Social Push and the Direction of Innovation[*]

Elias Einio, *VATT Institute for Economic Research*
Josh Feng, *University of Utah – Eccles School of Business*
Xavier Jaravel, *London School of Economics*

June 2022

Abstract

Innovators are intrinsically-motivated individuals who use ideas to create new goods and services. This raises the possibility that their social backgrounds may affect the direction of their innovative activity. Consistent with this "social push" channel, we document that innovators create products that are more likely to be purchased by customers similar to them along observable dimensions including gender, age, and socioeconomic status, both across and within detailed industries. Next, we provide causal evidence that social experience affects the direction of a person's innovative activity. Specifically, being exposed to peers from a lower-income group increases an entrepreneur's propensity to create necessity products, without affecting her rates of entrepreneurship and entrepreneurial income. We incorporate this channel into a general equilibrium model to assess its implications for cost-of-living inequality and long-run growth when there is unequal access to the innovation system.

I Introduction

What governs the direction of innovation? Much of the economics literature focuses on market size and financial incentives as the main endogenous drivers of the direction of innovation (e.g., Schmookler, 1966; Acemoglu, 2002, 2007). At the same time, the innovation literature has documented that innovators are intrinsically-motivated individuals who rely on ideas as inputs to their innovative activity.[1] Because innovators of different backgrounds will be exposed to different ideas and possess different motivations, they may pursue different innovation directions independent of financial incentives. Despite the apparent plausibility of this "social push" channel, little is known about its relevance for the direction of innovation in practice.

The social push channel is particularly relevant given recent research showing that innovators in many countries are not representative of society at large. For example, women, minorities and individuals from low-income backgrounds and certain geographic regions are under-represented among innovation leaders including startup founders, patent inventors, and venture capitalists (e.g., Aghion et al., 2017; Bell et al., 2017; Hvide and Oyer, 2019; Agarwal and Gaulé, 2020). Although a growing literature documents such gaps in access to innovation and entrepreneurship, much less is known about their impacts on the *direction* of innovation, and in turn on the distribution of the gains from innovation.

In this article, we present new stylized facts and quasi-experimental evidence on the relationship between innovators' backgrounds and the direction of innovation. We then assess the quantitative importance of the social push channel for general equilibrium growth and cost-of-living inequality using a standard growth model. Overall, we find that an innovator's background affects the direction of her innovative activity, which, combined with the under-representation of certain groups in the innovation system, can lead to reduced growth and greater cost-of-living inequality.

In the first part of the article, we use data from the United States and Finland to present new facts about the direction of innovation and innovators' socio-demographic backgrounds. We build a data set linking consumer characteristics to innovators' parental income and gender. Consumer characteristics are measured in comprehensive consumption surveys, in detailed scanner data for consumer packaged goods, and in a new data set covering mobile phone applications. Innovators and their backgrounds are identified from patent records, start-up and VC databases, registries of

---

[1]For example, the discovery of entrepreneurial opportunities depends on the distribution of information in society (e.g., Hayek, 1945) and often requires engagement with specific real-world problems and users (e.g., Von Hippel, 1986; Shane, 2000). Other findings in the literature have noted that scientists, inventors, and entrepreneurs are motivated by non-pecuniary benefits (Stern, 2004; Pugsley and Hurst, 2011).

firms, and administrative tax records.

We find that innovators from a high-income family are more likely to create products purchased by high-income consumers. For example, people from high-income families are less likely to get a patent or start a firm within a "necessity" industry like food, but are more likely to do so in a "luxury" industry like finance. Similar empirical patterns exist in terms of gender, age, and geography. These patterns hold across detailed industries as well as across firms within the same industry, in both the United States and Finland. We provide additional results showing that the background of venture capital partners is separately predictive of direction, more general directional differences exist across innovators from different backgrounds, and that directional differences have remained similar even as more women inventors have entered the system.

In the second part of the article, we provide direct evidence that innovators' social experiences have a causal impact on the direction of innovation. Using a study-peer design based on data from Finnish universities, we examine whether variation in the socioeconomic background of an individual's peer group has an impact on her direction of entrepreneurship.[2] We find that exposure to lower-income peers increases her probability of starting a business in a necessities industry (conditional on being an entrepreneur) but does not affect her probability of becoming an entrepreneur or her entrepreneurial income. These results provide direct evidence that social experience has a causal effect on the direction of innovation, which stands in contrast with the mechanisms based on economic incentives, such as market size.

In the final part of the article, we investigate the relevance of the social push channel for cost-of-living inequality and long-run growth in dynamic general equilibrium, building on standard growth models. The model features multiple sectors, heterogeneity in consumer tastes, presence of the "social push" channel, and barriers to entering the innovation system that vary across socio-demographic groups. We derive the balanced-growth path and steady-state cost-of-living inequality in this setting. Our analytical results formalize the conditions under which the social push channel affects long-run growth and cost-of-living steady-state inequality. In addition, we provide simulations to assess the quantitative effects of access barriers. In an economy with two equally-sized groups calibrated to reflect differences in participation rates in the innovation system by gender, reducing access barriers can increase long-run growth rates from 2 percent to 4 percent and reduce cost-of-living inequality by 12 percent.[3] If access barriers cannot be reduced in the

---

[2]The design is based on Hoxby (2000).

[3]Appendix A discusses a static, partial equilibrium approach to assessing the impact of unequal access on inequality, based on Feenstra (1994). This alternative quantitative framework yields similar estimates for inequality.

short-run, increasing innovators' propensity to innovate for other social groups can also reduce steady-state cost-of-living inequality, by about 10 percent.

Together, the three parts of the analysis provide a comprehensive answer to the question of whether the "social push" channel is an important driver of the direction of innovation. The descriptive evidence suggests that the social push channel could operate at a large scale, both across and within industries. The peer-effects design provides direct causal evidence that social experience has an impact on the direction of innovation. Finally, the structural model shows that the impact of the social push channel on the direction of innovation can be large even in general equilibrium. The descriptive, reduced-form and structural approaches thus paint a consistent picture, highlighting the social push channel as an important but understudied driver of the direction of innovation.

This article primarily contributes to the literature on unequal access to innovation. Recent research documented that certain groups are underrepresented in the innovation system, particularly in terms of gender, race, parental background, and geography (Cook and Kongcharoen, 2010; Bell et al., 2017; Toole et al., 2019). Several articles have studied potential mechanisms that can explain this under-representation, including funding barriers (Brooks et al., 2014; Malmström et al., 2017; Kanze et al., 2018; Guzman and Kacperczyk, 2018; Hvide and Oyer, 2019), preferences (Thébaud, 2010; Bönte and Piegeler, 2013; Caliendo et al., 2015), and intergenerational transmission (Dunn and Holtz-Eakin, 2000; Mishkin, 2017; Hvide and Oyer, 2019). Consistent with our findings, contemporaneous work by Koning et al. (2019) documents that biomedical patents with female first authors are more likely to mention female medical conditions. The idea that women are underserved in a wide variety of settings is the focus in Criado Perez (2019). Our work builds on and extends this literature in three ways: (i) we provide comprehensive, economy-wide evidence on the homophily between innovators and their consumers; (ii) we providequasi-experimental evidence on the role of social factors; (iii) we quantify the importance of this channel in general equilibrium with a structural model.

We also contribute to the long literature on the determinants of the direction of innovation. As noted earlier, the economics literature has focused on market size as the key driver of innovation direction (Schmookler, 1966; Acemoglu, 2002, 2007). A related concept in the innovation literature is user innovation. Von Hippel (1986) and a body of subsequent work has noted the importance of users in driving the direction of innovation. Our paper provides descriptive and quasi-experimental evidence of the importance of innovators' social background in determining the overall direction of innovative activity in an economy with unequal access to innovation.

Finally, this article also contribute to the literature on endogenous growth and inequality. Our model builds on the product variety literature, starting with Romer (1990), by adding heterogeneity in research productivity and multiple sectors. The model is also related to Rivera-Batiz and Romer (1991), who analyze the impact of economic integration of two countries, and to Foellmi and Zweimüller (2006), who study endogenous growth when preferences are non-homothetic. By focusing on the role of social factors in innovation, we contribute to a growing literature on interactions and innovation (Lucas and Moll, 2014; Akcigit et al., 2018). Hsieh et al. (2019) propose an analysis of the impact of misallocation of talent for welfare, but without entrepreneurs and innovator; our framework extends this analysis by developingan endogenous growth model. Our results raise the possibility that misallocation in the innovation sector can affect long-run growth rates. Finally, a long-standing literature examines the influence of sorting and social interactions on inequality (Kremer, 1997; Fernandez and Rogerson, 2001). We identify a novel channel through which sorting and peer exposure can affect inequality: the impact of peers on the direction of innovation.

The remainder of the article is organized as follows. Section II presents the data. Section III presents the correlations between the socio-demographic characteristics of innovators and their consumers. Section IV provides quasi-experimental evidence showing that peer background affects entrepreneurs' choice of target markets. Section V examines the implications of these findings for growth and inequality through the lens of standard quantitative models.

II Data, Variable Descriptions and and Summary Statistics

We describe the data sources, define the sample and key variables used in the analysis, and present summary statistics.

II.A *Data Sources*

We use two "micro" datasets with respondent-level information on usage of phone applications and consumption of products within consumer packaged goods. We supplement the analysis with industry-level data covering the full consumption basket. We match the consumer information to information on innovators, using patent records (USPTO database) and startup databases (Crunchbase).

Phone applications. The first micro dataset is the Nielsen's Electronic Mobile Measurement (EMM) panel, which to our knowledge has not been used for research purposes. The dataset tracks

the mobile application usage of a representative sample of ten thousand US consumers in every month. The data contains detailed information on the gender, race, income, education, and state of residence for each panelist from April 2017 to June 2019. The data also provide classifies apps into 58 subcategories and 15 high-level categories.[4]

The data provides the company name of the developer associated with each application, which we match to information on venture-backed startups in Crunchbase. Crunchbase is a crowdsourced dataset that began tracking information on venture-backed startups and funding events in 2007. For each startup Crunchbase contains data on the name, location, and founders. For each founder, it also records gender and LinkedIn URL, which we use to collect additional information on the founders.[5] We clean names of companies in both the phone app and Crunchbase datasets, and merge based on cleaned name. Because Crunchbase attempts to track both legacy companies and startups, we define startups as firms founded after 2007, partly to reflect release date of the iPhone and partly to reflect the start of Crunchbase's data collection. The results are generally robust to the cutoff year.

Consumer packaged goods.   To track consumption patterns for consumer packaged goods, we rely on Nielsen's Homescan Consumer Panel, which has been used extensively in prior work. The data tracks between 40k-60k consumers from 2004 to 2016, and records household-level purchase data, with goods identified by bar code. Each good is classified into 9 departments (e.g., non-food groceries), 118 product groups (e.g., alcoholic beverages), and 1305 product modules (e.g., light beer). For each household, Nielsen also records race, income, education, and family structure, including classifying the household as one of five types (single female, female-led, married, male-led, single male). The disadvantage relative to phone apps data is that household purchases are for all members of the household, whereas phones are generally used by one individual.

To identify the manufacturer associated with each bar code, we use manufacturer prefix data collected by GS1, the organization in charge of allocating bar codes. The data contains the universe of bar code prefixes as of February 2016, with information on the current owner of the prefix. We use the bar code to identify the manufacturer of each purchased good in the Nielsen Consumer Panel. The GS1 data links almost all purchases to a manufacturer (97.5 percent of total revenue and 98.5 percent of total quantities).[6]

---

[4]These drop to 54 and 14 within the matched Crunchbase sample.

[5]Crunchbase uses first name to first guess the gender of each person, and then manually checks for errors. We also verify the data accuracy when we manually collect additional demographic information on each founder.

[6]This calculation excludes bar code entries associated with A.C. Nielsen, which are used to record items without

We match the combined Nielsen-GS1 dataset to information on venture-backed startups in Crunchbase, using the same procedure as for the phone applications dataset. As an additional step, we also manually check pairs of companies that share the same city and first word to look for additional matches.[7] Finally, we also match all GS1 companies to patent data. This allows us to measure the gender and age composition of the inventors who patent at a given company.

Industry-level data. We leverage several data sets to examine the industry-level patterns in terms of similarity between consumer characteristics and the socio-demographic background of innovation leaders. First, we use data on patent inventors' socio-demographic backgrounds from (Bell et al., 2017) and from PatentsView. We link inventors to consumers by first mapping the primary patent class of an inventor's patents to six-digit NAICS industry code using the concordance created by Lybbert and Zolas (2014), and then linking industry to consumer characteristics based on data from the Consumer Expenditure Survey (CEX).[8] We also use the Panel Survey of Income Dynamics (PSID) to provide additional evidence on family background and the direction of entrepreneurship. Second, we use Compustat patenting data as an additional way to link inventors to consumers. This approach only covers inventors working at public companies and may be less precise in measuring the nature of an inventor's activity.

Finally, we examine whether similar patterns hold in Finland, using Finnish administrative data covering the full population of Finnish entrepreneurs. Section IV provides a complete description of the Finnish data.

II.B Summary Statistics

Table 1 provides basic summary statistics for the two micro datasets used in the analysis. We find that companies with at least one woman founder represent 14 percent of venture-backed startups in the phone app industry and 24 percent of startups in the consumer packaged goods industry. The rates of female venture capital partner involvement are even lower, at 6 percent and 4 percent, respectively.[9] We also find under-representation in the industry-level data: only 12 percent of patent

---

a bar code, and account for 16% of total revenue in the raw data.

[7]We also tried to match based on word similarity, but find almost all false positives, similar to the results noted in Guzman and Stern (2016).

[8]The CEX category to NAICS industry crosswalk is based on work in Borusyak and Jaravel (2018). We also produce results using a more direct link for inventors working at publicly traded firms by mapping their employer's NAICS to CEX. The limitation of this approach is that it covers a smaller sample of inventors and firm-based industries may be too coarse for capturing an inventor's direction of innovative activity.

[9]This is consistent with the observation that venture capital partners tend to be drawn from the pool of successful entrepreneurs, creating a lag in representation.

inventors are women, and individuals from families with below median income are significantly less likely to become entrepreneurs.

III   Descriptive Evidence on the Social Push Channel

This section documents new descriptive facts on the relationship between innovators and the consumers they serve, focusing on U.S. consumption data. After describing the regression specifications, we present the estimates from cross-industry analysis and the estimates within phone applications and consumer packaged goods.

III.A   Research Design

We document correlations at the product level and at the industry level between consumer characteristics and innovator characteristics. Consumer characteristics are measured as the share of sales to certain consumer groups, for example female consumer, young households, low-income households, or households residing in a certain state. Innovators' characteristics are defined similarly, with the exception of income, where we use parental income instead of own income.

Specifically, at the product level, we run regressions of the form

$$SalesShareConsX_{ij} = \alpha + \beta \times FractionInnovatorX_j + \mu_k + \epsilon_{ij}, \tag{1}$$

where i indexes the good sold by a startup j, k indexes the product category or subcategory, and X denotes the socio-demographic characteristic of interest. We run the analysis at the product level because some startups are active in multiple categories. Standard errors are clustered at the firm level.

At the macro level, we run a similar regression

$$SalesShareConsX_l = \alpha + \beta \times FractionInnovatorX_l + \mu_m + \varepsilon_l, \tag{2}$$

where l indexes the industry, and m is a higher-level industry fixed effects, for example a 2-digit NAICS code.

III.B   Cross-Industry Evidence

To document industry-level patterns, we first use the PSID to measure the relationship between parent income and the direction of entrepreneurship.

Panel A of Figure 1 is based on the 2017 sample of the PSID, restricted to all individuals who can be matched to their parents. The figure shows the relationship between parent income and the probability of being self-employed. The relationship is strongly upward sloping: the rate of self-employment is about three times as high for individuals whose parents are near the top of the income distribution, compared with individuals whose parents are close to the bottom of the income distribution. The relationship is similar when focusing on high-income entrepreneurs with annual earnings above $100,000 (not reported).

Panel B of Figure 1 restricts the sample to entrepreneurs and uses the sectoral income elasticity estimates measured in Borusyak and Jaravel (2018). Within this sample, we document the relationship between the income elasticity of sector in which the entrepreneur is active and the income of this entrepreneur's parents. The relationship is upward sloping, i.e., entrepreneurs from a higher-income background tend to cater to richer consumers. The graph indicates that a 10 percent increase in parent income leads to a comparable 10 percent increase in income elasticity of the sector the entrepreneur enters, conditional on being an entrepreneur.

Figure 2 examines corresponding cross-industry patterns for gender (panel A) and age (panel B). We find that female inventors are significantly more likely to work in industries catering to women, and, similarly, older patent inventors are more likely to work in industries with a higher share of sales to older households. Table 2 summarizes the evidence on gender and age. The effects are large and statistically significant in the sample of all industries using the Lybbert and Zolas (2014) concordance. An increase in female inventor fraction of 0.1 is associated with a 0.055 increase in share of sales to households headed by a woman, and an increase in inventor age by one is associated with an increase in consumer age of 0.40. We find weaker effects in the Compustat sample, although the positive association for gender is still statistically significant. The difference could reflect sample differences (e.g., less autonomy for inventors working at publicly-traded companies) and the coarseness of measuring innovation direction using firm-level industry classifications.

We also provide evidence in directional differences over time, using the case of female inventors. The fraction of inventors who are female has increased from 9 percent to 12 percent from 1995 to 2019, a slow but very significant proportional change. We find that the correlation between the gender of inventors and consumers at the industry level has remained very similar over time, suggesting that marginal inventors still exhibit directional differences. Appendix C.B provides detailed results.

III.C   Within-Industry Evidence

Next, we turn to the results from our micro datasets. Again, we find strong correlations between innovator characteristics and consumer characteristics.

III.C.1   Phone Applications

Table 3 presents the estimates. Panel A presents the estimates for phone application startups, which are depicted graphically in Figure 3. We find that female-founded startups have an 8.2pp higher female market share relative to their male counterparts in the same subcategory, on a baseline of about 54 percent. Furthermore, we also find that startups funded by female venture capital partners are also more likely to create products that are more likely to be used by female customers. In addition, we also find a very large home-state effect, even after controlling for state population and subcategory fixed effects. Apps created by companies in a given state have 8.6pp higher usage in their home state than expected, with the estimate dropping to 4.1pp when controlling for category fixed-effects.

We also document additional results within this sample in Appendix C.A. When both female founder and female funder are simultaneously used as predictors of female consumer usage, we find that both factors are positively predictive.

III.C.2   Consumer Packaged Goods

Panel B of Table 3 reports the estimates for consumer packaged goods, which are also shown in Figure 4. We find that female-founded consumer packaged goods startups are 4.7pp more likely to sell to female-led households, on a baseline rate of 25 percent.[10] This magnitude is similar to the phone apps setting, at about 20 percent of the baseline rate.

Next, we study the gender and age-based associations for patent inventors working at consumer packaged goods manufacturers. Here, we extend to the full sample of manufacturers in the GS1-Nielsen Consumer Packaged Goods dataset, rather than restricting attention to startups. For each manufacturer, we calculate the percentage of female inventors on their patents. We then compute the correlation between this measure and their sales to female households, again finding a significant effect, even within narrow categories. A change from male to female inventor is associated with a 2.7pp increase in sales to female-led households. We also find a significant age correlations in the

---

[10]We focus on female-led households as the outcome here, but in robustness check we obtain similar patterns when analyzing all households weighted by family member gender composition or when focusing single-person households.

consumer packaged goods sample. Entrepreneurs that are a year older sell to consumers who are 0.1 years older on average.

Finally, we do not find evidence of correlations along racial dimensions, although this may be due to the lack of statistical power. After manually collecting founder race for the set of consumer packaged goods startups, which is more feasible than for phone apps because of the smaller sample, we find that 11 percent of the founders in are data are Asian, 5.2 percent are Hispanic, and none are African-American. There is no robust correlation with consumer characteristics, likely due to the small number of minority entrepreneurs and the smaller number of minority consumers in the panel. As more minority-led companies are created and the data becomes more readily available, the patterns may become clearer.

We provide a series of additional results and robustness checks in Appendix C.A. The results show that the correlations we find are generally robust to methodological choices (e.g., how to measure female consumption when data is at the family level) and additional controls such as per-unit price.

III.D   General Differences in Direction – Environmental and Social Impact

Although we use consumer characteristics as the primary way to measure direction in our core analysis, our empirical approach can be applied to understanding broader issues in the direction of innovation.

An important area of research on the direction of innovation examines the factor that can encourage the innovation of "green patents" (e.g., Acemoglu et al. (2012), Aghion et al. (2016)). We find that female inventors as well as young inventors are more likely to invent "green" patents, i.e., to have positive environmental externalities. Using the data in Aghion et al. (2016), we study the differences in characteristics of inventors on "clean" versus "dirty" patents (13.3 percent of energy patents are classified as "clean"). We find that 6.5 percent of inventors of clean patents are female, as opposed to 2.8 percent for dirty patents. The difference is statistically significant. We also find evidence related to age. In the sample of patents with available information on inventor age, we find that younger individuals are more likely to patent in clean energy technologies (0.1pp less likely to work clean patents for one year's increase in age; p-value 0.083).

More broadly, women entrepreneurs appear to have very different focuses, even within narrow product categories. Appendix Table A6 presents additional results surrounding startups in the Nielsen CPG-Crunchbase sample and the broader Crunchbase sample. We find that female-led

startups are more likely to mention "healthy", "kids", "sustainability", and B-Corporation certification in their company descriptions. These results are consistent with recent evidence in the literature that women entrepreneurs are motivated by social impact (Guzman et al., 2019), but more systematic analysis is needed to confirm that the descriptions reflect real differences and to understand broader differences in direction and their implications for social welfare.

IV   Peer Background and the Direction of Innovation: Causal Estimates from Finland

The empirical evidence on similarity between innovators and consumers raises the possibility that a person's background affects the direction of her inventive activity. In this section, we present direct evidence supporting this conjecture by using peers as a source of variation in a person's background, using detailed data from Finland.

IV.A   Data and Descriptive Evidence from Finland

To start, we document descriptive patterns surrounding entrepreneurs and patent invnetors in the Finnish population panel, which covers the whole working age population in Finland.

Data.   The dataset is based on administrative registers compiled by Statistics Finland. It provides individual-level information on income, occupation, entrepreneurship, and industry of employment. The data set also includes information on family links, which allows us to measure the parental income of individuals.

A key variable in the population panel is entrepreneurship status, which is based on pension contribution and income tax records. We use the status for the last week of the year which allows for temporal consistency across variables.[11]   An individual is defined as an entrepreneur if she has received only entrepreneurial income, and no employee salary income, during the year and is associated with a private business in the entrepreneur pension insurance system in the last week of the year. She is also identified as an entrepreneur in the last week of the year if she has made entrepreneur pension contributions in that week. If an individual has both entrepreneur and employee pension contributions in the last week of the year, she/he will be defined as an entrepreneur if the entrepreneurial income associated with the contributions is larger.

---

[11]The data includes also codes for the company which is associated with the longest employment spell during the year.

A second key variable is the unique company identifier that provides information on the company for which a worker or entrepreneur worked. This information is also based on work spells reported in the national pension systems for entrepreneurs and employees. We use the code for the company an employee/entrepreneur is associated with in the last week of the year.

In terms of patent inventors, we link individuals in the PATSTAT database to the Finnish population panel by first name, family name, postcode, and company identifier. The company identifiers are drawn from the Business Information System web interface maintained by the Finnish Patent and Registration Office. We also use different combinations of the match variables to include inventors who are not associated with a company in the population panel, have different spelling of the first or family name in the two datasets, or have missing location information in PATSTAT. We include only exact unique matches to avoid match error.

We link industry income elasticities and CEX consumption share data to the population panel by the industry code of the company an individual is associated with in the last week of the year. The industry codes available in Statistics Finland are NACE codes, which are standard in Europe. We then use a crosswalk to connect NACE to NAICS and, in turn, consumption data. We focus on the 2007–2015 period, because the industry code in the population panel is less comprehensive in the pre-2007 period and 2015 is the last year for our linked inventor data. The match rate for the sample of individuals with industry codes in the population panel is 80 percent.[12]

Results. Panel A in Figure 5 shows the fraction of inventors by parent income for Finland. The figure indicates that individuals from high-income families are more likely to become inventors. 0.12 percent of individuals whose parents are in the top income decile become inventors, while the rate for the lowest decile is 0.05 percent. Panel B plots industry income elasticity by parent income for the inventor sample. The figure shows that inventors from high-income families are more likely to work in industries with high income elasticity. Figure 6 shows gender-based correlations for inventors in Finland, which reveals a significant and positive relationship.

Finally, we provide a series of additional results in the Appendix. Appendix Figures A2 and A3 present descriptive regressions related to income, gender, and age for Finnish entrepreneurs and inventors. We find similar results to the U.S. setting. Our data from Finland also allows

---

[12]There are 27,292,828 individual-year observations in the population panel over the years 2010–2016. Industry code for the company an individual worked in in the last week of the year is available for 15,930,880 observations. We link industry income elasticities for 6,638,780 individual-year observations by 4-digit industry code. For the remaining unmatched industry codes, we match 1,943,906 at the 3-digit and 4,171,209 at the 2-digit industry level. This result in 12,753,895 matches.

us to study additional associations. Appendix Figure A4 shows a positive relationship between entrepreneur's parent income and her employees' parental income measure, potentially speaking to additional benefits generated by entrepreneurs for individuals from similar backgrounds.

IV.B Peer Family Background and the Direction of Entrepreneurship

Research design. To directly test the social push channel, we construct a design based on variation in vocational and university study peers in Finland. We use data drawn from the student register maintained by Statistics Finland, covering the 1999–2013 period. The data provide annual individual-level information on the institution and study program a student is enrolled in and cover upper secondary, vocational, and university study programs.

We define study peers as individuals who start in the same institution and study program in the same year. For individuals who are observed in several programs, we use the last program entered.[13] Our baseline model is a standard linear-in-means peer regression for individual i who starts in program j of school k in year s, controlling for school-by-program fixed effects, school-by-start-year fixed effects, and own parent income:

$$y_{ijkt} = \gamma \overline{X}_{(i)jks} + \beta_1 X_{ijks} + \beta_2 W_{ijkt} + \alpha_{js} + \lambda_{ks} + \epsilon_{ijks}. \tag{3}$$

Here X_{ijks} is own parent income; $\overline{X}_{(i)jks}$ is the leave-own-out mean of parent income of study peers; and y_{idst} is the outcome of interest measured in year t. The terms λ and α are school-by-program and school-by-start-year fixed effects. The parameter of interest is γ, which is the coefficient on average parent income of study peers.[14] We also include a vector of control variables denoted by W including pre-peer-exposure characteristics for the student and her parents. Conditioning on school-by-program fixed effects means that the peer effect is identified from idiosyncratic within-school variation in peer composition across classes. This approach follows several previous studies that have estimated peer effects in education in settings where randomization of students to peer groups is not available (e.g., Hoxby, 2000; Hanushek et al., 2003; Carrell et al., 2018). The key identifying assumption is that, while there can be selection into schools and programs, the variation in peer parent income across entering classes of students in the same study program and school is uncorrelated with an individual's own characteristics.

---

[13]We use the full data when we calculate peer composition variables in order to avoid measurement error due to omitted peer observations.

[14]We note that controlling for own parent income eliminates the potential mechanical correlation between it and the peer mean, which may arise in a peer regression where an individual is allowed to be both the subject of peer effects and the peer (Angrist, 2014).

Falsification tests. In Table 4, we test for the validity of the empirical design by estimating the impacts of peer parent income on a predicted residual outcome and on pre-peer-exposure characteristics. We construct the predicted residual outcome by first calculating the residuals from a regression of expenditure elasticity on school-by-program fixed effects, school-by-start-year fixed effects, and dummies for age, gender, and year of outcome measurement. We then run a regression of these residuals on pre-peer-exposure characteristics and construct the predicted outcome as the linear prediction from this model. This variable captures the variation in expenditure elasticity within the same program and school associated with a linear combination of pre-peer-exposure characteristics, where weights are chosen to best predict expenditure elasticity of future industry of activity (see e.g., Carrell et al., 2018). If idiosyncratic variation in peer composition across classes within the same program and school is as good as random, this linear combination of pre-peer-exposure characteristics should not be correlated with the socioeconomic background of peers. Reassuringly, the coefficient for the predicted expenditure elasticity in Panel A is small and statistically insignificant. Panel B shows results separately for each pre-peer-exposure characteristic. The t-test statistics for individual coefficients are statistically insignificant, except for a binary outcome for primary language being Finnish. The likelihood that this arises by chance in 14 separate tests, even if all tested hypotheses were true, is likely higher than 5 percent. To account for this, we calculate the stepdown p-values adjusted for multiple hypothesis testing (Romano and Wolf, 2005). The adjusted tests do not reject the hypothesis that peer family income is uncorrelated with the pre-peer-exposure characteristics. Nevertheless, in our estimations, we include control variables for all pre-peer-exposure characteristics listed in Panel B to ensure that sampling variance in pre-peer-exposure characteristics does not affect our results.

Results. Panel A in Table 5 shows the estimates of peer effects on industry expenditure elasticity when an individual is at age 28 or older for all entrepreneurs and separately for entrepreneurs whose parent income is above and below the median. Column 1 shows a significant positive peer effect for all entrepreneurs. This estimate means that when parent income of peers increases by one standard deviation (equal to around 28 thousand euro or 25 thousand US dollars) the entrepreneur will enter an industry with around 2.3 percent higher expenditure elasticity compared to the sample mean of 1.1. Figure 7 shows a graphical presentation of this result. It plots binned averages of the residuals from separate regressions of industry expenditure elasticity and peer parent income on own parent income, school-by-program fixed effects, program-by-start-year fixed effects, and dummies for age,

14

gender, and year of outcome measurement. In Columns 2 and 3 of Table 5, we estimate the impact separately for individuals who have high and low own-parent income. The results indicate that the peer effects are driven by impacts on individuals who come from high-income families. For them, a standard deviation increase in peer parent income increases industry elasticity by around 4.8 percent from the sample mean.

Given that peer background influences the direction of entrepreneurial activity, we want to test whether it also impacts entrepreneurial income. The results on earnings can help distinguish between three possibilities. First, individuals could changing direction in a way that increases non-pecuniary benefits but decreases measured earnings. Second, peers could be providing more information, thereby improving the choice set and earnings of the entrepreneur. Finally, peers could just be affecting direction without affecting earnings, because there are many equally good business opportunities available to entrepreneurs. We find small and statistically insignificant impacts of peer family background on earnings and total income of entrepreneurs, as shown in Panels B and C of Table 5. In Panels D to F, we also document insignificant impacts on binary indicators for an entrepreneur's income being above the 50th, 90th, or 99th percentile in the population income distribution, in order to capture any tail outcomes. Overall, our findings support the third view.

Robustness. Appendix D provides analysis using an alternate research design based on dormitory roommates during conscription (Einiö, 2019). The setting provides a better measure of peer exposure, but also a more limited sample size. We find that peer effects on industry choice exist and are driven by individuals from higher-income backgrounds, but only after including managers in addition to entrepreneurs in our outcomes sample.

V   Implications for Growth and Cost-of-Living Inequality

We use a dynamic general equilibrium framework based on Romer (1990) to assess the implications of our empirical findings for growth and inequality. We characterize the conditions under which the "social push" channel affects growth and inequality, and then assess the channel's quantitative importance through a calibrated simulation. Online Appendix A presents quantitative estimates from a static framework based on Feenstra (1994), as a point of comparison.

V.A   A Two-Sector Product Variety Framework

We study a Romer-style economy that contains groups with different tastes, unequal access to innovation for some groups, and innovators who are more productive at creating new varieties for their own group. The quantities of interest are the long-run growth rate and steady-state inequality across groups along a balanced growth path. For clarity, our analysis focuses on an economy with two groups and two sectors, but the results can be generalized to multiple groups and sectors.

Specifically, demand is based on two groups with different preferences in terms of consuming from each sector. For example, group 1 individuals might want to spend 60 percent of their budget on sector 1 goods, whereas group 2 individuals might only want to spend 45 percent. Formally, individuals in the economy look to maximize lifetime discounted utility

$$\int_0^\infty e^{-\rho t} log(C(t))dt,$$

where the instantaneous utility is Cobb-Douglas over the two sectors

$$C(t) = C_1(t)^\alpha C_2(t)^{1-\alpha}$$

and the consumption indices for each sector depends primarily on the number of varieties in that sector

$$C_i(t) = [\int_0^{N_i(t)} c_i(\nu, t)^{\frac{\epsilon-1}{\epsilon}} d\nu]^{\frac{\epsilon}{\epsilon-1}}$$

We incorporate heterogeneity in tastes in the following way. Let there be two groups of individuals. Group 1 has size $1 - \delta$ and preference parameter α. Group 2 has size δ and preference parameter α'. Let $\alpha > \alpha'$, i.e., group 2 has a stronger preference for sector 2 compared with group 1.

On the supply side, individuals can either work on producing existing goods on the market or can choose to conduct research to create new varieties. Entry into research depends on a person's research productivity and the market wage for production work. Formally, individuals that invent new varieties receive perpetual patents, and produce the variety using a linear production function $y(\nu, t) = l(\nu, t)$, where each unit of labor is paid $w(t)$. The functional form makes it so that entrepreneurs earn a constant markup on top of the market clearing wage in each period and owning a patent on a variety in either sector is equally profitable.

We add two elements to this standard supply side setup. First, we model individuals as being more productive at creating varieties in the sector that that they relatively prefer as consumers, in order to incorporate the "social push" channel. Formally, if an individual from group j chooses to enter into research in sector i, they will create sector i varieties at the following rate:

$$\dot{N}_{ji}(t) = \begin{cases} \eta N(t) & i = j \\ \eta' N(t) & i \neq j \end{cases}$$

where $\eta' < \eta$ and $N(t) = N_1(t) + N_2(t)$.[15] This parametrization allows us to capture the idea that individuals are better at innovating in the sector that they or people similar to them prefer as consumers. For tractability, we only allow individuals to innovate in one sector.

Second, we add access barriers to research for one of the two groups. This can take the form of either a reduction in research productivity or a preference friction that only reduces the probability of entering into research but not productivity conditional on entry. Individuals from group 1 are willing to enter research in a sector based on whether returns from patents exceeds production wages:

$$\dot{N}_{1i}(t)V(t) > w(t)$$

where $V(t)$ is the net present value of profits from each patent that depends on the equilibrium interest rate. For individuals from group 2, we add an access barrier τ:

$$(1 - \tau)\dot{N}_{2i}(t)V(t) > w(t)$$

As noted earlier, the framework allows for either a real research productivity reduction or a preference friction. In the case of a real reduction in productivity, Group 2 individuals have productivity $(1-\tau)\dot{N}_{2i}(t)$ if they enter the innovation sector. In the case where the access barrier is a preference friction (e.g., discrimination, preference for production work), they would still produce varieties at the rate $\dot{N}_{2i}(t)$.

In this economy, we can solve for a balanced growth path where the growth rate, the wage, the implied interest rate, and the labor allocated to research are all constant. Appendix B provides detailed solutions for the model, starting with the basic model. Proposition 1 summarizes the key insight from the model, which is that the existence of the social push channel means that access barriers lead to lower growth and greater inequality.

---

[15]The use of total varieties is a departure from the standard framework, with the exception of the trade and innovation model in Rivera-Batiz and Romer (1991), but is necessary to avoid one sector having explosive growth.

Proposition 1. *In a two-sector model with preference heterogeneity and the social push channel, the presence of "preference" barriers to innovation creates greater inequality and lowers long-run growth relative to the case with no barriers under either of the following conditions:*

1. $\eta' > (1-\tau)\eta$ and $\delta L < \frac{\rho(\epsilon-1)}{(\alpha\eta+(1-\alpha)\eta')}$

2. $\eta' < (1-\tau)\eta$ and $(1-\delta)L > \frac{\alpha}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1+(1-\alpha)(1-\tau)\eta_2)}\right)$
   and $\delta L > \frac{(1-\alpha)(1-\tau)}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1+(1-\alpha)(1-\tau)\eta_2)}\right)$

The conditions ensure that the economy is not at a corner case where all members of one group are pursuing entrepreneurship.

To derive Proposition 1, two cases must be considered. In the first case, $\eta' > (1 - \tau)\eta$. Under this scenario, individuals from group 1 will perform all research in the economy, as long as there are enough people in group 1 (which corresponds to the second part of the boundary condition). Individuals from group 2 do not receive enough returns from conducting research. Under this scenario, the equilibrium growth rate is $g^* = \frac{1}{2}\left(\frac{1}{\epsilon-1}(\tilde{\alpha}\eta + (1-\tilde{\alpha})\eta')L - \rho\right)$ and the steady-state inequality between the two groups is $\frac{C^1}{C^2} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{\eta}{\eta'}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}$. The growth rate is smaller than the one without entry barriers, and the inequality is higher.[16]

The second case to consider features $\eta' < (1-\tau)\eta$. Under this scenario, group 2 individuals do enter research, but fewer do so than in a world without entry barriers. Intuitively, the outcome is in-between a world where their research productivity is $(1-\tau)\eta$ and a world where there are no taxes, because they are still mechanically productive, but behave according to an indifference equation with a wedge. The equilibrium growth rate and inequality under this scenario are $g^* = \frac{1}{2-(1-\tilde{\alpha})\tau}\left(\frac{1}{\epsilon-1}[1-(1-\tilde{\alpha})\tau]\eta L - \rho\right)$ and $\frac{C^1}{C^2} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{1}{(1-\tau)}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}$. In this case, an increase in entry barriers also leads to a lower equilibrium growth rate and an increase in inequality along the balanced growth path.

Finally, the results from Proposition 1 also hold in the case where the barrier directly affects the research productivity of individuals in group 2. If $\eta' > (1-\tau)\eta$, then individual from group 1 will perform all research again and we recover the same outcome as above. If $\eta' < (1-\tau)\eta$, then group 2 will perform research in their own sector. This case will be equivalent to the basic two-sector model where research productivity in sector 1 is η and research productivity in sector

---

[16]Appendix B shows that the corresponding quantities in the case without entry barriers are $g^* = \frac{1}{2}\left(\frac{1}{\epsilon-1}\eta L - \rho\right)$ and $\frac{C^1}{C^2} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}}$.

2 is $\tilde{\eta} = (1 - \tau)\eta$. The growth rate and inequality will be $g^* = \frac{1}{2} \left(\frac{1}{\epsilon-1} (1 - \tilde{\alpha}(1 - \tau)) \eta L - \rho \right)$ and $\frac{C^1}{C^2} = \left(\frac{(1-\delta)\alpha+\delta\alpha'}{(1-\delta)(1-\alpha)+\delta(1-\alpha')} \right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{1}{(1-\tau)} \right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}}$. The growth rate is lower and inequality is higher in this case versus the derivation with preference barriers because there is a real difference in research productivity on top of the preference-driven distortion in allocation.

As discussed in greater detail in Appendix B, the model also highlights when access barriers do not create reduced growth or increased inequality. In a one-sector model with no heterogeneity in research productivity, access barriers do not matter, because group 1 individuals essentially fill in for the individuals in group 2 who face barriers. In a two-sector model, if individuals have the same research productivity in both sectors, then access barriers again do not alter inequality. This can be seen by the growth and inequality expression in the case where $\eta' > (1 - \tau)\eta$, the only relevant case when $\eta' = \eta$.

V.B   Quantitative Results From Model Simulations

Next, we quantify the implications of access barriers and social factors for long-run growth rates and steady-state inequality, using the growth model framework presented above. To focus on the social push channel, we simulate an economy with two groups that have equal spending power, which is roughly true in the case of gender. One group is a "minority" group that faces access barriers to the innovation system. We focus on the case where barriers reflect real differences in research productivity.

The key change we make from the analytical model above is that we use a Pareto distribution in innovation productivity. One motivation is that it is more reflective of evidence in the literature but precludes the ability to derive closed-form solution. In addition, projections using the closed form solutions above yield much larger quantitative effects because the marginal and average individuals have the same research productivity. Under a Pareto distribution, the most productive researchers choose to enter the innovation sector regardless of barriers, so relaxing the barriers has a less dramatic effect.

To set up the simulation, we fix preference parameters based on values in the literature and then calibrate access barriers, productivity, and specialization to match moments in the data. We set $\delta = 0.5$ to create two equally-sized groups. Next, we set standard preference parameters, with a discount rate $\rho = 0.95$ and an elasticity of substitution between products $\epsilon = 3$ (DellaVigna and Gentzkow, 2019). To capture taste differences, we set $\alpha = 0.6, \alpha' = 0.4$.[17] On the innovation side,

---

[17]These parameters match the dissimilarity index of expenditure shares across products between the bottom of top

we start with a basic Pareto distribution in research production $x_m = 1, \alpha = 1.5$, which is scaled by a factor η to arrive at actual research productivity. The minority group suffers a multiplicative real productivity shock equal to $1 - \tau$. Finally, to capture specialization in a tractable matter, agents are randomly assigned a sector in which they can innovate. A parameter ϕ is used to capture the rate of specialization. Agents have a probability $\frac{1+\phi}{2}$ of being assigned to their "own" sector, the one they relatively prefer as a consumer.[18]

Given this setup, we solve for real productivity cutoffs in each sector. Individuals with productivities above the cutoff in the sector they are assigned to become entrepreneurs. The cutoffs have to satisfy: i) allocation of research labor to ensure equal growth rates of varieties; ii) the Euler equation condition; and iii) indifference of marginal agents in each sector between pursuing entrepreneurship and production work. Given the cutoffs, we can compute the BGP growth rate and inequality.

We calibrate the parameters τ, η, ϕ by matching to three moments: i) long-run growth rate $g^* = 0.02$; ii) fraction of inventors from the "minority" group 0.12; and iii) regression coefficient $\beta = 0.1$ from Equation 1, based on estimates from the phone app data. This is intended to be conservative, because the coefficient is larger in other settings. The regression coefficient within the simulation is calculated by assigning a "minority" consumer fraction to each variety (based on the taste difference parameters) and then attributing each variety to an inventor from one of the groups. Because "minority" group inventors are more likely to work in their own sector, this leads to a positive regression coefficient.

We compute outcomes for 200 ability and sector draws, with each draw representing an economy with 250,000 agents. On average, we recover the following estimates: $\tau = 0.8693, \eta = 0.1169, \phi = 0.7025$. This suggests high barriers for minority entrepreneurs and high rates of specialization. For each draw, we compute two counterfactuals. First, we reduce τ to zero: this reflects removing access barriers. Panel A of Figure 8 plots the average results. We find that removing barriers leads to an increase in BGP growth rate from 2 percent to 4 percent, and reduces cost-of-living inequality, as defined in Equation A2, by 12 percent. This value can be compared to the gender wage gap.

Second, we start at the original parameters and reduce ϕ to zero: access barriers still exist, so this change makes majority group innovators become more likely to innovate for the minority

---

income quintiles in the Consumer Expenditure Survey (CEX).

[18]This structure helps simplify the problem by eliminating sectoral choice.

group. This reflects the empirical findings in Section IV. Panel B of Figure 8 plots the average results. The long-run growth rate increases from 2 percent to 2.25 percent, because innovation is slightly more aligned with consumer tastes, and inequality decreases by about 10 percent, reflecting the creation of more varieties in the minority group's preferred consumption sector.

V.C   Policy Relevance

The results that precede document how the social push channel and unequal access to the innovation system combine to affect growth and cost-of-living inequality in general equilibrium. Our quantitative estimates show that policies that aim at promoting equal access to entrepreneurship across all socio-demographic groups may have a double dividend, achieving at the same time higher economic growth and reduced cost-of-living inequality.[19] Our framework shows that the magnitude of the effects is large, provided that certain empirically-grounded assumptions hold.

The theoretical results can also be restated as follows: unequal access significantly matters for growth and inequality if there is heterogeneity in individuals' research productivity. As investigated in greater detail in Appendix B, unequal access does not matter for growth and inequality in general equilibrium if individuals possess homogeneous research productivity. The intuition is that individuals who do not face access barriers fill in for those who do. Labor market clearing then ensures that individuals in both groups make the same income.[20] Deviations from this "null" model include differential ability to innovate for different sectors, as modeled above, and heterogeneous research productivities even in an economy with homogeneous consumers. In these cases, if a research-productive individual does not pursue research, because of either preferences or discrimination, the potential replacement individual is less productive, which reduces long-run growth. Empirically, research productivity is highly unequal across innovators, with thick Pareto tails (Bell et al., 2019).

Our quantitative analysis also allows us to assess potential differences between static and general equilibrium projections. The quantitative simulation above uses the estimates from the phone apps data, and projects about a 12 percent reduction in inequality from removing access barriers. Appendix A derives quantitative results using a framework based on Feenstra (1994), and finds that relaxing access barriers leads to a similar reduction in cost-of-living inequality. This finding provides

---

[19]The fall in cost of living inequality in our simulations, close to 12%, is large compared with observed inflation inequality in the United States. Inflation is about 50 basis point larger for the top income quintile, relative to the top income quintile, i.e. it takes about 23 years to generate an increase in cost-of-living inequality of 12% for the bottom income quintile relative to the top income quintile (Jaravel, 2019).

[20]If bargaining or discrimination were features of the production labor market, this result would not hold.

a justification for using simpler quantitative models, which appear to make accurate predictions and may be easier to use for policy in certain contexts.

VI   Conclusion

We have documented novel empirical facts on the relationship between innovators and the consumers they serve. We find significant evidence that innovators are more likely to create goods used by consumers similar to them along observable dimensions.[21] In addition, we provide evidence from a peer-effects design that personal experience has a causal impact of the type of business an entrepreneur starts, specifically in terms of the consumers targeted.

Given the empirical findings, we investigate the implications for long-run growth and cost-of-living inequality, taking into account general equilibrium effects. In a two-sector growth model with greater ability to innovate for one's own group, unequal access leads to both slower long-run growth and greater welfare differences across groups. Simulations confirm that these effects are quantitatively significant. Our empirical results and model also suggest that greater social interactions across groups could act as a second-best solution to increasing growth and reducing inequality in the presence of barriers to entrepreneurship.

Our findings speak to policies and initiatives aiming at promoting access to entrepreneurship for women and minorities. For example, the Small Business Innovation Research (SBIR) program in the U.S. aims to "foster and encourage" the participation of women, minorities, and businesses in underrepresented geographic areas. Another recent piece of legislation, the SUCCESS Act, encourages the USPTO to measure and encourage the participation of women, minorities, and veterans in invention. Beyond explicit policies, there has also been increased hiring of female partners in venture capital and initiatives in the tech industry to encourage more women and underrepresented minorities to participate in engineering. Our results suggest that these initiatives are likely to lead to a more diverse set of new goods and services, and to have a double dividend by simultaneously increasing growth and reducing inequality.

---

[21]We have also presented additional results on differences in the direction of innovation between innovators from different backgrounds, highlighting dimensions that are directly relevant for social welfare. These include green patenting and startup founders that address health, children, and sustainability. An interesting avenue for future work would be to confirm whether the differences in textual descriptions reflect real differences and whether these relationships are causal.

References

Acemoglu, Daron, "Directed technical change," *The Review of Economic Studies*, 2002, *69* (4), 781–809.

_ , "Equilibrium bias of technology," *Econometrica*, 2007, *75* (5), 1371–1409.

_ , *Introduction to Modern Economic Growth* 2009.

_ , **Philippe Aghion, Leonardo Bursztyn, and David Hemous**, "The environment and directed technical change," *American economic review*, 2012, *102* (1), 131–66.

Agarwal, Ruchir and Patrick Gaulé, "Invisible Geniuses: Could the Knowledge Frontier Advance Faster?," *American Economic Review: Insights*, 2020.

Aghion, Philippe, Antoine Dechezleprêtre, David Hémous, Ralf Martin, and John van Reenen, "Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry," *Journal of Political Economy*, 2016, *124* (1), 1–51.

_ , **Ufuk Akcigit, Ari Hyytinen, and Otto Toivanen**, "The Social Origins of Inventors," 2017.

Akcigit, Ufuk, Santiago Caicedo, Ernest Miguelez, Stefanie Stantcheva, and Valerio Sterzi, "Dancing with the Stars: Innovation Through Interactions," 2018.

Angrist, Joshua D., "The perils of peer effects," *Labour Economics*, 2014, *30*, 98–108.

Bell, Alex, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Do tax cuts produce more Einsteins? The impacts of financial incentives versus exposure to innovation on the supply of inventors," *Journal of the European Economic Association*, 2019.

Bell, Alexander, Raj Chetty, Xavier Jaravel, Neviana Petkova, and John Van Reenen, "Who Becomes an Inventor in America? The Importance of Exposure to Innovation," 2017.

Bönte, Werner and Monika Piegeler, "Gender gap in latent and nascent entrepreneurship: Driven by competitiveness," *Small Business Economics*, 2013, *41* (4), 961–987.

Borusyak, Kirill and Xavier Jaravel, "The Distributional Effects of Trade: Theory and Evidence from the United States," *SSRN Electronic Journal*, 2018.

Brooks, Alison Wood, Laura Huang, Sarah Wood Kearney, and Fiona E Murray, "Investors prefer entrepreneurial ventures pitched by attractive men," *PNAS*, 2014, *111* (12), 4427–4431.

Caliendo, Marco, Frank M. Fossen, Alexander Kritikos, and Miriam Wetter, "The gender gap in entrepreneurship: Not just a matter of personality," *CESifo Economic Studies*, 2015, *61* (1), 202–238.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka, "The long-run effects of disruptive peers," *American Economic Review*, 2018, *108* (11), 3377–3415.

Cook, Lisa D and Chaleampong Kongcharoen, "The Idea Gap in Pink and Black," 2010.

Criado Perez, Caroline, *Invisible Women: Data Bias in a World Designed for Men* 2019.

Deaton, Angus and John Muellbauer, "An Almost Ideal Demand System," *The American Economic Review*, 1980, *70* (3), 312–326.

DellaVigna, Stefano and Matthew Gentzkow, "Uniform pricing in us retail chains," *The Quarterly Journal of Economics*, 2019, *134* (4), 2011–2084.

Dunn, Thomas and Douglas Holtz-Eakin, "Financial Capital, Human Capital, and the Transition to Self-Employment: Evidence from Intergenerational Links," *Journal of Labor Economics*, 2000, *18* (2), 282–305.

Einiö, Elias, "Mixing the Rich and Poor: The Impact of Peers on Education and Earnings," 2019.

Feenstra, Robert C, "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84* (1), 157–177.

Fernandez, Raquel and Richard Rogerson, "Sorting and long-run inequality," *Quarterly Journal of Economics*, 2001, *116* (4), 1305–1341.

Foellmi, Reto and Josef Zweimüller, "Income Distribution and Demand-induced Innovations Income Distribution and Demand-induced Innovations," *Review of Economic Studies*, 2006, *73* (212), 941–960.

Guzman, Jorge and Aleksandra Kacperczyk, "Gender Gap in Entrepreneurship," 2018.

and **Scott Stern**, "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 15 US States, 1988-2014," 2016.

, **Jean Joohyun Oh, and Ananya Sen**, "What Motivates Innovative Entrepreneurs ? Evidence from Three Field Experiments," 2019.

Hanushek, Eric A., John F. Kain, Jacob M. Markman, and Steven G. Rivkin, "Does peer ability affect student achievement?," *Journal of Applied Econometrics*, 2003, *18* (5), 527–544.

Hayek, Friedrich August, "The use of knowledge in society," *The American economic review*, 1945, *35* (4), 519–530.

Hoxby, Caroline M., "Peer Effects in the Classroom: Learning From Gender and Race Variation," 2000.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," 2019.

Hvide, Hans K. and Paul Oyer, "Dinner Table Human Capital and Entrepreneurship," 2019.

Jaravel, Xavier, "The unequal gains from product innovations: Evidence from the us retail sector," *The Quarterly Journal of Economics*, 2019, *134* (2), 715–783.

Jones, Benjamin F., "The burden of knowledge and the "death of the renaissance man": Is innovation getting harder?," *Review of Economic Studies*, 2009, *76* (1), 283–317.

Kanze, Dana, Laura Huang, Mark A. Conley, and E. Tory Higgins, "We ask men to win and women not to lose: Closing the gender gap in startup funding," *Academy of Management Journal*, 2018, *61* (2), 586–614.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, "Female Inventors and Inventions," *SSRN Electronic Journal*, 2019, (799330), 1–43.

Kremer, Michael, "How much does sorting increase inequality?," *Quarterly Journal of Economics*, 1997, *112* (1), 115–139.

Lucas, Robert E. and Benjamin Moll, "Knowledge growth and the allocation of time," *Journal of Political Economy*, 2014, *122* (1), 1–51.

Lybbert, Travis J. and Nikolas J. Zolas, "Getting patents and economic data to speak to each other: An 'Algorithmic Links with Probabilities' approach for joint analyses of patenting and economic activity," *Research Policy*, 2014, *43* (3), 530–542.

Malmström, M, J Johansson, and J Wincent, "We Recorded VCs' Conversations and Analyzed How DifferentlyThey Talk About Female Entrepreneurs," *Harvard Business Review*, 2017, pp. 1–6.

Mishkin, Elizabeth, "Gender and Sibling Dynamics in the Intergenerational Transmission of Entrepreneurship," *Working Paper*, 2017.

Pugsley, BW and E Hurst, "What Do Small Businesses Do?," *Brookings Papers on Economic Activity*, 2011, (Fall).

Rivera-Batiz, Luis A and Paul M Romer, "Economic Integration and Endogenous Growth," 1991, *106* (2), 531–555.

Romano, Joseph P. and Michael Wolf, "Stepwise multiple testing as formalized data snooping," *Econometrica*, 2005, *73* (4), 1237–1282.

Romer, Paul M., "Endogenous technological change," *Journal of Political Economy*, 1990, *98* (5), S71–S102.

Schmookler, Jacob, *Invention and economic growth*, Harvard University Press, 1966.

Shane, Scott, "Prior knowledge and the discovery of entrepreneurial opportunities," *Organization science*, 2000, *11* (4), 448–469.

Stern, Scott, "Do scientists pay to be scientists?," *Management Science*, 2004, *50* (6), 835–853.

Thébaud, Sarah, "Gender and entrepreneurship as a career choice: Do self-assessments of ability matter?," *Social Psychology Quarterly*, 2010, *73* (3), 288–304.

Toole, Andrew A., Stefano Breschi, Ernest Miguelez, Amanda Myers, Edoardo Ferrucci, Valerio Sterzi, Charles A.W. DeGrazia, Francesco Lissoni, and Gianluca Tarasconi, "Progress and Potential: A Profile of Women Inventors on U.S. Patents," 2019.

Von Hippel, Eric, "Lead users - a source of novel product concepts," *Management Science*, 1986, *32* (7), 791–806.

Figure 1: Parent Income and the Direction of Entrepreneurship in the U.S.

Panel A: Probability of Self-Employment by Parent Income
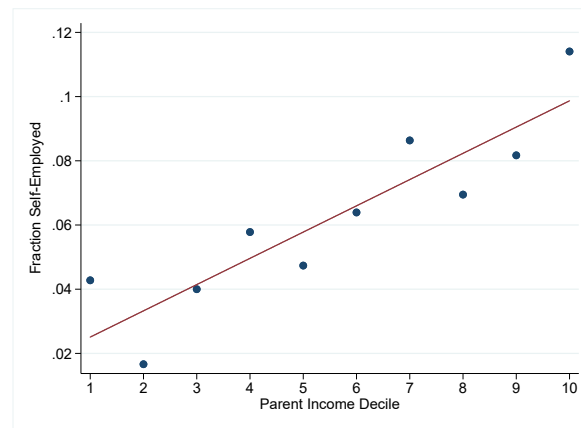
Panel B: Income Elasticities and Parent Income
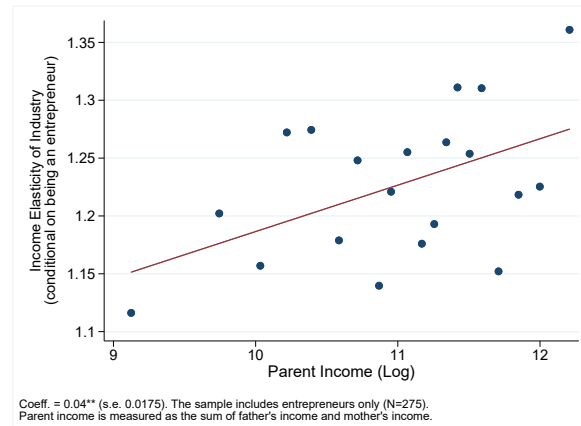
*Note*s: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 2: Gender and Age Similarity between Patent Inventors and Consumers across Industries

A. Gender

B. Age
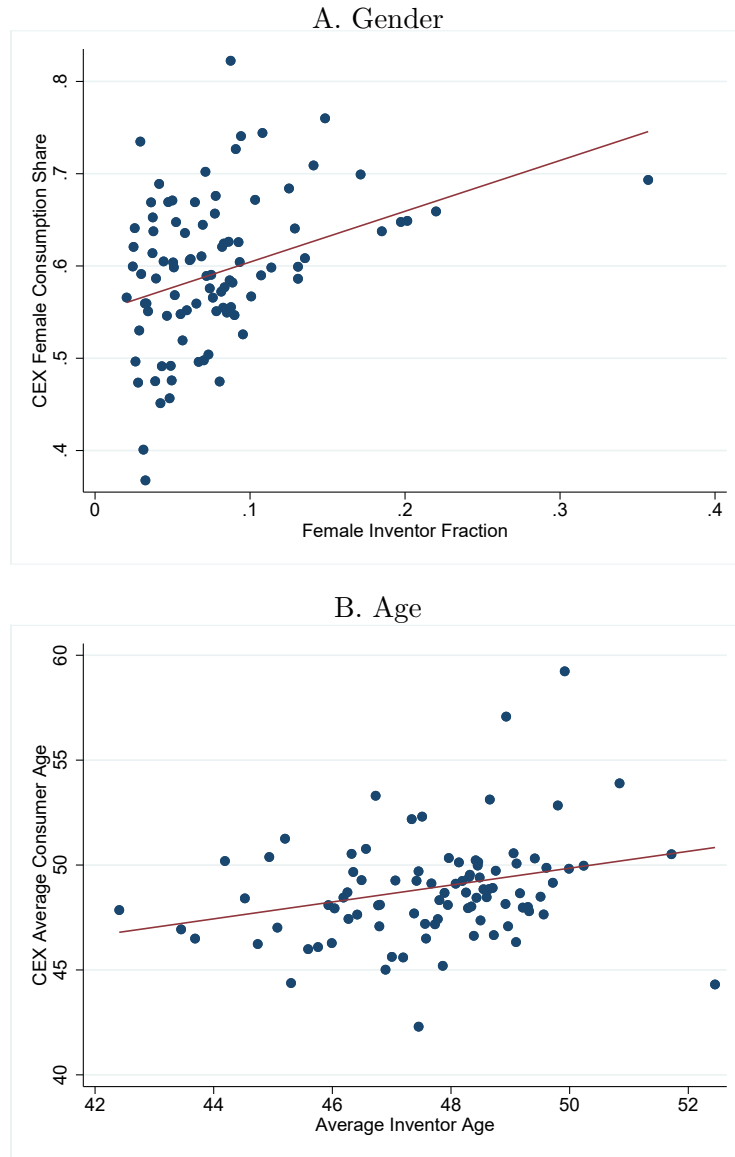
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 3: User-Innovator Relationships within Phone Applications

A. Female User Fraction vs. Founder Gender
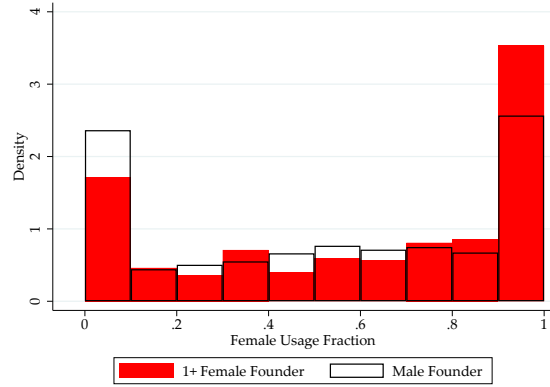
B. Female User Fraction vs. VC Gender
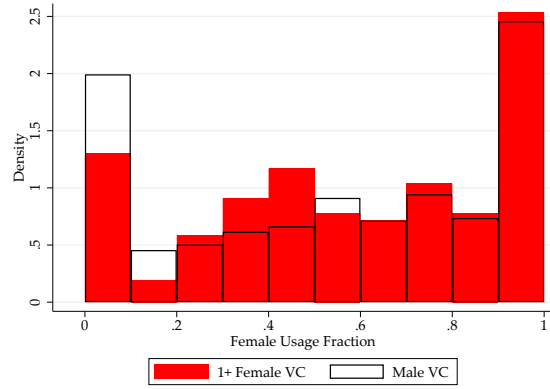
C. Home State Sales Share
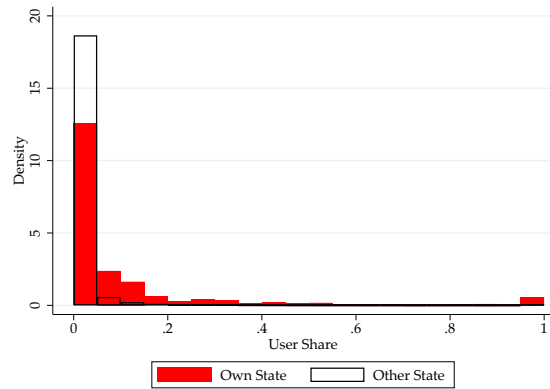
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 4: Consumer-Innovator Relationships within Consumer Packaged Goods

A. Sales Share to Female Consumers vs. Founder Gender (Startups)
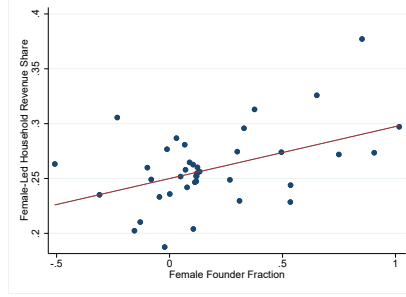
B. Sales Share to Female Consumers vs. VC Partner Gender (Startups)
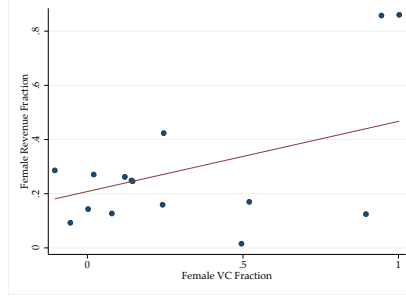
C. Sales Share to Female Consumers vs. Patent Inventor Gender (All Firms)
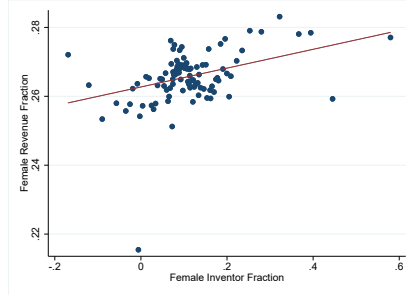
D. Sales-weighted Consumer Age vs. Founder Age (Startups)
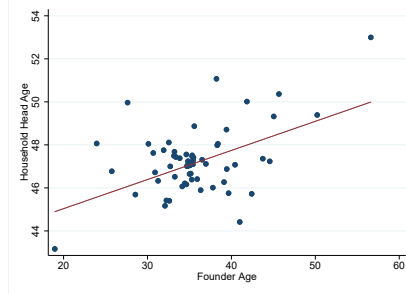
Notes: See Section 2 for a description of the data and Section 3 for the methodology.

Figure 5: Income elasticity of industry by own parent income

A. Inventors

Coeff. = .0044*** (s.e. 0.0005).
Parent income is measured as the sum of father's income and mother's income.
Data sources: Finnish FLEED and FOLK datasets and patent applications 2007-2015 (N=12448113).
The scatter plot uses sample averages by parent income decile.
The smallest bin includes 2217874 observations.

B. Income Elasticities and Parent Income

Coeff. = .0304** (s.e. 0.0125). The sample includes inventors (N=7962).
Parent income is measured as the sum of father's income and mother's income.
Data sources: Finnish FLEED and FOLK datasets and patent applications. Inventors in 2007-2015.
The scatter plot uses sample averages by 5-percentile bins.
The smallest bin includes 398 observations.

Notes:

Figure 6: Gender Similarity between Patent Inventors and Consumers across Industries

Coeff. = .103*** (s.e. 0.016). Aggregate data by 4-digit industry (1701 industry-year observations).
Data sources: Finnish FLEED and FOLK datasets and patent applications. Inventors in 2007-2015.
The scatter plot uses averages by 5-percentile bins for x-axis values above zero
and a separate category for zero values. The smallest bin is based on 89 individual-level
observations. All bins together are based on 21807 individual-level observations.

Figure 7: Study peer parent income and industry expenditure elasticity

Coeff. = .00089 (s.e. .00041). N = 197966.
Bin scatter plot with equal-sized bins
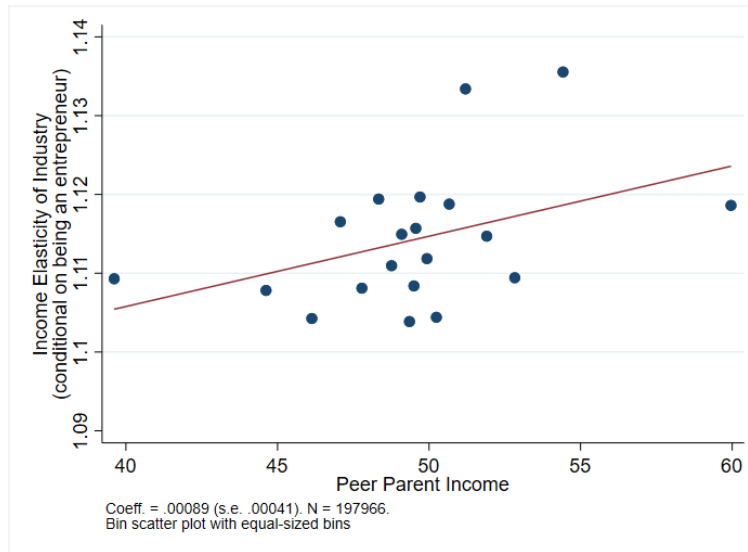
Notes: The figure displays the relationship between parent income of study peers and industry expenditure elasticity in a sample including entrepreneurs. The figure plots the residuals from separate regressions of industry income elasticity and study peer parent income on own parent income, school-by-program fixed effects, school-by-start-year fixed effects, and dummies for gender, age, and year of outcome measurement. Data sources: Statistics Finland FLEED and FOLK datasets.

Figure 8: Simulations of Multi-Sector Model

Panel A: Varying Barriers
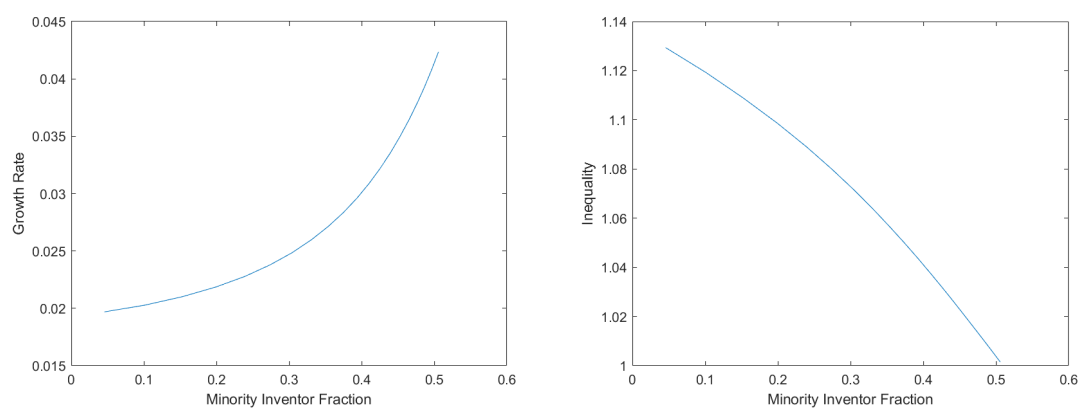
Panel B: Varying Innovators' Own-Sector Bias
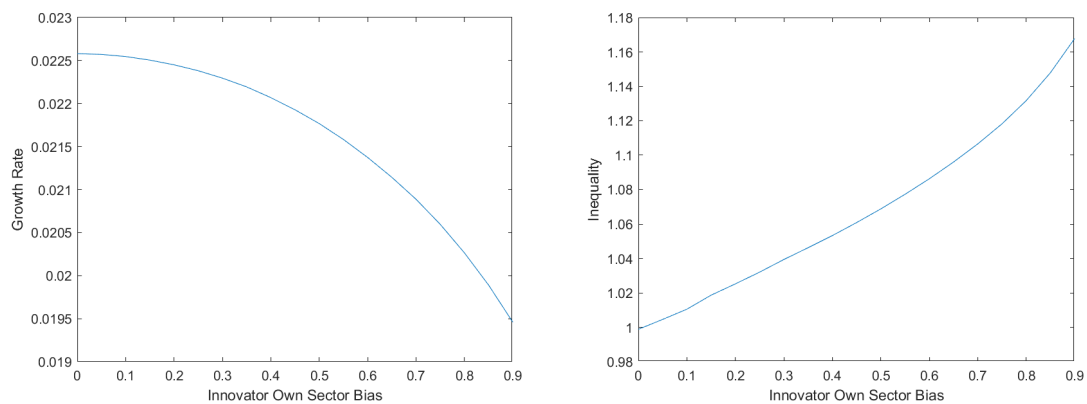
Table 1: Summary Statistics

	Phone Applications	Consumer Packaged Goods
# VC-Backed Startups	1,679	158
Female Founders ≥ 1	0.14	0.24
Female VC Partner ≥ 1	0.06	0.04
# New Products	3,380	4,058
# Categories/Subcategories	14/54	90/294
# Panelists	50,725	168,775
Timeframe	2017-2019	2004-2016

Notes: See Section 1 for a description of the datasets.

Table 2: Innovator-Consumer Similarity by Gender and Age across Industries

	Specification			OLS Estimates		
#	Consumer-based Outcome (LHS)	Patent Inventor Char. (RHS)		Coeff.	Constant	Sample
1.	Share of Sales to Households	Fraction of Female Inventors		0.551** (0.105)	0.549** (0.012)	All industries, $N = 325$
2.	Share of Sales to Households	Fraction of Female Inventors		0.100** (0.048)	0.549*** (0.012)	Compustat, $N = 206$
3.	Average Consumer Age, weighted by Sales	Average Age of Patent Inventors		0.402*** (0.147)	29.73** (7.002)	All industries, $N = 325$
4.	Average Consumer Age, weighted by Sales	Average Age of Patent Inventors		0.007 (0.014)	47.41*** (0.816)	Compustat, $N = 206$

Notes: See Section 2 for a description of the data and Section 3 for the methodology. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

35

Table 3: Innovator-Consumer Similarity within Industries

A. Within Phone Applications

	Specification			OLS Estimates		
#	Founder Char. (LHS)	User-based Outcome (RHS)	Baseline	With Subcat. F.E.	Mean of Dep. Var.	Sample
1.	Female	Share of Time Use by Female Users	0.092*** (0.034)	0.082** (0.032)	0.542	All apps, $N = 3,380$
2.	Home State	Share of Time Use in Founder Home State	0.086*** (0.007)	0.041*** (0.008)	0.02	All apps, $N = 3,380$

B. Within Consumer Packaged Goods

	Specification			OLS Estimates		
#	Innovator Char. (LHS)	Consumer-based Outcome (RHS)	Baseline	With Module F.E.	Mean of Dep. Var.	Sample
1.	Female Founder	Share of Sales to HHs with Female Head	0.030 (0.024)	0.047** (0.021)	0.25	New Goods, $N = 4,058$
2.	Age of Founder	Average Consumer Age, weighted by Sales	0.151*** (0.047)	0.135** (0.052)	47.2	New Goods, $N = 4,058$
3.	Fraction of Female Patent Inventors	Share of Sales to HHs with Female Head	0.039** (0.019)	0.027* (0.015)	0.265	All goods, $N = 1,094,229$

Notes: See Section 2 for a description of the data and Section 3 for the methodology. Both analyses are conducted at the product level, and standard errors are clustered by firm. $*p < 0.1$, $** p < 0.05$, $*** p < 0.01$.

Table 4: Falsification tests for the study peer design

Dependent variable	Own parent income		Parent income, study peers		Adjusted p-value	Dependent mean
	coef.	s.e.	coef.	s.e.		
	(1)	(2)	(3)	(4)	(5)	(6)
	A. At age 28-42 (N=198,710)					
Predicted expenditure elasticity	0.000442***	(0.000010)	0.000022	(0.000038)	-	1.11
	B. Pre-determined characteristics (N=50,446)					
Female	-0.0003***	(0.0001)	0.00009	(0.00037)	0.98	0.40
Earnings	0.0143***	(0.0025)	-0.0120	(0.0148)	0.87	10.89
Employed (%)	0.0573***	(0.0075)	-0.0101	(0.0385)	0.80	76.5
Years of schooling	0.0032***	(0.0004)	0.0021	(0.0021)	0.99	11.59
Married	-0.0066	(0.0059)	-0.0205	(0.0374)	0.92	80.47
Foreign	-0.0308***	(0.0025)	-0.0021	(0.00856)	0.98	0.84
Primary language Finnish	0.0417***	(0.0045)	0.0439**	(0.0212)	0.36	92.75
Unemployment benefits	-0.0016***	(0.0004)	-0.0023	(0.0024)	0.85	0.72
General housing allowance	-0.0027***	(0.0002)	-0.0008	(0.0009)	0.87	0.32
Number of parents employed	0.3583***	(0.0097)	-0.0574	(0.0363)	0.60	1.66
Parent years of schooling	0.0905***	(0.0013)	0.0077	(0.0050)	0.61	23.32
Parent pension income	0.0191***	(0.0033)	-0.0153	(0.0126)	0.71	5.19
Parent unemployment benefits	-0.0187***	(0.0007)	0.0045	(0.0034)	0.68	1.53
Parent general housing allowance	-0.0090***	(0.0004)	-0.0010	(0.0012)	0.87	0.37

Notes: The baseline estimation sample, consisting of 50,446 individuals who become entrepreneurs. Each row presents coefficients from a separate regression of the dependent variable indicated by the row label on own parent income and study peer parent income. All regressions include dummies for gender, year of outcome measurement, age of outcome measurement, school-by-programme fixed effects, and school-by-start-year fixed effects. Pre-determined characteristics are measured one year before the first study year. Income, earnings, benefits, allowances, and pensions are in thousand euro. In Panel B, predicted expenditure elasticity is the best linear prediction of expenditure elasticity at age 28-42 based on the pre-determined characteristics listed in Panel B. Column (5) shows the stepdown p-values (Romano and Wolf, 2005) adjusted for multiple hypthesis testing of 14 coefficient on peer parent income in Panel B. Standard errors robust for clustering by school and programme start year are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by *, **, and ***, respectively.

Table 5: Impacts of study peer parent income

	All conditional on being an entrepreneur	Below-median own parent income	Above-median own parent income	Dependent mean, all
A. Industry elasticity	0.00091** (0.00041)	-0.00008 (0.00067)	0.00189*** (0.00055)	1.11
B. Earnings	0.01481 (0.01524)	-0.01339 (0.02253)	0.03220 (0.02407)	11.1
C. Income	0.01447 (0.01799)	-0.01407 (0.02949)	0.01553 (0.02626)	28.18
D. Income above 50th percentile	0.000036 (0.000320)	-0.000401 (0.000518)	0.000424 (0.000469)	0.66
E. Income above 90th percentile	0.000115 (0.000278)	-0.000279 (0.000420)	0.000216 (0.000424)	0.20
F. Income above 99th percentile	-0.000041 (0.000159)	-0.000195 (0.000238)	-0.000060 (0.000246)	0.04
Observations	198,710	99,302	99,408	
Students	50,446	23,100	26,579	
Classes	20,267	12,703	13,743	
Schools	549	527	516	

Notes: The table displays the estimates of the impact of parent income of study peers on the dependent variable indicated by the row label. Outcomes are measured from age 28 onward. Parent income, income, and earnings are in thousands of euros. All regressions include dummies for gender, year of outcome measurement, age of outcome measurement, school-by-programme fixed effects, school-by-start-year fixed effects, and control variables for pre-determined characteristics listed in Panel B of Table 4. Income percentiles are calculated from the sample including all students. Standard errors robust for clustering by school and programme start year are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by *, **, and ***, respectively.

Table 6: Distributional Effects from "Equal Opportunity Innovator Pools"

Sample	Counterfactual Innovator Pool	Distributional Effects rel. to Baseline Pool	Uses estimate from
Within phone apps	Female founder instead of male founder	**46.4%** larger gains for female users	Table 3.A spec. 1
Within consumer packaged goods	Female founder instead of male founder	**27%** larger gains for HHs with female head	Table 3.B spec. 1

Notes: See Section 4 for a description of the methodology..

Appendix to "What Drives Inclusive Innovation? The Importance of Innovators' Backgrounds"

Elias Einio, *VATT Institute for Economic Research*

Josh Feng, *University of Utah – Eccles School of Business*

Xavier Jaravel, *London School of Economics*
June 2020

A   Assessing Distributional Effects Using a Static Framework

In the main text, we focus on dynamic general equilibrium effects. Here, we develop a static framework that can be used to quantify the distributional effects that may result from a more equal distribution of socio-economic backgrounds and gender among innovators. Throughout this section, we assume that marginal innovators exhibit the same average patterns found in Section III.

A.A   Quantitative Framework

We present calibrations that use the descriptive estimates to assess the distributional effects of unequal access to innovation.

Consumer preferences and welfare effects of innovation. Assume consumers have CES preferences over a set of goods index by $k \in \Omega_t$, within a product module. The set of available goods Ω_t may vary over time, for instance as startups introduce new goods in the market. Utility of agent i is CES:

$$U_i = \left(\sum_{k \in \Omega_t} \omega_{k,i} q_{k,i}^{1-\sigma} \right)^{1/(1-\sigma)},$$

where σ is the elasticity of substitution between products within the product category, $q_{k,i}$ is the quantity of good k consumed by agent i, and $\omega_{k,i}$ is a taste parameter reflecting the intensity of i's preference for k.

Feenstra (1994) showed that with CES preferences, the welfare gain from the introduction of new goods (i.e., the set of available products Ω_t increases in the product category) can be expressed as a percentage of i's current income (equivalent variation for household i) as follows:

$$\pi_i = \frac{1}{1-\sigma} log \left(\frac{1 + GrowthSpendingContinuedGoods_i}{1 + GrowthTotalSpending_i} \right).$$

Assuming inelastic labor supply and taking the wage as the numeraire,[22] the growth of total spending is effectively normalized to zero, and the growth in spending on continued products is mechanically related to the share of spending on new goods S_i^N, with

$$GrowthSpendingContinuedGoods_i = -S_i^N.$$

With a first-order Taylor expansion around $S_i^N = 0$, the formula becomes:

$$\pi_i \approx \frac{S_i^N}{\sigma - 1}$$

For example, with $\sigma = 6$, a spending share on new goods of 2% is equivalent to a fall in inflation of $\frac{10}{5} = 2\%$ in welfare terms.

Innovators' backgrounds. We now consider the welfare impact of two startups that cater to different types of consumers. We consider a startup drawn from the baseline distribution of entrepreneur background ("Baseline"), which is skewed toward rich parents and male innovators, compared with a hypothetical equalized distribution ("Equal"), which could match the population gender ratio and the population distribution of parental income.

Distributional effects across consumer groups. Next, we consider two representative households, denoted "Type 1" and "Type 2." We derive the welfare comparison between these two households when transitioning from the "Baseline" to the "Equal" distribution of innovator background. We then bring the formulas to the data, computing the distributional effects between high- and low-income households, as well as between male and female customers.

As discussed in Deaton and Muellbauer (1980), CES preferences for a representative agent can be interpreted as the aggregation of discrete-choice logit preferences from a population of underlying agents. We assume that the startups drawn from different distributions of innovator backgrounds have similar elasticities of substitution σ, but differ in their preferences $\omega_{i,k}$, such that they may have different spending shares on the new goods introduced by different startups. S_1^N is the spending share of the "Type 1" representative agent from the bottom income decile on the startup's products, while S_2^N corresponds to the spending share of the "Type 2" representative agent. Y_1 and Y_2 denote the total spending of the two household types.

Next, consider the entry of a new startup in the market. Each representative household buys products from this startup depending on its preferences, and the relative welfare gains are given by the following formula:

$$\frac{\pi_1}{\pi_2} \approx \frac{S_1^N}{S_2^N} = \frac{S_1^N \cdot Y_1}{S_2^N \cdot Y_2} \cdot \frac{Y_2}{Y_1} = \frac{R_1/R_2}{Y_1/Y_2}$$

where R_i denotes the total sales of the startup to representative household i. The ratio of sales to each of the representative agents is thus a sufficient statistic for the relative welfare effect, when appropriately normalized by the ratio of total spending of each of the agent. This result is intuitive: when agents have CES preferences with similar elasticities of substitution σ's, welfare differences can be reduced to differences in spending shares, and in turn to differences in the firm's revenue share from each agent, with a normalization for total purchasing power.

We wish to examine whether one of the household types benefits more from transitioning to a new distribution of innovator background. The unequal welfare effect across household types from the actual distribution of entrepreneur background ("Baseline") vs. a counterfactual distribution ("Equal") can be expressed as:

$$\Delta W \equiv \frac{\pi_1^{Equal}/\pi_1^{Baseline}}{\pi_2^{Equal}/\pi_2^{Baseline}} = \frac{R_1^{Equal}/R_2^{Equal}}{R_1^{Baseline}/R_2^{Baseline}}.$$

The relative welfare effect is thus governed by the share of sales to household groups of different types.

Linking the formula to regression specifications. We can link the formula for ΔW to our regression estimates from Tables 2 and 3, which document the revenue share of different types of startups for different types of households. Denoting by λ the share of sales to "Type 1" households, we can write $R_1/R_2 = \lambda/(1-\lambda)$. Tables 2 and 3 are directly informative about λ for startups with different innovator backgrounds.

A.B Results

Using the formula from the previous section, we find that equalizing access to innovation could substantially reduce inequality and increase real income growth at the median of the income distribution. We also find that increase the share of female innovators would benefit female consumer substantially more. The results are reported in Table 3.

Denoting by λ the share of revenue from female-led households (which is the outcome in Table 2), we can write $\lambda = R_F/(R_M + R_F) \implies R_F/R_M = \lambda/(1-\lambda)$. Our previous results are directly informative about the expected λ for a female-founded startup ($\lambda^F = 0.2974$) compared to a male-founded startup ($\lambda^M = 0.25$). Therefore our estimates imply:

$$\Delta W = \frac{\lambda^F/(1-\lambda^F)}{\lambda^M/(1-\lambda^M)} = \frac{0.2974/(1-0.2974)}{0.25/(1-0.25)} \approx 1.2698$$

Thus, in the consumer packaged goods sample our preferred specification indicates that the welfare gains from female-founded startups are 27 percent larger for the representative female household, relative to the representative male household.[23] This number increases to 46.4 percent in the sample of phone applications.[24]

Using these values, we can calibrate the relative effect of having a more representative pool of innovators. Currently, women represent about 12 percent of innovators and provide 46.4 percent more relative value to female consumers. Relative to a world where 50 percent of innovators are women, the current system provides 16.5 percent higher relative benefit to male consumers. Assuming that the current flow of new goods is comparable to the long-run flow, this estimate is comparable to the 12 percent cost-of-living inequality value simulated using the calibrated growth model.

B Growth Model Appendix

B.A Single-Sector Model

To start, we present a simple product variety model of growth. We then derive results on long-run growth with homogeneous and heterogeneous research productivity.

Baseline model The basic model follows Romer (1990) and Acemoglu (2009). There is a representative household with preference[25]

$$\int_0^\infty e^{-\rho t} log(C(t)) dt,$$

where

$$C(t) = [\int_0^{N(t)} c(\nu, t)^{\frac{\epsilon-1}{\epsilon}} d\nu]^{\frac{\epsilon}{\epsilon-1}}.$$

---

[23]Note that this difference is substantially larger than the difference in revenue shares from female households that arises between female-founded and male founded startup. As shown in Column (4) of Table 2, the revenue share from female-led households is 18.96% larger for female-founded startups compared with male-founded startups (29.74% vs 25%). The welfare calculation from CES utility indicates that the comparison of revenue shares if biased downward. Intuitively, a downward bias arises because the revenue from female consumers appears in both the numerator and the denominator in the revenue share approach, while it appears only in the numerator of the welfare-relevant formula.

[24]To be applied to the setting of free phone applications, the quantitative framework presented above can be re-cast using a time constraint instead of a budget constraint.

[25]The log form can be generalized to a CRRA function. This just alters the Euler equation by the inverse of the risk aversion parameter: $\frac{\dot{C}(t)}{C(t)} = \frac{1}{\theta}(r(t) - \rho)$.

The representative household maximizes lifetime discounted utility subject to the interest rate r. Entrepreneurs invent new varieties, receive perpetual patents, and produce the variety using a linear production function:

$$y(\nu, t) = l(\nu, t).$$

The research production function takes the form

$$\dot{N}(t) = \eta N(t) L_R(t),$$

where η is the research productivity of all individuals and L_R is the amount of labor allocated to research rather than production. All labor must be allocated to production or research, such that the labor market clears:

$$\int_0^{N(t)} l(\nu, t) d\nu + L_R(t) \leq L.$$

Throughout, we assume that there is a single market-clearing wage $w(t)$. This will be especially important when heterogeneity is added in the model.

Solving the Model   The intuition behind solving the model is that along a balance growth path, where the growth rate, the wage, the interest rate, and the labor allocated to research are all constant, we need to allocate labor to research and production in a way that creates

1. Balance, in the form of the Euler equation. Allocating too much to research does not make sense if consumers are impatient.

2. Correct incentives, in the form of entrepreneur indifference equations

These equations will pin down the labor allocation along a balanced growth path. The quickest way to arrive at the balanced growth path solution is to take the following steps:

1. Along a balanced growth path, there will be some equilibrium interest rate r^*

2. People are indifferent between production work and entrepreneurship: $\eta N(t)V(t) = w(t)$, where $V(t)$ is the NPV of profits per variety

3. Per-period profits from entrepreneurship is governed by consumer preferences (pins down the markup) and amount of production labor used (divided equally between the varieties):

A4

$\pi(t) = \frac{1}{\epsilon-1}\frac{L-L_R(t)}{N(t)}w(t) = \frac{1}{\epsilon-1}(L - L_R(t))\eta V(t)$. The second expression comes from plugging in the indifference condition.

4. $V(t) = \frac{\pi(t)}{r^*}$ based on balanced growth path interest rates. Plugging in allows us to express $r^* = \frac{\eta}{\epsilon-1}(L - L_R(t))$

5. Finally, we can plug everything into the Euler equation, which ensures that the entrepreneurial labor allocation and growth in varieties accords with time preferences. The main thing to note here is that we can use household preferences to express the relationship between the growth rate of consumption and varieties $C(t) \sim N(t)^{\frac{1}{\epsilon-1}}$

$$r(t) - \rho = \frac{1}{\epsilon-1}\frac{\dot{N}(t)}{N(t)}$$
$$\frac{\eta}{\epsilon-1}(L - L_R(t)) - \rho = \frac{1}{\epsilon-1}\eta L_R^*$$
$$\frac{\eta}{\epsilon-1}L - \rho = \frac{2}{\epsilon-1}\eta L_R^*$$
$$L_R^* = \frac{L}{2} - \frac{\epsilon-1}{2\eta}\rho$$

The equilibrium growth rate is then $g^* = \frac{1}{2}(\frac{\eta}{\epsilon-1}L - \rho)$.

Adding Barriers to Innovation Now, we add barriers to innovation for a subset of agents. This can be modeled in two ways. One way is to have a "real" barrier that directly impacts research productivity. Individuals facing barriers will be $1 - \tau$ times as productive, which could reflect lack of opportunity to develop skills or lack of access to adequate funding. A second way is to have a "preference" barrier, which only distorts the decision to enter into research, but does not impact research productivity conditional on entry. This could capture factors such as discrimination and preferences.

In both cases, the indifference equation becomes

$$(1 - \tau)\eta N(t)V(t) = w(t)$$

and the only difference between the two cases is the law of motion governing product variety growth.

Proposition 2. *In an expanding product variety model with access barriers, if the size of the unrestricted group is greater than L_R^*, there exists a BGP where aggregate consumption expenditure grows at $g^* = \frac{1}{2}(\frac{\eta}{\epsilon-1}L - \rho)$.*

Adding the wedge does not violate any equilibrium conditions in the baseline model. Only people from the unrestricted group participate in research production, because the equilibrium wage rate is higher than the effective returns to entrepreneurship for the restricted group. As long as the unrestricted group is larger than L_R^*, there is no impact on growth.[26] Intuitively, unrestricted individuals can fill in for individuals who are restricted and still produce varieties at the same rate. Thus, barriers to innovation do not matter for growth if there is homogeneity in preferences and research productivity.

Heterogeneous Research Productivity   Next, we show that entry barriers do matter for long-run growth if there is heterogeneous research productivity. To obtain closed-form solution, we consider a uniform distribution of ability.[27] We assume that research productivity in the population is equal to $\eta\kappa$, where $\kappa \sim U[0,1]$. In equilibrium there is a cutoff $\bar{\kappa}$ below which the agents work in production. This leads to the following varieties growth equation:

$$\frac{\dot{N}(t)}{N(t)} = \eta L \int_{\bar{\kappa}}^1 \kappa f(\kappa)d\kappa = \frac{1-\bar{\kappa}^2}{2}\eta L$$

The marginal agent is indifferent between production wages and the returns to entrepreneurship:

$$\bar{\kappa}\eta N(t)V(t) = w(t),$$

The monopolist profits per line per period is governed by the number of production workers:

$$\pi(\nu,t) = \frac{1}{\epsilon-1}\frac{\bar{\kappa}L}{N(t)}w(t) = \frac{1}{\epsilon-1}\bar{\kappa}^2\eta L V(t)$$

We can then solve for the balanced growth path interest rate

$$V(t) = \frac{\pi(t)}{r^*}$$
$$r^* = \frac{\bar{\kappa}^2\eta L}{\epsilon-1}$$

We can then plug everything into the Euler equation

---

[26]Note that the assumption that there is a single labor market clearing wage is important for this result. This prevents firms from offering lower wages to the restricted group, knowing that they do not want to pursue entrepreneurship.

[27]The solution can be easily generalized to other distributions, including a Pareto distribution. However, in the case of the Pareto distribution, the model doesn't generally have a closed form solution. In that case, the equation governing the number of entrepreneurs is a polynomial with an order that depends on the Pareto parameter.

$$\frac{\dot{C}(t)}{C(t)} = r(t) - \rho = \frac{1}{\epsilon - 1}\frac{\dot{N}(t)}{N(t)}$$

$$\frac{\bar{\kappa}^2 \eta L}{\epsilon - 1} - \rho = \frac{1}{\epsilon - 1}\cdot\frac{1 - \bar{\kappa}^2}{2}\eta L$$

Solving the model, we obtain $\bar{\kappa} = \sqrt{\frac{1}{3} + \frac{2\rho(\epsilon-1)}{3\eta L}}$ and $g^* = \frac{1}{3}\left(\frac{\eta L}{\epsilon - 1} - \rho\right)$.[28]

In terms of inequality in earnings, people working in production will earn the prevailing market wage w. So do the marginal entrepreneurs in each group. Inframarginal entrepreneurs earn at some multiple relative to the marginal entrepreneur $\frac{\kappa}{\bar{\kappa}}w$, where $\bar{\kappa}$ is the equilibrium cutoff. The total wages for a group, conditional on the equilibrium cutoff $\bar{\kappa}$, will be:

$$W = \int_0^{\bar{\kappa}} wf(\kappa)d\kappa + \int_{\bar{\kappa}}^1 \frac{\kappa}{\bar{\kappa}}wf(\kappa)d\kappa$$

$$= w\left(\kappa + \frac{1}{2\bar{\kappa}} - \frac{\bar{\kappa}}{2}\right)$$

$$= \frac{w}{2}\left(\bar{\kappa} + \frac{1}{\bar{\kappa}}\right)$$

Because $0 \leq \bar{\kappa} \leq 1$, the total wage will be decreasing in $\bar{\kappa}$:

$$\frac{\partial W}{\partial \bar{\kappa}} = \frac{w}{2}\left(1 - \frac{1}{\bar{\kappa}^2}\right) < 0$$

Given that $\bar{\kappa}_1 = (1 - \tau)\bar{\kappa}_2 < \bar{\kappa}_2$, the wedge will decrease the relative wages of the restricted group.

Heterogeneous Research Productivity with Access Barriers   Next, we add entry barriers. We focus here on the case of "preference" barriers.[29] Assume that there are two groups of size $\frac{L}{2}$, and group 2, the "restricted" group, faces barriers to entrepreneurship. Both groups have the same underlying ability distributions. Now, there will be two cutoffs $\bar{\kappa}_1, \bar{\kappa}_2$ that are governed by the following indifference equations:

$$\bar{\kappa}_1 \eta N(t)V(t) = w(t),$$

---

[28]Note that this expression is smaller than the growth rate in the baseline model with homogeneous research productivity, but this mechanically results from the fact that now the average productivity in the population is $\frac{\eta}{2}$. The productivity parameter can be normalized such that the growth rate remains unchanged.

[29]We can also solve the case where the barrier affects productivity. The solution is similar to the solution to the basic heterogeneous research productivity model above, but with a step function representing the PDF of the ability distribution.

$$(1 - \tau)\bar{\kappa}_2 \eta N(t) V(t) = w(t),$$

which implies that $\bar{\kappa}_1 = (1 - \tau)\bar{\kappa}_2$. In other words, the marginal agent in the restricted group has higher research productivity. This means that the overall innovation rate can be expressed as:

$$
\begin{aligned}
\frac{\dot{N}}{N} &= \frac{1 - \bar{\kappa}_1^2}{4}\eta L + \frac{1 - \bar{\kappa}_2^2}{4}\eta L \\
&= \frac{\eta L}{2}\left(1 - \frac{1 + (1 - \tau)^2}{2}\bar{\kappa}_2^2\right)
\end{aligned}
$$

We can solve out the model, expressing everything in terms of $\bar{\kappa}_2$:

$$
\begin{aligned}
c(\nu, t) &= l(\nu, t) = \frac{(\bar{\kappa}_1 + \bar{\kappa}_2)L}{2N(t)} = \frac{(2 - \tau)\bar{\kappa}_2 L}{2N(t)} \\
C(t) &= \frac{2 - \tau}{2}\bar{\kappa}_2 L N(t)^{\frac{1}{\epsilon - 1}} \\
\pi(\nu, t) &= \frac{1}{\epsilon - 1}\frac{(2 - \tau)\bar{\kappa}_2 L}{2N(t)}w(t) \\
&= \frac{1}{\epsilon - 1} \cdot \frac{(2 - \tau)(1 - \tau)}{2}\bar{\kappa}_2^2 \eta L V(t) \\
r^* &= \frac{1}{\epsilon - 1} \cdot \frac{(2 - \tau)(1 - \tau)}{2}\bar{\kappa}_2^2 \eta L
\end{aligned}
$$

We can then plug this back into the Euler equation to solve for $\bar{\kappa}_2$:

$$
\frac{1}{\epsilon - 1}\frac{(2 - \tau)(1 - \tau)}{2}\bar{\kappa}_2^2 \eta L - \rho = \frac{1}{\epsilon - 1} \cdot \frac{\eta L}{2}\left(1 - \frac{1 + (1 - \tau)^2}{2}\bar{\kappa}_2^2\right)
$$

$$
\bar{\kappa}_2^2\left((2 - \tau)(1 - \tau) + \frac{1 + (1 - \tau)^2}{2}\right) = \frac{2\rho(\epsilon - 1)}{\eta L} + 1
$$

$$
\bar{\kappa}_2^2 = \frac{1 + \frac{2\rho(\epsilon - 1)}{\eta L}}{(3\tau^2 - 8\tau + 6)}
$$

which leads to the growth rate presented in the following proposition:

Proposition 3. *In an expanding product variety model with uniformly distributed research productivity, $\eta\kappa, \kappa \sim U[0, 1]$, and a random half of the population facing barriers to entrepreneurship, the aggregate consumption expenditure along the unique BGP grows at*

$$
g^* = \frac{1}{\epsilon - 1}(2 - \tau)(1 - \tau) \cdot \frac{\eta L + 2\rho(\epsilon - 1)}{(3\tau^2 - 8\tau + 6)} - \rho
$$

Greater τ reduces growth rate, increases the number of entrepreneurs from the unrestricted group, decreases the number of entrepreneurs from the restricted group, and increases relative wages across the groups.

Note that g^* is equal to the solution without access barriers at $\tau = 0$ and is decreasing in τ on the interval from 0 to 1. Therefore, entry barriers for one group reduce long-run growth. In addition, it also creates differences in average income across the two groups, hence inequality, because there are fewer inframarginal entrepreneurs in the restricted group.

B.B Solving a Basic Two-Sector Model

Two-Sector Model with Homogeneous Consumers Next, we turn to solving a basic two-sector model, which provides a foundation for the core results in the main text. We first derive the BGP with two sectors and homogeneous consumers. Consumers maximize lifetime discounted utility and their instantaneous utility is Cobb-Douglas over two sectors:

$$\int_0^\infty e^{-\rho t} log(C(t)) dt,$$

$$C(t) = C_1(t)^\alpha C_2(t)^{1-\alpha}$$

where each aggregate consumption index depends on the number of varieties available in the sector

$$C_i(t) = [\int_0^{N_i(t)} c_i(\nu, t)^{\frac{\epsilon-1}{\epsilon}} d\nu]^{\frac{\epsilon}{\epsilon-1}}$$

Given the Cobb-Douglas form, $C_1(t) = \alpha C(t), P_2 C_2(t) = (1-\alpha)C(t)$. Sector 1 prices are the numeraire and the implied relative price $P_2 = \left(\frac{N_2}{N_1}\right)^{\frac{1}{1-\epsilon}}$. Let $N(t) = N_1(t) + N_2(t)$. The innovation productivity in each sector i is:[30]

$$\dot{N}_i(t) = \eta_i N(t) L_{iR}(t)$$

Let L_{iM}, L_{iR} be the production and research labor allocated to sector i, respectively. The labor market clearing is now

---

[30]If sectors had different productivities and varieties production only depended on varieties in the same sector, then there would be explosive growth in one sector and the equilibrium would not admit the existence of a balanced growth path.

$$L = L_{1R} + L_{2R} + L_{1M} + L_{2M}$$

Along a BGP, the growth rate of varieties in both sectors will be constant, along with the amount of labor allocated to each sector, and to research and production within each sector. We can derive the relationships between research and manufacturing across sectors, and then plug into the Euler equation to pin down the absolute levels.

On a balanced growth path,

$$r^* = \frac{1}{\epsilon - 1} \frac{N(t)}{N_i(t)} L_{iM}(t)\eta_i, \forall i$$

This implies that $\eta_1 L_{1M}/N_1(t) = \eta_2 L_{2M}/N_2(t)$.

The amount of labor needed to produce varieties to accommodate Cobb-Douglas tastes will be $\frac{L_{1M}}{L_{2M}} = \frac{\alpha}{1-\alpha}$. Put together, this means that along a BGP, varieties will satisfy

$$\frac{N_1(t)}{N_2(t)} = \frac{\alpha \eta_1}{(1-\alpha)\eta_2}$$

The BGP interest rate can then be written in terms of one of the production labor variables:

$$r^* = \frac{1}{\epsilon - 1}(\eta_1 + \frac{1-\alpha}{\alpha}\eta_2)L_{1M}$$

Because varieties grow at the same rate over time, this pins down the research allocation ratio across sectors $\frac{L_{1R}(t)}{L_{2R}(t)} = \frac{\frac{\alpha}{1-\alpha}\eta_1 + \eta_2}{\eta_1 + \frac{1-\alpha}{\alpha}\eta_2} = \frac{\alpha}{1-\alpha}$

We can then plug everything into the Euler equation to establish a relationship between manufacturing and research:

$$L_{1M} = \frac{\alpha \rho(\epsilon - 1)}{(\alpha \eta_1 + (1-\alpha)\eta_2)} + L_{1R}$$

And then we can solve everything out to arrive at allocations and equilibrium growth rate. The research allocation and growth rate are given by:

$$L_{1R}^* = \frac{\alpha}{2}\left(L - \frac{\rho(\epsilon - 1)}{(\alpha \eta_1 + (1-\alpha)\eta_2)}\right), L_{2R}^* = \frac{1-\alpha}{2}\left(L - \frac{\rho(\epsilon - 1)}{(\alpha \eta_1 + (1-\alpha)\eta_2)}\right),$$

$$g^* = \frac{1}{2}\left(\frac{1}{\epsilon - 1}(\alpha \eta_1 + (1-\alpha)\eta_2)L - \rho\right). \tag{A1}$$

The formula reduces to the growth rate in the one-sector model when $\alpha = 0$ or $\alpha = 1$.

Two-Sector Model with Taste heterogeneity We now examine the impact of taste heterogeneity across consumers. Following the assumptions in the main text, let there be two groups of consumers. Group 1 has size $1 - \delta$ and preference parameter α. Group 2 has size δ and preference parameter α'. Let $\alpha > \alpha'$, i.e. group 2 has a stronger preference for sector 2 compared with group 1. Let $\delta < \frac{1}{2}$, i.e., group 2 is the minority group.

In this model, the equilibrium growth rate is given by (A1), using $\tilde{\alpha} = (1 - \delta)\alpha + \delta\alpha'$ as the effective taste parameter. However, in this case, the model features inequality across consumers through an expenditure channel. Welfare inequality persists at a constant ratio along the BGP. Starting from the primitives, we can derive the ratio of consumption across the two groups:

$$
\begin{aligned}
\frac{C^1}{C^2} &= \frac{[L_{1M}N_1(t)^{\frac{1}{\epsilon-1}}]^\alpha [L_{2M}N_2(t)^{\frac{1}{\epsilon-1}}]^{1-\alpha}}{[L_{1M}N_1(t)^{\frac{1}{\epsilon-1}}]^{\alpha'} [L_{2M}N_2(t)^{\frac{1}{\epsilon-1}}]^{1-\alpha'}} \\
&= \left(\frac{L_{1M}}{L_{2M}}\right)^{\alpha-\alpha'} \left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}} \\
&= \left(\frac{(1-\delta)\alpha + \delta\alpha'}{(1-\delta)(1-\alpha) + \delta(1-\alpha')}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{\eta_1}{\eta_2}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}
\end{aligned}
\tag{A2}
$$

This expression shows that welfare inequality is determined by two terms, corresponding to exogenous determinants of technology and endogenous market size effects. The second term depends on the ratio of technological parameters, $\frac{\eta_1}{\eta_2}$, i.e. consumer groups who prefer the sector with higher research productivity have relatively higher welfare along the BGP. The ratio in the first term corresponds to the relative sales of the two sectors, $\frac{\tilde{\alpha}}{1-\tilde{\alpha}}$; it shows that welfare is higher for consumer groups who prefer the sector with a higher market size. Intuitively, the larger sector endogenously gets more varieties along the BGP. The market size imbalance is weaker when the minority consumer group is larger.

B.C Solving the Two-Sector Model with Access Barriers

Here, we derive the solution to a generalized version of the model in the main text with preference barriers. This can be applied specifically to the case with innovator specialization and access barriers, where the group facing barriers still ends up entering into entrepreneurship. We highlight the differences in the equations that pin down a balanced growth path solution.

Let there be a wedge $1 - \tau$ in facing innovators in sector 2. The labor market indifference equation becomes

$$w(t) = \eta_1 N(t) V_1(t) = (1 - \tau)\eta_2 N(t) V_2(t)$$

The production labor allocation remains the same, because tastes have not changed. On a balanced growth path, varieties will satisfy

$$\frac{N_1(t)}{N_2(t)} = \frac{\alpha\eta_1}{(1-\alpha)(1-\tau)\eta_2}$$

The equilibrium interest rate will be $r^* = \frac{1}{\epsilon-1}(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2)L_{1M}$. The ratio of research labor devoted to each sector is now $\frac{L_{1R}(t)}{L_{2R}(t)} = \frac{\frac{\alpha}{(1-\alpha)(1-\tau)}\eta_1+\eta_2}{\eta_1+\frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2} = \frac{\alpha}{(1-\alpha)(1-\tau)}$, based on equal growth rates in varieties and labor market indifference. Using the Euler equation, we can pin down the

$$\frac{\dot{C}(t)}{C(t)} = r(t) - \rho = \frac{\alpha}{\epsilon-1}\frac{\dot{N}_1(t)}{N_1(t)} + \frac{1-\alpha}{\epsilon-1}\frac{\dot{N}_2(t)}{N_2(t)}$$

$$\frac{1}{\epsilon-1}(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2)L_{1M} - \rho = \frac{1}{\epsilon-1}\left(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2\right)L_{1R}$$

We can then plug it into the labor budget constraint to solve out for research:

$$L_{1R} = \frac{\alpha}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2)}\right)$$

The equilibrium growth rate will be

$$
\begin{aligned}
g^* &= \frac{\dot{C}(t)}{C(t)} = \frac{1}{\epsilon-1}\left(\eta_1 + \frac{(1-\alpha)(1-\tau)}{\alpha}\eta_2\right)L_{1R} \\
&= \frac{1}{\epsilon-1}(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2) \cdot \frac{1}{1+\alpha+(1-\alpha)(1-\tau)}\left(L - \frac{\rho(\epsilon-1)}{(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2)}\right) \\
&= \frac{1}{1+\alpha+(1-\alpha)(1-\tau)}\left(\frac{1}{\epsilon-1}(\alpha\eta_1 + (1-\alpha)(1-\tau)\eta_2)L - \rho\right)
\end{aligned}
$$

This formula reduces back to the simple two-sector model when $\tau = 0$. An interesting aspect of the solution is that there are cases where the wedge is actually good for long-run growth rates, because the wedge skews allocation towards the more productive research sector. People move from innovating in sector 2 to innovating and producing in sector 1. However, in the case where research productivities are equal across sectors, which we focus on in the main text, the expression is clearly decreasing in the size of the wedge.

To arrive at the results in Proposition 1, we can evaluate the expressions under specific cases:

1. In the case when $\eta' > (1-\tau)\eta$, innovation in both sectors will be done by group 1 individuals as long as $L_{1R} + L_{2R} < (1-\delta)L$. This is equivalent to plugging in $\eta_1 = \eta, \eta_2 = \eta', \tau = 0$ into the core expression.

2. In the case when $\eta' < (1-\tau)\eta$, there will be specialization and group 2 faces barriers to entering innovation in sector 2. $\eta_1 = \eta, \eta_2 = \eta, \tau \neq 0$.

B.D   The role of interactions with peers

Motivated by the quasi-experimental evidence on the impact of peers on the direction of entrepreneurship, we introduce a new parameter to parsimoniously model sociological interactions. Starting from the same natural research productivities η and η' defined in the model above, we introduce the parameter ψ, which governs the intensity of interaction with individuals from the other group. We assume that when an agent interacts with another agent with a higher productivity, they will learn from them and increase their own productivity in innovating for the other group. Through these interactions, research productivity in the sector for which agents have a comparative disadvantage is given by $\eta'' = \psi\eta + (1-\psi)\eta'$. Thus η'' depends on the degree of interactions and gives the effective productivity in the sector where an agent has a comparative disadvantage.

Proposition 4. *In a two-sector model with preference heterogeneity, innovator specialization, and real barriers, higher ψ leads to a weakly higher growth rate and weakly lower inequality along the balanced growth path.*

To derive Proposition 4, several cases must be considered. Start with the case where $\eta' < (1-\tau)\eta$. Absent exposure, individuals specialize in research. As we increase the exposure parameter, η'' increases, which at first has no effect on outcomes. For a high enough ψ, $\eta'' > (1-\tau)\eta$. At this point, Group 1 individuals will enter entrepreneurship in sector 2, while Group 2 individuals drop out of entrepreneurship. As exposure grows further, the growth rate $g^* = \frac{1}{2}\left(\frac{1}{\epsilon-1}(\tilde{\alpha}\eta + (1-\tilde{\alpha})\eta'')L - \rho\right)$ will increase and the inequality across the two groups will decrease $\frac{C^{1*}}{C^{2*}} = \left(\frac{\alpha(1-\delta)+\alpha'\delta}{(1-\alpha)(1-\delta)+(1-\alpha')\delta}\right)^{\frac{\epsilon(\alpha-\alpha')}{\epsilon-1}} \cdot \left(\frac{\eta}{\eta''}\right)^{\frac{\alpha-\alpha'}{\epsilon-1}}$.

In the case of a "preference" barrier, Proposition 5 holds in the region where $\eta'' > (1-\tau)\eta$. At the point where $\eta'' = (1-\tau)\eta$, Group 1 individuals replace Group 2 individuals in innovating, causing a drop in long-run growth rate, because Group 1 individuals are less productive at innovating in Sector 2. Eventually, as ψ rises further, the growth rate will rise above the growth rates under specialization.

Table A1: Phone Apps – Female Usage Fraction, Founders, and Funders

Panel A – Core Analysis

	(1)	(2)	(3)
Female Founder Frac.	0.0917***	0.0852***	0.0818**
	(0.0337)	(0.0330)	(0.0323)
Fixed Effects	None	Category	Subcategory
Observations	1403	1401	1401
R^2	0.008	0.032	0.044

Panel B – Female Usage Fraction vs. Funders and Founders

	(1)	(2)	(3)
Female VC Frac.	0.160*	0.148*	0.153*
	(0.0635)	(0.0665)	(0.0670)
Female Founder Frac.	0.0928*	0.0736*	0.0480
	(0.0368)	(0.0367)	(0.0372)
Fixed Effects	None	Category	Subcategory
Observations	1403	1401	1401
R^2	0.008	0.032	0.044

C Additional Descriptive Results

C.A U.S. Setting

We provide additional results and robustness checks for the phone apps analysis and the consumer goods analysis. We also provide robustness results for our across-industry evidence and document results for green patenting.

For phone apps, we provide detailed regression results and additional results surrounding venture capital funding. First, Panel A of Table A1 shows that the relationship between founder and user gender is present across all apps and within categories and subcategories, and generally similar in magnitude. Panel B provides correlations when both female founder fraction and female venture capital partner fraction are used as predictors for female consumer usage. We find that both factors are positively associated with usage fraction, although the results for venture capital partner are stronger and more robust.

For consumer packaged goods, we provide robustness checks and additional results. As noted earlier, consumer packaged goods purchases are measured at the household level, unlike phone apps. Given the lack of precision, we focus on female-led households in the core analysis. Here, we provide additional robustness checks. In Panel A of Table A2, we present comparisons within different classification levels (specification 4 is the preferred result presented in the main text). The

results are strongest within narrow categories. Panel B presents results where we weight spending by gender composition of the entire household. We also find that female founders sell more to more "female" households. Panel C shows the results using a binary predictor instead of female founder fraction. We recover slightly smaller point estimates, albeit still statistically significant, when using this coarser measure. Panel D presents results within single households and also find significantly positive correlations, although the point estimates are again smaller relative to the baseline rate.

Beyond robustness checks, we also present additional results that are only relevant for this setting. Table A3 presents correlations between female fraction in the venture capital partners who fund a given startup and the female founder fraction. We find strong positive correlations both in the general Crunchbase sample and the sample of startups that sell goods tracked by Nielsen. Table A4 controls for per-unit price of goods. Unlike phone apps, where more than 95 percent of apps are free, consumer goods have associated prices. One concern is that the correlations we find are driven by price (e.g., female entrepreneurs create higher-priced goods and female consumers are less price sensitive). We find strong correlations even after controlling for price.

Next, we provide additional robustness related to the cross-industry patenting results. These correspond to the first estimate in Table 2. We document in Table A5 that the estimates are robust to a variety of methodological choices and controls. These include focusing on the most patent-intensive industries, controlling for high level industry fixed effects, and weighting by different transformations of the number of patents in the industry.

Finally, we document additional results surrounding the direction of innovation within the consumer packaged goods sample and Crunchbase startups more broadly. Our approach relies on the descriptions of companies in Crunchbase ("description" variable). We note whether the description mentions B-corp, healthy/health, sustainable/sustainability/environment, and kid/child/baby/toddler. Panel A of Table A6 presents evidence from the full Crunchbase sample. We find that female-founded companies are positively associated with all of the groups of outcomes noted above. The estimates for kids and B-corporation are particularly large relative to the baseline rate of mentions. We repeat the analysis in the set of companies matched between Nielsen consumer packaged goods and Crunchbase. We find similar effects, although there are no companies that mention B-corp in this subsample and the sustainability mentions are not statistically significantly different between male and female founded startups. We again find large positive associations in terms of mentions of healthy and kids. Further systematic analysis is needed to understand differences along various social impact dimensions generated by male and female entrepreneurs.

Table A2: Consumer Packaged Goods – Robustness Checks

Panel A – Core Results With Different Fixed Effects

	(1)	(2)	(3)	(4)
Female Founder Frac.	0.0302	0.0201	0.0436**	0.0474**
	(0.0240)	(0.0222)	(0.0197)	(0.0216)
Constant	0.252***	0.254***	0.251***	0.250***
	(0.00713)	(0.00464)	(0.00336)	(0.00365)
FE	–	Department	Group	Module
Observations	4054	4044	4044	4054
R^2	0.001	0.016	0.065	0.155

Panel B – Weighted by Family Member Gender

	(1)	(2)	(3)	(4)
Female Founder Frac.	0.0352**	0.0292**	0.0433***	0.0367**
	(0.0139)	(0.0133)	(0.0138)	(0.0159)
Constant	0.533***	0.534***	0.532***	0.533***
	(0.00431)	(0.00438)	(0.00287)	(0.00300)
FE	–	Department	Group	Module
Observations	4054	4044	4044	4054
R^2	0.0015	0.011	0.049	0.129

Panel C – Indicator for Any Female Founder

	(1)	(2)	(3)	(4)
Female Founder Indicator	-0.00260	-0.00565	0.0277*	0.0277*
	(0.0250)	(0.0202)	(0.0145)	(0.0146)
Constant	0.257***	0.257***	0.251***	0.251***
	(0.00681)	(0.00420)	(0.00336)	(0.00355)
FE	–	Department	Group	Module
Observations	4054	4044	4044	4054
R^2	0.000	0.016	0.064	0.154

Panel D – Within Single Households

	(1)	(2)	(3)	(4)
Female Founder Frac.	0.0607	0.0671*	0.0684*	0.0509
	(0.0387)	(0.0363)	(0.0370)	(0.0383)
Constant	0.687***	0.686***	0.686***	0.688***
	(0.0236)	(0.0122)	(0.00663)	(0.00557)
FE	–	Department	Group	Module
Observations	4054	4044	4044	4054
R^2	0.001	0.016	0.065	0.155

Table A3: Founder Gender vs. Venture Capital Partner

	(1)	(2)	(3)	(4)
Female VC Frac.	0.0911***	0.0830***	0.758***	0.876***
	(0.0121)	(0.0125)	(0.139)	(0.180)
Constant	0.0750***	0.155***	0.162***	0.578
	(0.00188)	(0.0228)	(0.0513)	(0.785)
Sample	All	All	Nielsen	Nielsen
Founding Year FE		x		x
Log Funding Control		x		x
Observations	15803	14289	51	47
R^2	0.007	0.025	0.175	0.460

Notes: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A4: Revenue from Female-Led Households (Controlling for Price)

	(1)	(2)	(3)	(4)
Female Founder Frac.	0.0308	0.0199	0.0446**	0.0485**
	(0.0238)	(0.0222)	(0.0195)	(0.0209)
Per-Unit Price	-0.000468**	-0.000322	-0.000395	-0.000770
	(0.000230)	(0.000222)	(0.000273)	(0.000597)
Constant	0.256***	0.257***	0.254***	0.257***
	(0.00723)	(0.00465)	(0.00447)	(0.00656)
Department FE		x		
Product Group FE			x	
Product Module FE				x
Observations	4054	4044	4044	4054
R^2	0.004	0.017	0.066	0.156

Notes: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Across-Industry Inventor Gender Robustness Checks

	Female Consumption Share					
Female Inventor Frac.	0.551***	0.546***	0.885***	0.417***	0.634***	0.948***
	(0.105)	(0.0931)	(0.178)	(0.122)	(0.110)	(0.180)
Constant	0.549***	0.546***	0.509***	0.560***	0.536***	0.492***
	(0.0123)	(0.0131)	(0.0201)	(0.0121)	(0.0136)	(0.0214)
Selection (no. patents)	None	≥ 25th	≥ median	None	None	None
Weighting	None	None	None	None	Log Pats	Sqrt Pats
Fixed Effects	None	None	None	NAICS-2	None	None
Industries	325	246	163	325	325	325

Table A6: Company Description Text vs. Female Founder

Panel A – All Crunchbase Companies

	B-Corp	Healthy	Sustainable	Kids
Female Founder Frac.	0.000682**	0.0386***	0.0126***	0.0532***
	(0.000319)	(0.00303)	(0.00257)	(0.00256)
Constant	0.000291***	0.0600***	0.0505***	0.0188***
	(0.0000551)	(0.000749)	(0.000692)	(0.000435)
Observations	115,554	115,554	115,554	115,554
R^2	0.000	0.002	0.000	0.009

Panel B – Nielsen Crunchbase Matched Sample

	B-Corp	Healthy	Sustainable	Kids
Female Founder Frac.	–	0.112***	0.0351	0.111***
		(0.0384)	(0.0297)	(0.0329)
Constant	–	0.115***	0.0751***	0.0517***
		(0.00959)	(0.00794)	(0.00668)
Observations	1,300	1,300	1,300	1,300
R^2		0.009	0.001	0.017

Notes: *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

We also discuss our approach to studying green patenting in the full sample of U.S. patents and provide formal results. As noted in the main text, we follow the approach described in Aghion et al. (2016) to classify green vs. other energy patents. Specifically, we classify patents as green if their international patent classification (IPC) falls under the following categories: B60L11, B60L3, B60L15, B60K1, B60W10/08, B60W10/24, B60W10/26, B60K6, B60W20, B60L7/10, B60L7/20, B60W10/28, B60L11/18, H01M8. The list corrects one typo in the original paper (B60L7/10 instead of B60L7/1). Patents are classified as dirty or "grey" based on the corresponding lists in the original paper. Then, at the patent level, we can compute the fraction of inventors who are female. We also use age data from Jones (2009). The age data has limited coverage, but we use the average age of inventors for which data is available.

Table A7: Green Patents – Full Sample of Patents Granted in the U.S.

Panel A – Clean Patents and Inventor Gender

	Clean Patent	
Female Inventor Fraction	0.326***	
	(0.015)	
Any Female Inventor		0.212***
		(0.007)
Observations	1403	1401
R^2	0.012	0.0316

Panel B – Clean Patents and Inventor Age

|	Clean Patent
Age	-0.001
	(0.0006)
Observations	1243
R^2	0.002

C.B   Trends Over Time

Next, we document trends over time in directional differences. We focus on gender and invention, given the increase in participation over time and data availability over a long period. As noted in Bell et al. (2017), there has been a slow but steady increase in the number of female inventors on U.S. patents. It is possible that marginal inventors do not exhibit directional differences, because the market for female products is already saturated. In this case, an increase in the number of female inventors would not necessarily provide greater benefit to female consumers, and would reduce the homophily coefficient.

We again use USPTO patent data and provide estimates corresponding to the inventor gender estimates in Table 2 but at an annual level, based on the application year for the granted patent. Figure A1 shows that homophily estimates are very similar over the 1995–2019 period, under a couple of specifications. In Panel A, we focus on industries with more than ten patents in a given year. In Panel B, we include all industries but add weights based on the square root of patents in the industry and year. The stability of the estimate holds in both cases despite an increase in female inventor fraction from below 9 percent to about 12 percent.

Figure A1: Trends over time in homophily estimate and female inventor participation

Panel A – Industry-years with more than ten patents
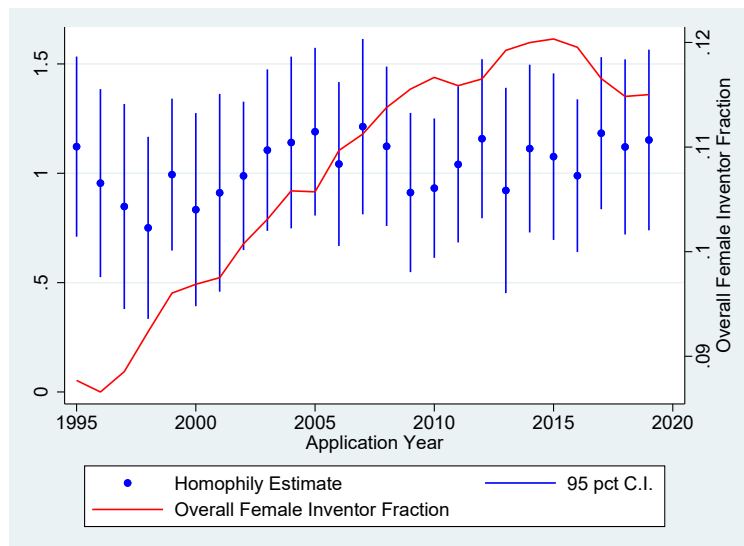
Panel B – All industries, weighted by square root of patents
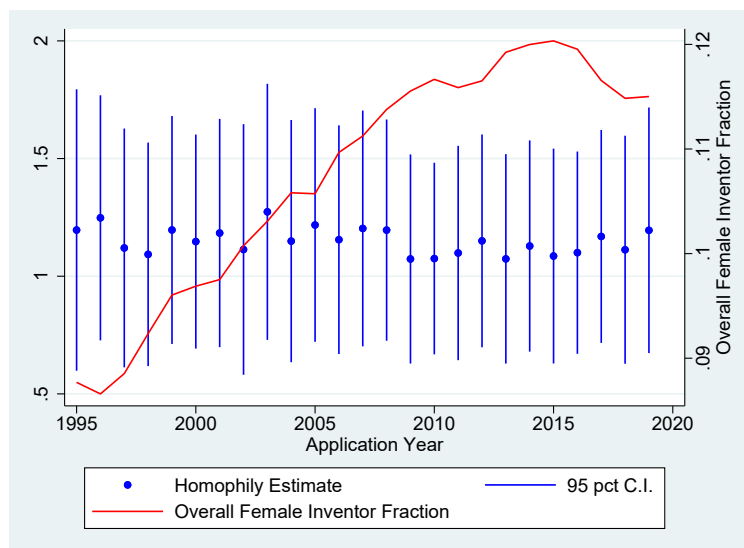
C.C Across-Industry Evidence from Finland

Finally, we present additional descriptive evidence from Finland to complement the descriptive evidence from the U.S. context. The data on entrepreneurship is more comprehensive than available in the PSID, which allows us to obtain more precise estimates.

Similar to the U.S. setting, we find a significant positive relationship between an entrepreneur's parental income and the income elasticity of the industry in which they start businesses. Figure A2 presents the results, which hold with and without the inclusion of agriculture. Next, we also

find a strong positive relastionship between an entrepreneur's gender and the female consumption share in the industry in which they start businesses. Figure A3 presents the results. Finally, we also document that the parental income of entrepreneurs is positively associated with the parental income of the workers they hire. Of course, this relationship can be driven by the industry choice or other unobservable factors beyond a direct homophily effect.

Figure A2: Income elasticity of industry and own parent income, entrepreneurs

A. All Entrepreneurs

Coeff. = .1416*** (s.e. 0.0034). The sample includes entrepreneurs (N=393596)
Parent income is measured as the sum of father's income and mother's income.
Data sources: Finnish FLEED and FOLK datasets. Entrepreneurs in 2007-2015.
The scatter plot uses sample averages by 5-percentile bins.
The smallest bin includes 19678 observations.
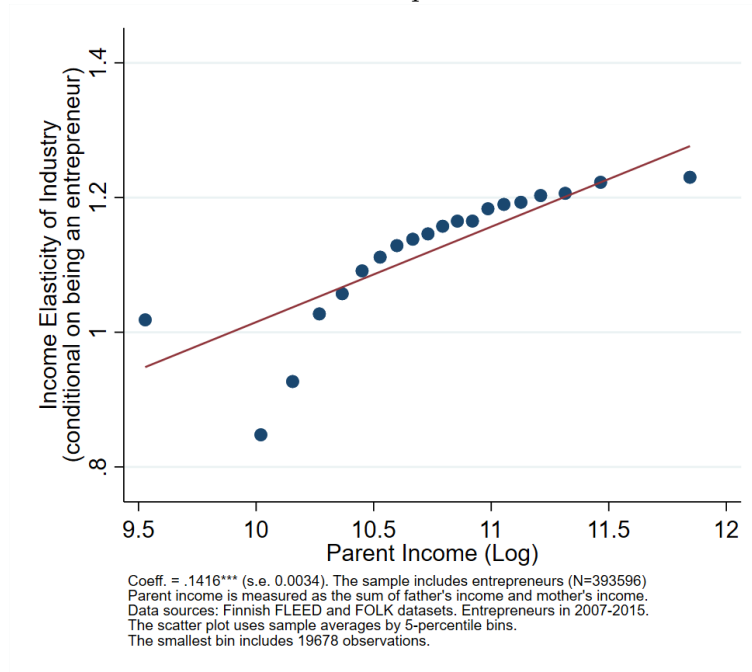
B. Excluding Agriculture

Coeff. = .0119*** (s.e. 0.0027). The sample includes entrepreneurs (N=316993)
Parent income is measured as the sum of father's income and mother's income.
Data sources: Finnish FLEED and FOLK datasets. Entrepreneurs in 2007-2015.
The scatter plot uses sample averages by 5-percentile bins.
The smallest bin includes 15846 observations.
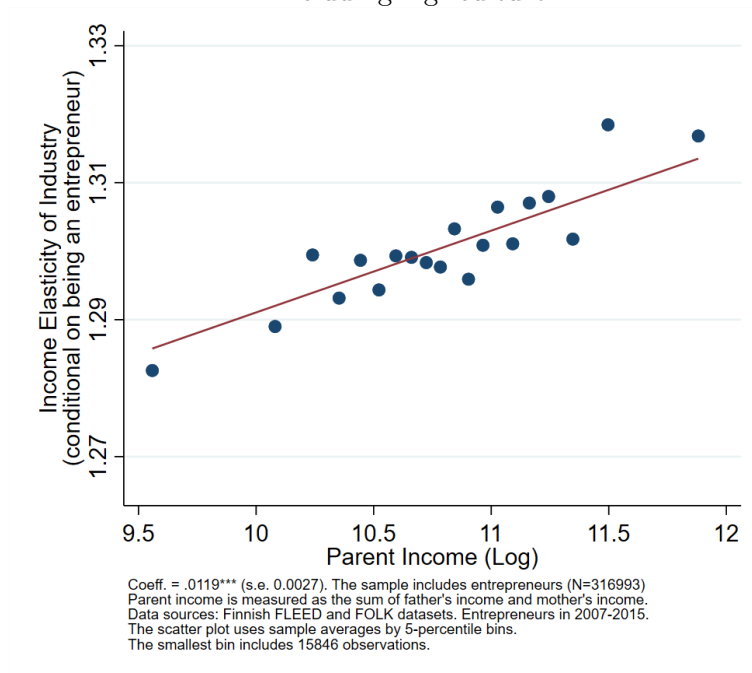
Figure A3: Gender Similarity between Entrepreneurs and Consumers across Industries

Coeff. = .12*** (s.e. 0.021). Aggregate data by 4-digit industry (3981 industry-year observations).
Data sources: Finnish FLEED and FOLK datasets. Entrepreneurs in 2007-2015.
The scatter plot uses averages by 5-percentile bins for x-axis values above zero
and a separate category for zero values. The smallest bin is based on 2538 individual-level
observations. All bins together are based on 1773426 individual-level observations.

Figure A4: Parent income of employees and own parent income

Coeff. = .004*** (s.e. 00).
The sample includes entrepreneurs who have at least one employee (N=167573).
The regression is weighted by the number of employees.
Parent income is measured as the sum of father's income and mother's income.
Data sources: Finnish FLEED and FOLK datasets. Entrepreneurs in 2007-2015.
The scatter plot uses sample averages by 5-percentile bins
The smallest bin includes 6751 observations.

D   Quasi-Experimental Evidence from Conscription Data

In this appendix, we present additional quasi-experimental evidence using an alternative setting from Finland. Relative to the college peers design, this design has a more precise measure of peer exposure, but offers less statistical power due to limited data availability.

We follow the approach in Einiö (2019), who estimates the effects of peer parent income among dormmates in the Finnish military conscription, finding positive impacts of peers from high-income families on earnings and wages. The basic approach is as follows. Within military squadrons of about eighty people, conscripts are assigned to dorms of about eight people. A common way to assign dorms within a squadron is alphabetically. Even within a dorm, individuals tend to bunk with those next to them in alphabetical order, for logistical reasons such as roll call. Therefore, variation in the backgrounds of the two closest alphabetical peers will often translate to variation in peer exposure.

Formally, Einiö (2019) uses the following IV procedures based on the alphabetic rule in assigning conscripts to dorms within squadrons:

$$y_{idst} = \gamma \overline{X}^{(2)}_{(i)ds} + \beta_1 X_{ids} + \alpha_s + \alpha_t + \epsilon_{idst} \tag{A3}$$

$$\overline{X}^{(2)}_{(i)ds} = \rho \overline{Z}^{(2)}_{(i)ds} + \theta_1 X_{ids} + \eta_s + \eta_t + \nu_{idst}. \tag{A4}$$

where $\overline{X}^{(2)}_{(i)ds}$ is the parent income of two alphabetically nearest dormmates of conscript i in dorm d and squadron s; X_{ids} is the conscript's own parent income; and y_{idst} is the outcome of interest measured in year t. The instrumental variable is the parent income of two alphabetically nearest squadmates of the conscript $\overline{Z}^{(2)}_{(i)ds}$. Conditioning on squad fixed effects, α_s and η_s, means that the model identifies γ from variation in the parent income of a conscript's two alphabetically nearest squadmates across the squad's alphabetic ordering. Einiö, 2019 shows that this instrument is uncorrelated with pre-service characteristics of a conscript and his parents, which indicates that the instrument generates within-squad variation in dormmate parent income that is as good as random.

We merge industry income elasticities with the conscript data by unique person identifiers and estimate the IV model in Equations (A3) and (A4) for individuals who are entrepreneurs. Sample size is a challenge here, because dorm data is only available for 24% of conscripts.[31] To increase the sample size, we also include individuals who are in managerial occupations.[32]

---

[31] As reported in Einiö (2019), squadron data is available for all conscripts. The individuals in the sample with dorm data available do not appear to be different to the general population.

[32] Information on managerial occupation is based on the occupation code in the population panel.

Results. Table A8 shows the estimates and Figure A5 the corresponding reduced-form plots. In panel A of table A8, the first-stage coefficient on the instrument is large and statistically significant. The reduced-form and IV estimates are all positive, but significant only for the income elasticity outcome. Panel B shows that these effects are concentrated in the sample of individuals who are in the highest own parent income tercile, similar to the results in the college peer context. When the average parent income of two alphabetically nearest dormmates is increased by ten thousand euro, which is equivalent to around one standard deviation, the income elasticity of the industry is increased by around 0.1, or is 8.5 percent higher compared to the sample mean of 1.22. The results are similar when we use the shares of industry consumption by rich households as the outcomes.

Table A8: The Effects of Peer Parent Income on Industry Expenditure Elasticity and Sales Share to High-Income Consumers

| Dependent variable	First stage	Reduced form	IV	Dependent mean	N
		A. Entrepreneurs and managers			
Expenditure elasticity	0.3358***	0.0014***	0.0043***	1.24	9755
	(0.0377)	(0.0004)	(0.0011)		
Rich share (60k)	0.3358***	0.0003	0.0008	0.63	9755
	(0.0377)	(0.0002)	(0.0008)		
Rich share (100k-30k)	0.3358***	0.0004	0.0011	0.70	9755
	(0.0377)	(0.0003)	(0.0010)		
		B. Entrepreneurs and managers with own parent income in the highest tercile			
Expenditure elasticity	0.3518***	0.0037**	0.0105**	1.23	4319
	(0.0661)	(0.0018)	(0.0051)		
Rich share (60k)	0.3518***	0.0008**	0.0023**	0.62	4319
	(0.0661)	(0.0004)	(0.0010)		
Rich share (100k-30k)	0.3518***	0.0009*	0.0025*	0.70	4319
	(0.0661)	(0.0005)	(0.0014)		

Notes: The figure displays the estimates of the impact of parent income of two alphabetically nearest dormmates on the outcome at age 28-42, using parent income of two alphabetically nearest squadmates as the instrument. Parent income is in thousand euro. All regressions include linear terms for parent income, dummies for calendar year, and fully interacted service start year and squadron dummies. Standard errors allowing for clustering at the level of management unit are in parentheses. The 90%, 95%, and 99% confidence levels are denoted by * , **, and ***, respectively.

Figure A5: The Causal Effect of Peers on the Direction of Entrepreneurship: Peer Parent Income vs. Entrepreneur's Industry Income Elasticity

A. Entrepreneurs and managers
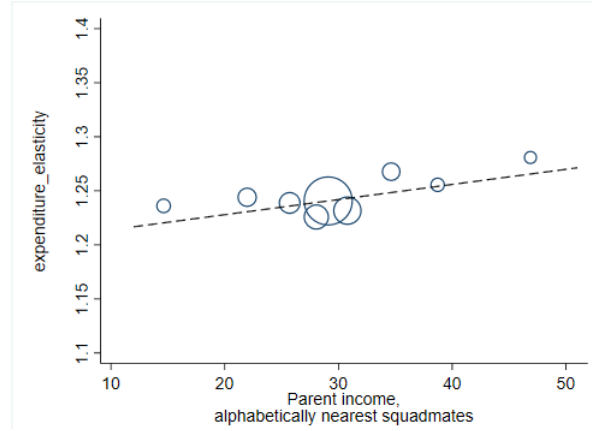
B. Entrepreneurs and managers with
own parent income in the highest tercile
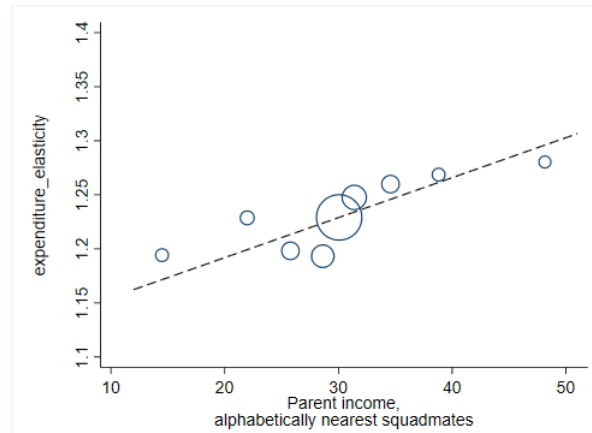
Notes: The figure displays the reduced-form relationship between parent income of two alphabetically nearest squadmates and industry expenditure elasticity in a sample including entrepreneurs and managers (panel A) and a sample restricted to entrepreneurs and managers in the highest own parent income tercile (panel B). The figure plots the residuals from separate regressions of industry income elasticity and squadmate parent income on own parent income, squadron fixed effects, and year dummies. The size of the circle represents the number of conscripts within each bin. Data sources: Finnish FLEED and FOLK datasets and FDF conscript registry.