Information (A)		0.871	
Monitor Processes, Materials, or Surroundings (A)		0.645	
Identifying Objects, Actions, and Events (A)		0.877	
Inspecting Equipment, Structures, or Material (A)		0.458	0.599
Estimating the Quantifiable Characteristics of Products, Events, or Information (A)		0.856	
Judging the Qualities of Things, Services, or People (A)		0.849	
Processing Information (A)		0.811	-0.478
Evaluating Info to Determine Compliance with Standards (A)		0.851	
Analyzing Data or Information (A)		0.879	
Making Decisions and Solving Problems (A)		0.920	
Thinking Creatively (A)		0.784	
Updating and Using Relevant Knowledge (A)		0.878	
Developing Objectives and Strategies		0.796	
Scheduling Work and Activities (A)	0.786	0.424	
Organizing, Planning, and Prioritizing Work (A)		0.837	
Performing General Physical Activities (A)			0.829
Handling and Moving Objects (A)	-0.442		0.634
Controlling Machines and Processes (A)	-0.428		0.678
Operating Vehicles, Mechanized Devices, or Equipment (A)			0.521
Interacting With Computers (A)		0.671	-0.499
Drafting, Laying Out, and Specifying Technical Devices, Parts, and Equipment (A)		0.553	
Repairing and Maintaining Mechanical Equipment (A)			0.660
Repairing and Maintaining Electronic Equipment (A)			
Documenting/Recording Information (A)		0.795	-0.428
Interpreting the Meaning of Information for Others (A)		0.807	
Communicating with Supervisors, Peers, or Subordinates (A)	0.876		
Communicating with Persons Outside Organization (A)	0.486	0.627	-0.448
Establishing and Maintaining Interpersonal Relationships (A)	0.596	0.636	
Assisting and Caring for Others (A)	0.624		
Selling or Influencing Others (A)	0.549	0.470	
Resolving Conflicts and Negotiating with Others (A)	0.616	0.589	
Performing for or Working Directly with the Public (A)	0.737		
Coordinating the Work and Activities of Others (A)	0.848		
Developing and Building Teams (A)		0.852	
Training and Teaching Others (A)	0.701	0.443	
Guiding, Directing, and Motivating Subordinates (A)	0.798		

Table A.1 (Continued) Rotated Factor Loadings of First Rotation (<0.40 is blank)

ONET Item	People	Brains	Brawn
Contact With Others (C)	0.744		
Deal With External Customers (C)	0.665		
Coordinate or Lead Others (C)	0.735		0.450
Responsible for Others' Health and Safety (C)	0.413	0.506	
Responsibility for Outcomes and Results (C)			0.701
Frequency of Conflict Situations (C)	0.639		0.479
Deal With Unpleasant or Angry People (C)	0.746		
Deal With Physically Aggressive People (C)	0.525		
Indoors, Environmentally Controlled (C)			-0.539
Outdoors, Exposed to Weather (C)			0.643
Sounds, Noise Levels Are Distracting or Uncomfortable (C)			0.710
Very Hot or Cold Temperatures (C)			0.820
Extremely Bright or Inadequate Lighting (C)			0.779
Exposed to Contaminants (C)			0.826
Cramped Work Space, Awkward Positions (C)			0.797
Exposed to Whole Body Vibration (C)			0.644
Exposed to Radiation (C)			
Exposed to Disease or Infections (C)			
Exposed to High Places (C)			0.638
Exposed to Hazardous Conditions (C)			0.770
Exposed to Hazardous Equipment (C)			0.763
Exposed to Minor Burns, Cuts, Bites, or Stings (C)			0.772
Spend Time Sitting (C)			-0.684
Spend Time Standing (C)			0.630
Spend Time Climbing Ladders, Scaffolds, or Poles (C)			0.728
Spend Time Walking and Running (C)	0.434		0.462
Spend Time Kneeling, Crouching, Stooping, or Crawling (C)			0.746
Spend Time Keeping or Regaining Balance (C)			0.801
Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls (C)	-0.401		0.466
Spend Time Bending or Twisting the Body (C)			0.783
Spend Time Making Repetitive Motions (C)	-0.448		
Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets (C)			0.854
Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection (C)			0.646
Consequence of Error (C)		0.588	
Degree of Automation (C)	-0.459		
Importance of Being Exact or Accurate (C)		0.567	
Importance of Repeating Same Tasks (C)			
Pace Determined by Speed of Equipment (C)	-0.539		

Notes: Blanks indicate an item has a loading <0.40 in absolute value on that factor.

Table A.2 Rotated Factor Loadings of Final Latent Factors (<0.40 is blank)

ONET Item	People	Brains	Brawn
Monitoring and Controlling Resources (A)		0.717	
Staffing Organizational Units (A)		0.720	
Performing Administrative Activities		0.674	
Provide Consultation and Advice to Others (A)		0.855	
Coaching and Developing Others (A)	0.802		
Getting Information (A)		0.854	
Monitor Processes, Materials, or Surroundings (A)		0.672	
Identifying Objects, Actions, and Events (A)		0.867	
Inspecting Equipment, Structures, or Material (A)			
Estimating the Quantifiable Characteristics of Products, Events, or Information (A)		0.890	
Judging the Qualities of Things, Services, or People (A)		0.856	
Processing Information (A)		0.750	
Evaluating Info to Determine Compliance with Standards (A)		0.855	
Analyzing Data or Information (A)		0.862	
Making Decisions and Solving Problems (A)		0.933	
Thinking Creatively (A)		0.795	
Updating and Using Relevant Knowledge (A)		0.865	
Developing Objectives and Strategies (A)		*0.868*	
Scheduling Work and Activities (A)		0.834	
Organizing, Planning, and Prioritizing Work (A)		0.852	
Performing General Physical Activities (A)			0.839
Handling and Moving Objects (A)			0.557
Controlling Machines and Processes (A)			0.638
Operating Vehicles, Mechanized Devices, or Equipment (A)			0.546
Interacting With Computers (A)			0.658
Drafting, Laying Out, and Specifying Technical Devices, *Parts, and Equipment (A)*			0.570
Repairing and Maintaining Mechanical Equipment (A)			0.649
Repairing and Maintaining Electronic Equipment (A)			
Documenting/Recording Information (A)		0.749	
Interpreting the Meaning of Information for Others (A)		0.778	
Communicating with Supervisors, Peers, or Subordinates (A)		0.883	
Communicating with Persons Outside Organization (A)		0.619	
Establishing and Maintaining Interpersonal Relationships (A)			
Assisting and Caring for Others (A)	0.492		
Selling or Influencing Others (A)			
Resolving Conflicts and Negotiating with Others (A)			
Performing for or Working Directly with the Public (A)	0.610		
Coordinating the Work and Activities of Others (A)	0.915		
Developing and Building Teams (A)		0.886	
Training and Teaching Others (A)	0.754		
Guiding, Directing, and Motivating Subordinates (A)	0.879		

Table A.2 (Continued) Rotated Factor Loadings of Final Latent Factors (<0.40)

ONET Item	People	Brains	Brawn
Contact With Others (C)	0.592		
Deal With External Customers (C)	0.522		
Coordinate or Lead Others (C)	0.803		
Responsible for Others' Health and Safety (C)			
Responsibility for Outcomes and Results (C)		0.785	
Frequency of Conflict Situations (C)			
Deal With Unpleasant or Angry People (C)	0.586		
Deal With Physically Aggressive People (C)	0.620		
Indoors, Environmentally Controlled (C)			-0.591
Outdoors, Exposed to Weather (C)			0.721
Sounds, Noise Levels Are Distracting or Uncomfortable (C)			0.710
Very Hot or Cold Temperatures (C)			0.869
Extremely Bright or Inadequate Lighting (C)			0.817
Exposed to Contaminants (C)			0.840
Cramped Work Space, Awkward Positions (C)			0.818
Exposed to Whole Body Vibration (C)			0.626
Exposed to Radiation (C)			
Exposed to Disease or Infections (C)			
Exposed to High Places (C)			0.705
Exposed to Hazardous Conditions (C)			0.783
Exposed to Hazardous Equipment (C)			0.757
Exposed to Minor Burns, Cuts, Bites, or Stings (C)			0.772
Spend Time Sitting (C)			-0.689
Spend Time Standing (C)			0.622
Spend Time Climbing Ladders, Scaffolds, or Poles (C)			0.701
Spend Time Walking and Running (C)			
Spend Time Kneeling, Crouching, Stooping, or Crawling? (C)			0.753
Spend Time Keeping or Regaining Balance (C)			0.830
Spend Time Using Your Hands to Handle, Control, or Feel Objects, Tools, or Controls (C)			0.550
Spend Time Bending or Twisting the Body (C)			
Spend Time Making Repetitive Motions (C)	-0.470		
Wear Common Protective or Safety Equipment such as Safety Shoes, Glasses, Gloves, Hearing Protection, Hard Hats, or Life Jackets (C)			0.856
Wear Specialized Protective or Safety Equipment such as Breathing Apparatus, Safety Harness, Full Protection Suits, or Radiation Protection (C)			0.680
Consequence of Error (C)		0.617	
Degree of Automation (C)	-0.558		
Importance of Being Exact or Accurate (C)	-0.411	0.491	
Importance of Repeating Same Tasks (C)			
Pace Determined by Speed of Equipment (C)	-0.487		

Notes: Italics indicates that an item has been dropped either because it loaded weakly on all factors (<0.40 in absolute value) or it cross-loaded on more than one factor (>0.40 in absolute value on more than one factor). Blanks indicate an item has loaded <0.40 in absolute value on that factor.

Table A.3 Job Satisfaction Regressions
Uncorrelated PBB Factors/All ONET Items

	Samples							
	US – NLSY		Britain – BHPS		Britain – BHPS		Russia	
	Females	Males	Females	Males	Females	Males	Males	Females
Dependent Variable	Overall Job Satisfaction		Overall Job Satisfaction		Satisfaction with Work Itself		Overall Job Satisfaction	
People	0.038	0.015	0.017	0.018	0.056	0.030	0.040	0.011
	(0.006)	(0.006)	(0.011)	(0.010)	(0.014)	(0.011)	(0.018)	(0.018)
Brains	0.076	0.045	0.041	-0.002	0.046	-0.008	0.003	0.028
	(0.008)	(0.008)	(0.013)	(0.012)	(0.018)	(0.012)	(0.014)	(0.015)
Brawn	-0.041	-0.010	-0.045	-0.015	-0.055	-0.008	-0.053	-0.028
	(0.038)	(0.015)	(0.014)	(0.012)	(0.017)	(0.012)	(0.016)	(0.017)
Number of Observations	91234	97638	49606	46099	49606	46099	35443	27117

Notes: PBB factors are derived from first rotation, retaining cross loaders and weak loaders. All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are created by confirmatory factor analysis on the first rotation of the exploratory factor analysis described in the methods. The extracted factors are uncorrelated. For all countries three factors are extracted that can loosely be labeled ‘people’ ‘brains’ and ‘brawn.’

Table A.4 Stayer Regressions
Uncorrelated PBB Factors/All ONET Items

	Samples					
	US – NLSY		Britain – BHPS		Russia	
	Females	Males	Females	Males	Females	Males
People	0.013	0.012	0.034	0.018	0.007	-0.010
	(0.004)	(0.004)	(0.011)	(0.011)	(0.007)	(0.011)
Brains	0.038	0.023	0.030	-0.010	-0.001	-0.003
	(0.006)	(0.006)	(0.018)	(0.012)	(0.006)	(0.008)
Brawn	-0.004	0.002	-0.044	0.012	-0.015	0.004
	(0.005)	(0.005)	(0.013)	(0.011)	(0.007)	(0.009)
Number of Observations	91234	97638	48116	44863	23449	16792

Notes: PBB factors are derived from first rotation, retaining all cross loaders and weak loaders as depicted by Table A.1. All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are created by confirmatory factor analysis on the first rotation of the exploratory factor analysis described in the methods. The extracted factors are uncorrelated. For all countries three factors are extracted that can loosely be labelled ‘people’ ‘brains’ and ‘brawn.’

Appendix B: Robustness Analyses

Our baseline analysis of the main text incudes individual fixed effects. Table B.1 and B.2 document the same estimates with the individual fixed effects removed. Results tend to be somewhat stronger without the fixed effects but general patterns are similar.

Focus in this study is on overall job satisfaction and satisfaction with work itself, however the BHPS does collect data on other satisfaction domains. These are satisfaction with hours, satisfaction with pay and satisfaction with overall security. Estimates using these alternative dependent variables are documented in Table B.3. None of the PBB factors are particularly strongly related to these satisfaction domains. On the other hand, satisfaction with hours is lower in high hours occupations and satisfaction with pay is higher in high paying occupations, so the satisfaction variables do carry information. These results are in line with our interpretation that the PBB coefficients for overall satisfaction (or satisfaction with work itself) pick up preferences for job content rather than something else.

Tables B.4, B.5 and B.6 document analyses which add 'own' characteristics to the baseline models for the US, UK and Russia, respectively. Specifically we add, log of own wages, own hours, a college graduate dummy variable, number of children and marital status (married =1, 0 otherwise). These variables are often included in satisfaction equations but we are worried about interpretability with some of these variables, particularly own job outcomes like wages and hours, included. However, results are very similar except for Russia where samples are smallest.

We also create a variable to proxy the wage-hours elasticity used by Goldin (2014). Goldin interprets this occupation specific elasticity as capturing the wage penalty arising from working shorter hours: high elasticities imply a penalty for workers seeking short hours and indicate a lack of flexibility. Specifically, 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 a number of other controls[22] separately for each country. Of interest are the coefficients on the interaction between

[22] These are gender, age, age squared, age to the power of three, age to the power of four, education, ethnicity and year dummies.

occupation and log hours. Table B.7 reports estimates which add the wage-hours elasticity to the baseline model. The PBB coefficients do not change greatly when this variable is added. The coefficients on the wage-hours elasticity itself have different signs for different countries. We computed this elasticity once for the entire time period we analyse. This may be problematic if flexibility has changed in some occupations over time, which is why we don't stress these results more. On the other hand, we believe that changes over time in the broad job content measures we use are unlikely important.

A lot of the recent literature on gender differences has focused on more educated women or professional and managerial jobs. Table B.8 compares our results for women to those with a college education only. The job satisfaction results tend to be stronger for college educated women while the results for retention in the occupation are more similar to the whole sample. This suggests that our results are not simply driven by blue collar or low skilled jobs where male-female differences in occupational choice may be most pronounced.

Appendix B: Alternative Specifications

Table B.1: Cross Sectional Job Satisfaction Regressions

	Samples							
	US – NLSY		Britain – BHPS		Britain – BHPS		Russia – RLMS	
	Females	Males	Females	Males	Females	Males	Females	Males
Dependent Variable	Overall Job Satisfaction		Overall Job Satisfaction		Satisfaction with Work Itself		Overall Job Satisfaction	
People	0.048	0.062	0.055	0.048	0.102	0.060	0.150	-0.005
	(0.006)	(0.007)	(0.017)	(0.012)	(0.015)	(0.014)	(0.020)	(0.019)
Brains	0.106	0.077	0.050	0.019	0.072	0.041	0.033	-0.023
	(0.010)	(0.010)	(0.022)	(0.018)	(0.022)	(0.018)	(0.018)	(0.020)
Brawn	-0.037	0.026	-0.006	0.031	-0.006	0.041	-0.049	-0.024
	(0.009)	(0.008)	(0.019)	(0.014)	(0.014)	(0.016)	(0.020)	(0.019)
Number of Observations	91234	97638	49606	46099	49606	46099	35443	27117

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using ivreg2. PBB factors are derived using the approach described in Heckman et al (2012). For all countries three factors are extracted that can loosely be labeled 'people' 'brains' and 'brawn.'

Table B.2: Cross Sectional Stayer Regressions

	Samples					
	US – NLSY		Britain – BHPS		Russia – RLMS	
	Females	Males	Females	Males	Females	Males
People	-0.006	-0.003	0.036	0.041	0.009	0.006
	(0.005)	(0.004)	(0.010)	(0.010)	(0.014)	(0.015)
Brains	0.050	0.040	0.013	-0.021	0.007	0.001
	(0.007)	(0.006)	(0.014)	(0.013)	(0.010)	(0.009)
Brawn	0.007	0.013	-0.023	0.028	-0.009	-0.003
	(0.006)	(0.005)	(0.014)	(0.013)	(0.012)	(0.011)
Number of Observations	91234	97638	48116	44862	23449	16792

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using ivreg2. PBB factors are derived using the approach described in Heckman et al (2012). For all countries three factors are extracted that can loosely be labeled 'people' 'brains' and 'brawn.'

Table B.3: Other Satisfaction Domains in BHPS

	Samples					
	Females	Males	Females	Males	Females	Males
Dependent Variable	Satisfaction with Hours		Satisfaction with Pay		Satisfaction with Security	
People	-0.022	-0.025	-0.004	0.006	0.037	0.029
	(0.010)	(0.009)	(0.014)	(0.008)	(0.011)	(0.009)
Brains	0.013	-0.023	-0.016	-0.004	0.003	0.008
	(0.012)	(0.011)	(0.013)	(0.010)	(0.014)	(0.011)
Brawn	-0.007	-0.009	0.003	0.022	0.002	-0.020
	(0.013)	(0.013)	(0.018)	(0.013)	(0.014)	(0.011)
Average Hours	-0.749	-0.389	-0.241	-0.134	0.010	0.048
	(0.168)	(0.198)	(0.208)	(0.187)	(0.060)	(0.053)
Log Average Income	0.023	0.134	0.226	0.281	-0.284	-0.103
	(0.059)	(0.054)	(0.060)	(0.052)	(0.215)	(0.187)
Number of Observations	49606	46099	49606	46099	49606	46099

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn.’

Table B.4 US Regressions with Own Characteristics

	Samples							
	Females	Males	Females	Males	Females	Males	Females	Males
Dependent Variable	Overall Job Satisfaction				Stayers			
People			0.020	0.014			0.004	0.008
			(0.006)	(0.007)			(0.003)	(0.003)
Brains			0.069	0.050			0.031	0.002
			(0.009)	(0.009)			(0.004)	(0.005)
Brawn			-0.038	0.005			0.001	0.014
			(0.008)	(0.007)			(0.004)	(0.004)
Log Own Wage	-0.004	0.031	-0.002	0.033	0.027	0.027	0.027	0.027
	(0.006)	(0.008)	(0.006)	(0.008)	(0.003)	(0.004)	(0.003)	(0.004)
Own Hours	-0.000	0.002	-0.004	0.001	0.018	0.015	0.017	0.016
	(0.009)	(0.010)	(0.009)	(0.010)	(0.004)	(0.005)	(0.004)	(0.005)
College Graduate	-0.010	0.010	-0.009	0.012	-0.032	-0.007	-0.031	-0.009
	(0.014)	(0.021)	(0.014)	(0.021)	(0.010)	(0.015)	(0.010)	(0.015)
Number of Children	0.018	0.008	0.018	0.008	0.008	0.001	0.008	0.001
	(0.008)	(0.008)	(0.008)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
Married	0.025	-0.016	0.023	-0.016	0.011	0.016	0.011	0.016
	(0.012)	(0.014)	(0.012)	(0.014)	(0.006)	(0.008)	(0.006)	(0.008)
Number of Observations	79312	63546	79312	63546	79312	63546	79312	63546

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled 'people' 'brains' and 'brawn.'

Table B.5 British Regressions with Own Characteristics

	Samples											
	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females	Males
Dependent Variable	Overall Job Satisfaction				Satisfaction with Work Itself				Stayers			
People			0.029	0.039			0.062	0.048			0.035	0.030
			(0.012)	(0.011)			(0.016)	(0.011)			(0.011)	(0.010)
Brains			0.027	-0.012			0.034	-0.020			0.014	-0.004
			(0.015)	(0.013)			(0.020)	(0.016)			(0.019)	(0.011)
Brawn			-0.065	-0.025			-0.066	-0.026			-0.048	0.014
			(0.016)	(0.014)			(0.019)	(0.014)			(0.015)	(0.011)
Log Own Wage	0.070	0.184	0.071	0.186	0.013	0.094	0.016	0.096	0.004	0.014	0.005	0.015
	(0.023)	(0.026)	(0.023)	(0.026)	(0.021)	(0.027)	(0.021)	(0.026)	(0.011)	(0.014)	(0.011)	(0.014)
Own Hours	-0.002	-0.002	-0.002	-0.003	-0.002	-0.002	-0.002	-0.002	0.001	0.000	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
College Graduate	0.044	-0.001	0.041	0.007	0.096	0.102	0.083	0.111	0.044	0.029	0.039	0.034
	(0.080)	(0.077)	(0.079)	(0.078)	(0.093)	(0.058)	(0.091)	(0.059)	(0.040)	(0.054)	(0.040)	(0.053)
No of Children	0.040	0.019	0.041	0.019	0.030	0.017	0.030	0.016	0.003	0.010	0.003	0.010
	(0.012)	(0.011)	(0.012)	(0.011)	(0.013)	(0.011)	(0.013)	(0.011)	(0.005)	(0.006)	(0.005)	(0.006)
Married	0.042	-0.020	0.044	-0.021	-0.029	-0.062	-0.025	-0.063	0.010	0.013	0.013	0.013
	(0.024)	(0.026)	(0.024)	(0.026)	(0.024)	(0.028)	(0.024)	(0.028)	(0.015)	(0.016)	(0.015)	(0.016)
Number of Obs.	36213	30098	36213	30098	36213	30098	36213	30098	35007	29178	35007	29178

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn.’

Table B.6 Russian Regressions with Own Characteristics

	Samples							
	Females	Males	Females	Males	Females	Males	Females	Males
Dependent Variable	Overall Job Satisfaction				Stayers			
People			0.042	-0.028			0.005	0.004
			(0.017)	(0.015)			(0.008)	(0.014)
Brains			-0.019	-0.052			-0.011	-0.012
			(0.018)	(0.018)			(0.006)	(0.010)
Brawn			-0.084	-0.026			-0.015	-0.005
			(0.018)	(0.017)			(0.009)	(0.012)
Log Own Wage	0.198	0.183	0.196	0.185	0.001	0.008	0.001	0.009
	(0.014)	(0.014)	(0.014)	(0.014)	(0.007)	(0.011)	(0.007)	(0.011)
Own Hours	-0.002	0.002	-0.000	0.002	0.000	-0.001	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
College Graduate	0.138	0.031	0.138	0.031	0.027	0.054	0.027	0.056
	(0.089)	(0.072)	(0.089)	(0.072)	(0.023)	(0.058)	(0.023)	(0.058)
No of Children	0.085	-0.012	0.083	-0.012	-0.042	0.016	-0.042	-0.016
	(0.035)	(0.030)	(0.035)	(0.030)	(0.026)	(0.021)	(0.026)	(0.021)
Married	-0.007	0.069	-0.007	0.070	0.005	0.033	0.005	0.032
	(0.032)	(0.049)	(0.032)	(0.049)	(0.018)	(0.026)	(0.018)	(0.026)
Number of Observations	24116	19104	24116	19104	18714	14152	18714	14152

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn.’

Table B.7 Regressions with Wage-hours elasticity

	US – NLSY		Britain – BHPS		Britain – BHPS		Russia – RLMS	
	Females	Males	Females	Males	Females	Males	Males	Females
Dep. Variable	Overall Job Satisfaction		Overall Job Satisfaction		Satisfaction with Work Itself		Overall Job Satisfaction	
Wage-hours elasticity	-0.134	-0.180	0.105	-0.033	0.146	-0.001	-0.004	-0.025
	(0.020)	(0.023)	(0.051)	(0.036)	(0.068)	(0.041)	(0.017)	(0.016)
People	0.019	0.007	0.036	0.021	0.074	0.036	0.040	0.002
	(0.006)	(0.006)	(0.010)	(0.009)	(0.015)	(0.010)	(0.017)	(0.016)
Brains	0.064	0.027	0.028	-0.010	0.026	-0.012	-0.005	0.022
	(0.009)	(0.008)	(0.013)	(0.011)	(0.018)	(0.013)	(0.015)	(0.019)
Brawn	-0.034	-0.008	-0.041	-0.018	-0.047	-0.011	-0.068	-0.045
	(0.008)	(0.006)	(0.015)	(0.013)	(0.018)	(0.013)	(0.018)	(0.017)
Number of Obs.	91234	97638	49606	46099	49606	46099	35443	27117
Dep. Variable	Stayers		Stayers				Stayers	
Wage-hours elasticity	-0.072	-0.070	-0.024	-0.058			-0.001	0.002
	(0.011)	(0.013)	(0.050)	(0.034)			(0.007)	(0.008)
People	0.001	0.005	0.033	0.018			0.004	-0.003
	(0.003)	(0.003)	(0.011)	(0.009)			(0.007)	(0.012)
Brains	0.027	-0.008	0.022	-0.011			0.002	0.002
	(0.004)	(0.004)	(0.018)	(0.012)			(0.006)	(0.008)
Brawn	-0.002	0.010	-0.047	0.012			-0.011	-0.002
	(0.004)	(0.003)	(0.015)	(0.012)			(0.008)	(0.010)
Number of Obs.	91234	97638	48116	44862			23449	16792

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). For all countries the three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn.’

Table B.8 Regressions for College Educated Women

	US – NLSY		Britain – BHPS		Britain – BHPS		Russia – RLMS	
	All	College Educated	All	College Educated	All	College Educated	All	College Educated
Dep. Variable	Overall Job Satisfaction		Overall Job Satisfaction		Satisfaction with Work Itself		Overall Job Satisfaction	
People	0.021	0.026	0.028	0.020	0.063	0.055	0.050	0.048
	(0.006)	(0.013)	(0.010)	(0.015)	(0.014)	(0.019)	(0.019)	(0.025)
Brains	0.072	0.102	0.029	0.045	0.032	0.063	0.025	0.003
	(0.008)	(0.022)	(0.013)	(0.013)	(0.018)	(0.015)	(0.013)	(0.019)
Brawn	-0.031	-0.008	-0.046	-0.031	-0.053	-0.059	-0.044	-0.027
	(0.008)	(0.019)	(0.014)	(0.021)	(0.017)	(0.022)	(0.020)	(0.024)
Number of Obs.	91234	15606	49606	19839	49606	19839	35443	9203
Dep. Variable	Stayers		Stayers				Stayers	
People	0.002	0.012	0.033	0.016			0.008	0.009
	(0.003)	(0.006)	(0.010)	(0.015)			(0.007)	(0.013)
Brains	0.033	0.035	0.022	0.022			-0.001	-0.005
	(0.004)	(0.009)	(0.017)	(0.019)			(0.006)	(0.013)
Brawn	0.000	0.009	-0.044	-0.030			-0.002	-0.018
	(0.004)	(0.008)	(0.013)	(0.019)			(0.008)	(0.014)
Number of Obs.	91234	15606	48116	19208			23449	7841

Notes: All regressions also include age and age squared of the individual, a dummy denoting if the female is in blue collar work, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). For all countries the three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn’

Appendix C: Robustness to Chosen Weights

Our US analysis of the NLSY utilizes sampling weights that reflect that the NLSY79 oversampled Blacks, Hispanics, and the economically disadvantaged. In this appendix we show the corresponding unweighted results.

Our British analysis uses all 18 waves of the original sample of the British Household Panel Survey (BHPS), a longitudinal study of around 5,050 households and approximately 10,000 individuals that began in 1991. This sample was nationally representative of the population in Great Britain. We combine this with the Welsh extension from 1999 (about 1500 households), the Scottish extension from 1999 and the Northern Ireland sample, which was added in 2001 (about 1900 households). We make this decision to preserve as many data points as possible, however we document in this appendix results which are based on responses from the original nationally representative sample only. Additionally, we documented results from weighted regressions of the main BHPS sample, where the weights are the longitudinal weights described in Taylor et al (2010). These are the weights recommended for use in longitudinal analysis. This results in a significantly smaller sample size since these weights are only provided when an individual is present in all waves.

Our RLMS regressions use weights that allow for the complex design of the RLMS where many observations are derived from following the housing unit rather than the person, as well as having oversamples from the first wave to allow for forecasted attrition. In this appendix we document unweighted regressions.

The general pattern of results is not very much affected by the weights in all cases.

Table C.1 Regressions for Overall Job Satisfaction

	Samples							
	US – NLSY		Britain – BHPS		Britain – BHPS		Russia – RLMS	
	Unweighted		Original Sample Unweighted		Longitudinal Weights		Unweighted	
	Females	Males	Females	Males	Females	Males	Males	Females
People	0.023	0.012	0.023	0.025	0.030	0.029	0.040	-0.037
	(0.005)	(0.005)	(0.012)	(0.010)	(0.016)	(0.015)	(0.025)	(0.036)
Brains	0.080	0.049	0.013	-0.004	0.032	-0.011	0.013	0.012
	(0.007)	(0.006)	(0.016)	(0.014)	(0.023)	(0.022)	(0.018)	(0.021)
Brawn	-0.044	-0.006	-0.039	-0.014	-0.032	-0.015	-0.039	-0.016
	(0.006)	(0.005)	(0.016)	(0.014)	(0.023)	(0.023)	(0.026)	(0.027)
Number of Observations	91234	97638	34024	32147	18863	15950	35443	27117

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn’

Table C.2 Regressions for Satisfaction with Work Itself (BHPS)

	Original Sample Unweighted		Longitudinal Weights	
	Females	Males	Females	Males
People	0.064	0.046	0.072	0.057
	(0.016)	(0.011)	(0.019)	(0.017)
Brains	0.016	-0.012	0.028	-0.014
	(0.019)	(0.015)	(0.026)	(0.020)
Brawn	-0.046	-0.004	-0.040	-0.010
	(0.019)	(0.015)	(0.026)	(0.023)
Number of Observations	34024	32147	18863	15950

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled ‘people’ ‘brains’ and ‘brawn’

Table C.3 Stayer Regressions

	Samples							
	US – NLSY		Britain – BHPS		Britain – BHPS		Russia – RLMS	
	Unweighted		Original Sample Unweighted		Longitudinal Weights		Unweighted	
	Females	Males	Females	Males	Females	Males	Females	Males
People	0.005	0.010	0.032	0.016	0.026	0.022	-0.008	-0.007
	(0.002)	(0.002)	(0.011)	(0.010)	(0.012)	(0.012)	(0.006)	(0.009)
Brains	0.029	-0.000	0.016	-0.008	0.027	-0.001	-0.004	0.006
	(0.003)	(0.003)	(0.018)	(0.012)	(0.018)	(0.015)	(0.004)	(0.005)
Brawn	-0.003	0.014	-0.048	0.013	-0.052	0.025	-0.002	0.003
	(0.003)	(0.003)	(0.014)	(0.013)	(0.016)	(0.014)	(0.006)	(0.007)
Number of Observations	91234	97638	34024	32147	18863	15950	35443	27117

Notes: All regressions also include age and age squared of the individual, the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, time and area as well as individual fixed effects. Standard errors are two-way clustered (by individual and their occupation) and shown in parentheses. Models are estimated using xtivreg2. PBB factors are derived using the approach described in Heckman et al (2012). The three factors extracted can loosely be labeled 'people' 'brains' and 'brawn'

Appendix D: Details on our Schools Survey

We conducted a survey among students in Year 11 (about age 15 – 16) in two secondary schools in Greater London. The survey satisfied the ethics committee at the author's home institution. Schools were advised of the survey two months ahead of time. The students completed the surveys in an assembly hall on a day when one of us visited the school. All students who were present on the day participated, but were given the option to opt out or skip any question they like. No identifying information was gathered and the students were aware of this. We received 311 responses and dropped four who provided no sex information. The resulting dataset contains 157 males and 150 females.

After collecting some basic demographic information we presented students with a list of six pairs of occupations. We picked these occupations as follows: We started with the 303 SOC 2000 occupation codes from the 2001-2012 QLFS data we use in conjunction with the BHPS. We limited these to 76 occupations, which employ at least 20 % university graduates, have an average income above £22,000, with a share of men between 10% and 90%, and at least 1,000 observations for each occupation (to eliminate uncommon occupations). We eliminated a further 40 occupations by hand, which we felt the students would unlikely be familiar with, which closely duplicated another occupation, or which had unusual characteristics which might have dominated student's choices (like priest, fashion model, musician, firefighter).

We split the remaining 36 occupations into three groups by earnings, and then each of these into two further groups with high or low average hours. We chose pairs of six occupations by maximizing the determinant of the $X'X$ matrix consisting of the 'people' and 'brawn' factors within in the six groups, using 10,000 Monte Carlo draws of different occupation combinations.[23] We did not use the 'brains' factor in the optimization analysis since we found no strong differences in terms of the sorting of men and women for that factor in Table 3, so care less about its contribution to the

[23] This is a D-optimal design. An A-optimal design, minimizing the trace of the $X'X$ matrix, yields the same choices. Either one should lead to favorable standard errors.

variance of the estimates. We also start with a small group of occupations and are worried about the degrees of freedom for the optimization analysis.

The logit models we run in Table 8 can be motivated by a standard random utility model with multiple choices. Suppose individual i chooses from J occupations. Utility for choosing occupation j is given by

$$y_{ij}^* = \beta' x_j + \mu_{ij}$$

where x_j is a vector of occupation characteristics. Define

$$P_{ij} = prob[y_{ij}^* = \max(y_{i1}^*, y_{i2}^*, ...y_{iJ}^*)]$$

Under IIA, for any pair of occupations, say 1 and 2, the probability that the individual will choose the first occupation, denoted as $y_i = 1$, is

$$prob[y_i = 1] = \frac{P_{i1}}{P_{i1} + P_{i2}} \ .$$

If the μ_{ij} have a type 1 extreme value distribution, then

$$P_{ij} = \frac{e^{\beta' x_j}}{\sum e^{\beta' x_j}}$$

and hence

$$prob[y_i = 1] = \frac{e^{\beta' x_1}}{e^{\beta' x_1} + e^{\beta' x_2}} = \frac{e^{\beta'(x_1 - x_2)}}{1 + e^{\beta'(x_1 - x_2)}}.$$

The multinomial choice problem of ranking the J occupations gives rise to a binary logit model for any pair of choices, with the difference in occupation characteristics $x_1 - x_2$ as regressors.

Skill Match Variables

In the initial demographic section of the survey, we asked students to list the subjects they are currently taking (the students are preparing for their GCSE exams, typically taking about ten subjects, and have some choice which subjects they study for these exams). We also asked the students to tell us which is their best subject. We classified the best subject into 10 categories (Design and Technology, Geography/History/Politics, PE, Arts and Music, Business, Computing, Languages including English, Math, Sciences, and Psychology), in addition to recording all, none, and don't know.

In order to relate the answers to these questions to the subject knowledge required in a specific occupation, we extracted all individuals with a college degree or higher from the American Community Survey (ACS) for 2009 to 2015. In these years, the ACS asked college graduates for the field of their degree. We classified degrees into the same 10 categories. We think of the distribution of degree fields within an occupation as reflecting the skill requirements of that occupation. In order to create variables reflecting the match between students' talents and the skill requirements of an occupation, we start by calculating the distribution of degree fields among employees in all occupations, not just our 12.We then produce the analogous distribution in each of the 12 occupations. Next, we calculate the log-odds ratio of an occupation's field frequency compared to the field frequency in overall employment. For example, 28% of all employees have science degree. If 35% of employees in occupation j have a science degree, then the log-odds ratio for science in the occupation is ln(0.35/0.28) = 0.22. Positive values indicate that a field is over-represented in an occupation compared to the workforce at large. We do this in order to deal with the fact that some degree fields are common (Business, Sciences) while others are rare (PE, Arts and Music). In particular, few individuals obtain a college math degree (1.4%) but math is an important subject in secondary school and also an important skill in many occupations. Despite the fact that mathematicians are rare, we basically claim that the relative abundance of mathematicians in an occupation tells us something about the math intensity of this occupations. Sensibly, we find, for example, that insurance underwriters are more math oriented (0.85) than public relations specialists (-1.03), and that both math and science graduates are common among physicists (1.15 for math and 1.14 for sciences).

We define two measures of a skill match for a student-occupation pair, a continuous one and a discrete one. The continuous measure is simply the log-odds ratio for the field which is the student's best subject. The discrete measure is 1 if the log-odds ratio is positive, i.e. when individuals with degrees in this field are overrepresented in the occupation. We use the differences between the two occupations in the pair as regressors just as we do with the factor scores for job content.

Table D.1 repeats the sorting regressions from the British QLFS for all occupations from Table 3 in the main text. It compares these to similar regressions when the sample is restricted to the twelve occupations (six occupation pairs) which we presented to the students (column 2). The coefficient on ‘brawn’ turns from positive to negative in this subsample of occupations. The switch is driven by the occupation pair accountant and health services manager. Column (3) omits this pair from the six, showing results for the five remaining pairs. The ‘brawn’ coefficient turns positive again.

Table D.1: The Relationship Between the Share of Males and People, Brains, and Brawn in the British QLFS

	Samples		
	All Occupations (1)	Six Occupation Pairs (2)	Five Occupation Pairs (3)
People	-0.057	-0.178	-0.077
	(0.013)	(0.068)	(0.071)
Brains	-0.029	-0.057	0.065
	(0.022)	(0.074)	(0.075)
Brawn	0.102	-0.059	0.245
	(0.018)	(0.114)	(0.121)
Number of Observations	4266356	94391	57042

Notes: All regressions also include the averages of the log hourly wage, hours, fraction college graduates, and age in the occupation, as well as time and area effects. Standard errors are clustered by occupation. The six occupations pairs are: Public relations specialist and Mental Health Worker; Architectural Drafter and Retail Buyer; Town Planner and Insurance Underwriter; Physicist and Art Director; Accountant and Health Services Manager; Architect and Human Resources Manager. The five pairs exclude Accountant and Health Services Manager.

Appendix E: Cross-Walking Across Samples

US Analysis

We pool the 1980, 1990, and 2000 Census and 2001-2014 American Community Survey (ACS) IPUMS samples to calculate the occupational averages. We draw on the crosswalks created by Autor and Dorn (2013) and Dorn (2009) in order to create a consistent set of occupations across the 1980, 1990, and 2000 Census codes. Call this consistent code occ1990dd. We then calculate occupation averages for each occ1990dd.

In the NLSY 1980 and 2000 Census codes are used for the 1982-2000 and 2002-2014 samples respectively. We assign an occ1990dd code to each observation using the crosswalks created by Autor and Dorn (2013) and Dorn (2009) and match the occupational averages derived from the Census based on these codes.

Our main analysis uses ONET version 5, whose items on activities and context are linked to Standard Occupation Codes (SOC) 2000. We start by using a Bureau of Labor Statistics (BLS) cross walk to assign a three-digit Census 2000 occupation code to each ONET occupation. We then use the Autor and Dorn (2013) and Dorn (2009) crosswalk and match the three-digit Census 2000 occupation codes to occ1990dd. Based on occ1990dd we match the ONET variables to the 1980, 1990, and 2000 Census and 2001-2014 ACS samples. We calculate the three latent factors 'people,' 'brains,' and 'brawn' (PBB) in this sample. Subsequently, we match the PBB variables to the NLSY data by occ1990dd.

British Analysis

We calculate the occupation averages at a three-digit occupation level using the 1993-2012 Quarterly Labor Force Survey (QLFS). 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 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. We then calculate the occupation averages based on SOC00. Using the cross walk, we also assign a SOC00 code to the BHPS data for years before 2000.

We then match the occupation averages to the BHPS data based on SOC00.

Our main analysis uses ONET version 5, whose items on activities and context are linked to US Standard Occupation Codes (SOC) 2000. We match the US SOC00 codes in the ONET data directly to the British SOC00 using a crosswalk provided by Anna Salomons. We then match the ONET items to the QLFS using the British SOC00 codes. The three latent factors 'people,' 'brains,' and 'brawn' (PBB) are calculated using this data. We then match the PBB factors for each occupation to the BHPS data using the British SOC00 codes.

Russian Analysis

Pooling the ISSP 1995-2011, the ESS 2002-2012 and the RLMS 1994-2012, we calculate the occupation averages based on the three digit ISCO 2000 codes. We match the items from ONET version 5 to ISCO 2000 utilizing a crosswalk provided by the BLS between SOC 2000 and ISCO 2000.