Using mobility networks to understand occupational sex segregation¹

Per Block

Department of Sociology and Leverhulme Centre for Demographic Science, University of
Oxford, United Kingdom.

per.block@sociology.ox.ac.uk.

Abstract

Decades after the beginning of the gender revolution, most women and men still work in sex-typed occupations. This is a primary driver of the gender wage gap. Research describing the patterns of occupational sex segregation focuses on supposedly innate job characteristics that match gender stereotypical abilities and preferences, such as the use of mathematical skills or social skills, on income and status differences between occupations, and on organizational job characteristics, for example, the need to work long hours. However, beyond such occupational attributes, sex segregation is hypothesized to exhibit emergent patterns that are linked to the interdependent job mobility of women and men, in particular, men selectively leaving feminizing

¹ The author thanks the attendants of the RC28 meeting in Bern, the EUSN conference in Mainz, and the attendants and organizers of the research seminars at the University of Stavanger, Nuffield College, UC Irvine, UC Los Angeles, and UC Santa Barbara, as well as the members of the network research groups in Zürich, Groningen and Melbourne for their valuable suggestions on the project. The author further thanks Emily Murphy for initial discussions on the conception of this study, as well as Garry Robins and Christoph Stadtfeld for collaboration on the method and software implementation. Zsofia Boda, Greg Goldsmith, James Hollway, Patricia Lopes, Isabel Raabe, and Andras Vörös provided valuable feedback on earlier versions of this manuscript. Per Block is supported by the Leverhulme Centre for Demographic Science. The majority of this work was conducted during postdoctoral employment at the ETH Zurich.

occupations. Developing new tools inspired by statistical network research, and using representative, longitudinal data that contain detailed occupational mobility from the UK between 2000 and 2008, this replacement mechanism is analyzed. I show that 19-28% of observed sex segregation is linked to this emergent phenomenon in a statistical model that disentangles the various predictors of the allocation of women and men to different occupations. This makes it the most important predictor of segregation in contrast to concurrently modelled explanations based on occupational characteristics.

Introduction

While gender equality in the labor market has advanced significantly since the 1960s, women and men still tend to work in different sectors and occupations. Indeed, contemporary labor markets are commonly characterized as ‘hyper-segregated’; around 50% of employed women would need to switch occupations to achieve a balanced sex composition (Charles and Grusky, 2005; Levanon and Grusky, 2016). The most important consequence of this striking level of occupational sex segregation is that it forms the basis of the large differences in earnings between women and men – the gender wage gap (Blau and Kahn, 2017; Charles, 2011), which is currently around 15% of median income in most western economies (OECD, 2019). Sex segregation also has consequences on the level of occupations; work that is done mainly by women is devalued in terms of income and prestige (Cohen and Huffman, 2003). Finally, taking an economic perspective, sorting individuals to tasks based on sex rather than ability is inefficient and leads to productivity loss for companies and economies at large.

A considerable literature discusses the causes of occupational sex segregation. Neoclassical economics, of course, views sex segregation as a consequence of rational, utility maximizing behavior (Becker, 1985). Due to initial differences rooted in women’s biological ability to bear children, women and men specialize in different tasks over the life-course. This leads, first, to

women spending more time on care work outside the labor market, and second, to women working in different occupations than men when they do paid work. Sociologists, of course, have a more nuanced view.

One broad source of sex segregation proposed in the sociological literature is discrimination against women in hiring, firing, and promotion decisions (Ridgeway, 2011; Rivera, 2020). Discrimination can have different reasons. Statistical discrimination arises from the presumption that women on average have a higher tendency to leave the labor market – even temporarily – to raise children or do other unpaid care work, making them less attractive as employees (Bielby and Baron; 1986). Employers might also have a “taste” for hiring men into better paying, more prestigious, and powerful positions, following their view of the role women and men should play in society, or following stereotypical views of innate abilities and status worthiness by sex (Ridgeway, 2001). Discrimination can also be viewed as a means to maintain power relations between women and men within and outside of the workplace. One particular approach that elucidates the link between discrimination, power, and sex segregation in occupations is the queuing model of the labor market (Reskin and Roos; 1990). The queuing model describes how preferences of employers for hiring male workers and relative desirability of positions determine who gets what kind of jobs. This allows to understand the allocation of jobs as a conflict between groups, enabling to describe, for example, the occupations in which changes in sex composition are likely to occur.

Next to the various forms of discrimination that sorts women into *worse* occupations than men, there are also essentialist mechanisms that sort women into qualitatively *different* occupations. They are based on the belief that women and men have innate interests and abilities that allow them to excel at different skills (Ridgeway, 2011; Levanon and Grusky, 2016). This makes them suitable for occupations in which such skills are used, for example, women are believed to be better at caring tasks while men are better at mathematical tasks. These beliefs are instilled into individuals by various socialization processes throughout the life course. Sources of socialization

include role models in the family, interactions with teachers in school and with peers in the free-time, exposure to popular culture and media that perpetuate stereotypes, and broader institutional frames that facilitate sex-typed entrance into specific occupations and social roles (e.g., Lauzen et al., 2008; Hilliard and Liben, 2010; Wood and Eagly, 2012). Furthermore, held stereotypes about female and male abilities are reinforced by cognitive processes that selectively make individuals remember and focus on evidence in line with stereotypes and discount or forget counterevidence (Brown, 2010). Once individuals hold such beliefs, sorting mechanisms into occupations work on the supply and demand side. Employers preferably hire individuals of the sex that matches the gender stereotype of an occupation. At the same time, women and men make educational and occupational choices throughout their entire life course that allow them to do tasks typically associated with their gender (Cech, 2013). This life-long aspect of the effect socialization has on careers is demonstrated by the “revolving doors” metaphor (Jacobs, 1989). It describes how even women that already work in occupations dominated by men continue to leave these occupations at high rates throughout their careers. This leads to stability in sex composition despite relative permeability on the individual level.

Regardless of their explanatory approach, all introduced mechanisms causing sex segregation attempt to explain empirically observed correlates of the gender composition of occupations. They explain why female-dominated and male-dominated occupations have different characteristics; for example, men work in better paying occupations, women work in occupations that use care skills, men work in self-employed occupations, and women work in occupations that facilitate part-time arrangements. At the same time, empirical patterns are used to support or contradict theoretical propositions – for example, if we find that women are under-represented in high-skilled, high-prestige occupations in a particular country despite few human capital differences, we could use this as evidence in line with discrimination. In consequence, advances in theories of occupational sex segregation need detailed and comprehensive descriptions of what characterizes occupations that are mainly done by women or by men.

Such a descriptive task of describing the characteristics of jobs women and men do is the empirical goal of the current study. I go beyond current literature that is limited to occupational characteristics in its description of sex segregation and focus on the emergent sorting of women and men into occupations and the resulting patterns that were neglected in statistical studies so far. Such emergent sorting patterns tend to represent the social interdependence of female and male occupational choices. In particular, I analyze the longitudinal movement of women and men through occupations and show that men tend to leave occupations that are entered mainly by women. Colloquially, the pattern can also be described by what seems like a tautology: men work in male-dominated occupations because other men work in this occupation². I find empirical support for this self-reinforcing process by looking at the complementary case: men leave occupations that are shifting towards being female dominated.

I show that this pattern is present even after controlling for a large battery of occupational characteristics that are traditionally invoked to describe female- and male-type occupations. Furthermore, it is a powerful statistical predictor of women's and men's movement through the labor market. Thus, it stands in contrast to established explanations of sex segregation that mostly focus on occupational characteristics and rewards. These emergent patterns help to explain the seemingly idiosyncratic gender composition of some occupations that deviate from predictions based on occupational characteristics, pertaining, for example, to swings in sex composition over short periods of time or different sex ratios of objectively similar occupations.

With this in mind, I note that the goal of the study is not to disentangle the different supply- and demand-side mechanisms that lead to sex segregation. A broad literature has engaged in this goal already. Instead, I provide the statistical predictors of the sex composition of detailed occupations beyond the current state of the literature. In the next paragraphs, I first identify the different types of occupational characteristics that correlate with gender composition, and subsequently outline

² Note that I describe the outcome of such a process without making the associated causal claim.

why patterns of emergence between women's and men's movement might be additionally necessary to statistically model the sex composition of occupations.

Correlates of occupational sex composition

There are three types of characteristics to describe occupations that are linked to the sex ratio of occupations: vertical predictors, horizontal predictors, and organizational predictors.

Distinguishing occupations *vertically*, men are systematically found in jobs with a higher income and jobs that have a higher status (Charles and Grusky, 2005; England, 2010). This is predicted by the various forms of explanations that describe discrimination against women in hiring and promotion based on stereotypes and role prescription. Additionally, a vertical segregation is predicted by the arguments rooted in the neoclassical economists view of sex segregation.

Horizontal segregation describes that women and men are found in occupations that have different work content, i.e., use different skills. This is based in the belief of a systematic difference in their abilities and preferences based on socialization. In particular, men work in jobs that require male-typed skills, including physical strength, mathematical ability, robustness, technical skills, authority, and problem-solving skills. The area of manual occupations is also more associated with male-typed occupations. Women work in jobs that need female-typed skills, including sociability, fine motor skills, and creativity (Levanon and Grusky, 2016; Cech, 2013; Leslie *et al.*, 2015).

Finally, the stereotypical division of labor in heterosexual couples often results in women doing a majority of housework and child rearing, making employment in jobs with particular *organizational arrangements* more suited. This division of labor is predicted by various discussed mechanisms, including stereotypical beliefs about the roles women and men should have in society, beliefs about the innate talent that women are better at caring tasks, and the economic argument that it is rational for women to specialize in household work. Consequently, potential employment of women is limited to jobs with organizational arrangements that are compatible with family life (Becker, 1985; Eccles, 1994; England, 2010). This includes occupations that do

not expect to work long hours or overtime, that can be done in part-time, that are not self-employed, and preferably in the public sector. The level and extent of education required for an occupation is related to the gender composition of an occupation for similar reasons.

These three types of occupational characteristics – vertical, horizontal, and organizational – are widely analyzed and supported by empirical studies that use representative survey data and correlate job characteristics with sex composition (Charles and Grusky, 2005; Percheski, 2008; Levanon and Grusky, 2016).

However, striking empirical regularities cannot be described sufficiently by these characteristics. First, specializations within disciplines that are characterized by very similar income, skill requirements, and work conditions often have hugely disparate sex compositions, e.g., some pairs of specialties among medical doctors or lawyers (Ku, 2011; Seron *et al.*, 2011). While small differences in work content exist between specializations, for example, in medicine, these are minute in comparison to the similarities of these specializations that arise from similar education, labor relations and social standing, which would predict very similar sex compositions³.

Second, sex ratios in occupations sometimes change considerably over time. For example, primary school teachers, pharmacists, and accountants have shifted from male- to female-dominated in recent decades (i.e., feminized; Reskin and Roos, 1990; Murphy and Oesch, 2015; Pan, 2015). These swings in sex composition do not co-occur with changes in skill profiles or entry requirements and, thus, cannot be explained by changes in occupational characteristics. Consequently, other patterns beyond sex-typed occupational characteristics are likely needed for a comprehensive description of sex segregation.

One mechanism that can explain these regularities has been described in previous theoretical work. It is suggested that men avoid and leave occupations and specializations in case the sex

³ Note that in some contexts there are considerable differences between income of medical specializations; however, even the worst paying medical specializations tend to be at the top of income distribution in all contexts.

composition swings towards being female dominated, named a pollution theory of discrimination (Reskin and Roos, 1990; England, 2008; Goldin, 2014). Discussions of this theory propose different potential mechanisms. On the one hand, men leaving feminizing occupations might be a result of men's psychological discomfort to work in what is perceived as a 'woman's job'. This is related to the essentialist differentiation of some tasks as female-typed and other tasks as male-typed. Men that have internalized such a differentiation might start to perceive that they are working in the 'wrong', meaning sex atypical, occupation in case the sex composition of others doing their work changes and they search for a different line of work. On the other hand, female dominated occupations tend to have lower prestige and bestow less social status on their incumbents. Further, wage penalties often occur when occupations feminize, especially when female proportion crosses thresholds of substantially more than 50% (England, 2008; Goldin, 2014; Murphy and Oesch, 2015; Pan, 2015). Thus, men might leave feminizing occupations, because they want to avoid work that has the reduced income and prestige associated with a sex-ratio leaning towards female dominated. In anticipation of decreasing wage and status, or the fear that they could be underpaid already, they move away from these occupations.

At this point we are not so much interested in which of the outlined mechanisms underlies occupational sex segregation, but rather, whether it can be generalized and observed at scale that men do selectively leave feminizing occupations. Regardless of which mechanism might be at work, the interest in this study is whether there is a direct, contemporaneous relation between female inflow and male outflow of occupations. Thus, the interest is *not* whether a female inflow changes occupational characteristics (such as devaluation or changing work content) which leads to a male outflow. This is because the contemporaneous relation describes an emergent phenomenon that has social basis and materializes before occupations change. In contrast, the lagged, mediated relation can be largely reduced to changes in occupational characteristics. In other words, I analyze what happens in *feminizing* occupations, not in *feminized* occupations.

In case a relation between female inflow and male outflow can be observed, a second question arises concerning the extent to which this produces sex segregation. How much of the observed segregation can be statistically attributed to this replacement pattern in comparison to traditional occupational characteristics used in the literature? For both questions – if replacement exists, and if yes, how important it is – we need to use advanced, non-standard statistical methodology.

The outlined research questions require a move to longitudinal data since the main pattern of interest is explicitly about sex-specific mobility that inherently unfolds over time. This deviates from contemporary descriptions of female-typed and male-typed occupations that rely on a cross-sectional correlation of occupational characteristics with sex composition. The operationalization of the research question for our purpose is then: do men leave occupations beyond what is predicted by occupational characteristics and only based on changes in the sex composition? A first, intuitive approach towards this question could be to consider how the changes in female ratio in an occupation between time-points t_1 and t_2 predict the mobility decision of men in the period t_2 to t_3 . This would satisfy the requirement of temporal contiguity (Granger causality) linking female inflow to male outflow. However, this approach has two important limitations compared to establishing a contemporaneous outflow of women and inflow of men over the same period t_1 to t_2 . First, as outlined above, the pattern that I intend to describe is indeed about the contemporaneous, emergent replacement, not the substitution over time that might be related to intervening changes in occupational characteristics. Second, a temporal contiguity approach requires twice as much data, because changes over the period t_2 to t_3 would need to be predicted by changes over the period t_1 to t_2 . Thus, for theoretical and practical reasons I analyze whether the descriptive pattern that men leave occupations that are feminizing in the same period can be observed. Notably, this question goes beyond what is possible using standard statistical methodology. This is because answering these questions regarding emergent, contemporaneous patterns requires analyzing the movement of men through the labor market *statistically dependent*

on women's movement, which violates a cornerstone of modern statistical methods: the assumption of independent observations (Snijders, 2011).

The current study

In this study, I overcome this crucial limitation by developing and using statistical methods and concepts originating in network research that do not require independent observations. I use these to analyze whether, and to what extent men leave occupations when more women enter, controlling for a comprehensive set of occupational characteristics. In that, I am able to test both, the importance of occupational characteristics – following existing research – and the prevalence of endogenous mobility patterns – extending previous research. To empirically relate the mobility of different sexes to each other, I define the concept of a mobility network, which represents the structure of occupational changes within a society. In a mobility network, nodes represent occupations and the edges denote the number of individuals that change their occupation from the origin to the target node within a given period. Mobility networks further include self-ties (individuals that remain in their occupation). An important feature of mobility networks [compared to classic network conceptions, e.g. in Müller *et al.* (2018), or Lewis *et al.* (2012)] is that edges are characterized by attributes – in this study by the sex of mobile individuals.

I construct a mobility network based on data describing occupational changes in Great Britain from 2000-2008 (N = 5967, 50.9% female). Great Britain shares important labor market characteristics with both the US and continental Europe and provides data of outstanding richness and quality. The mobility network here is enriched with harmonized information about the occupational nodes from two data sources on work conditions (e.g., income, part-time work) and on detailed skill use in the occupations. Together, this data can provide insights into all facets of sex-specific occupational mobility that relate to *vertical*, *horizontal* and *organizational* characteristics of occupations, as well as the *interdependent mobility* of women and men through occupations.

The mobility network is analyzed with a newly developed variant of an exponential family model, which applies recent advances in statistical network analysis to the study of mobility. While exponential family models are commonly used in the analysis of occupational mobility⁴, the proposed network model enables answering new questions: it allows for the analysis of how the mobility decisions of different individuals are *inter*-dependent. In mainstream statistical models this is prohibited by the explicit assumption that observations are *in*-dependent. In particular, I model men's tendency to move out of occupations *dependent* on the share of women moving in, therefore representing the interdependent replacement pattern of interest. I simultaneously model how women and men differentially move between, or stay in certain occupations, based on occupational characteristics.

The outlined, interdisciplinary network-based research design combines key advantages of classical designs from sociology and of new developments in network science for the current question. First, as common in quantitative sociological analyses the statistical approach allows the simultaneous testing of multiple competing explanations in one model. Thus, any tendency of men to leave feminizing occupations found in the analyses is *net of* modelled occupational characteristics. Second, formulating a comprehensive model of female and male occupational mobility that includes all known predictors of sex segregation enables the *quantification* of how much segregation is statistically predicted by men selectively leaving feminizing occupations using counterfactual simulations. Furthermore, the importance of this pattern can be assessed relative to the predictive power of different occupational characteristics. Third, applying new concepts and methods inspired by statistical network modelling to the sociological study of occupational mobility allows to analyze patterns in occupational mobility that result from social processes. I overcome independence assumptions routinely made in mainstream quantitative

⁴ The log-linear models classically used in mobility analyses are exponential family models. The relation of these models to the proposed one becomes clear when either is reformulated as a multinomial logit model, see Breen (1994), and Butts (2007).

sociological studies that reduce individuals to atomistic actors. Modelling the dependence between occupational transitions of different individuals enables large-scale testing of empirical patterns related to exciting theoretical propositions about the relations between sexes.

Data

The outlined analysis uses mobility data from the United Kingdom (UK), which is a substantively interesting case for which comprehensive, high quality data sources are available. Three datasets are combined to obtain detailed data on occupational mobility, as well as comprehensive data on occupational characteristics.

Mobility data between the years 2000 and 2008 on occupational transitions comes from the British Household Panel Survey (BHPS, 2010), a nationally representative, longitudinal study. The first mobility transition of all respondents of working age (25-63) that are part of the active work force is recorded in the mobility network. Individuals remaining in the same occupation for more than three consecutive observations are coded as staying in their occupation. Picking the outcome of the first recorded labor market spell (i.e., a transition or immobility) of BHPS respondents over a defined period provides (almost⁵) a random sample of independent labor market transitions of British workers at the beginning of the century. For the current analysis, moves to retirement and entries to the labor market are excluded from the analysis for practical purposes. However, robustness checks show that this exclusion does not alter the substantive results. The resulting mobility network contains 5967 (3037 female) individuals, of which 3727 (1843) are mobile. Details for data preparation of the mobility network are presented in the appendix.

Data on the occupations in the mobility network are enriched with a comprehensive battery of 25 occupational characteristics using the British Labour Force Survey (LFS, 2009) and the O*NET

⁵ within the randomness limits provided by the BHPS.

database (Peterson *et al.*, 1999). These characteristics represent the current theoretical knowledge on occupational characteristics that influence sex composition (Becker, 1985; Eccles, 1994; Kaufman, 2002; Charles and Grusky, 2005; Percheski, 2008; Gorman and Kmec, 2009; Blau *et al.*, 2010; England, 2010; Cech, 2013; Cha, 2013; Koellinger *et al.* 2013; Leslie *et al.*, 2015; Levanon and Grusky, 2016; see appendix). The LFS is a representative, quarterly survey designed specifically to provide insights into the UK labor market. It is part of the European Labour Force Survey. Data from 186,605 respondents between July 2005 and June 2006 provides information on the following dimensions for each occupation: average hourly income; percentage of workers in this occupation (%) working in the public sector; % working self-employed; % working full-time; % regularly working more than 40h per week; % regularly working more than 50h per week; % regularly working more than 60h per week; % regularly working any overtime; % regularly working more than 5h overtime per week; % regularly working more than 10h overtime per week; % that completed an apprenticeship; % that completed higher education; % that completed a university degree; and % that received job-related training in the previous three months.

The Occupational Information Network (O*NET) data is the most comprehensive data on skill use in detailed occupations currently available (Peterson *et al.*, 1999). The O*NET data combines information of occupational experts and job incumbents and is constantly extended. Using scales developed previously to capture relevant dimensions of sex-typing of occupations (Levanon and Grusky, 2016), the O*NET data provides occupation wise information on the occupation specific use of strength, sociability, fine motor skills, mathematical skills, robustness, technical skills, authority, problem solving skills, creative skills and whether it is a manual occupation. Additional data of occupational status measured by the Standard International Occupational Prestige Scale (SIOPS) comes from (Ganzeboom *et al.*, 1992). The appendix presents details for data preparation of all 25 occupational characteristics, variable choice and usability.

All data sources are harmonized to the same occupational classification. I rely on so-called micro classes, modified from Jonsson *et al.* (2009). They represent meaningful, small groups of

occupations that are well represented in the UK labor force. The final categorization scheme includes 59 occupations (for full coding schemes and cross-walks, see appendix). These harmonizing steps result in representative occupational mobility data of 5967 individuals between 59 occupations, including detailed characteristics of the occupations under analysis on 25 theoretically relevant dimensions.

Methods

The mobility network is analyzed with a novel variant of an exponential family model that allows incorporating flexible predictors that can be based on individual and occupational characteristics, as well as emergent structures. In its formulation it combines desirable properties of log-linear models classically used in the analysis of occupational mobility with advances from statistical network models. Compared to mainstream log-linear models, the proposed approach allows for the inclusion of parameters that represent the *inter*-dependence of observations. Thus, the probability to observe an emergent mobility system is more than the sum of independent parts. The developed methodology goes beyond what is currently possible in established statistical network methods (Butts, 2007; Lusher *et al.*, 2013). In comparison to exponential random graph models (ERGMs), in particular, the model accommodates (i) weights on network ties without specifying a distribution of counts in dyads *a priori*, (ii) weighted self-ties, and (iii) attributes of edges, which represent the number of mobile and non-mobile individuals, as well as their sex, respectively, neither of which is possible in standard ERGM analyses.

Intuitively, when estimating parameters from the model these provide insights into empirical regularities in occupational mobility, for example ‘are men more likely than women to be immobile?’, ‘are mobile women more likely to move to public sector occupations than mobile men?’, or ‘are men more likely to leave occupations that have a higher inflow of women compared to similar occupations with a lower inflow of women?’. The workings of the model are presented

in the short technical introduction that follows; it is extended in the appendix. However, readers with less interest in the detailed technicalities of the model might skip the rest of the section and can interpret the model results presented in the subsequent section in a first approximation as if it they came from a log-linear model with some unusual independent variables.

For those interested in the statistical formulation, the model works as follows. In line with previous approaches that focus on modelling data with dependencies between observations (Butts, 2007; Lusher *et al.*, 2013), the probability to observe a specific mobility network $\mathbf{x}$ is given by

$$Pr(X = \mathbf{x}) = \frac{\exp(Q(s(\mathbf{x}); \boldsymbol{\theta}))}{\kappa}$$

where X is a random state, and Q is a quality function that depends on the sufficient statistics $s(\mathbf{x})$ of the model and a statistical parameter $\boldsymbol{\theta}$. The normalizing constant κ ensures that all probabilities sum to one (see appendix). This is the standard formulation of exponential family models (including ERGMs and logistic regression models); the defining model features lie in the details outlined below. The quality function Q is a linear predictor. It allows incorporating the diverse variable-based and endogenous predictors (commonly termed independent variables) in the mobility model. In the analysis at hand it is defined as

$$\begin{aligned} Q(s(\mathbf{x}); \boldsymbol{\theta}) = & \theta_1 s_{loop}(\mathbf{x}) + \theta_2 s_{loop * fem}(\mathbf{x}) + \theta_3 s_{loop * fem.in}(\mathbf{x}) + \theta_4 s_{loop * male * fem.in}(\mathbf{x}) \\ & + \sum_v \left(\theta_{v,1} s_{loop * v}(\mathbf{x}) + \theta_{v,2} s_{loop * v * fem}(\mathbf{x}) + \theta_{v,3} s_{target * v * fem}(\mathbf{x}) + \theta_{v,4} s_v * homo.(\mathbf{x}) \right) \\ & + \sum_B \theta_B s_B * demand(\mathbf{x}). \end{aligned}$$

The first row contains three statistics associated to parameters $\theta_1 - \theta_3$ that represent (i) the general tendency of individuals to remain in occupations as opposed to leaving, i.e. a basic proclivity to mobility or immobility; (ii) whether this is different for women compared to men, and (iii) the general tendency to remain in occupations depending on the proportion of women moving into this occupation, respectively. The statistic associated to parameter θ_4 is of focal interest in this

study, modelling whether men specifically leave occupations depending on the proportion of women entering. The second row of the equation contains the parameters related to occupational characteristics v , such as whether it is a manual occupation, whether it requires physical strength, or whether it is particularly well paying. For each characteristic v , four parameters $\theta_{v,1} - \theta_{v,4}$ are included that model

- the tendency of individuals to stay in an occupation higher on v , i.e. whether certain occupational characteristics to generally discourage leaving (*stickiness*, $\theta_{v,1}$);
- whether this tendency differs for women compared to men, that is, whether women are more likely than men to stay in occupations with particular characteristics v ($\theta_{v,2}$);
- if women preferentially move to occupations higher in v in case they are mobile, i.e. the tendency of women that change jobs to move to occupations with different characteristics than men's mobility targets ($\theta_{v,3}$);
- and, whether individuals move between occupations similar on v , that is, between jobs with similar characteristics (*homophily*, $\theta_{v,4}$).

The non-technical interpretation of these parameters is illustrated in the Results section and the appendix. Finally, the last row of the equation contains fixed effects for all occupations B , ensuring correct representation of the column marginals of the mobility table. The model is conditional on the marginal row distribution and, consequently, needs no parameters to represent row margins (see appendix). Full interpretation for all model parameters and their mathematical formulation of the statistics $s_{...}(\mathbf{x})$ is presented in the appendix.

A note on estimation and simulation of this network model: given that the probability to observe a specific mobility transition depends on its relation to other mobility transitions, the likelihood of an observed network cannot be decomposed into a product of observation-wise probabilities as in traditional exponential family models (e.g. logistic regression). This makes the use of computationally expensive MCMC-based simulation methods necessary (Snijders, 2002). How

MCMC procedures are used to generate samples under a model, i.e., for simulation, and how these are used in maximum likelihood estimation is elaborated in the appendix.

Results

Descriptively, the analyzed mobility network presented in Figure 1 clearly shows that most women and men work in different occupations (denoted by relative node size) and switch between vastly differing sets of occupations (edge weight). While the labor market is largely segregated, there is still a broad distribution of sex composition over occupations, and a considerable number of occupations that have fairly balanced sex ratios. Furthermore, visual inspection suggests that men are especially likely to leave occupations that women move into (node coloring). Figure S1 gives an intuitive understanding of the overall mobility structure by showing the aggregate mobility network including labels that denote occupation titles.

----- Insert Figure 1 here -----

This network is analyzed with the statistical network model to detect regular patterns in women's and men's mobility. A forward model selection strategy determines which of the comprehensive battery of 25 occupational characteristics introduced in the Data section and denoted v in the Methods section known from previous research influences mobility (see appendix). Fourteen characteristics are retained in the model as significantly influencing at least one of four choices (indicated by the parameters $\theta_{v,1} - \theta_{v,4}$). These 14 characteristics are average income in occupation; % working in the public sector; % working self-employed; % working full-time; % regularly working more than 40h per week; % that completed an apprenticeship; % that completed higher education; % that received job-related training in the previous three months; strength; sociability; fine motor skills; mathematical skills; creative skills; and whether it is a manual occupation. Differences in supply and demand of workers in different occupations is accounted for using fixed effects. The final model is estimated to determine whether men selectively leave

occupations dependent on the proportion of women entering, conditional on the effect of these occupational characteristics. For full results, including all fixed effects, see Table S3.

The results show a substantial and clear tendency that men have a higher tendency to leave occupations when they are entered by more women ($\theta = -2.08$, $p < 0.001$). Combined with the main parameter, this allows to calculate changes in men's probability to leave the occupation they work in at t_1 dependent on female inflow. For example, an increase of 10% in the proportion of women moving into an occupation is associated to a reduction in the probability of men to stay by ~12%. Men are approximately only half as likely to stay in an occupation with 75% female inflow compared to 25% female inflow, *ceteris paribus* – a substantial effect compared to other model parameters⁶.

Aside this focal parameter, the model shows occupational characteristics that are associated with women or men moving into, or staying within particular occupations (Figure 2). Consistent with the current literature (Charles and Grusky, 2005; Percheski, 2008; Levanon and Grusky, 2016), more male-typed occupations – denoted by a decreased tendency of women to move into, or stay within these occupations – tend to require apprenticeships or higher education degrees, high rates of full-time work, and the use of physical strength. Female-typed occupations are rather in the public sector, and use fine-motor, and social skills. For women, moving into and staying in the category of manual occupations shows opposite patterns. This is likely to indicate the transitory nature of manual jobs for women (e.g. in the service-sector), as opposed to male-typed manual occupations (e.g. crafts) that tend to be long-term careers. Estimates for homophily (Figure S2) show that occupational mobility tends to be overwhelmingly between similar jobs. This can contribute to the perpetuation of sex segregation by facilitating mobility between occupations of

⁶ This probability of men to stay in occ. with 25% female inflow relative to 75% inflow can be approximated by $\exp((-2.082 + 0.753) * (0.75 - 0.25)) \approx 0.5$; considering the usual caveats when converting exponential family parameters into relative probabilities (see Materials and Methods). Analogous calculations for any differences in female inflow are possible, e.g. for a 10% increase of female inflow the relative probability of men to stay is $\exp((-2.082 + 0.753) * (0.1)) \approx 0.88$.

similar sex-typed characteristics. Stickiness estimates show that multiple characteristics are related to staying long-term in the same occupation, for example being stuck in low-income occupations or having reached a permanent high-skilled occupation, such as medical doctor (see appendix).

----- Insert Figure 2 here -----

While the estimate of the focal parameter clearly indicates that men are substantially more likely to leave occupations when the proportion of women that enter is higher, it is not clear to what extent this statistically predicts macro-level sex segregation in the labor market. To answer this question, I employ a counterfactual network simulation design using the initially observed data (Block, 2018), in which likely mobility networks are simulated under the condition that men (and women) pay no attention to the sex composition in their occupation when deciding to stay or leave. This means I simulate the target occupations of women and men as if their movements would only depend on occupational characteristics, reflecting the dominant paradigm in the literature. In practice, I simulate likely outcomes using the estimated model parameters, *except* that parameters θ_3 and θ_4 are artificially set to a value of zero. This mirrors conditions in which individuals are mobile identically to the full model, except that their propensity to leave an occupation is unrelated to the sex composition of individuals moving to their occupation. Theoretically, this formulation of a counterfactual simulation can be translated to individual choices using the model definition, as the parameters can be interpreted as guiding individuals' multinomial choice of occupation, *given* the mobility of everybody else (see appendix). To extrapolate to longer-term consequences, I extend the simulation a further time-step into the future, resulting in an additional set of hypothetical segregation outcomes. These simulated outcomes are compared to simulated data under the full model, which provides our best representation of the empirically observed mobility data. The metric along which the different outcomes are compared is the index of dissimilarity (D); a standard segregation measure (see appendix).

The results of the counterfactual simulation are presented in Figure 3. The analysis shows that in case women's and men's mobility would not be predicted by the sex composition of occupations but only by occupational characteristics, sex segregation in the case under analysis would be 19% lower over the modelled period and a total of 28% lower when simulating one period beyond the recorded data (represented by blue triangles in Figure 3). Notably, the full model including all modelled parameters (green squares) accurately represents the empirical data (red circles) at time 2, and predicts stable levels of sex segregation at time 3. Thus, a substantial portion of sex segregation in this statistical simulation model can be attributed men having a tendency to move out of occupations that are feminizing. This indicates that male and female mobility are strongly interdependent in the analyzed data. Definitions and details of simulation procedures are presented in the appendix.

----- Insert Figure 3 here -----

To further illustrate the magnitude of the effect, I perform analogous counterfactual simulations for the fourteen occupational characteristics in the model. In each case, simulated data is obtained under the condition that women and men pay no attention to this occupational characteristic in their mobility decision. In the statistical model this is represented by parameters $\theta_{v,2} - \theta_{v,3}$ for the relevant characteristic being set to a value of zero. The use of fine-motor skills has the strongest individual effect of any variable on overall sex segregation. However, fine-motor skills predict only 12.6% of observed segregation and, therefore, is a much lower predictor of sex segregation than the replacement mechanism analyzed above. Thus, men leaving feminizing occupations is the strongest determinant of sex segregation under my research design, compared to all occupational characteristics under analysis.

Additional analyses

The outlined results suggest that men tend to leave occupations when a higher proportion of women enter. This tendency is *net off* all occupational characteristics that are included in the

model. In this section I perform additional analyses that test specific pathways that might drive the statistical results that are not included in the main analysis.

First, I test whether the statistical results are driven by few outlier occupations that swing in sex composition substantially over the analyzed period, i.e., occupations that show an unusually a large outflow of men and large inflow of women given the initial composition. The descriptive statistics in Table S1 of the appendix show that this is not the case. Swings in sex composition over the period are moderate and in the vast majority lie between -10% and +10%. Only four comparatively small occupations that show more stochastic variability change sex composition to a larger extent. As the statistical method derives its power from the individuals as the unit of analyses, these small outlier occupations do not play an important role in driving the results.

Second, the results might be partly driven by age-specific mobility patterns. While occupational characteristics such as income, prestige, and part-time work are included in the model, they are assumed to influence all age-groups homogeneously. However, life-cycle patterns of mobility and employment differ between women and men. The division of labor in couples intensifies in early- to mid-career, which can lead to women of particular age-groups adjusting occupational choices. Similarly, employment favoritism, such as glass ceiling phenomena, can hinder women's progress in their career ladder predominantly at specific ages, which can result in career adjustment at these points in the life-course. Thus, the division of labor in couples and vertical segregation can lead to occupational changes of specific age-groups. However, the model only accounts for age-homogeneous mobility patterns that relate to these mechanisms. In case the tendency of men to leave feminizing occupations is only found for specific age-groups, these age-specific phenomena provide preliminary evidence for specific mechanisms underlying the found pattern.

Additional analyses find no evidence for age-specific patterns. First, Figure S4 in the appendix shows that rates of mobility do not develop differently between women and men over the life-course. Thus, heightened rates of career adjustment at specific ages for women compared to men

of the same age are not observed. Second, I perform additional score-type analyses that model whether the replacement process is visible only in specific age-groups (see appendix). I create four age groups, 27-35, 36-44, 45-53, and 54-63⁷ and test whether any group significantly differs from the estimated age homogenous parameter in the tendency of men to leave feminizing occupations. Results show that for neither age-group results differ significantly (p-values of 0.68, 0.37, 0.17, and 0.25, respectively). Thus, the tendency of men to leave feminizing occupations is not specific to certain age-groups.

Third, a linear operationalization of female inflow into an occupation might not represent the actual process appropriately. Previous work suggests that crossing a particular threshold between 50% and 60% female share in an occupation is related to wage penalties (Murphy and Oesch, 2015). A second study suggests that men selectively move out of occupations once the female share is above a tipping point that lies between 25% and 45% (Pan, 2015). Using score-type tests (see appendix), I analyze whether the initial approximation with a linear parameter represents the data appropriately, or whether further breaking points are visible. I test whether men are, additional to the linear effect, more likely to leave occupations once the female inflow exceeds thresholds of 20%, 25%, ..., 55%, and 60%. The p-values for these analyses range from $p = 0.89$ to $p = 0.31$. Thus, there is no indication that tipping points play a role beyond the linear effect of female inflow into an occupation.

Discussion and conclusion

In this study I use a novel mobility network approach for a comprehensive description of the predictors of occupational sex segregation that extends previous research by including emergent replacement patterns. Taken together, the results suggest that sex segregation in the labor market

⁷ The ages 25-26 are not included in the robustness analyses, as, by construction of the data, all individuals of these ages must be mobile (see Methods and appendix), which distorts the age-specific analyses if included.

must be described by endogenous patterns beyond occupational characteristics, mirroring segregation processes of, e.g., neighborhoods, in which ethnic composition can determine residential mobility of minority and majority members (Müller *et al.*, 2018). Thus, the variable-based description of mobility processes dominant in current literature is insufficient and inter-individual dependencies must be taken seriously; especially given the large substantive effect found in the study. Men leaving feminizing occupations can complete the empirical description of the sex composition of detailed occupations and specializations within occupations that were previously categorized as idiosyncratic (Charles and Grusky, 2005). Furthermore, the interdependence of men leaving work in feminizing occupations can be a credible process underlying the comparatively rapid swings in sex composition of various occupations (Pan, 2015).

While there is no direct claim to causality in an observational study like this, the observed patterns are in line with the causal mechanism that men leave occupations *because* they are entered mainly by women. Suggesting this as a mechanism underlying the empirical results is by no means an outlandish claim. Different streams of literature discuss why men could prefer working in sex segregated occupations. One argument discusses power relations in society: the integration of occupations is one instance of questioning the gender hierarchy that benefits men. The more women and men do the same tasks in society, inside and outside of the labor market, the shakier the grounds on which men can claim preferential wages and power in society. A second rationale argues on a more micro-level, labelled the pollution model of discrimination: an inflow of women into an occupation can be seen as a signal that the status of an occupation is declining. This is because women entering an occupation verifies that the skills needed for the respective tasks are not exclusive to a select group (men) – and the more common a skill is, the lower the rewards for it in the labor market. The third argument is on the affective level resulting from socialization: men do not want to do or be perceived as doing ‘girl things’. This is exemplified by the pattern that moves towards gender equality were mainly by women moving into male-dominated activities, not vice versa. Regardless of the reason for a male tendency of being unwilling to work in

integrated occupations, once the mechanisms of social control that keep women out of male occupations fail, a potential male response is to leave feminizing occupations. This exit of men from feminizing occupations has been shown in case studies and simulations; the current study adds generalizable, statistical evidence of this pattern while taking alternative occupation-based characteristics into account.

As a limitation of the current study it may be raised that the interpretation of the results does not shed light on deeper mechanisms underlying the found patterns but relies mostly on previous case studies and theoretical work. For example, with this design it is not possible to further understand whether men leave feminizing occupations voluntarily, or whether they are pushed out by potentially cheaper or more productive labor. Similarly, more complicated age dynamics might be at play: hypothetically, young women entering an occupation might selectively push out middle-aged men, while men at the beginning or the end of their careers could be less affected. With the current design these and other questions either cannot be treated or are beyond the scope of one article. However, I believe that this complementarity of work, when statistical analysis of representative data, studies of meaningfully selected cases, and theoretical work jointly provide a comprehensive picture of a phenomenon of interest is a model for knowledge generation in Sociology that is superior to trying to achieve everything by just one of these approaches.

A second, empirical limitation of this study lies in the representativeness of data. The analyzed case presents a period of relative employment stability before the financial crisis (2000-2008) in a country (UK) that shares similarities with economies in both, western Europe and the US. Nevertheless, at this moment we have no robust evidence that the found empirical patterns predicting occupational sex segregation and its magnitude generalize to other periods or countries.

A third limitation is that individual characteristics other than sex are omitted from the study. For example, education, social background, cohort, and age are known to influence the occupations one can and does aspire to. Omitting them means that the presented statistical model cannot claim

to represent mobility through the labor market faithfully. However, omitting individual differences also means that the analysis provides a clear, direct description of what predicts the sex composition of occupations. For the current study, clearly establishing whether men leave feminizing occupations controlling for occupational characteristics took precedence over formulating a comprehensive, but difficult to interpret model of individual mobility that includes sex as one variable among many.

Regardless of these discussion points, if we see the study's results as indicative of social mechanisms, the findings have interesting implications for public debates on gender roles in society. They question common narratives that justify prevailing sex segregation by referring to inherent differences in women's and men's interests and abilities in the labor market. Following such arguments, occupational sex segregation is the outcome of voluntary choices and individual freedom. Therefore, it is declared as not a matter of concern for policy makers, as discussed in Charles (2011) and Cech (2013). In contrast to these essentialist narratives, this study provides evidence suggesting that segregation is substantially produced by individuals' active tendencies against working in mixed-gender occupations regardless of work content. Thus, sex differences in labor market positions are not purely a reflection of individual interests in types of work tasks.

Finally, in case the sorting of individuals into occupations is an emergent process in which women and men react to each other's movements, perceived occupational characteristics might be consequence, rather than a cause of sex-typing of occupations (Goldin, 2014). Qualitative evidence supports this notion. For example, typists as an occupation swung from male to female dominated; concurrently, perceived job requirements shifted from requiring (male-typed) physical stamina to (female-typed) finger dexterity (Goldin, 2014). Similarly, remuneration in occupations at least partly follows changes in sex composition (England *et al.*, 2007; Murphy and Oesch, 2015). This suggests that occupational segregation is a cause, rather than a consequence of some occupational characteristics. This appears a promising research avenue to explanations how occupational status,

income or perceived skill requirements change over time, to remain in line with changing sex compositions.

References

- Becker, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 3(1), S33-S58.
- BHPS: University of Essex, Institute for Social and Economic Research (2010). *British Household Panel Survey: Waves 1-18, 1991-2009*. UK Data Service.
- Bielby, W. T., & Baron, J. N. (1986). Men and women at work: Sex segregation and statistical discrimination. *American Journal of Sociology*, 91(4), 759-799.
- Blau, F. D., Ferber, M. A., & Winkler, A. E. (2010). *The economics of women, men and work*. Boston: Pearson, 6th Edition.
- Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789-865.
- Block, P. (2018). Network evolution and social situations. *Sociological Science*, 5, 402-431.
- Breen, R. (1994). Individual level models for mobility tables and other cross-classifications. *Sociological Methods & Research*, 23(2), 147-173.
- Brown, R. (2010). *Prejudice: Its social psychology (2nd ed.)*. Wiley-Blackwell.
- Butts, C. T. (2007). Models for Generalized Location Systems. *Sociological Methodology*, 37(1), 283-348.
- Cech, E. A. (2013). The self-expressive edge of occupational sex segregation. *American Journal of Sociology*, 119(3), 747-789.

- Cha, Y. (2013). Overwork and the persistence of gender segregation in occupations. *Gender & society*, 27(2), 158-184.
- Charles, M. (2011). A world of difference: international trends in women's economic status. *Annual Review of Sociology*, 37, 355-371.
- Charles, M., Grusky, D. B. (2005). *Occupational ghettos: The worldwide segregation of women and men*. Stanford: Stanford University Press.
- Cohen, P. N., & Huffman, M. L. (2003). Individuals, jobs, and labor markets: The devaluation of women's work. *American Sociological Review*, 68(3), 443-463.
- Eccles, J. S. (1994). Understanding women's educational and occupational choices. *Psychology of women quarterly*, 18(4), 585-609.
- England, P. (2008). Toward gender equality: Progress and bottlenecks. In Blau, F., Brinton, M., Grusky, D. B. (Eds) *The declining significance of gender*, Sage.
- Goldin, C. (2014). A pollution theory of discrimination. In Boustan, L.P., Frydman, C., Margo, R.A. (Eds), *Human capital in history: The American record*. Chicago: University of Chicago Press.
- England, P. (2010). The gender revolution: Uneven and stalled. *Gender & society*, 24(2), 149-166.
- England, P., Allison, P., & Wu, Y. (2007). Does bad pay cause occupations to feminize, Does feminization reduce pay, and How can we tell with longitudinal data?. *Social Science Research*, 36(3), 1237-1256.
- Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21(1), 1-56.
- Gorman, E. H., & Kmec, J. A. (2009). Hierarchical rank and women's organizational mobility: Glass ceilings in corporate law firms. *American Journal of Sociology*, 114(5), 1428-1474.

- Hilliard, L. J., & Liben, L. S. (2010). Differing levels of gender salience in preschool classrooms: Effects on children's gender attitudes and intergroup bias. *Child development*, 81(6), 1787-1798.
- Jacobs, J.A. (1989). *Revolving doors: Sex segregation and women's careers*. Stanford: Stanford University Press, 1989.
- Jonsson, J. O., Grusky, D. B., Di Carlo, M., Pollak, R., & Brinton, M. C. (2009). Microclass mobility: Social reproduction in four countries. *American Journal of Sociology*, 114(4), 977-1036.
- Kaufman, R. L. (2002). Assessing alternative perspectives on race and sex employment segregation. *American Sociological Review*, 547-572.
- Koellinger, P., Minniti, M., & Schade, C. (2013). Gender differences in entrepreneurial propensity. *Oxford bulletin of economics and statistics*, 75(2), 213-234.
- Ku, M. C. (2011). When does gender matter? Gender differences in specialty choice among physicians. *Work and Occupations*, 38(2), 221-262.
- Lauzen, M. M., Dozier, D. M., & Horan, N. (2008). Constructing gender stereotypes through social roles in prime-time television. *Journal of Broadcasting & Electronic Media*, 52(2), 200-214.
- Leslie, S. J., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, 347(6219), 262-265.
- Levanon, A., Grusky, D. B. (2016). The persistence of extreme gender segregation in the twenty-first century. *American Journal of Sociology*, 122(2), 573-619.
- Lewis, K., Gonzalez, M., & Kaufman, J. (2012). Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences*, 109(1), 68-72.

- LFS: Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social and Vital Statistics Division (2009). *Labour Force Survey Annual Eurostat Dataset, 2005,2006. [data collection]*. UK Data Service.
- Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.
- Müller, T. S., Grund, T. U., & Koskinen, J. H. (2018). Residential segregation and ‘ethnic flight’vs.‘ethnic avoidance’ in Sweden. *European Sociological Review*, 34(3), 268-285.
- Murphy, E., & Oesch, D. (2016). The feminization of occupations and change in wages: A panel analysis of Britain, Germany, and Switzerland. *Social Forces*, 94(3), 1221-1255.
- OECD (2019). *Gender wage gap (indicator)*, doi:10.1787/7cee77aa-en.
- Pan, J. (2015). Gender segregation in occupations: The role of tipping and social interactions. *Journal of Labor Economics*, 33(2), 365-408.
- Percheski, C. (2008). Opting out? Cohort differences in professional women's employment rates from 1960 to 2005. *American sociological review*, 73(3), 497-517.
- Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., Fleishman, E.A. (1999). *An Occupational Information system for the 21st Century: The Development of O*NET*. APA Books.
- Reskin, B. F., & Roos, P. A. (1990). *Job queues, gender queues: Explaining women's inroads into male occupations*. Temple University Press.
- Ridgeway, C. L. (2001). Gender, status, and leadership. *Journal of Social Issues*, 57(4), 637-655.
- Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. Oxford University Press.

Rivera, Lauren A. (2020). Employer Decision Making. *Annual Review of Sociology* 46(1), 215-32.

Seron, C., Silbey, S. S., Cech, E., & Rubineau, B. (2016). Persistence is cultural: Professional socialization and the reproduction of sex segregation. *Work and Occupations*, 43(2), 178-214.

Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1-40.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, 37, 131-153.

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. *Advances in Experimental Social Psychology*, 46, 55-123.

Data and materials availability: The BHPS data and the LFS data are available from the UK Data Service (<https://www.ukdataservice.ac.uk/>). The O*NET data is available from the O*NET homepage (<https://www.onetcenter.org/>). Software implemented in the environment R and code for data analysis are available upon request from the authors. Publication of the software as an R library is planned.

Figures

Figure 1: Mobility Networks omitting self-ties disaggregated by sex for (A) women and (B) men. In each network nodes represent occupations; placement on x-axis denotes relative share of females in occupation; y placement uses a stress minimization algorithm. Edge weight denotes number of transitions. Node coloring indicates relative probability of women and men to leave occupation (green: men more likely to leave, red: women more likely to leave). Squares are manual occupations, circles non-manual occupations.

Figure 2: Occupational characteristics influencing mobility. Green triangles and blue circles show women's tendency (relative to men) to stay within occupations higher on the characteristic indicated by the row, and to move to such occupations in case of mobility, respectively. Estimates are normalized, i.e., the raw coefficients from the model are multiplied by the standard deviation of the underlying variable to make parameters comparable. Normalized parameter estimates indicate the effect of one s.d. in the independent variable. Error bars indicate 95% confidence intervals.

Figure 3: Predicted sex segregation after one and two periods for different models. Red circles: index of dissimilarity (see appendix) in observed mobility network; green squares: data simulated using full model, error bars = 2 SD; blue triangles: counterfactual simulation data, error bars = 2 SD.

Appendix to

Using mobility networks to understand occupational sex segregation

Content

Data preparation	35
Statistical analysis	39
Statistical model – simulation and estimation	42
Selection of model specification	44
Estimation results	46
Procedure for counterfactual simulation	47
Procedure for robustness checks	49
Statistics definition	50
References	52
Appendix Figures and Tables	55

Data preparation

The analysis undertaken in the article requires representative, longitudinal data on occupational mobility, as well as comprehensive data on occupational characteristics that are known to influence female and male mobility between jobs and occupations. Such occupational characteristics identified consistently in previous research fall into four broad categories. First, gradational characteristics of occupations rank these on continuous scales and include (a) income, and (b) occ. status (Charles and Grusky, 2005; Levanon and Grusky, 2016). Second, work conditions known to influence sex composition include whether jobs (c) are in the public or private sector, (d) require working self-employed, (e) require full-time work, (f) regularly entail working long hours, or (g)

working overtime (Kaufman, 2002; Cha, 2013; Koellinger et al. 2013; Blau and Kahn, 2017). Third, educational requirements that guide sex-typed occupational choices include the necessity of (h) an apprenticeship, (i) higher education, (j) a university degree, and (k) whether an occupation requires or facilitates continuous training (Becker, 1985; Blau *et al.*, 2010). Fourth, occupational skills used in occupations related to sex-typing include (l) strength, (m) sociability, (n) fine motor skills, (o) mathematical skills, (p) creative skills, (q) robustness, (r) technical skills, (s) authority, (t) problem solving skills, and (u) whether it is a manual occupation (Levanon and Grusky, 2016; Cech, 2013).

One country for which comprehensive data is available, as well as presenting an interesting case that shares many similarities with mainland European and Northern American major economies is the United Kingdom (UK). In this article I analyze mobility in the UK during the period between 2000 and 2008. The chosen period ends just before the financial crisis, which caused labor market turbulences, as well as austerity politics that influenced labor mobility. Occupational mobility data was taken from the British Household Panel Survey (BHPS, 2010); data on gradational job characteristics, work conditions, and education is taken from the British Labour Force Survey (LFS, 2009); skill use in different occupations was obtained from the O*NET database (Peterson *et al.*, 1999). Data extraction and preparation of the different data sources is as follows.

The BHPS is a nationally representative annual panel survey of British households that started in 1991 to gather longitudinal data for a wide variety of social science disciplines. It started with approx. 10,000 individuals in 5,500 households, being continuously extended in the following years. The BHPS is publicly available for researchers. The extraction of the mobility network from the BHPS proceeds in two steps. First, the data is reduced to the analysis period, and only retains observations of individuals of working age (25-63) that are part of the active work force. Second, for each individual in the reduced data the first observed occupational transition is recorded to denote a tie in the mobility network from the occupation of origin to the destination occupation. In case an individual remains in the same occupation for more than three years starting from his/her

first observation, he/she is coded as staying in this occupation (a self-tie in the network). Robustness checks with using more than two, and more than five years to delineate staying in occupations show that results and conclusion remain substantially unaffected by this choice. In case an individual changes occupation multiple times in the data, only the first transition is retained for analysis to ensure that each observed person only contributes one observation. The resulting mobility network under analysis contains 5967 (3037 female) individuals, of which 3727 (1843) are mobile. The occupational coding scheme used in the BHPS is the International Standard Classification of Occupations 1988 (ISCO-88).

The UK Labour Force Survey (LFS) provides data on gradational characteristics of occupations, as well as work conditions and average educational requirements. These are used to create characteristics of the nodes in the mobility network. The LFS is conducted by the UK Office for National Statistics specifically to provide information on the UK labor market. It is conducted four times a year and forms the UK component of the internationally standardized European Labour Force Survey. The research at hand uses combined data from four quarters between July 2005 and June 2006. The LFS data was subset to only include individuals of working age (25-63), that are an active part of the labor force, for whom an occupation is recorded, and that work regularly and at least 8 hours a week, providing a sample of $N = 186,605$. Occupational status in the LFS is coded using the UK version of the Standard Occupational Classification 2000 (SOC2000). For each occupation the recorded observations provide (a) the average hourly income. Occupational status (b) measured by the Standard International Occupational Prestige Scale (SIOPS) for each occupation comes from (Ganzeboom *et al.*, 1992). Extracted work conditions for each occupation are the percentage of workers (c) in the public sector, (d) working self-employed, (e) working full-time, (f) regularly working more than 40 hours, more than 50 hours, and more than 60 hours, and (g) regularly working any overtime, more than 5 hours overtime per week, and more than 10 hours overtime per week. Furthermore, for each occupation average education levels and types of workers are recorded denoting the percentage of workers that completed (h) an apprenticeship, (i)

higher education, and (j) a university degree. Finally, for each occupation the percentage of observations that (k) received job-related training in the previous three months is coded.

The most comprehensive data on skill use in detailed occupations is the Occupational Information Network (O*NET) data, collected under the sponsorship of the US Department of Labor. This data is necessary to code comprehensive occupational characteristics concerning skill use in the mobility network. No comparable resource collected specifically in the UK context is available; by using the O*NET data, I assume that the use of sex-specific skills in occupations is comparable in the US and the UK. The O*NET data combines information of occupational experts and job incumbents and is continuously updated and extended. Previous research (Levanon and Grusky, 2016) established and validated the relevant skill dimensions that guide sex-specific occupational sorting. These scales were used to code occupation specific use of (l) strength, (m) sociability, (n) fine motor skills, (o) mathematical skills, (q) robustness, (r) technical skills, (s) authority, (t) problem solving skills, and (u) whether it is a manual occupation. As I use a more current version of the O*NET data compared to (Levanon and Grusky, 2016), I could additionally create a scale for the use of (p) creative skills. It contains three items that denote level of inductive reasoning, creativity and originality used in each occupation. Occupations are classified using the US version of the Standard Occupational Classification (SOC). Given that the O*NET database is continuously improved, I faced the trade-off between using data temporarily aligned with the mobility data or contemporary data with more information. I opted for more recent data (version 21.2 dated Feb 2017) to get higher quality occupational descriptions; this implies the assumption that skill use within particular occupations did not change significantly in recent years.

Finally, the different data sources must be harmonized, as each one uses a different occupational coding scheme. A theoretically plausible and contemporarily used occupational classification scheme to analyze occupational mobility are the so-called micro classes. I start from the classification defined in Jonsson et al. (2009) but collapse some micro-classes that include very low numbers in the mobility data (e.g. farmers and farm laborers are combined into one category).

The final categorization scheme includes 59 occupations that are listed in Table S2 (first column). Each of the occupational categorizations used in the different data sources is recoded to match the proposed micro-class scheme with the cross-walk in Table S2, drawing from resources in Jonsson et al. (2009) and Jarvis and Song (2017). Data from the LFS was recoded into micro-classes prior to calculating occupational characteristics from the observations of individuals. For the skill use taken from the O*NET data, the average of occupations combined in a micro class was used to generate micro class characteristics. These harmonizing steps result in representative occupational mobility data of 5967 individuals between 59 occupations including detailed characteristics of the occupations under analysis in various dimensions.

Statistical analysis

The inferential task in the analysis of the mobility network is to model the extent to which men selectively leave occupations depending on the proportion of women among those that enter this occupation. This tendency is estimated next to a wide range of occupational characteristics including the relative demand for labor that determines the number of individuals in different occupations. This requires a statistical model for the interdependent movement of individuals between occupations that allows flexible specifications. Building on previous work in the statistical modelling of interdependent data (such as networks), I develop a variant of an exponential family model that fulfills these conditions. While the model shares important characteristics with, for example exponential random graph models (ERGMs), multiple features go beyond previous models for networks, including the incorporation of edge characteristics, the ability to model networks of count data without specifying a distribution of counts in dyads *a priori*, and the explicit consideration of self-ties in the statistical model.

I define the probability of the observed mobility network being in state $\mathbf{x}$ by

$$Pr(X = \mathbf{x}) = \frac{\exp(Q(\mathbf{x}))}{\kappa}, \quad (1)$$

where X is a random state, Q is a quality function that includes the sufficient statistics of the model and κ is a normalizing constant, defined two paragraphs below. This is the standard formulation of exponential family models (including ERGMs and logistic regression models); the defining model features lie in the details outlined below.

The quality function Q is a linear predictor that contains the network structure of interest – whether men leave feminizing occupations – as well as the control parameters formalized as statistics, weighted by statistical parameters θ_n . In particular, Q is defined as

$$\begin{aligned} Q(\mathbf{x}) = & \theta_1 s_{loop}(\mathbf{x}) + \theta_2 s_{loop * fem}(\mathbf{x}) + \theta_3 s_{loop * fem.in}(\mathbf{x}) + \theta_4 s_{loop * male * fem.in}(\mathbf{x}) \\ & + \sum_v \left(\theta_{v,1} s_{loop * v}(\mathbf{x}) + \theta_{v,2} s_{loop * v * fem}(\mathbf{x}) + \theta_{v,3} s_{target * v * fem}(\mathbf{x}) + \theta_{v,4} s_{v * homo.}(\mathbf{x}) \right) \\ & + \sum_B \theta_B s_{B * demand}(\mathbf{x}). \end{aligned} \quad (2)$$

The mathematical definitions of the statistics are given in the final section of this SI entitled statistics definition; substantively they model the following, in the order presented in the equation: The $s_{loop}(\mathbf{x})$ statistic models the intercept that denotes the general tendency to stay in an occupation, as opposed to changing to a specific other one; $s_{loop * fem}(\mathbf{x})$ models whether women are generally more or less likely to change occupations than men. The statistics $s_{loop * fem.in}(\mathbf{x})$ and $s_{loop * male * fem.in}(\mathbf{x})$ represent the additional tendencies of individuals to stay in occupations (as opposed to leaving them) depending on the proportion of women among those that enter it, and whether this tendency differs for men, respectively. The parameter associated to the latter statistic is of focal interest in the article. The next row indicates the four parameters that depend on the characteristics of occupations v . The statistic $s_{loop * v}(\mathbf{x})$ models whether workers in occupations higher on characteristic v are more likely to remain in their occupation, as opposed to being mobile

(called stickiness in the main text); $s_{loop * v * fem}(\mathbf{x})$ represents whether this tendency to stay in such occupations is more or less prevalent for women. The statistic $s_{target * v * fem}(\mathbf{x})$ models whether women that are mobile change to occupations higher on characteristic v than men that are mobile (there is no main effect for this parameter, as the marginal distribution of destination occupations is fully modelled by the fixed effects described below). Estimates for the two parameters just described are reported in the main article. Finally, $s_{v * homo.}(\mathbf{x})$ represents the tendency of individuals to switch between jobs similar on characteristic v , called homophily. The sum over all $s_{B * demand}(\mathbf{x})$ in the last row of equation (2) indicates that for every occupation B one fixed effect is included to represent the relative demand for workers in each occupation (i.e. occupation size).

To obtain the value of the Quality function $Q(\mathbf{x})$, each statistic is multiplied by a statistical parameter θ_n ; the sign of this parameter indicates whether the associated statistic is present in the observed network more or less than expected by chance, which refers to a model with the same specification except the focal statistic. The exact interpretation of the parameters is analogous to other exponential family models (e.g. logistic regression or ERGMs) and is elaborated for similar models e.g. in (Butts, 2007; Lusher *et al.*, 2013; Block *et al.*, 2018).

Finally, the normalizing constant κ from equation (1) is defined as

$$\kappa = \sum_{\mathbf{x}' \in \mathbb{C}} \exp(Q(\mathbf{x}')), \quad (3)$$

where $\mathbf{x}'$ is an alternative state the mobility network could be in. The set of all possible states of the network that could have been observed is denoted by $\mathbb{C}$. This set includes all mobility networks that have the same number of occupations (nodes), the same number of mobile individuals (edges), and the same distribution of individuals to origin occupation. Thus, the dimension that varies to give rise to different realization of $\mathbf{x}'$ is the destination occupation of individuals. This is directly aligned with my research goal, which is the analysis of the mobility of individuals, conditional on

their origin occupation. In other words, I am interested what makes people stay in their occupations compared to leaving, and if they leave, where do they move to. These questions concern the extent and target of mobility, given an individual's origin, directly linking to the model definition.

A useful feature of this definition of $\mathbb{C}$ is that the distribution of the origin is conditioned on; in reference to classical mobility models this means that the model is by definition row-conditioned. Given this definition of $\mathbb{C}$, this set includes all combinations of each of the 5967 individuals in one of 59 target occupations. For the analyzed mobility network, the size of this set is 59^{5967} , an enormous number with >10,000 digits that makes the calculation of the probability defined in equation (1) intractable. Thus, specialized methodology for the estimation of model parameters (and simulation of data under the model) is needed.

Statistical model – simulation and estimation

The aim of the analysis in this article is to find the set of parameters θ that make the model correspond closely to the observed data (i.e. the aim of modeling in the current case is the estimation of parameters). In many of the most used exponential family models (e.g. logistic regression), the likelihood of data under a model can be calculated by decomposition into case-wise likelihoods and their aggregation. Finding the parameters that maximize this likelihood – which is the common goal of parameter estimation – is a convex optimization problem to which many solutions exist. In the proposed model, as observations are not independent, observation-wise decomposition is not possible. As the likelihood is not tractable due to the size of the normalizing constant κ , alternative estimation methods are necessary. Fortunately, estimation routines for very similar models have been developed (Snijders, 2002; Butts, 2007) that I can adopt with minor changes.

For exponential family models, the maximum likelihood estimate (MLE) is also the solution of the moment equation (Lusher *et al.*, 2013)

$$s(\mathbf{x}_{obs}) = E_{\theta}(s(\mathbf{x})),$$

where $E_{\theta}(s(\mathbf{x}))$ denotes the expected values of the statistics $s(\mathbf{x})$ under the model given θ and $s(\mathbf{x}_{obs})$ are the statistics of the observed network. Thus, instead of requiring the calculation of the entire likelihood for a given θ using equation (1), the MLE can be found using $E_{\theta}(s(\mathbf{x}))$, which can be approximated with a sample of mobility networks under the model. With a reliable method to calculate $E_{\theta}(s(\mathbf{x}))$, which I outline in the following paragraph, finding the MLE reduces to an optimization problem that I tackle using the Robbins-Monro algorithm, which is a form of stochastic approximation. The procedure, including assessment of convergence is described in detail in Snijders (2002) and Lusher *et al.* (2013).

Drawing a representative sample from the distribution specified in equation (1) to calculate $E_{\theta}(s(\mathbf{x}))$ can be done using Markov Chain Monte Carlo (MCMC) procedures. A general introduction the MCMC can be found in Gilks *et al.* (1995), the initial application to the estimation of network models is presented in Snijders (2002) and use in a closely related statistical model in Butts (2007). As a short summary, the idea behind MCMC applied to the current case is to create a chain of network realizations taken from $\mathbb{C}$, where the transition between subsequent elements of the chain is independent of the previous states of the chain (the Markov assumption). In the long run, this means that the initial state of the chain is irrelevant (or ‘forgotten’). The transition probabilities between adjacent states is calculated from the relative probabilities to observe either state given in equation (1). In the calculation of these relative probabilities, the intractable denominator cancels out. This means transition probabilities are only dependent on the value of the quality function (2) of the respective states. Under common specification of transition probabilities, the stationary state of the chain converges to the distribution in equation (1), regardless of its initial state. This procedure is computationally expensive and requires the use of large-scale computing facilities for mobility networks of the size analyzed in this study. However, estimation using this approach is magnitudes faster than trying to directly evaluate equation (1).

The ability to simulate draws from the distribution under any given θ allows the approximate calculation of $E_{\theta}(s(\mathbf{x}))$. In turn, this enables estimation of parameters, as outlined above.

Selection of model specification

The comprehensive data on occupational characteristics contains 25 indicators categorized under (a) to (u), as outlined above, with past literature suggests their relevance for sex compositions of occupations. However, given the expected co-linearity between pairs of characteristics it is likely that not all of them are significant predictors of women's and men's mobility through the labor market when included all at once in a model. To avoid an over-specified model and only retain predictors significantly related to sex composition of occupations, I employ a forward model selection strategy to obtain a meaningful model specification. As the estimation of the statistical model requires long computation times, it is not feasible to estimate multiple candidate models in a step-wise fashion to build a model step by step, as is usually practiced in step-wise model selection. Instead, I use computationally more efficient score-type tests (Schweinberger, 2012) for the model selection that make use of the networks simulated in the estimation procedure [see Snijders (2002)]. The intuition behind these score-type tests is as follows:

Suppose we have two models, $P(\mathbf{x})$ and $P'(\mathbf{x})$, where the sufficient statistics in the latter model contain all statistics of the former model, but additionally contain candidate effect $s_c(\mathbf{x})$. We are interested whether a parameter estimate related to statistic $s_c(\mathbf{x})$ is likely to be significant when added to model $P(\mathbf{x})$, without estimating the resulting model $P'(\mathbf{x})$. To answer this question, we make use of the fact that the estimation of $P(\mathbf{x})$ requires simulating a large number of likely states under this model. From these simulations we can extract the expected distribution of the statistic $s_c(\mathbf{x})$ under model $P(\mathbf{x})$, including the 90% confidence range of the expected value. Subsequently, we can test whether the observed value of the statistics $s_c(\mathbf{x}_{\text{obs}})$ falls within this confidence range – if it does not, we take this as an indication that it is likely to be significant in model $P'(\mathbf{x})$ and

that we should include it in the model. Under this logic, $s_c(\mathbf{x})$ is an auxiliary statistic that is not modelled and the described procedure tests whether our model $P(\mathbf{x})$ can represent the observed outcome formalized by $s_c(\mathbf{x})$ adequately without modelling it (Hunter *et al.*, 2008). One advantage of this approach is that when estimating a relatively small model $P(\mathbf{x})$, one can test multiple candidate statistics at the same time, making this model selection strategy efficient enough to be employed in empirical research.

Using this approach in practice, I first specify an initial model that includes a set of statistics that are related to sex-typed mobility most consistently in past studies and are, for skill dimensions, least correlated to one another. The characteristics v for which I include the statistics $s_{loop * v}(\mathbf{x})$, $s_{loop * v * fem}(\mathbf{x})$, $s_{target * v * fem}(\mathbf{x})$ and $s_{v * homo.}(\mathbf{x})$ in the initial model are (a) income, (c) % working in the public sector, (e) % working full time, (f) % working more than 40 hours, (g) % working any overtime, the use of (l) strength, (m) sociability, (t) problem solving skills, and (u) whether it is a manual occupation. Characteristics for which estimates related to any of the four statistics are significant are retained in the model. For the characteristics not included in the model, all four statistics are score-tested according to the procedure outlined above.

In the resulting second model the parameters estimating statistics related to (t) problem solving skills are dropped from the model as none of the parameters was significant. At least one statistic that had a significant score-type test was obtained for the occupational characteristics of (d) the % self-employed, (h) % of workers with a finished apprenticeship, (i) % workers with higher education, (k) % workers that had recent training, and (n) the use of fine motor skills, (o) mathematical skills and (p) creativity. A new model that now included these parameters was estimated and all remaining, non-modelled parameters were score-tested once again to specify the next model in the forward-selection process.

In the third model, the four parameters related to (g) the % working any overtime were dropped from the model as none were significant under this specification. No further parameters were

included in the model as the score-type test had no significant parameters. To this third model the endogenous replacement parameters of interest were added, which provides the final model specification. Parameter estimates presented in the article are for this final model.

Estimation results

The full model results including all fixed effect estimates are presented in Table S3. Convergence checks following Snijders (2002) indicate good convergence of the estimation algorithm with all convergence t-ratios < 0.2 . Parameters and their interpretation are discussed in the order of equation (2). The tendency to stay in the current occupation (θ_1) is positive, significant, and large. This parameter can be understood as an intercept. Its interpretation is that in case a worker chooses between staying in their current occupation and moving to a particular other occupation, there is a strong tendency to stay in the current occupation, *ceteris paribus*. The estimate for the parameter that models whether women are more or less likely to switch occupations compared to men (Tendency to stay in current occ * Female worker, θ_2) is non-significant. Parameter θ_3 (Tendency to stay in current occ * % female inflow) is significant, meaning that, *ceteris paribus*, women tend to stay in occupations if more women move in; parameters $\theta_3 + \theta_4$ (Tendency to stay in current occ * Male worker * % female inflow) models the respective tendency for men. As this sum is negative and significantly different from zero, it shows that men tend to leave occupations in case mostly women move in, which is the main interest of this article.

Interpretation of parameters related to occupational characteristics is illustrated on the case of the binary characteristic manual occupations; interpretation of other characteristics is analogous. The positive stickiness parameter ($\theta_{manual,1}$) indicates that (male) workers in manual occupations are more likely to stay in their current occupation than workers in non-manual occupations (i.e. are less likely to change). Parameter $\theta_{manual,2}$ (Women staying in occ (ref: men): Manual occ) indicates that this tendency is lower for women than for men; the sum $\theta_{manual,1} + \theta_{manual,2}$ gives

the overall tendency of women to stay in manual occupations. In case women are mobile, they are more likely to move to manual occupations than men, as indicated by the positive estimate $\theta_{manual,3}$ (Women moving to occ (ref: men): Manual occ.). Finally, the positive estimate $\theta_{manual,4}$ (Homophily: Manual occ.) indicates that, in case workers are mobile, they tend to move between occupations that have the same manual occupation status, that is, between two manual or between two non-manual occupations. The parameters presented in Table S3 are multiplied by the standard deviation of the underlying variables to make their size comparable and these standardized parameters are presented in Figure 2 in the main text and Figure S2 in this supplementary material. Main results related to Figure 2 were discussed in the main text. I refrain from a detailed discussion of each parameters in Figure S2, as this goes beyond the scope of this article. In summary, mobility tends to be between occupations that are similar on the modelled dimensions, as homophily parameters are overwhelmingly positive. Stickiness parameters are mixed in their sign and would require deeper discussion than relevant for the purpose of this already sizeable appendix. Finally, Table S3 shows estimates for all fixed effects that account for relative occupational sizes.

Procedure for counterfactual simulation

The size of the focal parameter θ_4 – whether men selectively leave occupations in case more women move in – is difficult to interpret on its own for two reasons. First, it is problematic to translate estimates from exponential family models into differences in underlying choices for well-known reasons (Mood, 2010) without using constructs such as average marginal effects. Second, the interest in the analysis is not primarily concerning individual choices that the model parameters primarily allude to, but overall segregation outcomes. To tackle these challenges, I use a counterfactual simulation design [see Block (2018)].

The principle idea behind this analysis is to answer the question: How much segregation would we observe if individuals would move between occupations as described by the model, *except* that they do not pay attention to the sex composition of workers moving into their current occupations? A different way of framing the question is: How much segregation would we observe if individual mobility would only be guided by variables in the way found in the estimated model and not by interdependencies between individuals? As I can simulate data from my model under any given parametrization (as outlined above) I can simulate mobility networks under counterfactual ('alternate') model specifications; i.e. I can generate likely states of the world under the described conditions. This can be done by artificially setting the parameters that represent the interdependence of choice under the model to zero.

This counterfactual simulation design makes use of some inherent properties of this type of exponential family models. These models offer a direct interpretation of the model parameters on the micro-level, i.e. guiding individual choice. Analogous to the models described in Butts (2007) and Lusher *et al.* (2013), the model can be reformulated as a multinomial choice, in which individuals decide which occupation they want to move to (or stay in the current occupation). At each step, this multinomial choice is conditional on the choices of everybody else, as mobility decisions are inter-dependent. When setting some parameters to zero in the counterfactual simulation, this represents how individual choices would change if other workers do not influence occupational choice anymore. Thus, the use and interpretation of this counterfactual simulation is fully in line with the formulation of the statistical model.

In the simulation, I use the observed network at the first observation as the starting point and simulate 2400 outcomes with the parameterization found in Table S3, except that θ_3 and θ_4 (the two endogenous parameters) are set to zero. For each outcome of a simulated mobility network under the adjusted model, I calculate the index of dissimilarity D , a well-known measure of segregation defined as

$$D = \frac{1}{2} \sum_i |m_i - f_i|,$$

where m_i and f_i are the proportion of men or women, respectively, that are employed in the i -th occupation (Charles and Grusky, 2005). This simulation provides a distribution of likely segregation outcomes under the counterfactual model at a second time point. From the 2400 outcomes, 25 cases are sampled that are used as the starting conditions for a simulation a further time-point into the future. From each sampled starting point, another 384 outcomes under the counterfactual model specification are simulated, resulting in 9600 simulated outcomes at a hypothetical third time-point, for which a distribution of segregation outcomes is calculated. The described procedure is sketched in Figure S3. The outcomes of the simulations under the counterfactual model parametrization are compared to simulated outcomes under the full model at the second and hypothetical third time-point, as well as the observed data at the second time-point, as described in the main article.

For the additional analyses that compare the effect of different variables on overall segregation outcomes, an analogous procedure was employed. For each of the fourteen variables a set of 2400 counterfactual outcomes as the second time-point was simulated. The counterfactual condition is that the parameters guiding the differential tendency of men and women to stay in occupations high on this characteristic and the tendency of women, when mobile, to move to occupations high on this characteristic are set to zero. The results are as discussed in the main article, with fine motor skills being the most important characteristic in determining sex segregation; however, the effect is less strong than that of the replacement parameter of focal interest in this article.

Procedure for robustness checks

The robustness checks presented in the main text analyze whether alternative explanations related to the age distribution or non-linearities in the sex composition are likely to influence the observed

process. To test these propositions I use score-type tests (Schweinberger, 2012), following a similar procedure as the selection of the model specification. In the forward model selection analyses I employed a statistical test whether candidate model parameters would be significant if included in the model. The robustness checks employ the same logic. Using simulations, I obtain a set of likely mobility networks under the estimated model. These simulations represent a distribution of probable outcomes of the hypothesized process. Using this distribution, I can test whether candidate statistics are not represented well under this model, where the candidate statistics relate to the age-specific mobility tendencies of women and men and the tendency of to leave occupations that cross a certain threshold of female inflow, which offer alternative explanations for the observed data. If these statistics are not represented well in the simulated mobility networks under the full model, it is a sign that they would be likely to be statistically significant under a model that includes them. After performing these tests, neither of the candidate parameters as outlined in the main text were significant under a conservative threshold of $p < 0.1$, which would take very weak evidence for a candidate parameter as sufficient to necessitate including it in the model.

Statistics definition

The statistics in equation (2) use the following notation in their definition:

- B refers to a specific occupation,
- A to the set of individuals,
- o_i to the origin occupation of individual i ,
- d_i to the destination occupation of individual i ,
- w_i to an attribute of individual i ,
- v_k to a characteristic of occupation k , and

- $I(\dots)$ to an indicator function that takes the value 1 in case the condition in parentheses is fulfilled and 0 otherwise.

The statistics are defined as follows:

$$s_{loop}(\mathbf{x}) = \sum_{i \in A} I(o_i = d_i)$$

$$s_{loop * fem}(\mathbf{x}) = \sum_{i \in A} I(w_i = female) * I(o_i = d_i)$$

$$s_{loop * fem.in}(\mathbf{x}) = \sum_{i \in A} \frac{\sum_{h \in A} I(d_h = d_i) * I(w_h = female)}{\sum_{h \in A} I(d_h = d_i)} * I(o_i = d_i)$$

$$s_{loop * male * fem.in}(\mathbf{x}) = \sum_{i \in A} \frac{\sum_{h \in A} I(d_h = d_i) * I(w_h = female)}{\sum_{h \in A} I(d_h = d_i)} * I(w_i = male) * I(o_i = d_i)$$

$$s_{loop * v}(\mathbf{x}) = \sum_{i \in A} v_{o_i} * I(o_i = d_i)$$

$$s_{loop * v * fem}(\mathbf{x}) = \sum_{i \in A} I(w_i = female) * v_{o_i} * I(o_i = d_i)$$

$$s_{target * v * fem}(\mathbf{x}) = \sum_{i \in A} I(w_i = female) * v_{d_i}$$

$$s_{v * homo.}(\mathbf{x}) = \sum_{i \in A} 1 - \frac{|v_{o_i} - v_{d_i}|}{\Delta},$$

where Δ is the maximum difference between characteristics v of any pair of occupations.

$$s_{B * demand}(\mathbf{x}) = \sum_{i \in A} I(d_i = B)$$

References

- Becker, G. S. (1985). Human capital, effort, and the sexual division of labor. *Journal of Labor Economics*, 3(1, Part 2), S33-S58.
- BHPS: University of Essex, Institute for Social and Economic Research (2010). *British Household Panel Survey: Waves 1-18, 1991-2009*. UK Data Service.
- Blau, F. D., Ferber, M. A., & Winkler, A. E. (2010). *The economics of women, men and work*. Boston: Pearson, 6th Edition.
- Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789-865.
- Block, P. (2018). Network evolution and social situations. *Sociological Science*, 5, 402-431.
- Block, P., Koskinen, J., Hollway, J., Steglich, C., & Stadtfeld, C. (2018). Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, 52, 180-191.
- Butts, C. T. (2007). Models for Generalized Location Systems. *Sociological Methodology*, 37(1), 283-348.
- Cech, E. A. (2013). The self-expressive edge of occupational sex segregation. *American Journal of Sociology*, 119(3), 747-789.
- Cha, Y. (2013). Overwork and the persistence of gender segregation in occupations. *Gender & society*, 27(2), 158-184.
- Charles, M., Grusky, D. B. (2005). *Occupational ghettos: The worldwide segregation of women and men*. Stanford: Stanford University Press.
- Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social science research*, 21(1), 1-56.

- Gilks, W.R., Richardson, S., Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall.
- Hunter, D. R., Goodreau, S. M., & Handcock, M. S. (2008). Goodness of fit of social network models. *Journal of the American Statistical Association*, 103(481), 248-258.
- Jarvis, B. F., & Song, X. (2017). Rising intragenerational occupational mobility in the United States, 1969 to 2011. *American sociological review*, 82(3), 568-599.
- Jonsson, J. O., Grusky, D. B., Di Carlo, M., Pollak, R., & Brinton, M. C. (2009). Microclass mobility: Social reproduction in four countries. *American Journal of Sociology*, 114(4), 977-1036.
- Kaufman, R. L. (2002). Assessing alternative perspectives on race and sex employment segregation. *American Sociological Review*, 547-572.
- Koellinger, P., Minniti, M., & Schade, C. (2013). Gender differences in entrepreneurial propensity. *Oxford bulletin of economics and statistics*, 75(2), 213-234.
- Levanon, A., Grusky, D. B. (2016). The persistence of extreme gender segregation in the twenty-first century. *American Journal of Sociology*, 122(2), 573-619.
- LFS: Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social and Vital Statistics Division (2009). *Labour Force Survey Annual Eurostat Dataset, 2005,2006. [data collection]*. UK Data Service.
- Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2013). *Exponential random graph models for social networks: Theory, methods, and applications*. Cambridge: Cambridge University Press.
- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, 26(1), 67-82.

- Peterson, N.G., Mumford, M.D., Borman, W.C., Jeanneret, P.R., Fleishman, E.A. (1999). *An Occupational Information system for the 21st Century: The Development of O*NET*. APA Books.
- Schweinberger, M. (2012). Statistical modelling of network panel data: Goodness of fit. *British Journal of Mathematical and Statistical Psychology*, 65(2), 263-281.
- Snijders, T. A. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2), 1-40.

Appendix Figures and Tables

Figure S1. Full mobility network. Nodes represent occupations; placement on x-axis denotes relative share of females in occupation; y placement uses a stress minimization algorithm. Edge weight denotes number of transitions. Squares are manual occupations, circles non-manual occupations.

Figure S2. Occupational characteristics influencing mobility. Green triangles and blue circles show the tendency of individual to stay in occupations higher on the characteristic indicated by the row (stickiness) and to move between occupations similar on this characteristic (homophily), respectively. Estimates are normalized, i.e. the raw coefficients from the model are multiplied by the standard deviation of the underlying variable to make parameters comparable. Normalized parameter estimates indicate the effect of one s.d. in the independent variable. Error bars indicate 95% confidence intervals.

Figure S3. Simulation procedure for counterfactual simulation. Schematic representation for simulating data from observed time 1 to hypothetical time 3.

Figure S4. Proportion of women and men that change occupation by age. Scatter plot shows the raw proportion of mobile individuals and a smoothed trend line. Error bars take the size of the different age groups into account; thus, the confidence band at larger ages is wider.

Table S1. Descriptive statistics of sex composition and sex changes of analyzed occupations.

Table includes the distribution of individuals over the occupational groups at origin (t1) or target (t2) occupation, the % female that have the respective occupational group as origin (t1) or target (t2), the percentage of all women and men that have the occupational group as their origin and do not change occupation over the recorded interval, and the change in sex composition between origin and target.

Occ. Code	Occupation Name	Size t1	Size t2	t1 % female	t2 % female	% men staying	% women staying	% women inflow	swing sex ratio
1101	Jurists	19	25	57.9	52.0	75.0	81.8	40.0	-5.9
1102	Health professionals	22	21	59.1	61.9	77.8	69.2	80.0	2.8
1103	Professors, instructors, authors, journalists	94	83	57.4	60.2	60.0	46.3	73.5	2.8
1104	Natural, statistical and social scientists	44	46	45.5	39.1	41.7	20.0	43.8	-6.3
1106	Architects, Engineers	128	121	6.3	9.1	33.3	0.0	13.6	2.8
1107	Accountants	58	63	32.8	46.0	20.5	5.3	51.9	13.3
1201	Officials (government and non-profit organizations)	148	177	56.1	56.5	29.2	22.9	58.3	0.4
1202	Managers (General)	187	181	44.4	38.1	26.9	18.1	39.1	-6.3
1203	Managers (Production)	137	136	10.2	10.3	30.9	7.1	13.4	0.1
1204	Managers (internal; e.g. HR, finance)	212	283	49.5	54.1	31.8	22.9	57.3	4.5
1205	Managers (Retail, service, etc.)	94	133	45.7	42.1	19.6	25.6	40.2	-3.6
1206	Managers (Sales, Marketing etc.)	115	150	33.9	35.3	26.3	20.5	36.9	1.4
1301	Systems analysts and programmers	107	87	16.8	23.0	33.7	50.0	22.9	6.2
1303	Personnel and labor relations workers	18	30	88.9	80.0	0.0	18.8	77.8	-8.9
1304	Elementary and secondary school teachers	221	216	71.5	69.4	66.7	63.3	67.6	-2.0
1306	Creative artists	36	30	41.7	36.7	42.9	26.7	41.2	-5.0
1308	Professional and technical, NEC	32	41	71.9	51.2	33.3	26.1	46.9	-20.7

Occ. Code	Occ. Name	Size t1	Size t2	t1 % female	t2 % female	% men staying	% women staying	% women inflow	swing sex ratio
1309	Social, welfare & religious workers	169	206	72.8	77.7	56.5	38.2	85.0	4.9
1311	Nonmedical technicians	110	119	34.5	26.9	30.6	21.1	27.0	-7.7
1312	Health semiprofessionals	62	64	80.6	81.3	33.3	60.0	73.3	0.6
1313	Hospital attendants	150	158	93.3	92.4	60.0	82.1	83.8	-0.9
1314	Nursery school teachers and aides	171	144	95.9	97.2	14.3	59.8	93.3	1.3
3102	Agents, NEC	96	105	43.8	41.0	31.5	16.7	44.4	-2.8
3103	Insurance agents	23	23	30.4	47.8	31.3	14.3	58.8	17.4
3104	Cashiers	109	116	81.7	84.5	15.0	34.8	81.7	2.8
3105	Sales workers	361	342	70.1	69.9	18.5	39.9	62.4	-0.2
3106	Sales workers Stalls and Telephone	29	22	48.3	50.0	13.3	28.6	43.8	1.7
3202	Bookkeepers	211	185	79.1	73.0	18.2	29.9	66.9	-6.2
3203	Office workers	631	610	66.7	66.1	30.0	35.4	63.8	-0.7
3204	Postal clerks	40	29	22.5	17.2	61.3	33.3	28.6	-5.3
3205	Secretaries	109	122	97.2	100.0	0.0	50.9	100.0	2.8
3206	Computer Operators	28	22	64.3	68.2	10.0	11.1	68.4	3.9
4101	Craftsmen, NEC	63	53	12.7	17.0	32.7	25.0	21.2	4.3
4103	Electronics service and repair	61	58	14.8	8.6	32.7	0.0	12.2	-6.1
4104	Printers, Tailors and related workers	37	35	29.7	42.9	30.8	36.4	47.8	13.1
4106	Electricians	34	35	0.0	5.7	11.8	-	6.5	5.7
4108	Vehicle mechanics	54	39	5.6	5.1	45.1	0.0	12.5	-0.4
4109	Blacksmiths and machinists	32	32	0.0	0.0	34.4	-	0.0	0.0
4111	Other mechanics	102	98	2.9	6.1	28.3	33.3	7.2	3.2
4114	Bakers and butchers	34	28	23.5	21.4	46.2	37.5	23.1	-2.1
4115	Welders	44	41	2.3	4.9	41.9	0.0	8.7	2.6
4119	Bricklayers and carpenters	151	139	2.0	0.7	49.3	0.0	1.5	-1.3
4120	Heavy machine operators	19	22	0.0	0.0	57.9	-	0.0	0.0
4201	Truck drivers	102	111	4.9	3.6	66.0	20.0	6.5	-1.3
4202	Chemical processors, miners and related workers	34	25	11.8	20.0	13.3	0.0	23.8	8.2

Occ. Code	Occ. Name	Size t1	Size t2	t1 % female	t2 % female	% men staying	% women staying	% women inflow	swing sex ratio
4204	Longshoremen / warehousemen	22	30	9.1	6.7	25.0	0.0	8.0	-2.4
4205	Food processing and textile workers	51	44	66.7	56.8	29.4	32.4	50.0	-9.8
4208	Metal processors	31	22	0.0	0.0	32.3	-	0.0	0.0
4209	Operatives and related, NEC	200	155	24.5	20.6	25.8	28.6	17.6	-3.9
4211	Manual Labourers	107	130	29.0	20.8	18.4	12.9	20.5	-8.2
4301	Protective service workers	163	142	34.4	35.2	60.7	44.6	48.1	0.9
4302	Transport conductors and mass transportation operators	35	35	2.9	5.7	55.9	100.0	6.7	2.9
4304	Food service workers	153	108	73.9	77.8	35.0	44.2	77.3	3.9
4306	Service workers, NEC	158	179	88.0	81.6	47.4	43.2	78.2	-6.4
4307	Hairdressers	26	24	92.3	87.5	50.0	58.3	77.8	-4.8
4309	Launderers and housekeeping workers	50	52	92.0	96.2	0.0	23.9	95.1	4.2
4311	Janitors and cleaners	145	137	72.4	73.0	35.0	40.0	71.6	0.6
4313	Elementary Service Workers	45	58	22.2	24.1	45.7	20.0	30.0	1.9
5101	Farmers, fishermen and related labourers	43	45	20.9	22.2	55.9	11.1	36.0	1.3

Table S2. Cross tabulation for recoding of occupations (separate file). Matching of codes for the used occupational classification, the ISCO-88 codes (BHPS), the UK SOC2000 codes (LFS), and the US SOC2010 codes (O*NET).

Table S3. Full results table of statistical model (separate file). Table includes all estimates as reported in Fig.2, Fig. S2, standard control parameters, focal crowding out parameters and fixed effects for each occupational category to control for the marginal distribution of individuals to occupations. All parameters are unstandardized.