VATT INSTITUTE FOR
ECONOMIC RESEARCH

VATT WORKING PAPERS 160

Social Push and the Direction of Innovation

Elias Einiö
Josh Feng
Xavier Jaravel

VATT WORKING PAPERS 160

VATT WORKING PAPERS 160

Social Push and the Direction of Innovation

Elias Einiö
Josh Feng
Xavier Jaravel

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Helsinki 2023

VATT WORKING PAPERS 160

Elias Einiö:
VATT Institute for Economic Research

Josh Feng:
University of Utah – David Eccles School of Business

Xavier Jaravel:
London School of Economics and CEPR

VATT Working Papers:
https://doria.fi/handle/10024/147862

Valtion taloudellinen tutkimuskeskus
VATT Institute for Economic Research
Arkadiankatu 7, 00100 Helsinki, Finland

Helsinki, November 2023

Social Push and the Direction of Innovation*

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

October 2023

Abstract

What are the implications of unequal access to innovation careers for the direction of innovation and inequality? Leveraging novel linked datasets in the United States and Finland, we document that innovators create products more likely to be purchased by consumers like them in terms of gender, socioeconomic status, and age. We find that a key explanatory channel is that social exposure causes a shift in the direction of innovation, independent of financial incentives. Incorporating this "social push" channel into a growth model, we estimate that unequal access to innovation careers has a large effect on cost-of-living inequality and long-run growth.

*For thoughtful comments, we are grateful to Daron Acemoglu, Ruchir Agarwal, Philippe Aghion, Victor Bennett, Corina Boar, Jen Brown, Raj Chetty, Louphou Coulibaly, Gary Dushnitsky, Louisa Gagliardi, Matthew Jackson, Ben Jones, Kieran Larkin, Katja Mann, Kyle Myers, Tommaso Porzio, Stefanie Stantcheva, Jason Snyder, Johannes Spinnewijn, Peter Thompson, John Van Reenen, Fabian Waldinger, and Heidi Williams. The article also benefited from comments from seminar participants at the AOM meetings, ASSA meetings, BAIC, Bank of Finland, BYU-Utah Joint Research Day, Female Entrepreneurship Seminar Series, IIOC, Kauffman Showcase, Max Planck Institute, NBER Summer Institute, Society for Economic Dynamics, Society of the Study of Economic Inequality, SMS meetings, Strategy Science, UCLA Anderson, and VMACS. We also thank Gabriel Leite Mariante for excellent research assistance, and PatentsView for providing data on the gender of patent inventors. The results reflect our own analyses, calculated based in part on data from Nielsen Consumer LLC and marketing databases provided through the Nielsen IQ Datasets at the Kilts Center for Marketing Data Center at the University of Chicago Booth School of Business. The conclusions drawn from the Nielsen IQ data are ours and do not reflect the views of Nielsen. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein. Feng acknowledges support from the National Institute on Aging, Grant Number T32-AG000186, the Ewing Marion Kauffman Foundation, and Sorenson Center for Discovery and Innovation at the University of Utah. This paper replaces and supersedes an earlier working paper that was first circulated in May 2019 under the title "What Drives Inclusive Innovation? The Importance of Innovators' Backgrounds".

"If nobody else is going to invent a dishwashing machine, I'll do it myself."

Josephine Cochrane, inventor of the dishwasher (U.S. patent no. 355,139)

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., Linder, 1961; Schmookler, 1966; Aghion and Howitt, 1992; Acemoglu, 2002; Acemoglu and Linn, 2004; Acemoglu, 2007; Jaravel, 2019). At the same time, the discovery and pursuit of entrepreneurial opportunities depends on the distribution of information in society (e.g., Hayek, 1945), often requires engagement with specific real-world problems and users (e.g., Von Hippel, 1986; Shane, 2000), and is incentivized by non-pecuniary benefits (Stern, 2004; Hurst and Pugsley, 2011). Because innovators from different socio-demographic backgrounds are likely to be exposed to different ideas and to be responsive to different intrinsic motivations, they may pursue different types of innovations, independent of financial incentives. Despite the apparent plausibility of this "social push" channel, little is known about its economy-wide importance for the direction of innovation, its relevance in terms of innovators' family backgrounds, and its implications for economic inequality.[1]

Understanding 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 regions are under-represented among innovation leaders including startup founders, patent inventors, and venture capitalists (e.g., Bell et al., 2019b; Agarwal and Gaulé, 2020; Hvide and Oyer, 2020; Aghion et al., 2023).

Several historical examples suggest that innovators' personal experience may shape their entrepreneurial vision, and in turn inequality across the socio-demographic groups who benefit from these innovations. Despite the lack of opportunities for most American women in the

[1]Contemporaneous research focusing on gender, specific sectors, and early stages of innovation documents that patents with a female lead inventor are more likely to focus on women's health (Koning et al., 2021), provides evidence of gender homophily between scientific authors and commercializing inventors (Koffi and Marx, 2023), and shows how undergraduate gender diversity affects the direction of scientific research (Truffa and Wong, 2022).

19th century, Josephine Cochrane managed to receive her U.S. patent for her "Dish Washing Machine" in 1886; she wanted to protect her fine china and avoid having to hand-wash them herself.[2] Madam C.J. Walker, the first female self-made millionaire in America, made her fortune in the late nineteenth century by developing and marketing a line of cosmetics and hair care products for Black women; she suffered severe dandruff and other scalp ailments herself from a young age. Louis Braille invented a world-famous reading and writing system for visually impaired people; he was himself blinded at the age of three as a result of an accident in his father's workshop.

How important are inventors' background and social experience for the direction of modern innovations, and what could the implications be for inequality? We address this question in three steps: we uncover new stylized facts on the relationship between innovators' backgrounds and the direction of innovation; we present quasi-experimental evidence on the role of social factors, independent of financial incentives; and we assess the quantitative importance of the social push channel for economic growth and cost-of-living inequality across consumer groups in general equilibrium. Overall, we find that an innovator's background affects the direction of innovation, which, combined with the under-representation of certain groups in the innovation system, leads 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. The key challenge is to push the data frontier to paint a comprehensive portrait of the relationship between consumer and innovator characteristics. We do so by building several datasets linking consumer characteristics to innovators' gender, parental income, and age. 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 venture capital databases, registries of firms, and administrative tax records.

Using this comprehensive dataset, we document that innovator-consumer homophily is a

[2] Josephine Cochrane was posthumously inducted into the National Inventors Hall of Fame in 2006 for patent 355,139 issued on December 28, 1886, for her invention of the first hand-powered dishwasher. The United States Patent Office writes: "she struggled against society's limits on women, working tirelessly to build a successful prototype, sell her invention, and ultimately turn a tedious task into an iconic American appliance."

very common feature of modern innovation systems, holding in both the United States and Finland for all types of innovations we study (new phone applications, new consumer goods, patents, new firms). For example, we find that innovators from a high-income family are more likely to create products purchased by high-income consumers: they 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 homophily patterns exist in terms of gender and age. These patterns hold across detailed industries as well as across firms within the same industry.

To assess whether these directional differences are likely to persist as under-represented groups gain better access to the innovation system, we examine the relationship between the secular increase in the share of female inventors and homophily. We find that gendered inventor-consumer homophily has remained stable over the past 25 years, even though the share of women among patent inventors has increased by 30 percent during this period. Overall, we find that innovator-consumer homophily is a striking empirical regularity.

In the second part of the article, we provide direct evidence that innovators' social experiences have a causal impact on the direction of innovation, independent of financial incentives. Assessing the role of non-financial incentives as a determinant of innovation direction is crucial for understanding equilibrium effects on inequality and growth.[3] Using a quasi-experimental study-peer design in Finnish education programs, we examine whether variation in the gender and socioeconomic composition of an individual's peer group has an impact on the direction of entrepreneurship.[4] We find that exposure to lower-income peers increases the probability of starting a business in a necessities industry (conditional on becoming an entrepreneur) but does not increase entrepreneurial income or affect the probability of becoming an entrepreneur. Likewise, exposure to female peers leads to an increase in entrepreneurial activities targeting female consumers. These results provide direct evidence that social factors affect the direction of innovation, which stands in contrast with mechanisms based on financial incentives, such as market size.

In the final part of the article, we investigate the relevance of the social push channel for

[3]Differences in exposure to innovation and the role of non-financial incentives combine to determine the allocation of innovators across sectors. We formalize these ideas in the third part of the article.

[4]This research design is similar to Hoxby (2000).

cost-of-living inequality and economic growth. Our model extends Romer (1990) to include multiple sectors, heterogeneity in consumer tastes, social exposure to innovation, and barriers to entering the innovation system that vary across socio-demographic groups, including by gender. Importantly, this modeling framework allows for directional differences in the choice of target markets by innovators' socio-demographic backgrounds. Modeling choices and model parameters are disciplined by both the stylized facts about innovator-consumer homophily and the quasi-experimental estimates of social exposure effects on the direction of innovation, as well as by other findings in the literature. We analyze counterfactual scenarios reducing access barriers and shutting down the social push channel, thus changing the equilibrium size and composition of the pool of innovators.

We find that, in a world with social push, access barriers for potential female inventors create a 18.20% difference in cost-of-living between men and women, which is almost as large as the gender pay gap. Lowering access barriers induces some highly productive female inventors to start innovating, and reduces the cost-of-living gap due to the disproportionate impact on female consumers through homophily. We also find large impacts on overall growth rates: eliminating access barriers leads to an increase in long-run growth rates from 2% to 3.4% per year. We obtain comparably large effects, for both inequality and growth, when analyzing counterfactuals by income groups. In particular, we find that innovator-consumer homophily by socioeconomic status has a larger impact on cost-of-living inequality across the income distribution than market size effects.

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 homophily evidence provides support to the idea that the social push channel operates at a large scale, both across and within industries, in countries as different as the United States and Finland, and for many different types of innovations. The peer-effects design provides direct causal evidence that social experience has an impact on the direction of innovation, independent of financial incentives. Finally, the structural model shows that the economic implications of the social push channel can be significant even in general equilibrium. The descriptive, quasi-experimental, and structural approaches thus paint a consistent picture, highlighting the social push channel as an important but understudied driver of the direction of innovation.

Prior work. This article primarily contributes to the literature on unequal access to innovation. Recent research documents that certain groups are underrepresented in the innovation system, particularly in terms of gender, race, parental background, and geography (Bell et al., 2019b; Toole et al., 2019; Cook et al., 2022). 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; Hvide and Oyer, 2020; Guzman et al., 2020), preferences (Thébaud, 2010; Bönte and Piegeler, 2013; Caliendo et al., 2015), intergenerational transmission (Dunn and Holtz-Eakin, 2000; Mishkin, 2021; Hvide and Oyer, 2020), social exposure (Markussen and Røed, 2017; Calder-Wang and Gompers, 2021), frictions in social networking (Howell and Nanda, 2023), and discrimination (Howell et al., forthcoming). Consistent with our findings, Koning et al. (2021) document that biomedical patents with female first authors are more likely to mention female medical conditions. 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 by several socio-demographic factors (including but not limited to gender), as well as between entrepreneurs and their employees; (ii) we provide quasi-experimental evidence on the role of social exposure effects; and (iii) we quantify the macroeconomic importance of this channel in general equilibrium using a structural model. Furthermore, we focus on products that make it to market (i.e., are economically viable), which complements contemporaneous work documenting gender homophily at earlier stages in the research process (Koning et al., 2021; Koffi and Marx, 2023; Truffa and Wong, 2022).

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 for a distinct mechanism, showing the importance of innovators' social background and peers in determining the overall direction of innovative activity in an economy with unequal access to innovation.

Finally, this article contributes to the literature on endogenous growth and inequality. Our model builds on the product variety literature, starting with Romer (1990) and 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 on welfare, but without entrepreneurs and innovators; our framework extends this analysis by developing an endogenous growth model. Our results show that misallocation in the innovation sector can have a sizable impact on long-run growth *rates* and cost-of-living inequality, while the reallocation channel studied by Hsieh et al. (2019) affects the level of GDP. Our analysis also contributes to strands of research that examine the influence of sorting and social interactions on career choice (e.g., Jones and Kofoed, 2020; Dahl et al., 2021; Michelman et al., 2022) and on inequality (Kremer, 1997; Fernandez and Rogerson, 2001). We highlight 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 documents the homophily between innovators and their consumers. Section IV provides quasi-experimental evidence showing that peer background affects entrepreneurs' choice of target markets. Finally, Section V examines the implications of these findings for growth and inequality with a quantitative growth model.

II Data and Summary Statistics

This section presents our data sources, samples, key variables, and summary statistics.

II.A Data Sources and Variable Descriptions

To document the extent and generality of the "social push" channel, we compile complementary datasets in the United States and Finland, allowing us to study different types of innovations in different settings. We use two "micro" datasets for the United States. We start with sector-specific analysis employing respondent-level information on usage of phone applications and purchases of consumer packaged goods. We then provide economy-wide evidence for the full consumption basket based on measures of consumer characteristics at a detailed industry level. We draw socio-demographic information on innovators – including gender, parent income, and age – from startup databases (Crunchbase), patent records (USPTO and PATSTAT databases) and administrative tax data (for inventors and entrepreneurs in Finland).

Phone applications. The first micro dataset is the Nielsen's Electronic Mobile Measurement (EMM) panel, which tracks the mobile phone application usage of a representative sample of ten thousand U.S. consumers in every month. To our knowledge, this database has never been used for research purposes. The data contains detailed information on the gender, income, and state of residence for each panelist from April 2017 to June 2019. Apps are classified into 58 categories (e.g., gaming, social networking, photography, etc.). A key advantage of this dataset is that it measures consumption at the individual level.

The data provide the company name 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. It contains data on the name, location, and founders of each startup. For each founder, it also records gender and LinkedIn URL, which we use to collect information on their age.[5] We clean company names in both the phone app and Crunchbase datasets prior to merging. Because Crunchbase attempts to track both legacy companies and startups, we define startups as firms founded after 2007, partly to reflect the start of the smartphone era and partly to reflect the start of Crunchbase's data collection.

[5] Crunchbase uses first names to 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 in the consumer packaged goods sample. In our linked dataset, the number of available phone app categories falls to 54, compared to 58 in the original phone application data.

Consumer packaged goods. To track consumption patterns for consumer packaged goods in the United States, we rely on Nielsen's Homescan Consumer Panel.[6] The data are based on a panel of 40k-60k consumers and provide information on household-level expenditure by products identified by barcodes from 2004 to 2016. Products are classified into 9 departments (e.g., non-food groceries), 118 groups (e.g., alcoholic beverages), and 1305 modules (e.g., light beer). These categories account for about 15% of aggregate consumption expenditures, and about 40% of expenditures on goods. For each household, Nielsen also records income, family structure, and household type (single female, female-led, married, male-led, single male).

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

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 unmatched companies in GS1 and Crunchbase that share the same city and first word of the company name. Finally, we also match all GS1 companies to patent data by company name. This allows us to measure the gender and age composition of the inventors who patent at a given company.

Industry-level data in the United States. To extend our analysis to cover the whole consumption basket, we leverage several data sets in the United States. To characterize consumers, we employ the Consumer Expenditure Survey (CEX), which provides information on consumption patterns by different consumer groups at a detailed industry level. We use these data to construct industry-level measures for the gender, socio-economic, and age composition of consumers. To characterize patent inventors, we use individual-level data on gender from PatentsView and age from Jones (2009). We link these datasets by six-digit NAICS industry. Specifically, we use the CEX category to NAICS industry crosswalk and

[6] These data have been widely used in economic research (for an overview, see, e.g., Dubois et al., 2022).

the data processing steps in Borusyak and Jaravel (2018). For inventors, we link the NAICS code by primary patent class, using the concordance created by Lybbert and Zolas (2014). Furthermore, we obtain information on entrepreneurship and parental income from the Panel Survey of Income Dynamics (PSID), which we link to the CEX at the level of the broader industries available in PSID.

Industry-level data in Finland. Finally, we use Finnish administrative data covering the full working-age population. The dataset is based on administrative registers compiled by Statistics Finland. It provides individual-level information on income, entrepreneurship, and industry. The data set also includes information on family links from the Finnish Population Information System, which allows us to measure the parental income of individuals. We study in turn entrepreneurs and patent inventors.

Information on entrepreneurship status is based on pension contribution and tax records. We use the status for the last week of the year, which allows for temporal consistency across variables.[7] The other key variable is the unique company identifier, which is 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.

To identify 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 study consumption patterns and the direction of innovation, we use measures of consumer characteristics from the U.S. CEX because it is more granular than available Finnish consumption surveys.[8] We link the CEX to the Finnish population panel by the

[7] 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.

[8] The Finnish consumption survey describes expenditure patterns at the level of 89 categories 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; to link the consumption data, we use a crosswalk between NACE and NAICS.

II.B Summary Statistics

Panel A of Table 1 provides summary statistics on the characteristics of innovators and consumers for the two sector-specific micro datasets available in the United States, phone applications and consumer packaged goods.

We find that female founders are underrepresented in both sectors. Companies with at least one female founder represent only 14 percent of venture-backed startups in the phone app industry and 24 percent of startups in consumer packaged goods. The rates of female venture capital partner involvement are even lower, at 6 percent for phone applications and 4 percent for consumer packaged goods. There is substantial variation in consumer gender composition across startups. For phone applications, the average time share of female users is 54% and the standard deviation is 38pp across applications. For consumer packaged goods, the average share of purchases coming from families with a female head-of-household is 26%, with a standard deviation of 25pp. Heterogeneity by age is lower.

Panel B of Table 1 reports the industry-level patterns from the inventor-CEX data covering the full consumption basket, for both the United States and Finland. The first part of the panel reports statistics on innovators, considering both entrepreneurs (Col. (1) and (3)) and patent inventors (Col. (2) and (4)). The table shows that there is substantial variation in terms of innovators' gender, age, and parent income. The second part of the table documents that there is also significant heterogeneity in consumer characteristics, including age, gender, and income elasticities – our summary measure of the variation in consumer incomes.

---

purpose of consumption (COICOP), which is not sufficiently detailed to characterize innovator-consumer homophily accurately. Linking the Finnish and U.S. consumption surveys at this level of product aggregation, we find that the income elasticities and consumption shares are very similar in both datasets. Specifically, we rank the broad COICOP categories by average consumer income, and we find that the rank-rank correlation coefficient between the Finnish and U.S. datasets is 0.73 (s.e. 0.12). Furthermore, using OECD data on household consumption at the 3-digit level of the COICOP classification, we find that U.S. consumption shares predict Finnish shares well: in a regression of Finnish on U.S. consumption shares, the slope is 0.83 (s.e. 0.27).

Table 1: Summary Statistics

Panel A: Phone Applications and Consumer Packaged Goods in the United States

		Phone Applications	Consumer Packaged Goods
Innovator statistics	# VC-backed startups	1,679	158
	# Manufacturers with patents	N/A	1,191
	Female founders ≥ 1	0.14 (0.35)	0.24 (0.43)
	Female VC partner ≥ 1	0.06 (0.24)	0.04 (0.19)
	Founder age at founding	N/A	35.56 (9.00)
	Female patent inventor ratio	N/A	0.11 (0.13)
Consumer statistics	# Startup Products	3,380	4,058
	# Product categories	54	294
	Female consumer share	0.54 (0.38)	0.26 (0.25)
	Consumer average age	43 (14.82)	47.18 (7.62)
	# Panelists	50,725	168,775
	Data sources	Nielsen EMM, Crunchbase	Nielsen Homescan, Crunchbase, USPTO
	Timeframe	2017–2019	2007-2016

Panel B: Industry-level Data in the United States and Finland

		United States		Finland	
		Entrepreneurs (1)	Patent Inventors (2)	Entrepreneurs (3)	Patent Inventors (4)
Innovator Statistics	# Innovators	325	2,219,193	344,698	9,643
	Fraction female	0.27 (0.30)	0.12 (0.32)	0.35 (0.48)	0.078 (0.26)
	Fraction parent income in top 20%	0.345 (0.21)	N/A	0.11 (0.31)	0.26 (0.44)
	Age	43.03 (6.2)	47.0 (13.9)	46.5 (10.4)	41.7 (9.4)
	# Industries	19	325	476	342
Consumer Stats.	Female consumer share	0.60 (0.071)	0.57 (0.09)	0.63 (0.10)	0.56 (0.09)
	Industry income elasticity	1.21 (0.30)	1.07 (0.36)	1.12 (0.53)	1.27 (0.35)
	Consumer average age	48.9 (5.1)	47.4 (2.2)	51.19 (6.19)	49.11 (4.13)
	# Panelists	20,700		20,700	
	Data source	CEX, PSID	CEX, USPTO	CEX, Admin. data	CEX, Admin. data, PATSTAT
	Timeframe	2017	1976-2015	2007-2015	

Notes: This table provides summary statistics on innovators and consumers for the micro-datasets in the U.S. (Panel A) and the industry-level analysis in both the U.S. and Finland (Panel B). In Panel B, the first row shows the number of innovators in the largest available samples. For some variables the sample is restricted to a smaller number of individuals due to data limitations. The number of innovators with available information on parent income is 275 in Column 1; 99,189 in Column 3; and 3,812 in Column 4. Age in Column 2 uses data from Jones (2009), available for 48,156 inventors. Standard deviations are reported in parentheses and computed across the most detailed available product category and across industries.

In sum, we find large variation in the characteristics of both innovators and consumers, at all levels of analysis. Next, we estimate the extent to which these characteristics covary.

III Estimating Innovator-Consumer Homophily

This section estimates the extent of innovator-consumer homophily, which we find to hold across all measures of innovations and sectors we consider, in both the United States and Finland.

We report the homophily estimates separately for phone applications (Subsection III.A), consumer packaged goods (Subsection III.B), and at the industry level (Subsection III.C), using regression specifications of the form:

$$\text{ConsumerType}_{ij} = \alpha + \beta \cdot \text{InnovatorType}_{ij} + \mu_k + \varepsilon_{ij}, \tag{1}$$

where i indexes a product sold by company j, $\text{InnovatorType}_{ij}$ is a characteristic of the inventor of the product, and ConsumerType_{ij} is a measure of consumer characteristics in the firm's market for product i. We study alternative socio-demographic characteristics to separately estimate homophily by gender, parent income, and age. For example, for gender homophily, "InnovatorType" is a binary indicator for female inventor and "ConsumerType" is the share of sales of product i to women in the innovator's company j. μ_k is a fixed effect for product category k, which we use in certain specifications to assess whether homophily arises primarily within or across product categories. For industry-level regressions, which do not have the product level, j indexes industry and i is suppressed.

As discussed in Section V, the regression coefficient β has a direct consumer welfare interpretation under the assumption of Constant Elasticity of Substitution (CES) consumer demand, allowing us to assess the magnitude of the distributional effects across consumers that arise from unequal access to innovation careers through innovator-consumer homophily. Our homophily estimates will thus discipline the structural analysis of cost-of-living inequality in the last part of the paper.

III.A Innovator-Consumer Homophily for Phone Applications

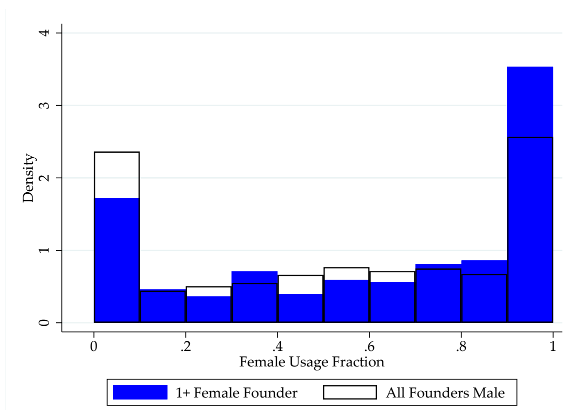
We now estimate innovator-consumer homophily for phone applications, which are one of the iconic forms of innovation of the past decade. Table 2 presents the estimates.

Table 2: Innovator-Consumer Homophily for Phone Applications

	Female User Share		
	(1)	(2)	(3)
Female Founder Fraction	0.092***	0.082**	0.092*
	(0.034)	(0.032)	(0.036)
Female VC Fraction			0.160*
			(0.063)
Fixed Effects	None	Category	None
Sample Size		$N = 3,380$	

Notes: The sample used in this table includes all phone applications for VC-backed startups. The outcome variable is the fraction of time usage by female users, with a sample mean of 0.542. The level of observation is a firm-application. Standard errors are clustered at the startup level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Share of Female Usage of Phone Applications by Founder Gender Composition



Notes: The sample used in this figure includes all phone applications for VC-backed startups. The histograms depict the distribution of time use by gender for phone apps from a VC-backed startup with either at least one female founder (blue histogram) or all founders being male (white histogram). For example, a value above 0.9 for "Female usage fraction" on the x-axis covers apps for which more than 90% of time use is accounted for by female users.

We focus on gender in a regression of the fraction of time usage of an application accounted for by female users on measures for the gender composition of the startup founders as well as of the venture capitalists who fund the startup. While female users account for

54.2% of total time spent on phone applications created by VC-backed startups, the share is 9.2pp higher (or 17 percent of the baseline rate) when moving from an all-male founding team to an all-female founding team (Column 1). Figure 1 illustrates these homophily patterns non-parametrically by showing the distribution of female user shares separately for male and female entrepreneurs. The figure shows that homophily primarily stems from gender-specialized applications that have more than 90% or less than 10% female users.

Column (2) of Table 2 shows that the homophily estimate remains similar in magnitude when the specification includes fixed effects for the 54 application categories (i.e., within productivity apps, social media apps, etc.). The stability of the coefficient across Columns (1) and (2) indicates that gender homophily occurs to a similar degree within and across these detailed application categories. In Column (3), we augment the specification to study gender homophily between consumers and both founders and venture capitalists. The level of founder homophily remains stable, while homophily by the gender composition of venture capitalists is even stronger. For applications created by companies funded only by female venture capitalists, female users account for 70.2% of total time usage, i.e., a 16pp higher share (or 29.5 percent of the baseline rate) relative to companies funded only by male venture capitalists. With both an all-female founding team and a female venture capitalist, the share of female users is 25.3pp higher (or 46.6 percent of the baseline rate).

In Appendix Table C1, we show that the patterns remain similar when we give more weight to the most impactful phone applications: with weighted regressions, using the logarithm of time use as weights for each application, the homophily coefficients remain similar in magnitude, which shows that homophily is not driven by marginal innovations.[9]

[9]Additional results are reported in the Online Appendix. While we focus on expenditure shares in the main text, as this specification can be linked to welfare (see Section V), in Appendix Table C2 we analyze how the *level* of usage from female users depends on innovator gender. We find that the level of female usage is positively correlated with the female founder indicator in our sample, but the precision of the estimation for this specification is low. Furthermore, in Appendix Table C3, we document homophily by place of residence, studying the fraction of time usage of an application by users located in the same U.S. state as the founder of the application. We estimate Equation (1) in a sample where each observation is the time usage for application i in state j, i.e., "ConsumerType$_{ij}$" is now the share of application i's usage by consumers in state j, and "InnovatorType$_{ij}$" is an indicator of whether the startup that created the application is based in state j. We find a large "home bias": the time share of users in the same state as the founder is 8.6pp larger than for users from other states, a fourfold increase relative to the average state share.

III.B Innovator-Consumer Homophily for Consumer Packaged Goods

We now turn to product innovations within consumer packaged goods sector, a segment which has been widely examined in the literature on creative destruction at the micro level (e.g., Broda and Weinstein, 2010). Our specification includes fixed effects for detailed product categories (product modules), so that we isolate homophily arising at a granular level, before turning to differences arising across industries in the next section.

We first analyze gender homophily, measuring consumer gender composition by the share of sales to households with a female head. Column (1) of Table 3 reports that startups founded by female entrepreneurs are more likely to sell to female consumers, with a 4.7pp higher share of sales to households with a female head. This constitutes an increase of 18.8% relative to the baseline rate of 25%. Founder-consumer gender homophily for consumer packaged goods is thus quantitatively similar to our estimates for phone applications.

Table 3: Innovator-Consumer Homophily for Consumer Packaged Goods

	Share of Sales to Women		Average Consumer Age, Sales-weighted
	(1)	(2)	(3)
Female Founder Fraction	0.047**		
	(0.021)		
Female Patent Inventor Fraction		0.027*	
		(0.015)	
Founder Age			0.135**
			(0.052)
Product Module F.E.	Yes	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 1,094,229$	Startups, $N = 4,058$

Notes: In columns (1) and (2), the outcome variable is the fraction of sales to households with a female head. The sample means are 0.256 in column (1) and 0.265 in column (2). In column (3), the outcome variable is the average age of consumers, using sales weights, with a sample mean of 47.2. The level of observation in this table is a firm-product. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Column (2) conducts a similar analysis in the sample of manufacturers with at least one patent. Manufacturers with a higher share of female patent inventors sell more to female consumers: a change from all-male to all-female inventor composition in a firm is associated with a 2.7pp increase in sales to households with a female head (or 10.8% of the baseline rate). This homophily estimate is smaller than for founders in column (1), consistent with the hypothesis that the founder may set the "entrepreneurial vision" of the startup, whereas patent inventors implement firm-level goals.

In addition, the consumer packaged goods dataset allows us to study homophily by age, as reported in column (3) of Table 3. For each startup in our sample, we compute the average, sales-weighted age of consumers. We find that entrepreneurs that are one year older (at the time of founding) sell to consumers that are on average 0.135 years older. Thus, a one standard deviation increase in founder's age (9 years) is associated with an increase in average consumer age of 1.22 years, about 2.7% of the average consumer age of 47.2 years.

Several robustness checks are reported in the Online Appendix. First, we obtain similar results when we repeat the analysis with an alternative measure of consumer gender, weighting sales by the fraction of female members in the household (Table C4). Second, the homophily coefficients also remain similar in weighted regressions, using the logarithm of product sales as weights, which shows that the results are not driven by marginal product innovations (Table C5). Third, we show that homophily remains similar when controlling for prices (Table C6), which rejects classes of mechanisms where homophily occurs through prices (e.g., female entrepreneurs could introduce higher-priced goods which attract more female consumers because they are less price sensitive). Finally, we present an alternative analysis of age-based homophily, using age group bins, in Table C7.

III.C Industry-Level Estimates of Innovator-Consumer Homophily

We now turn to the industry-level homophily estimates, which are reported in Table 4 for both the United States and Finland. In these homophily regressions, innovator characteristics vary at the individual level, while the outcomes (e.g., the share of sales to female consumers) only vary at the industry level. The industry-level homophily estimates thus only capture the "between-industry" component of homophily and must be added up to the "within-industry" homophily patterns documented in the previous sections.[10]

[10] With Y_i the outcome for innovator i, we can write $Y_i \equiv Y_{j(i)} + \Delta Y_i$, where $Y_{j(i)}$ is the average outcome in i's industry, indexed by j, and ΔY_i is the deviation between i's outcome and the industry average.

Table 4: Innovator-Consumer Homophily across Industries

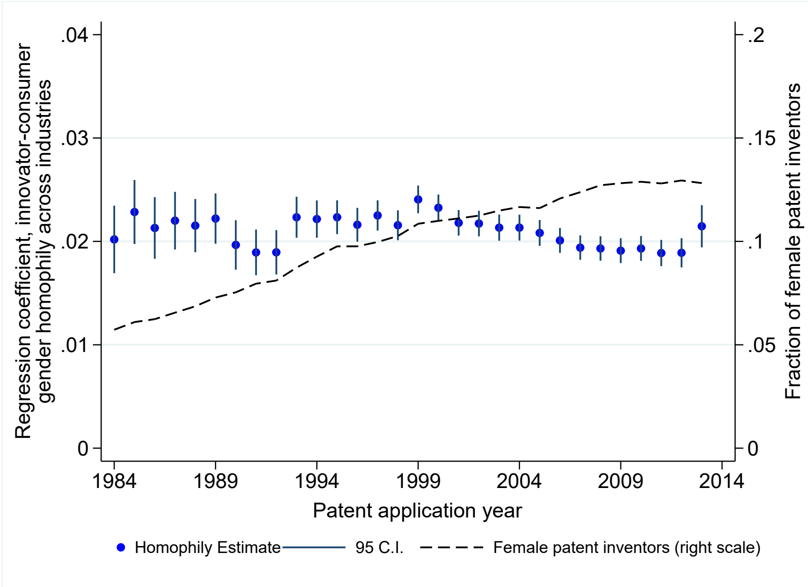
	Share of Industry Sales to Women			Industry Income Elasticity			Average Consumer Age, Sales-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Patent Inventor	0.0222***	0.0484**							
	(0.000194)	(0.0037)							
Female Entrepreneur			0.0302***						
			(0.0004)						
Patent Inventor's Log Parent Income					0.0304**				
					(0.0125)				
Entrepreneur's Log Parent Income				0.0394***		0.1416**			
				(0.0129)		(0.0034)			
Patent Inventor Age							0.00502***	0.0535***	
							(0.000739)	(0.0050)	
Entrepreneur Age									0.0122***
									(0.0010)
Country	U.S.	Finland	Finland	U.S.	Finland	Finland	U.S.	Finland	Finland
Mean	0.573	0.569	0.6367	1.2259	1.2766	1.1205	47.39	49.11	51.19
N industries	325	342	476	17	253	441	323	342	476
N individuals	2,219,193	9,643	344,698	275	3,812	99,189	48,156	9,643	344,698

Notes: All regressions are run at the level of an individual innovator, with outcomes measured at the industry level. Standard errors are clustered at the individual level. $^*p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

We first analyze gender homophily patterns, finding again that female innovators are more likely to cater to female consumers. Column (1) of Table 4 shows that, in the United States, female patent inventors work in industries where the share of sales to households with a female head is 2.2pp higher, or a 3.9% increase relative to the baseline rate. This "between-industry" homophily adds up to the within-industry homophily, which we summarize as a 17.9% increase relative to the baseline rate by taking the average of our results for phone applications and consumer packaged goods. Thus, our estimate of the overall gender homophily coefficient for the United States is 21.8%.

By linearity of OLS, the overall homophily regression coefficient, obtained by regressing Y_i on innovator characteristic X_i, can be equivalently obtained by adding the industry-level regression coefficient (which we analyze in Table 4) to the within-industry coefficient (which we analyze in Tables 2 and 3 for specific industries).

Figure 2: Innovator-Consumer Gender Homophily across Industries over Time



Notes: The figure reports the innovator-gender homophily coefficient across industries, estimated in each year for patent applications submitted between 1984 and 2014 (left scale). The figure also reports the fraction of female patent inventors over time (right scale). The vertical lines represent 95 percent confidence intervals, with heteroskedasticity-robust standard errors.

Figure 2 shows that gender homophily across industries has been very stable over time, despite the large increase in the share of female patent inventors. The figure reports year-specific homophily coefficients, based on the application date of patents. The homophily coefficient hovers close to 2pp from 1984 to 2014. During this period, the fraction of female patent inventors more than doubled, rising from 5% in 1984 to 13% in 2014. This finding supports the idea that gender homophily would persist if access barriers to the innovation system were significantly reduced for women. Our results stand in contrast to those found in Koning et al. (2021), who find that homophily has weakened over time in life sciences patenting. They note that their results may be driven by the fact that scientific funding has increasingly promoted research into women's health, a feature not present in the broader economy.

Columns (2) and (3) of Table 4 document industry-level gender homophily for Finland. The patterns are similar to the United States, with slightly larger magnitudes. Finnish female patent inventors work in industries where the share of sales to women is 8.5% larger than average (col. (2)). The corresponding increase is 4.7% for Finnish female entrepreneurs (col. (3)).

Next, we document homophily between innovators' parent income and consumers' income. We find that innovators from more affluent backgrounds cater to richer consumers. For our baseline specification, we use an industry's income elasticity as our summary measure of consumer income.[11] Column (4) reports the patterns for entrepreneurs in the United States, showing that an increase in parent income of one log point is associated with an increase of 0.039 in the industry's income elasticity. Column (5) documents similar patterns for patent inventors in Finland, with an increase of 0.0304 in the industry's income elasticity when parent income increases by one log point. Column (6) shows that the relationship is stronger for entrepreneurs in Finland, with an increase of 0.14 in the industry's income elasticity. We find that this large coefficient is driven by entrepreneurs in agriculture; when excluding agriculture, the coefficient falls to 0.0119 (Appendix Table C8).

While the income elasticity is a convenient summary measure for depicting the patterns across the income distribution, we also provide result for the share of sales by household income groups (Appendix Table C9). We examine whether an entrepreneur's family background systematically varies with the ratio of sales to households earning over \$100k ("high-income") to sales to those earning less than \$30k ("low-income"). We find that, in the United States, the fraction of sales to high-income households increases by 4.15% relative to the baseline when an entrepreneur comes from the top 20% of the family income distribution, compared with the bottom 20%. Importantly, this number only reflects the "between industry" component of homophily by social background, ignoring any within-industry patterns. This magnitude of "between-industry" homophily by income groups is very similar to gender homophily, as we found a 3.9% increase relative to the baseline rate for gender.[12] Using sales fractions as outcomes in Finland yields similar magnitudes: the fraction of sales to high-income households increases by 3.85% relative to the baseline when an entrepreneur comes from a family in the top income quintile instead of the bottom quintile. Using a threshold of \$60k for both low- and high-income households leads to similar conclusions.

[11] We use the sectoral income elasticity estimates measured in Borusyak and Jaravel (2018).

[12] If we assume that the relative magnitudes of gender homophily and income homophily are the same within and between industries, then by rescaling our overall gender homophily estimate (= 21.8) by the ratio of between-industry income and gender homophily estimates (= 4.15/3.9), we obtain that the overall income homophily coefficient for the United States is 23.8%.

Finally, columns (7) to (9) of Table 4 show that there is also homophily by age across industries. Older patent inventors are more likely to work in industries selling to older consumers in both the United States (col. (7)) and Finland (col. (8)). In column (9), we report that older entrepreneurs are also more likely to cater to older consumers.

Several robustness results are reported in the Online Appendix. First, we show that the estimates remain similar with weighted regressions, giving higher weights to more impactful innovations (using patent counts or proxies for firms' private returns), i.e., the results are not driven by marginal innovators (Table C10). Second, in Figure C1, we depict graphically the main regressions from Table 4, showing that the linear specifications provide a good fit to the data and that the estimates are not driven by outliers. Third, for completeness, in Appendix Table C11 we run industry-level regressions after averaging worker characteristics by industry; as expected, the coefficients are much larger than in Table 4, where the independent variables vary within industries while the dependent variables do not.

III.D Extensions

Appendix A reports two extensions, documenting homophily between the socio-demographic characteristics of entrepreneurs and their employees, and relating innovators' backgrounds to the environmental and social impacts of their innovations.

IV Social Exposure and Homophily: Quasi-experimental Estimates

Motivated by the pervasive pattern of homophily between innovators and their consumers documented in Section III, we now examine a potential explanatory channel. Homophily could in principle stem from demand- or supply-side factors. We conjecture that social exposure may cause a shift in the direction of innovation through the supply side. Indeed, social peers may shape innovators' motivations, aspirations, and ideas. Such a social mechanism can operate independently of financial incentives, in contrast with the channels that have been the focus of the directed technological change literature (e.g., Acemoglu, 2002).

To test the relevance of this channel, we employ a quasi-experimental research design isolating the impact of a shift in an innovator's social environment in the educational program she attends during early adulthood. We present the research design in Section IV.A, the falsification tests in Section IV.B, the main results in Section IV.C, and robustness checks in Section IV.D.

IV.A Research Design

We estimate social exposure effects among over fifty thousand vocational and university students in Finland. Young adults at this formative life stage are a useful population for studying the effects of social exposure on the direction of innovation, because they are making decisions that determine their future careers.

Following Hoxby (2000) and the subsequent literature, we exploit idiosyncratic variation in the composition of peers across student cohorts within the same program and school to assess whether peers have a causal effect on the direction of innovation. The data are drawn from the student register maintained by Statistics Finland. Our sample covers the 1999–2013 period and includes individual-level information on the educational institution (hereafter, school) and program for upper secondary, vocational, and university students.[13] We identify study peers as individuals who start to study in the same study program and school in the same year. For individuals who are observed in several programs, we define study peers as co-students in the last program the student enters. We use all co-students in the last program in the full student population data when constructing measures of peer composition.

For individuals who become entrepreneurs, we examine in turn the effect of: (i) peer gender and (ii) peer parent income composition on the direction of innovation. In addition, we examine entrepreneurs' incomes to further assess the role of financial incentives. We can thus investigate both the extent to which the innovator-consumer homophily documented in Section III arises from differences in social exposure and the importance of financial incentives. While we focus on the direction of innovation among students who become entrepreneurs, we also present auxiliary results about the effects of peers on the probability of becoming an entrepreneur.

[13] Our sample excludes programs targeted at older individuals, the goal of which is to update, complement, and advance an existing degree. A large fraction of these programs are based on out-of-class studies.

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 α_{jk} and school-by-start-year fixed effects λ_{ks}:

$$Y_i = \beta \overline{X}_{(i)jks} + \gamma_1 X_i + \gamma_2 W_i + \alpha_{jk} + \lambda_{ks} + \varepsilon_{ijks}. \tag{2}$$

We examine the impact of peer composition on the outcome Y_i, characterizing consumers in the market the entrepreneur caters to. Our key measures of consumer characteristics are the share of sales to women and income elasticity of the industry in which the entrepreneur operates.[14] The main regressor of interest is the peer mean, $\overline{X}_{(i)jks}$, which is the average characteristic X_i of the peers of individual i. For example, if the characteristic X_i is a dummy for female or parent income, $\overline{X}_{(i)jks}$ is the leave-own-out fraction of female peers or the leave-own-out mean parent income of her co-students. In addition, we include the control X_i for i's own characteristics (own gender or parent income).[15] To account for sampling variation in background characteristics and reduce noise, we also include a vector of control variables, denoted by W_i, which comprises a rich set of predetermined characteristics for the student and her parents, and which are all measured one year before the first study year. For inference, we report standard errors clustered by school and program start year.

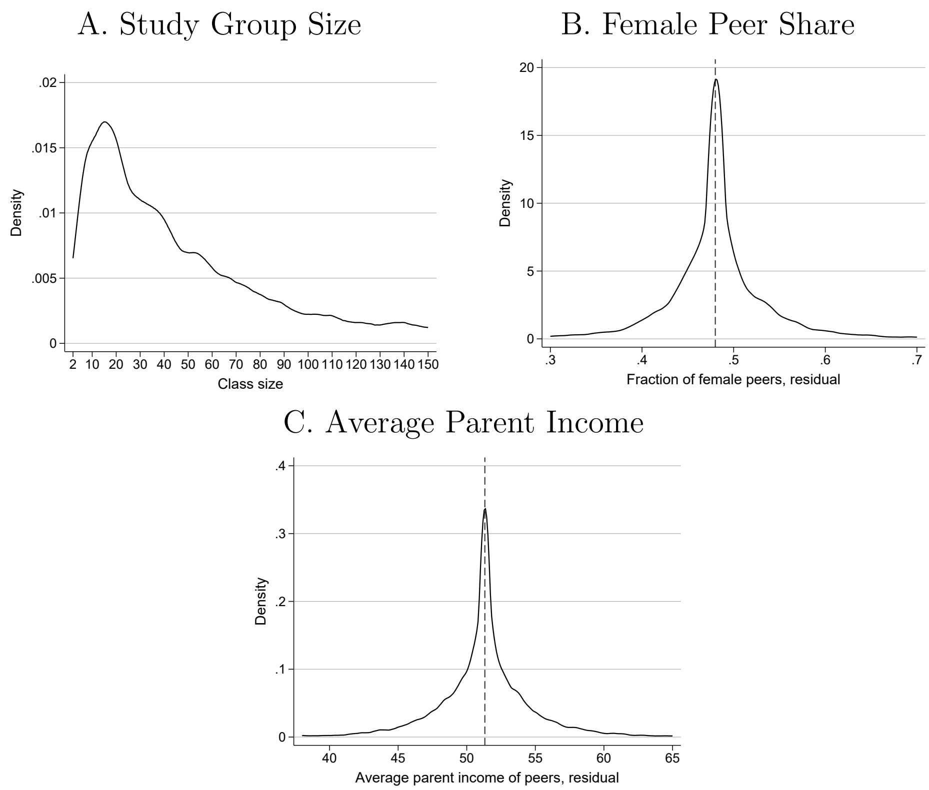
The parameter of interest in Equation (2) is β, the coefficient on the average characteristic of study peers. Conditioning on school-by-program fixed effects means that the peer effect is identified from idiosyncratic variation in peer composition across student cohorts within the same program and school. We also control for school-by-start-year fixed effects to account for common shocks at the school level. 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, variation in peer gender or parent income composition across student cohorts in the same study program and school is uncorrelated with other determinants of the direction

[14] We focus on outcomes from age 28 onwards, and average them across years for entrepreneurs who are active over several years.

[15] We note that controlling for own characteristic is standard practice to eliminate the mechanical correlation between this variable 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).

of entrepreneurship. We believe that this assumption is likely to be valid in our context, as it appears unlikely that year-to-year variation in peer characteristics in a specific program and school is correlated with unobserved factors that drive the choice of industry. We provide empirical support for the credibility of the design by implementing falsification tests, demonstrating that peer attributes are uncorrelated with pre-determined characteristics of the student and her parents.

Figure 3: Study Group Size Distribution and Identifying Variation in Peer Attributes



Notes: Panel A depicts the variation in study group size in our baseline sample, using data on 21,009 study groups in which at least one individual becomes an entrepreneur. Study groups with more than 150 students are not displayed. Panels B and C illustrate the identifying variation in our design, showing the degree of variation in the female peer share and the average peer parent income across student cohorts within a school-by-program cell. The figures display the distributions of residuals from separate regressions of female peer share and average peer parent income on pre-determined characteristics listed in footnote 16, dummies for calendar year, and school-by-start-year and school-by-program fixed effects. We add to these residuals the sample means of the corresponding variables. Parent income is in thousand euros.

Our sample includes 556 schools and 21,009 study peer groups. Panel A in Figure 3 displays the size distribution of the study groups. The median size is 26 students and the mean is around 45 with a standard error of around 60. We also consider a sample with restricted group size, which is of specific interest because peer estimates can have poor reliability in settings with large peer groups (see, e.g., Angrist, 2014). When restricting study group size to 25 students, both the median and mean are 12 students and the standard error is 6.5. The estimates for this sample turn out to be similar compared to the estimates from the baseline sample without size restriction, as we later show in the robustness checks. In the main text, we report results for the full data, which have higher statistical precision due to larger sample size.

Panels B and C in Figure 3 illustrate the identifying variation that we exploit by showing the distributions of peer attributes, conditional on pre-determined characteristics and school-by-start-year and school-by-program fixed effects. All variation occurs over time across student cohorts within a school-by-program cell. The standard deviation is 0.07 for the residual female peer share and 4.28 thousand euros for the residual average peer parent income. The standard deviation of the residual female peer share corresponds roughly to replacing one male student with one female student in a class of 15 students ($1/15 \approx 0.07$).

IV.B Falsification tests

To examine the plausibility of the identification assumption underlying Equation (2), we run an analogous specification to assess whether peer characteristics are predictive of pre-determined characteristics of individual i:

$$\widetilde{Y}_i = \widetilde{\beta}\overline{X}_{(i)jks} + \widetilde{\gamma}_1 X_i + \widetilde{\alpha}_{jk} + \widetilde{\lambda}_{ks} + \widetilde{\varepsilon}_{ijks}. \tag{3}$$

The dependent variables in our primary falsification regressions are linear predictions of the measures of the direction of innovation (the share of sales to women and the industry income elasticity), based on predetermined characteristics of the student and her parents.[16] The advantage of this approach is that it assigns larger weights to the characteristics that best

[16]The predicted outcomes are constructed by first running separate regressions of the share of sales to women and industry income elasticity on predetermined characteristics (excluding X_i, which is on the right-hand side of eq. 3) and school-by-start-year, school-by-program, and year of outcome measurement fixed effects. We then use the coefficients from this regression to calculate the predicted values. The full set of predetermined characteristics includes: labor earnings, years of education, unemployment benefits, housing allowance, parent income, parents' years of education, number of employed parents, unemployment benefits of parents, housing allowance of parents, pension income of parents, and dummies for gender, age, marital status, foreign, and Finnish as primary language.

predict future outcomes. For completeness, we also provide results for each predetermined outcome separately.

Table 5: Falsification Tests, Study Peer Design

Dependent variable	Fraction female among study peers		Average parent income of study peers	
	Coeff.	s.e.	Coeff.	s.e.
A. Predicted share of sales to women	-0.00003	(0.00029)	-0.000003	(0.000006)
B. Predicted expenditure elasticity	0.00249	(0.00291)	-0.000052	(0.000043)
C. Predicted income	-0.343	(0.268)	-0.0034	(0.0047)

Notes: The baseline estimation sample consisting of 51,186 individuals who become entrepreneurs. Each cell presents a coefficient from a separate regression, corresponding to Equation (3) and using the variable indicated by the row label as the outcome. The outcomes are the best linear predictions at age 28-42 based on the predetermined characteristics listed in footnote 16. All specifications include program-by-school and school-by-start-year fixed effects. Standard errors are clustered at the school-by-start-year level and reported in parentheses. The sample includes 556 schools and covers 15 start years. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 reports the results for the sample of 51,186 individuals who become entrepreneurs in our data. We report the results separately for regressions where the independent variable is the predicted fraction of female peers and predicted average parent income of peers. If variation in peer composition across peer groups within a school-by-program cell is as good as random, these peer attributes should not be correlated with the direction of innovation predicted by variables that are realized before peer assignment. In Panels A and B, all coefficients for both measures of predicted direction are close to zero and statistically insignificant. This result provides support for our assumption that there is no confounding selection of students to peer groups by background characteristics that predict our key outcomes. Panel C shows that coefficients are also small and insignificant for predicted income, indicating that endogenous peer group assignment by earnings potential is not a concern in our setting. Taken together, these results lend credibility to the assumption that within-program-and-school variation in peer composition in our data is as good as random in terms of student characteristics that best predict the future direction of innovation and income.[17]

[17] For completeness, Appendix Table C12 reports results separately for each predetermined characteristic and both peer variables (30 separate regressions). After adjusting p-values for multiple hypotheses testing,

IV.C Results

Main estimates. We now report the estimates of the impacts of study peers on the direction of innovation, conditional on being an entrepreneur. Columns (1) to (3) in Row A of Table 6 report the estimated effect of exposure to female peers on the share of sales to women. We find that, for male students, there is a statistically significant increase in the propensity to sell to women (row A, col. (3)). For male entrepreneurs, a one standard deviation increase in the fraction of female study peers (0.34) leads to an increases of 0.64pp ($= 0.0187 \times 0.34$) in the female consumption share of the industry in which they operate a business at age 28 or above.[18] In contrast, an increase in the fraction of female peers has no impact on the direction of innovation for female entrepreneurs (row A, col. (2)). In the full sample including both female and male entrepreneurs, the effect is not statistically significant either (row A, col. (1)). The fact that the exposure effect is gender-specific is consistent with the social exposure channel: exposure to additional female peers should not matter as much among female students, who have already interacted with other female peers throughout their lives, whereas there can be an effect for male students who are more likely to be "under-exposed" to female peers.[19]

Columns (4) to (6) in Row B of Table 6 report the estimated impact of exposure to peers from different parts of the parent income distribution. The outcome is the industry income elasticity where the entrepreneur is active at age 28 or above. We consider, in turn, all entrepreneurs (col. (4)) and separately those whose parent income is above and below the median (col. (5) and (6)). According to the point estimate in Column (4), a

none of the coefficient are significant at the conventional risk levels. Nevertheless, in our primary peer effect estimations, we control for all predetermined characteristics to account for the sampling variation associated with them.

[18] We base our assessment of magnitudes on sample standard deviation rather than residual standard deviation reported for Panel B of Figure 3, because the unconditional distribution describes the full variation in the sample and is more informative about the potential implications of changing exposure in the relevant population (rather than within school-by-program cells).

[19] The fact that the regression coefficient is not significant for women may still seem surprising, in that female entrepreneurs with a lower fraction of female study peers mechanically have more male peers, which could in principle lead them to envision products targeting male consumers. We conjecture that female entrepreneurs already have substantial exposure to male tastes throughout the entrepreneurial process, where male collaborators are over-represented across the board (among venture capitalists, business angels, etc.). For this reason, it seems plausible that exposure to study peers of the opposite sex matters more for male than for female entrepreneurs, leading to the heterogeneity in coefficients on peer gender composition found in the first row of Table 6.

one standard deviation increase in peer parent income (€12,319) leads to an increase in the income elasticity of the industry in which the entrepreneur operates a business of 0.0098 ($= 0.00080 \times 12.319$). Columns (5) and (6) show that this effect is driven by peers from high-income background. The point estimate is not statistically significant when considering entrepreneur with parent income below median in Column (5). The point estimate for entrepreneurs with parent income above median, in Column (6), is twice as large as in the full sample. For these individuals, a one standard deviation increase in peer parent income leads to an increase in the industry income elasticity of 0.017 ($= 0.00141 \times 12.495$). We obtain similar results when we use the share of sales to high-income versus low-income consumer groups instead of the industry income elasticity as the outcome, as reported in Appendix Table C13.

In sum, our key results in terms of heterogeneity indicate that: (i) entrepreneurs from high-income backgrounds target more lower-income households when they are exposed to peers from low-income backgrounds; and (ii) exposure to high-income peers does not affect the type of targeted industry in terms of consumer income among entrepreneurs from low-income families.[20] These findings are consistent with the view that entrepreneurs from all backgrounds are already exposed to the needs and preferences of the "majority consumer group." For example, it appears plausible that the behaviors of high-income households are well understood by low-income households through the media. In contrast, individuals from high-income backgrounds may be less exposed to the low-income group and as a result change their target market towards lower-income consumers when they are exposed to peers from low-income backgrounds.[21]

[20]Note that the positive coefficient on average peer parent income among high-income group means that when they are exposed to a peer group with low average parent income, they will target lower-income consumers in the future. Similarly, exposure of male students to majority-female peer groups makes them target more female-intensive markets in the future.

[21]A complementary explanation is that entrepreneurs from low-income backgrounds likely already have substantial exposure to high-income tastes throughout the entrepreneurial process, where high-income groups are over-represented among venture capitalists, angel investors, etc.

Table 6: Impacts of Study Peers on the Direction of Innovation and Income

	Fraction female among study peers			Average parent income of study peers		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
A. Sales share to women	0.0013	-0.0080	0.0187**			
	(0.0056)	(0.0110)	(0.0082)			
B. Industry income elasticity				0.00080**	-0.00020	0.00141**
				(0.00040)	(0.00071)	(0.00055)
C. Income	-2.165*	-0.735	-2.873*	0.0016	-0.0171	0.0198
	(1.117)	(1.847)	(1.710)	(0.017)	(0.0277)	(0.0284)
Sample	All	Women	Men	All	Own parent income below median	Own parent income above median
Students	51,186	20,714	30,472	51,186	23,889	27,297
Study groups	21,009	11,212	13,884	21,009	13,485	14,468
Schools	556	516	518	556	539	526

Notes: The baseline estimation sample consisting of 51,186 individuals who become entrepreneurs. The table displays the estimates of the impact of study peers on the dependent variable indicated by the row label. We consider two sets of study peer characteristics: gender (columns (1)-(3)) and parent income (columns (4)-(6)). Each cell presents a coefficient from a separate regression, following Equation (2). Outcomes are means from age 28 onward. All specifications include program-by-school and school-by-start-year fixed effects and control for dummies for the year of outcome measurement and pre-determined characteristics listed in footnote 16. All control variables are measured one year before entering the study program. Standard errors clustered at the school-by-start-year level are in parenthesis. The sample includes 556 schools and covers 15 start years. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

The role of financial incentives. Having established that there is a causal effect of an entrepreneur's social environment on the direction of innovation, we now examine the role of financial incentives as an explanatory channel. For example, it could be the case that, by interacting with peers from the opposite gender or different socioeconomic backgrounds, entrepreneurs may come to realize there exist untapped market opportunities. Market discovery and the associated financial incentives could thus be an important mechanism. Alternatively, it could be that exposure to peers shifts an entrepreneur's intrinsic motivation or ideas to target specific consumers, independently of financial incentives.

To assess the role of financial incentives, in row C of Table 6, we estimate the impacts of peer attributes on long-term income. Columns (1) to (3) present suggestive evidence that exposure to female peers may lead to a small fall in income, with statistically significant estimates at the 10% level for male entrepreneurs. The point estimate in Column (3) of row C indicates that a one standard deviation increase in the fraction of female study peers (0.34) leads to a fall in annual income of about €975 (= $0.34 \times 2.87k$), a 3.5 percent reduction compared to the sample mean of €28,200. In Appendix Table C14, we show that the effect is insignificant for the binary indicators for income being above the 50th or 99th percentile, while it is negative and statistically significant for the 90th percentile. Thus, the negative impact on income appears to be driven by the top decile, but not the top percentile, of the income distribution.

For peer parent income, there is no evidence of any response of income. In columns (4) to (6) of row C in Table 6, exposure to peers from different parts of the parent income distribution leads to no significant impact on average income. Appendix Table C14 confirms this finding, with no statistically significant responses across the income distribution.

Overall, these findings indicate that the causal effects of social exposure on the direction of innovation are not driven by profit incentives. The results are consistent with the dominant role of entrepreneurs' intrinsic motivations and ideas, which are shaped by social factors.

Comparing peer effects to homophily estimates. To gauge the quantitative importance of the "social push" channel, we compare the estimated peer effects in Table 6 to the descriptive homophily estimates from Table 4. The comparisons are reported in Table 7, starting with gender in Column (1). The difference in the share of sales to women is 3.02pp between female and male entrepreneurs in Finland, as reported in Column (3) of Table 4. By comparison, the difference in average exposure to female peers between female and male entrepreneurs leads to a difference in the share of sales to women of 0.80,[22] i.e., 26.6% (=0.80/3.02) of the overall difference. Thus, differences in peer exposure can explain a sizable fraction of the observed gender homophily. We use these findings to guide some of our counterfactuals when we assess the implications of our results for cost-of-living inequality

[22]We multiply the female study peer coefficient for male students, equal to 1.87pp according to Row A of Column (3) of Table 6, by the difference in the average fraction of female peers (i.e., mean exposure) between male and female students, equal to $0.74 - 0.31 = 0.43$ in our sample.

and growth in Section V.

Table 7: Study Peer Effects vs. Homophily Estimates

	Share of sales to women	Industry income elasticity
	(1)	(2)
A. Difference b/w female and male entrepreneurs	3.04pp	
B. Effect of difference in average exposure to female peers b/w female and male entrepreneurs	0.80pp	
⇒ Ratio A/B	26.6%	
C. Difference b/w top and bottom quintiles of parent income		0.091
D. Effect of difference in average study peer parent income b/w top and bottom quintiles of own parent income		0.0091
⇒ Ratio C/D		10%

Notes: This table compares the magnitudes of study peer effects from Table 6 to our homophily estimates from Table 4. The calculation steps are provided in the text.

Column (2) of Table 7 focuses on industry income elasticities. The industry income elasticity is higher by 0.091 on average for entrepreneurs coming from a family in the top 20% of the income distribution, compared with those from the bottom 20%.[23] According to our peer effects estimates, the difference in peer parent income between these two groups leads to an increase in the industry income elasticity of 0.0089,[24] i.e., 10% (=0.0091/0.091) of the overall difference. This sizable effect is plausible: while study peers in vocational and university programs constitute a subset of all social interactions of an individual, they can be expected to be particularly important for the direction of innovation.[25]

[23] We use the homophily regression coefficient of 0.1416 in Column (6) of Table 4. Given that average parent income in our sample of Finnish entrepreneurs is equal to €24,325 for the bottom 20% and €108,680 for the top 20%, we obtain that the average difference in income elasticities for entrepreneurs from these family income groups is $0.14 \times \log(108,680/24,325) = 0.091$.

[24] We multiply the peer effect coefficient capturing the causal effect of the parent income of study peers (in thousands of euros) on the industry income elasticity, equal to 0.00080 in Row B of Column (4) of Table 6, by the difference in the average peer parent income for entrepreneurs across the parent income distribution. Specifically, we compare entrepreneurs in the bottom 20% of the parent income distribution, with an average peer parent income of €48,150, to those in the top 20%, with an average peer parent income of €59,542.

[25] We obtain similar results when studying the share of sales to high- vs. low-income consumer groups (Table C15).

IV.D Robustness

We now present several robustness checks. First, Figure C2 presents binned scatter plots of the main estimates from Tables 5 and 6. The figure shows that the falsification tests and main results are not driven by outliers and are appropriately summarized by linear regressions.

Second, Table C16 presents evidence on extensive margin effects. We implement this analysis on the full sample of students, with a dummy variable for being an entrepreneur as outcome. We find that the average parent income of study peers has a precisely-estimated zero effect on the probability of becoming an entrepreneur. For gender, our estimates in Table C16 indicate that being exposed to more female peers reduces entrepreneurship for women. The coefficient associated with male entrepreneurs, however, is small and statistically insignificant. Given that the causal effect on the direction of innovation is driven by men in Table 6, these results convey that selection effects are unlikely to drive our key finding that exposure to female peers leads men to target more female-consumer-intensive markets.

Third, Table C17 investigates the sensitivity of our results to various specification and sampling choices. We find that the estimates remain similar when we implement the specifications from Table 6 without controls, as well as when we use weighted regressions and study groups with 25 students or less.

V Implications for Cost-of-Living Inequality

Guided by the empirical estimates from the previous sections, we now develop a quantitative model to assess the equilibrium implications of the "social push" channel for cost-of-living inequality between men and women, due to unequal access to innovation careers.

V.A Motivation

Incorporating micro estimates into a model. Our microdata estimates yield several insights that help discipline the macroeconomic analysis. Innovator-consumer homophily is strong, including for the most impactful innovations, and these homophily patterns appear to be invariant to changes in the composition of innovators. Indeed, we found that innovator-consumer gender homophily has been stable over the past 25 years, even though the fraction of female inventors more than doubled (Figure 2). Section IV highlighted that social interactions, which are determined by a large set of slow-moving social, institutional, and cultural factors, are a key driver of innovator-consumer homophily. The key role of social interactions provides a simple theoretical explanation for the observed stability of homophily over time.

We build and calibrate an equilibrium growth model to incorporate features of innovator behavior found here and in the literature, in order to assess equilibrium implications. First, we model differences in the direction of innovation by innovators' socio-demographic backgrounds as being driven by social factors independent of financial incentives, based on the quasi-experimental evidence presented in Section IV. Second, we model barriers to access innovation careers as differences in innovation exposure, based on findings in Bell et al. (2019b). Third, we allow for innovation productivity to have a Pareto distribution, based on Bell et al. (2019a).

Our primary focus in counterfactual analysis are policies reducing access barriers for women in light of recent experimental estimates showing that simple interventions can be very effective at attracting talented female students into science-related careers and STEM fields. For example, Breda et al. (2023) show in a large-scale field experiment that a brief exposure to female role models working in scientific fields has a large impact on high school students' perceptions and choice of undergraduate major, especially for high-achieving girls. We thus focus on gender in Section V.C, and report a complementary analysis by socio-economic status in Section V.D. We also examine counterfactuals varying innovator-consumer homophily, integrating the results from Table 7 into our structural model.[26]

Back-of-the-envelope macro estimates. Before presenting the full-fledged macro analysis, we motivate the potentially large impact of broadening the pool of talent on cost-of-living inequality by gender with a simple love-of-variety framework (Appendix B.A provides

[26]We note that some caution is in order when interpreting counterfactuals changing the social structure. Although our causal analysis indicates significant peer effects in the Finnish education system, the results might not generalize to other settings. In particular, the findings in Carrell et al. (2013) suggest that redirecting social interactions with policies changing peer composition can be difficult.

details). Using a simple static framework with CES consumer preferences, we find that the ratio of sales from each innovator type (male vs. female) to each consumer type (male vs. female) is a sufficient statistic for the relative welfare effect of innovations created by male and female innovators. This sufficient statistic is directly obtained from our homophily regressions in Section III. The result is intuitive: when agents have CES preferences with similar elasticities of substitution, welfare differences can be reduced to differences in spending shares, and in turn to differences in the firm's revenue shares from each type of innovator.

Using this insight, we find that our homophily estimates from Section III imply that the welfare gains from female-founded startups are 27 percent larger for the representative female household, relative to the representative male household. Considering a scenario where the fraction of female innovators increases from 12% to parity, we find that welfare increases by 10% more for female consumers, compared to male consumers. This back-of-the-envelope calculation highlights the possibility that access barriers combined with innovator-consumer homophily may have a significant impact on cost-of-living inequality between men and women.

V.B Model and Estimation

Model We build a model staying as close as possible to workhorse models of endogenous growth (e.g., Romer, 1990) but allowing for unequal access to innovation, heterogeneity in tastes, and differences in the direction of innovation stemming from social push rather than financial incentives. The quantities of interest are the long-run growth rate and steady-state inequality across groups along a balanced growth path. We consider an economy with a unit mass of agents indexed by i belonging to either of two equally-sized groups, men and women, indexed by $g \in \{M, W\}$.

Preferences and consumption. Agents in the economy maximize lifetime discounted utility $\int_0^\infty e^{-\rho t} \log\left(C_i(t)\right)\, dt$,where the aggregate utility is a composite over the two sectors, $C_i(t) = C_{1i}(t)^{\alpha_{g(i)}} \cdot C_{2i}(t)^{1-\alpha_{g(i)}}$. Preference parameters α_g are specific to each group, allowing for potential cost-of-living inequality in equilibrium. The consumption indices for each sector are determined by the consumption of sectoral varieties, with $C_{ji}(t) = \left(\int_0^{N_j(t)} c_{ji}(\nu, t)^{(\varepsilon-1)/\varepsilon} d\nu\right)^{\varepsilon/(\varepsilon-1)}$, where $N_j(t)$ denotes the number of varieties available in sector j at time t and ε is the elasticity of substitution between varieties. The corresponding price index for each sector is $P_j(t) = N_j(t)^{1/(1-\varepsilon)}$.[27] Because preferences only vary based on group membership g, an agent's price index is given by $P_{it} = P_{1t}^{\alpha_g} \cdot P_{2t}^{1-\alpha_g}$, i.e., the cost of living is lower for consumers with a stronger taste for sectors where product variety is higher.

Agents allocate consumption optimally across sectors to maximize lifetime discounted utility subject to their intertemporal budget constraint, $\int_0^\infty e^{-rt} P_{it} C_i(t)\, dt \leq \int_0^\infty e^{-rt} w_{it}\, dt$, where r denotes the interest rate and w_{it} earnings per period.

Technology and production. Agents supply labor inelastically and decide whether to work in the production of existing varieties or innovate, i.e., create new varieties. All labor is allocated to production or innovation and the market clears: $\sum_j (L_{jI}(t) + L_{jM}(t)) \leq 1$. $L_{jI}(t)$ denotes the quantity of labor allocated to innovation in sector j, while $L_{jM}(t) = \int_0^{N_j(t)} l_j(\nu, t) d\nu$ gives the quantity of labor allocated to production. $l_j(\nu, t)$ denotes the quantity of labor allocated to the production of existing variety ν in sector j, which is paid at the wage rate $w(t)$. For tractability we assume that the wage rate is the same in all sectors, as when production workers are perfectly mobile.

Agents differ in their innovation productivity η_i, which follows a Pareto distribution with scale parameter $\bar{\eta}$ and shape parameter λ and is identical across groups.[28] The innovation production function, featuring knowledge spillovers, takes the form $\dot{N}_{ji}(t) = \eta_i N(t)$, with $N(t) = \sum_j N_j(t)$.[29] Agents inventing new varieties in sector j receive perpetual patents generating profits $\pi_j(t)$ in each period, with a total value $V_j(t) = \frac{\pi_j(t)}{r}$ based on the balanced growth path interest rate r.

Occupational choice, homophily, and exposure frictions. Two factors govern each agent's choice of occupation: financial payoffs and exposure frictions. The financial incentives are standard: the agent is indifferent between innovating in sector j and producing varieties when $w(t) = \eta_i N(t) V_j(t)$. In addition, motivated by empirical evidence, we specify two layers of exposure frictions, such that each agent has the opportunity to innovate at most in

[27] Without loss of generality, we normalize the price of individual varieties, $p(\nu)$, to be one.

[28] The assumption of a Pareto distribution for productivity is in line with the empirical evidence on the citations and wages of inventors in Bell et al. (2019a).

[29] The use of total varieties when specifying spillovers in the innovation production function is a necessary assumption to avoid explosive growth in a single sector. This assumption is similar to the trade and innovation model of Rivera-Batiz and Romer (1991). Intuitively, this means that innovators' research productivity benefits from all innovations in the economy, not just those in the sector on which they focus.

a single sector j.[30]

First, motivated by evidence from Bell et al. (2019b), we assume that agents' decisions are influenced by whether they have received exposure to innovation, denoted by τ_{gi}. Specifically, we assume that all agents in group M are exposed ($\tau_{Mi} = 1$), while those in group W are exposed based on a binary variable τ_{Wi} that follows a Bernoulli distribution, $\tau_{Wi} B(\tau)$. Agents who do not receive exposure to innovation ($\tau_{gi} = 0$) never pursue innovation, whereas those who receive exposure ($\tau_{gi} = 1$) can. We further assume the probability of exposure to innovation is uncorrelated with agents' abilities to innovate.[31]

Second, motivated by the evidence on innovator-consumer homophily presented earlier in this paper, we refine the notion of innovation exposure. We specify that, among agents with $\tau_i = 1$, agent $i \in g$ is exposed with probability $\phi \in (0, 1)$ to the sector for which group g has a stronger relative taste preference, as governed by α_g in the agent's utility function, and with probability $1 - \phi$ to the other sector. This sectoral exposure mechanism generates "social push" towards a sector to which an individual is most exposed to and governs the strength of innovator-consumer homophily. It incorporates factors driving homophily, including the social exposure effects identified in Section IV. Without loss of generality, we assume that men have a stronger relative preference for sector 1.

Agents can decide whether to innovate only for the specific sector for which they have received innovation exposure. They choose whether to innovate in this sector or to produce existing varieties by maximizing expected lifetime utility, comparing the returns to innovation in the sector they are exposed to, $V_j(t)$, with production earnings, w_{it}.

Inequality. We introduce a parameter, δ, to model the gender pay gap. Specifically, while $w(t)$ is the average wage in the economy, we assume that women earn $w_{Wt} = \delta \cdot w(t)$, while men earn $w_{Mt} = (1 - \delta) \cdot w(t)$. This parameter allows us to match the observed gender

[30] This exposure structure greatly simplifies the solution algorithm but these choices do not drive our quantitative results. We obtain similar results when specifying alternative models of career choice, as described in extensions and robustness tests below.

[31] Bell et al. (2019a) document the causal effects of exposure to innovation on the probability of becoming an inventor, and show that individuals who are less exposed to innovation— for example, women, minorities, and children from low-income families—do not appear to have large differences in their latent abilities to innovate, as measured for instance by their math test scores early in childhood.

earnings ratio.[32] Consumption inequality is given by:

$$\underbrace{\frac{C_{Mt}}{C_{Wt}}}_{\text{gender consumption gap}} = \underbrace{\frac{w_{Mt}}{w_{Wt}}}_{\text{gender pay gap}} \times \underbrace{\frac{P_{Wt}}{P_{Mt}}}_{\text{gender cost-of-living gap}} . \tag{4}$$

The gender wage gap is given by the ratio $(1-\delta)/\delta$. Using the expression for sectoral price indices and consumer preferences, we obtain that the cost-of-living gender gap is $\frac{P_{Wt}}{P_{Mt}} = \left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha_M - \alpha_W}{\varepsilon - 1}}$.

Equilibrium and counterfactuals. We solve for the steady-state equilibrium along a balanced growth path, where consumption growth, wage growth, the interest rate, and the growth rate of varieties in both sectors are constant, along with the amount of labor allocated to each sector, and to research and production within each sector. The economy is closed, i.e., $y(v,t) = c(v,t)$, and the equilibrium interest rate is consistent with the Euler equation, the financial incentives governing the optimal career choice between production work and innovation, and zero net savings in equilibrium. All first-order conditions and constraints characterizing the equilibrium are provided in Appendix B.B.1, along with our numerical solution algorithm. In particular, we solve for the research productivity cutoffs in each sector j, above which agents exposed to innovation in that sector decide to pursue an innovation career.

In equilibrium, the allocation of innovators across sectors is governed by two forces, leading to cost-of-living gender inequality. First, profit incentives must be equalized across sectors, such that there are more innovators in larger sectors. Due to the nominal gender wage gap, the sector preferred by men is larger, which attracts more innovators. Second, women have lower exposure to innovation ($E(\tau_{Wi}) < E(\tau_{Mi})$), which has a disproportionate impact on innovation in the sector preferred by women due to asymmetric sectoral exposure directing an inventor's innovative efforts towards consumers who are similar, generating innovator-consumer homophily.

Finally, two features are worth highlighting about the counterfactual equilibrium we study, relaxing exposure barriers τ_i. First, higher-ability innovators enter innovation activities in equilibrium, but only if they have received exposure to innovation ($\tau_i = 1$). In the

[32] We assume the parameter δ scales in the same way the returns to innovation $V_j(t)$ for each group g, so that it leaves occupational choice unaffected.

baseline equilibrium, society is losing some high-impact innovators, which does not occur in a standard Roy rational sorting model without exposure effects. In the counterfactual without exposure barriers, there can potentially be a large impact on growth by attracting these "Lost Marie Curies" into innovation. Second, a key offsetting effect in general equilibrium operates through interest rates. When all individuals receive exposure to innovation, there is additional entry into innovation careers. The resulting increase in the growth rate of varieties requires an increase in interest rates in equilibrium so that the Euler equation holds, which reduces the value of innovation as discounted profits are now reduced; this mechanism dampens the general equilibrium impact of reducing exposure frictions.[33]

Table 8: Baseline Parameters

Panel A: Parameters calibrated outside of the model

Model Parameter	Parameter Definition	Source	Value
$\lvert\alpha_M - \alpha_W\rvert$	Expenditure dissimilarity index by gender	Consumer Expenditure Survey, Nielsen data, phone applications data (cf. Appendix Table C18, first row)	0.24
ε	Elasticity of substitution between varieties	DellaVigna and Gentzkow (2019)	1.9
ρ	Discount rate, annual	Kaplan et al. (2018)	0.051
λ	Pareto shape parameter of innovators' productivity	Bell et al. (2019a)	1.26
$\frac{1-\delta}{\delta}$	Male to female earnings ratio	U.S. Department of Labor	1.205

Panel B: Jointly estimated model parameters

Model Parameter	Parameter Definition	Value
τ	Exposure to innovation careers	0.111
ϕ	Sectoral exposure	0.725
$\bar{\eta}$	Pareto scale parameter of of innovators' productivity	0.011

Panel C: Targeted moments and model fit

Targeted Moment [Source]	Data	Model
Share of female patent inventors [Toole et al. (2019)]	0.128	0.128
Gender homophily regression coefficients [first columns of Tables 2, 3 and 4]	0.218	0.218
Annual growth rate of labor productivity, 1990-2020 [Saint Louis Fed]	0.02	0.02

Notes: This table presents the baseline parameters of the growth model, for the analysis by gender. Panel A shows the model parameters which are directly set to the value observed in data or taken from the literature. In Panel B, the three parameters are estimated jointly to match the moments from the model with moments observed in the data, displayed in Panel C.

[33]We follow endogenous growth models (e.g., Romer, 1990; Aghion and Howitt, 1992), and allow long-run growth rates to vary in our counterfactuals. With a semi-endogenous growth model (Jones, 1995), there would be no impact on long-run growth, but we conjecture that (i) the effects on steady-state cost-of-living inequality would be similar; (ii) the impacts on growth rates would be similar over the transition to the new steady state, which could take several decades given that we are far from parity among innovators.

Estimation Table 8 summarizes our baseline parameters for the analysis by gender. Using values from the literature, in Panel A we directly set five parameters of the model: the expenditure dissimilarity index by gender $|\alpha_M - \alpha_W|$, the elasticity of substitution between varieties ε, the discount rate ρ, the Pareto shape parameter λ for the innovators' productivity distribution, and the male to female earnings ratio $\frac{1-\delta}{\delta}$. In Panel B, we report the last three parameters of the model, which jointly target the three moments reported in Panel C: the observed fraction of female inventors, the empirical innovator-consumer gender homophily documented in Section III, and the average growth rate of labor productivity. The model matches these three moments exactly. Next, we use these parameters to analyze several counterfactuals and document the sensitivity of the results to alternative parameters.

V.C Counterfactuals by Gender

Main results Figure 4 present our main results on the effects of the under-representation of women on cost-of-living inequality and economic growth. Panel A of Figure 4 shows that the cost-of-living gender gap is 18.7% in our baseline model, which is close to the nominal pay gap between men and women of 20.5%. Therefore, per Equation (4), the real consumption of men is 43% larger compared to women; without the cost-of-living gender gap, the consumption gender gap would only be half as large.[34]

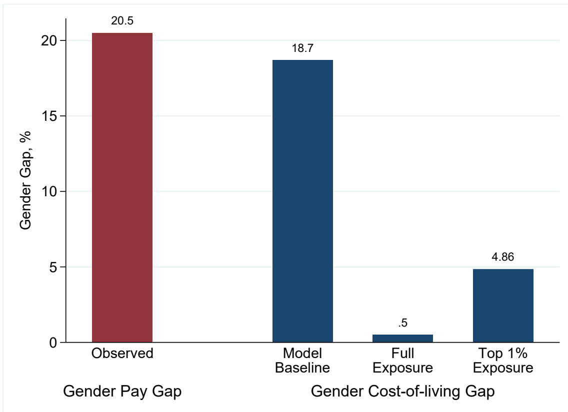
We start by studying a counterfactual equilibrium in which women are not under-represented among innovators. Specifically, we first examine a "full exposure" counterfactual, in which all women are exposed to innovation careers, i.e., we set $\tau_i = 1$ for all women. Panel A.i. shows that, in this counterfactual scenario, the cost-of-living gender gap falls to 0.50%. As reported in the figure, we find that the fall is also large when considering an alternative "exposure" policy targeting only women in the top 1% of the inventor ability distribution, i.e., we set $\tau_i = 1$ for all women in this group. In this case, the cost-of-living gender gap falls to 4.86%. Intuitively, given the skewness of the distribution of innovation abilities, it is sufficient to attract the most talented individuals to generate most of the gains.

[34]The large cost-of-living gender gap uncovered by our model is consistent with reduced-form evidence about product variety across gender groups. Measuring time use by gender for phone applications, we find that the Herfindahl–Hirschman Index (HHI) is 60% larger for women. Specifically, the HHI is 0.034 for women and 0.021 for men, consistent with there being a much larger variety of phone applications catering to the tastes of men.

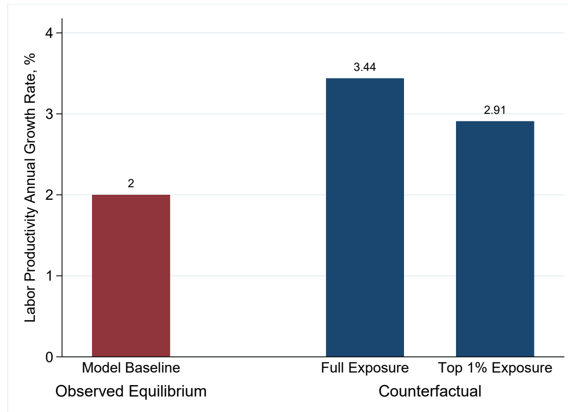
Figure 4: Main Counterfactual Estimates

Panel A: Exposure Counterfactuals

A.i. Cost-of-living

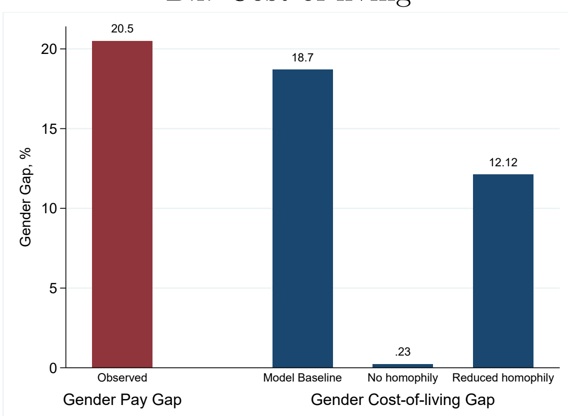


A.ii. Growth

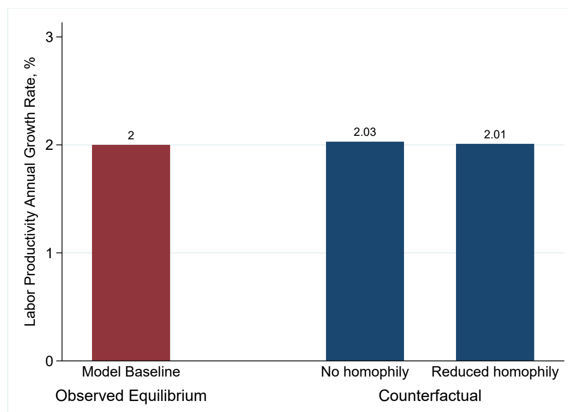


Panel B: Homophily Counterfactuals

B.i. Cost-of-living



B.ii. Growth



Notes: Panel A of this figure reports counterfactuals varying "exposure", i.e. access barriers to innovation careers. Panel A.i. focuses on cost-of-living gender gaps in different scenarios. The baseline specification uses the model parameters summarized in Table 8. The "full exposure" counterfactual sets $\tau_i = 1$ for all women. The "top 1% exposure" scenario sets $\tau_i = 1$ for all women in the top 1% of the innovation productivity distribution. The observed gender gap is shown for comparison. Panel A.ii. reports labor productivity growth at the observed equilibrium and in the two counterfactual scenarios, with full exposure or top 1% exposure. Panel B presents the counterfactuals varying homophily and social interactions, holding exposure frictions fixed to their baseline level. We study in turn a "no homophily" scenario, with no sectoral exposure bias ($\phi = 0.5$), and a "reduced homophily" scenario, reducing the targeted gender homophily coefficient by the fraction of homophily explained by the difference in the average fraction of female study peers between male and female students ($\phi = 0.661$, using Col. 1 of Table 7).

Even with full exposure of women to innovation careers, the cost-of-living gender gap does not completely disappear because of market size effects. Due to the gender pay gap, it is more profitable for innovators to target men instead of women. However, Panel A shows that this effect is relatively small quantitatively. The cost-of-living gender gap falls from 0.50% in the "full exposure" counterfactual with market size effects to 0 in the scenario eliminating differences in market size (i.e., $\delta = 0.5$).

Thus, in a model with social push we find that access frictions for female inventors explain 97% (= 18.2/18.7) of the cost-of-living gender gap, and the implication for consumption inequality by gender are just as large as the gender pay gap, around 20% of consumption.[35] In contrast, the market size channel — which has been the focus of the literature on the direction of innovation — plays a limited role, representing about 0.5% of consumption.[36] This is the first key takeaway of our quantitative analysis.

Panel A.ii. of Figure 4 turns to the growth of labor productivity across counterfactual scenarios. At baseline, our estimated model matches the 2 percent observed growth rate of labor productivity in the United States. The panel shows that growth increases to 3.44% per year under the "full exposure" counterfactual, in which women are fully exposed to innovation careers and account for 50% of inventors. Such a large increase in growth rates (+72% relative to baseline) is intuitive, since our counterfactual relaxes barriers to innovation for half of the population. In this counterfactual, the fraction of individuals exposed to innovation increases by 79% relative to baseline; however, the equilibrium number of inventors increases by 37% only, and the average productivity of inventors increases by 23%, as less productive inventors are displaced.[37] This large growth impact is the second key takeaway of our quantitative analysis.

Panel A.ii. also reports the results when exposure targets only the top 1% of women: the growth rate increases to 2.91 percent, a 46% increase relative to the baseline. In this

[35] Without cost-of-living inequality, consumption inequality between men and women is $\frac{C_{Mt}}{C_{Wt}} = \frac{w_{Mt}}{w_{Wt}} = 1.205$. With cost-of-living inequality stemming from innovator-consumer homophily, our estimates yield $\frac{C_{Mt}}{C_{Wt}} = 1.205 \times 1.182 = 1.424$.

[36] The finding that market size effects have a relatively small impact on cost-of-living inequality by gender is intuitive and can be seen in a back-of-the-envelope calculation, stepping back from our model. Reduced-form work has documented that a 1% increase in market size leads to a fall in the price index ranging between 0.1% and 0.3% (Jaravel, 2019; Costinot et al., 2019; Faber and Fally, 2022). Due to the pay gap between men and women, market size is 17.1% smaller for women (= 1/1.205). Because there is a 76% overlap in the expenditure patterns of men and women (cf. Table 8, first row), a back-of the-envelop calculation suggests market size effects lead to a modest increase in the price index for women ranging between 1.2% $(= (-17.1\%) \times 0.24 \times (-0.3))$ and 0.41% $(= (-17.1\%) \times 0.24 \times (-0.1))$.

[37] For completeness, Appendix Table C19 reports the number of inventors and their productivity at baseline and in the counterfactual equilibrium.

case, the equilibrium number of inventors falls by 9% and the average productivity of inventors increases by 66%, as high productivity female inventors displace less productive male inventors.

Thus, our quantitative model shows that policies expanding the pool of talent can have a very large macroeconomic impact, even when they focus on a relatively small number of top-talent individuals. This macroeconomic quantification complements mounting microeconomic evidence showing that simple policies can attract women into innovation careers in practice (e.g., Breda et al., 2023). Our results imply that such policies are a powerful approach both to increase growth rates and reduce consumption inequality between men and women.

We can also use our model to study the role of homophily and social interactions, holding access barriers fixed at current levels. The results are reported in Panel B. First, we study a “no homophily” counterfactual with no sectoral exposure bias. We find that the gender cost-of-living gap falls by 18.47pp (Panel B.i.), with small effects on growth (Panel B.ii). Policies reducing sectoral exposure bias, e.g., by fostering certain social interactions (Dahl et al., 2021), may thus have the potential to significantly reduce gender inequality.[38] We also connect our structural model to the quasi-experimental peer effects estimates from Table 7: we reduce the targeted gender homophily coefficient by 26.6%, i.e., by the fraction of homophily explained by the difference in the average fraction of female study peers between male and female students (see Column (1) of Table 7). In this “reduced homophily” scenario, we obtain a meaningful fall inequality (-6.58pp), with close to no impact on growth (+0.01pp).

Robustness and extensions Next, we analyze the extent to which our quantitative counterfactual results are sensitive to parameter choices. The results are reported in Table 9. Row A summarizes the results from our baseline specification, which are identical to those in Panel A of Figure 4.

[38] This counterfactual also highlights the idea that both social push (directional choices that are independent of financial incentives) and access frictions are necessary for generating quantitatively large cost-of-living inequality in equilibrium.

Table 9: Alternative Parameters and Scenarios

Specification	Cost-of-Living Gender Gap			Change in Labor
	Before Policy Shock	After	Change	Productivity Growth
	(1)	(2)	(3)	(4)
A. Main: full exposure policy, $\varepsilon = 1.9$, $\beta = 0.218$, $\lvert\alpha_M - \alpha_F\rvert = 0.24$	18.70%	0.50%	-18.20pp	+1.44pp
B. Lower epsilon, $\varepsilon = 1.34$	57.50%	1.33%	-56.17pp	+1.44pp
C. Higher epsilon, $\varepsilon = 2.5$	10.86%	0.30%	-10.56pp	+1.44pp
D. Lower innovator-consumer homophily, $\beta = 0.169$	12.25%	0.36%	-11.89pp	+1.37pp
E. Higher innovator-consumer homophily, $\beta = 0.24$	22.58%	0.44%	-22.14pp	+1.50pp
F. Less skewed ability distribution, $\lambda = 1.5$	15.51%	0.80%	-14.71pp	+1.33pp
G. Less skewed ability distribution, $\lambda = 2$	11.73%	0.61%	-11.12pp	+1.10pp
H. Less skewed ability distribution, $\lambda = 3$	8.10%	0.80%	-7.30pp	+0.80pp
I. Roy rational sorting model, with productivity wedge	18.44%	0.41%	-18.03pp	+1.45pp
J. Roy rational sorting model, no productivity wedge	4.44%	0.57%	-3.87pp	+0.38pp
K. Full exposure top 0.1%	18.70%	9.17%	-9.53pp	+0.53pp
L. Full exposure top 0.5%	18.70%	6.27%	-12.43pp	+0.78pp
M. Full exposure top 1%	18.70%	4.86%	-13.84pp	+0.91pp

Notes: This table presents the results of the models for alternative specifications, considering alternative values for certain parameters of the model, or alternative counterfactual scenarios. Unless otherwise noted, the parameters are identical to the baseline specification in the first row. Unless specified otherwise, we study the "full exposure" counterfactual, setting $\tau_i = 1$ for all women.

Focusing on the "full exposure" counterfactual, we assess in turn the sensitivity of our results to changes in the elasticity of substitution between varieties, ε, and to alternative target moments for innovator-consumer gender homophily, β.[39] Rows B to E show that the magnitudes remain large across all specifications, demonstrating the robustness of our results. Rows F to H illustrate the key role of the skewness of the ability distribution, parametrized by λ. The effects become smaller with less skewed ability distributions (with a lower λ), because female innovators displace male innovators who are closer in ability.

We also present results obtained in a Roy rational sorting model. Instead of using exposure friction, we specify that women face a wedge reducing their entrepreneurial income

[39]For ε we use the range from DellaVigna and Gentzkow (2019); for β we take the range from Tables 2, 3 and 4.

relative to men. As a result, there are fewer female inventors in equilibrium, but women at the very top of the innovation ability distribution always pursue innovation careers, in contrast to our baseline model with exposure frictions. Appendix B.B.3 provides additional discussion of this version of the model.

Using the Roy rational sorting setup, we first consider a case with "on the job discrimination", i.e., the wedge reducing entrepreneurial income for women also applies to their productivity for innovation. This specification thus reduces the aggregate innovation productivity in the economy, and row I shows that the counterfactual with no barriers for women yields large impacts for inequality (-18.03pp) and growth (+1.45pp). These magnitudes are close to our baseline model with exposure frictions. We also consider a specification where the wedge only reduces entrepreneurial income for women but leaves their innovation productivity unaffected (in the spirit of Hsieh et al. (2019)), i.e., they get lower private returns but generate as impactful innovations for society. With this specification, row J shows that relaxing frictions yields smaller effects for inequality and growth.[40]

Rows K to M in Table 9 also report the results with targeted counterfactuals, eliminating exposure frictions for women in the top 0.1% and 0.5% of the innovation ability distribution. The effects are smaller than in the "full exposure" or "top 1% exposure" counterfactuals, but remain sizable.

V.D Counterfactuals by Income Groups

We now repeat the analysis by considering low- and high-income groups, defined as individuals in the top and bottom income quintiles. The model is unchanged, except that g now indexes these two income groups. The estimation strategy is similar to our approach for the model by gender; the parameters are reported and discussed in Appendix Table C20.

Table 10 reports our main result. We find that cost-of-living inequality between the top and bottom income quintiles is 14.25% at baseline. Row A shows that, in a counterfactual imposing full exposure to the agents from the bottom income quintile, cost-of-living inequality falls by 9.74pp to 4.51%. Thus, innovator-consumer homophily by socioeconomic status

[40]We take the model with exposure frictions as our baseline in light of prior work showing the importance of exposure effects in the context of innovation (Bell et al., 2019b,a).

explains 68% of cost-of-living inequality across the distribution in our model. In contrast with the model focusing on gender, sizable cost-of-living inequality remains in the counterfactual scenario, because of market size effects. Indeed, the difference in market size between the top and bottom income quintiles is much larger than between genders, and leads to a larger set of varieties preferred by the rich in equilibrium.[41]

Table 10: Counterfactuals for Top and Bottom Income Quintiles

Specification	Cost-of-Living Inequality			Change in Labor
	Before Policy Shock	After	Change	Productivity Growth
	(1)	(2)	(3)	(4)
A. Main: full exposure policy, $\varepsilon = 1.9$, $\beta = 0.238$, $\lvert\alpha_H - \alpha_L\rvert = 0.34$	14.25%	4.51%	-9.74pp	+1.34pp
B. No homophily counterfactual, $\phi = 0.5$	14.25%	4.20%	-10.05pp	+0.00pp
C. Reduced homophily counterfactual (Col. 2 of Table 7), $\phi = 0.586$	14.25%	13.15%	-1.10pp	+0.00pp
D. Lower homophily, $\beta = 0.184$	11.75%	4.30%	-7.45pp	+1.34pp
E. Industry-level homophily, $\beta = 0.0415$	5.93%	4.48%	-1.45pp	+1.35pp
F. Lower epsilon, $\varepsilon = 1.34$	42.27%	12.38%	-29.89pp	+1.34pp
G. Higher epsilon, $\varepsilon = 2.5$	8.32%	2.68%	-5.64pp	+1.34pp
H. Less skewed ability distribution, $\lambda = 1.5$	15.11%	6.73%	-8.38pp	+1.21pp
I. Less skewed ability distribution, $\lambda = 2$	16.57%	10.10%	-6.47pp	+0.98pp
J. Less skewed ability distribution, $\lambda = 3$	18.08%	13.67%	-4.41pp	+0.70pp
K. Full exposure top 0.1%	14.25%	8.81%	-5.44pp	+0.50pp
L. Full exposure top 0.5%	14.25%	6.86%	-7.39pp	+0.74pp
M. Full exposure top 1%	14.25%	5.92%	-8.33pp	+0.86pp

Notes: This table presents the results of the models when studying high- and low-income groups, defined as individuals in the top and bottom income quintiles, considering alternative values for certain parameters of the model. Unless otherwise noted, the parameters are identical to the baseline specification in the first row. In all rows, we study the "full exposure" counterfactual, setting $\tau_i = 1$ for all agents from the low-income group.

Thus, equalizing access to innovation across the income distribution would lead to a

[41] Following the same steps as in footnote 36, a back-of the-envelope calculation based on reduced-form estimates suggests market size effects alone lead to an increase in the price index for the bottom income quintile ranging between 8.6% ($= (-84.5\%) \times 0.34 \times (-0.3)$) and 2.9% ($= (-84.5\%) \times 0.34 \times (-0.1)$). The result from the model, 4.51%, is in the middle of this range.

11.3% increase in purchasing power for the bottom income quintile, i.e., about $4,400 per year for each household in this group. This represents a transfer of about $135 billion,[42] which comes close to the budget of the Supplemental Nutrition Assistance Program ($182.5 billion) and represents about 2% of total U.S. federal spending in 2021 ($6.4 trillion). The macroeconomic relevance of the channel we study is also shown by considering growth rates, which increase from 2% a year to 3.34%, a 67% increase. This finding highlights again the large macroeconomic potential of policies broadening access to innovation careers.

Next, we present the sectoral exposure counterfactuals, again holding fixed existing access barriers. In row B, with no sectoral exposure bias, we obtain a large fall in cost-of-living inequality of 10.05pp, while growth is unaffected. In row C, we reduce the targeted income homophily coefficient by 10%, i.e., by the fraction of homophily explained by the difference in average peer parent income for entrepreneurs whose parents are in the bottom or top quintiles of the income distribution (see Column (2) of Table 7). We obtain a modest fall in inequality (-1.10pp), with no impact on growth.

For completeness, rows D to M of Table 10 document the sensitivity of the exposure results to alternative parameter values, with effects similar to the robustness analysis by gender.

VI Conclusion

In this paper, we have established three complementary results to characterize how the "social push" channel determines the direction of innovation. First, we documented a widespread pattern of innovator-consumer homophily. Second, we provided causal evidence that social factors (specifically, peer effects) affect the direction of innovation, independent of financial incentives. Finally, we developed a quantitative growth model incorporating the two preceding facts and other findings in the literature to show the macroeconomic relevance of broadening access to innovation and reducing inventor-consumer homophily, both for the overall growth rate and for cost-of-living inequality.

These findings highlight the importance of policies and initiatives aimed at promoting

[42]For the 30.75 million households in the bottom income quintile, we apply the 11.3% increase in purchasing power to their average income of $39,000 per household.

access to entrepreneurship for women and individuals from disadvantaged socio-economic backgrounds. Such policies have great potential to lead to a more diverse set of new goods and services, and to yield a double dividend by simultaneously increasing growth and reducing inequality.

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.

_ **and Joshua Linn**, "Market size in innovation: theory and evidence from the pharmaceutical industry," *The Quarterly journal of economics*, 2004, *119* (3), 1049–1090.

_ **, 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 and Peter Howitt, "A Model of Growth Through Creative Destruction," *Econometrica*, 1992, *60* (2), 323–351.

_ **, 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**, "Parental education and invention: The Finnish enigma," *International Economic Review*, 2023.

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

_ **,** _ **,** _ **,** _ **, and** _ , "Who becomes an inventor in America? The importance of exposure to innovation," *The Quarterly Journal of Economics*, 2019, *134* (2), 647–713.

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," *Working paper*, 2018.

Breda, Thomas, Julien Grenet, Marion Monnet, and Clémentine Van Effenterre, "How Effective are Female Role Models in Steering Girls Towards Stem? Evidence from French High Schools," *The Economic Journal*, 2023, *133* (653), 1773–1809.

Broda, Christian and David E Weinstein, "Product creation and destruction: Evidence and price implications," *American Economic Review*, 2010, *100* (3), 691–723.

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.

Calder-Wang, Sophie and Paul A Gompers, "And the children shall lead: Gender diversity and performance in venture capital," *Journal of Financial Economics*, 2021, *142* (1), 1–22.

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, Bruce I Sacerdote, and James E West, "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 2013, *81* (3), 855–882.

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, Janet Gerson, and Jennifer Kuan, "Closing the Innovation Gap in Pink and Black," *Entrepreneurship and Innovation Policy and the Economy*, 2022, *1* (1), 43–66.

Costinot, Arnaud, Dave Donaldson, Margaret Kyle, and Heidi Williams, "The more we die, the more we sell? A simple test of the home-market effect," *Quarterly Journal of Economics*, 2019, *134* (2), 843–894.

Dahl, Gordon B., Andreas Kotsadam, and Dan Olof Rooth, "Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams," *Quarterly Journal of Economics*, 5 2021, *136*, 987–1030.

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.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell, "The Use of Scanner Data for Economics Research," *Annual Review of Economics*, 2022, *14* (1), 723–745.

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.

Faber, Benjamin and Thibault Fally, "Firm heterogeneity in consumption baskets: Evidence from home and store scanner data," *The Review of Economic Studies*, 2022, *89* (3), 1420–1459.

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, Jean Joohyun Oh, and Ananya Sen, "What motivates innovative entrepreneurs? evidence from a global field experiment," *Management Science*, 10 2020, *66*, 4808–4819.

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.

Howell, Sabrina T. and Ramana Nanda, "Networking Frictions in Venture Capital, and the Gender Gap in Entrepreneurship," *Journal of Financial and Quantitative Analysis*, 2023, pp. 1–56.

_ , Theresa Kuchler, David Snitkof, Johannes Stroebel, and Jun Wong, "Lender Automation and Racial Disparities in Credit Access," *Journal of Finance*, forthcoming.

Hoxby, Caroline M, "Peer Effects in the Classroom: Learning from Gender and Race Variation," *NBER Working Paper 7867*, 2000.

Hsieh, Chang-Tai, Erik Hurst, Charles I Jones, and Peter J Klenow, "The allocation of talent and us economic growth," *Econometrica*, 2019, *87* (5), 1439–1474.

Hurst, Erik and Benjamin Wild Pugsley, "What do small businesses do?," *Brookings Papers on Economic Activity*, 2011, pp. 73–143.

Hvide, Hans K. and Paul Oyer, "Who Becomes a Successful Entrepreneur? The Role of Early Industry Exposure," *Working paper*, 2020.

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.

Jones, Charles I, "R & D-based models of economic growth," *Journal of political Economy*, 1995, *103* (4), 759–784.

Jones, Todd R and Michael S Kofoed, "Do peers influence occupational preferences? Evidence from randomly-assigned peer groups at West Point," *Journal of Public Economics*, 2020, *184*, 104154.

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.

Kaplan, Greg, Benjamin Moll, and Giovanni L Violante, "Monetary policy according to HANK," *American Economic Review*, 2018, *108* (3), 697–743.

Koffi, Marlène and Matt Marx, "Cassatts in the Attic," *Working paper*, 2023.

Koning, Rembrand, Sampsa Samila, and John-Paul Ferguson, "Who do we invent for? Patents by women focus more on women s health, but few women get to invent," *Science*, 2021, *372* (6548), 1345–1348.

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

Linder, Staffan Burenstam, *An essay on trade and transformation*, Almqvist & Wiksell Stockholm, 1961.

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 Differently They Talk About Female Entrepreneurs," *Harvard Business Review*, 2017, pp. 1–6.

Markussen, Simen and Knut Røed, "The gender gap in entrepreneurship - The role of peer effects," *Journal of Economic Behavior & Organization*, 2017, *134*, 356–373.

Michelman, Valerie, Joseph Price, and Seth Zimmerman, "Old boys' clubs and upward mobility among the education elite," *Quarterly Journal of Economics*, 2022, *137* (2), 845–909.

Mishkin, Elizabeth, "Gender and sibling dynamics in the intergenerational transmission of entrepreneurship," *Management Science*, 2021, *67* (10), 6116–6135.

Rivera-Batiz, Luis A and Paul M Romer, "Economic integration and endogenous growth," *The Quarterly Journal of Economics*, 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," *USPTO, Office Of The Chief Economist. IP Data Highlights*, 2019.

Truffa, Francesca and Ashley Wong, "Undergraduate gender diversity and direction of scientific research," *Working paper*, 2022.

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

For Online Publication

Appendix to "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*
October 2023

A Extensions

In this appendix we report two extensions, documenting homophily between the socio-demographic characteristics of entrepreneurs and their employees, and relating innovators' backgrounds to the environmental and social impacts of their innovations.

Entrepreneur–employee homophily. Using the Finnish administrative dataset, we document strong patterns of homophily between entrepreneurs and their employees. Figure C3 depicts these patterns for gender, parent income, and age. The share of female employees is 35pp larger in a firm headed by a female entrepreneur. The elasticity of the parent income of employees to the entrepreneur's parent income is 0.15, while the employee-entrepreneur age elasticity is 0.19. These findings suggest an additional channel whereby broadening access to innovation can reduce inequality, by stimulating labor demand for women and individuals from low-income backgrounds.

Environmental Impacts Prior work has emphasized the role of financial incentives to steer innovation toward "green innovations" (e.g., Acemoglu et al., 2012; Aghion et al., 2016). Motivated by the idea that an inventor's social identity and intrinsic motivations matter for the direction of innovation, we document whether female inventors have a different propensity to invent "green" patents, with positive environmental externalities. Specifically, using the data in Aghion et al. (2016), we study the differences in characteristics of inventors of "clean" versus "dirty" patents.[1] In this sample of energy patents, where 13.3 percent of energy patents are classified as "clean", we find that moving from an all-male energy patent to an all-female patent is associated with a 32.6pp increase in the probability of the patent being clean, as reported in Appendix Table C21. This difference is large: women are 2.5 times more likely to work on green patents than men. Expanding access to innovation careers for women may thus be a powerful tool to direct innovation toward cleaner technologies.[2]

B Theory Appendix

In this appendix, we first present a simple love-of-variety framework (Section B.A) and then proceed to the full-fledged growth model (Section B.B).

B.A A Simple Love-of-Variety Framework

We present a simple love-of-variety framework to assess the effects of unequal access to the innovation system on cost-of-living inequality.

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

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

where σ is the elasticity of substitution between products, $q_{k,i}$ 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.

In this setting, we model innovation as the introduction of new goods, i.e., an increase in the set of available products Ω_t. Following Feenstra (1994), the welfare gains as a percentage

[1]Following Aghion et al. (2016), we classify patents as "green" depending on their their international patent classification (IPC). We then compute the fraction of female inventors at the patent level.

[2]Using the Nielsen CPG-Crunchbase sample and the full Crunchbase sample, we also find that female-led startups are more likely to mention "healthy", "kids", "sustainability", and B-Corporation certification in their company descriptions (not reported). These results are consistent with recent evidence that women entrepreneurs are motivated by social impact (Guzman et al., 2020).

of i's current income, i.e., the equivalent variation for household i, are given by:

$$\pi_i = \frac{1}{\sigma - 1} \log \left(\frac{1 + \text{Growth of spending on continued goods}_i}{1 + \text{Growth of spending on all goods}_i} \right).$$

Assuming inelastic labor supply and taking the wage as the numeraire,[3] the equilibrium growth of total spending must be zero (by normalization), and the growth in spending on continued products is mechanically related to the share of spending on new goods, S_i^N, with

$$\text{Growth of spending on continued goods}_i = -S_i^N.$$

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 10% is equivalent to a welfare increase of 2% (= 10/5). This number can equivalently be interpreted as the fall in the cost-of-living brought about by innovation.

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."[4] We derive the welfare comparison between these two households when transitioning from the "Baseline" to the "Equal" distribution of innovator backgrounds. We then discuss how to bring these formulas to the data, computing the distributional effects between high- and low-income households, as well as between male and female consumers.

[3] We assume there is only one wage rate in the economy but that different households are endowed with different efficiency units of labor, such that they can have different income and spending levels.

[4] 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 terms of their consumer base. In other words,through preference parameters $\omega_{i,k}$, different groups of consumers may have different spending shares on the new goods introduced by different startups. S_1^N denotes 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.

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:

$$\frac{\pi_1}{\pi_2} \approx \frac{S_1^N}{S_2^N} = \frac{S_1^N \cdot Y_1}{S_2^N \cdot Y_2} \cdot \frac{Y_2}{Y_1} = \frac{R_1/R_2}{Y_1/Y_2}, \tag{A1}$$

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, Y_1/Y_2. 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.

A simple calibration. We wish to examine the extent to which one of the household types may benefit more from transitioning to a new distribution of innovator background. When moving from the baseline distribution of entrepreneur background ("Baseline") to a counterfactual distribution ("Equal"), from Equation (A1) we obtain that the unequal welfare effect across household types can be expressed as:

$$\Delta W \equiv \frac{\pi_1^{\text{Equal}}/\pi_1^{\text{Baseline}}}{\pi_2^{\text{Equal}}/\pi_2^{\text{Baseline}}} = \frac{R_1^{\text{Equal}}/R_2^{\text{Equal}}}{R_1^{\text{Baseline}}/R_2^{\text{Baseline}}}. \tag{A2}$$

The relative welfare effect is thus governed by the share of sales to household groups of different types. We can directly connect this expression to the homophily regression coefficients from specification (1) in the main text. Denoting by λ the share of sales to "Type

1" households, we can write $R_1/R_2 = \lambda/(1-\lambda)$. Tables 3 and 4 are directly informative about λ for startups with different innovator backgrounds.

Consider female ("Type 1") and male ("Type 2") entrepreneurs. Using the notation for the regression coefficients in Equation (1), then for a female entrepreneur we have $\lambda^F = \alpha + \beta$, while for a male entrepreneur $\lambda^M = \alpha$. For example, for consumer packaged goods, $\lambda^F = 0.297$ and $\lambda^M = 0.25$ (Table 3, col. (1)). Our homophily estimates are thus directly informative about changes in revenue shares, which govern the welfare gains from different types of startups across consumers. For consumer packaged goods, Equation (A2) yields:

$$\Delta W = \frac{\lambda^F/(1-\lambda^F)}{\lambda^M/(1-\lambda^M)} = \frac{0.297/(1-0.297)}{0.25/(1-0.25)} \approx 1.27$$

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.[5] This number increases to 46 percent in the sample of phone applications.[6]

Using these estimates, we can assess the relative effect of having a more representative pool of innovators. We start from the observed distribution of entrepreneur backgrounds, where women represent about 12 percent of innovators (Toole et al. (2019)), i.e., $\lambda^{\text{Baseline}} = 0.12 \times \lambda^F + 0.88 \times \lambda^M = 0.2556$. We consider a counterfactual distribution with parity between male and female inventors, where $\lambda^{\text{Counterfactual}} = 0.5 \times \lambda^F + 0.5 \times \lambda^M = 0.2735$. According to Equation (A2), moving to a world where fifty percent of innovators are women would yield 9.6% larger relative benefits to female consumers ($= \frac{0.2735/(1-0.2735)}{0.2556/(1-0.2556)}$), i.e., cost-of-living inequality would fall by this amount.[7]

[5] Note that this difference is larger than the difference in revenue shares from female households that arises between female-founded and male founded startup. As shown in Column (1) of Table 3, the revenue share from female-led households is 18.8% larger for female-founded startups compared with male-founded startups (29.47% 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.

[6] 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.

[7] We find larger effects in the full-fledged growth model, because the back-of-the-envelope calculation from the simple love-of-variety framework does not take into account that reducing access barriers may bring into the innovation system some highly productive female inventors, who will have a disproportionate impact on female consumers through homophily.

Note that this simple framework is easy to take to the data to document the relative welfare effects from changes in the distribution of innovator characteristics, but it does not provide estimates of the welfare effects in levels. This limitation is addressed by the endogenous growth model we develop.

B.B Endogenous Growth Model

This section presents the main derivation steps for the growth model, as well as our numerical solution algorithm.

B.B.1 Derivations

In this section, we present derivations related to the model setup in Section V.B.

Our goal is to derive the first-order conditions that characterize the research productivity cutoff in each sector in a balanced growth path equilibrium. Given the setup in the main text, we have two sets of conditions that pin down the cutoffs. First, we have indifference conditions at the cutoff in each sector, where the marginal agent is indifferent between production work and innovation work. Second, we have the Euler equation, which ensures that the allocation of labor to innovation in the economy accords with consumer preferences. For example, if consumers are very impatient, then the economy cannot sustain high rates of innovation in equilibrium.

We make one simplifying assumption that allows us to vary market size in a tractable way. We assume that the earnings of inframarginal entrepreneurs is negligible and therefore do not impact market size. Under this assumption, δ governs the relative market size of the two sectors.

Given preference parameters $(\rho, \varepsilon, \alpha_M, \alpha_W)$ and research-related parameters $(\bar{\eta}, \lambda, \tau, \phi)$, the model is solved as follows:

1. We express the distribution of research productivity in each sector, $f_1(x)$ and $f_2(x)$, as a function of the exogenous parameters of the model:

    - Let $f(x)$ be the research productivity distribution in the population, a Pareto distribution with scale parameter $\bar{\eta}$ and shape parameter λ.

- Taking into account exposure and sector assignment, we can write the two sector-level innovation productivity distributions in terms of parameters as follows:
  - $f_1(x) = \frac{1}{2}(\phi \cdot f(x) + (1-\phi) \cdot \tau \cdot f(x))$ and $f_2(x) = \frac{1}{2}((1-\phi) \cdot f(x) + \phi \cdot \tau \cdot f(x))$.
  - Note that at $\tau = 1$, when there are no differences in exposure, the two distributions will be identical. The distributions will also be equal at $\phi = \frac{1}{2}$ for any value of τ.

2. We define some additional variables:

   - Let $\hat{\eta}_1$ and $\hat{\eta}_2$ represent the cutoffs in each sector. Agents assigned to the sector pursue an innovation career if their productivity exceeds the sector-specific cutoff.
   - Let $\tilde{\alpha} = \delta\alpha_W + (1-\delta)\alpha_M$ represent the market-size-weighted preference, or "effective taste", for sector 1 across the whole population.
   - Let L_{sM} denote the measure of agents devoted to production work in sector s.

3. Using this notation, we have the following in equilibrium:

   (a) Production workers will be split between sectors based on effective tastes in the economy: $\frac{L_{1M}}{L_{2M}} = \frac{\tilde{\alpha}}{1-\tilde{\alpha}}$.

   (b) The labor marker clears, so with a mass one of agents the number of production workers relates to the number of innovators as follows: $L_{1M} + L_{2M} = 1 - \int_{\hat{\eta}_1}^{\infty} f_1(x)dx - \int_{\hat{\eta}_2}^{\infty} f_2(x)dx$

4. In equilibrium, the marginal agent in each sector is indifferent between production work and innovation:

   (a) The indifference condition is $w(t) = \hat{\eta}_s N(t) V(t) = \hat{\eta}_s N(t) \frac{\pi(t)}{r^*} = \frac{1}{\varepsilon - 1} \cdot \hat{\eta}_s N(t) \frac{w(t)}{r^*} \cdot \frac{L_{sM}}{N_s(t)}, \forall s$.

   (b) This indifference condition implies $\hat{\eta}_1 L_{1M}/N_1(t) = \hat{\eta}_2 L_{2M}/N_2(t)$.

   (c) Furthermore, we know that varieties grow at the same rate across the two sectors in a balanced growth path equilibrium: $\frac{N(t)}{N_1(t)} \int_{\hat{\eta}_1}^{\infty} x f_1(x)dx = \frac{N(t)}{N_2(t)} \int_{\hat{\eta}_2}^{\infty} x f_2(x)dx$

(d) Combining steps 3(a), 4(b), and 4(c), we obtain the first equilibrium condition **(FOC1)**:

$$\frac{\hat{\eta}_1}{\hat{\eta}_2} = \frac{1-\tilde{\alpha}}{\tilde{\alpha}} \cdot \frac{\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx}{\int_{\hat{\eta}_2}^{\infty} x f_2(x)dx}.$$

This first-order condition pins down the relative levels of innovation productivity cutoffs across sectors.

5. Next, we use the Euler equation to pin down the level of these cutoffs:

- The standard Euler equation holds: $\frac{\dot{C}(t)}{C(t)} = r(t) - \rho = \frac{\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_1(t)}{N_1(t)} + \frac{1-\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_2(t)}{N_2(t)}$.
- From step 4(a) we have $r^* = \frac{1}{\epsilon-1} \cdot \hat{\eta}_s N(t) \frac{L_{sM}}{N_s(t)}, \forall s$. Plugging this expression on the LHS of the Euler equation yields: $\frac{1}{\varepsilon-1}\hat{\eta}_1 \frac{N(t)}{N_1(t)} L_{1M} - \rho = \frac{\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_1(t)}{N_1(t)} + \frac{1-\tilde{\alpha}}{\varepsilon-1}\frac{\dot{N}_2(t)}{N_2(t)}$.
- On the RHS of the Euler equation, we plug in the growth in varieties using the research production function: $\frac{1}{\varepsilon-1}\hat{\eta}_1 \frac{N(t)}{N_1(t)} L_{1M} - \rho = \frac{\tilde{\alpha}}{\varepsilon-1}\frac{N(t)\cdot\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx}{N_1(t)} + \frac{1-\tilde{\alpha}}{\varepsilon-1}\frac{N(t)\cdot\int_{\hat{\eta}_2}^{\infty} x f_2(x)dx}{N_2(t)}$.
- Using the equations in steps 3(a) and 4(b), we obtain the ratio of varieties across sectors, $\frac{N_1(t)}{N_2(t)} = \frac{\int_{\hat{\eta}_1} x f_1(x)dx}{\int_{\hat{\eta}_2} x f_2(x)dx}$. We use this expression to plug in for $\frac{N(t)}{N_1(t)}$ and $\frac{N(t)}{N_2(t)}$ on the RHS of the Euler equation, and we arrive at the second equilibrium condition **(FOC2)**:

$$\frac{1}{\varepsilon-1}\hat{\eta}_1 \frac{\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx + \int_{\hat{\eta}_2}^{\infty} x f_2(x)dx}{\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx} L_{1M} - \rho = \frac{1}{\varepsilon-1}\left(\int_{\hat{\eta}_1}^{\infty} x f_1(x)dx + \int_{\hat{\eta}_2}^{\infty} x f_2(x)dx\right).$$

Together, FOC1 and FOC2 pin down the two innovation productivity cutoffs. Note that in FOC2, L_{1M} also depends on $\hat{\eta}_1$, so the entire system is non-linear and does not permit a closed form solution.

Having solved the model, we compute the cost-of-living gender gap as $\frac{P_{Wt}}{P_{Mt}} = \left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha_M-\alpha_W}{\epsilon-1}}$, using the fact that the sector CES price index is $P_s = \left(\int_0^{N_s} p(x)^{1-\varepsilon}dx\right)^{1/(1-\varepsilon)} = N_s^{1/(1-\varepsilon)}$, which we plug into each consumer group's price index, $P_g = P_1^{\alpha_g} P_2^{1-\alpha_g}$.

B.B.2 Simulation Algorithm

We now describe our numerical simulation algorithm. The estimation of the model proceeds in three steps:

First, for each agent we draw a research productivity value $n_i \sim PI(1, 1.26)$ (type-I Pareto distribution). We also draw a sector assignment number$\widetilde{\phi}_i \sim U[0, 1]$ for all individuals. For agents in the minority group, we draw an exposure value from $\widetilde{\tau}_i \sim U[0, 1]$. All individuals in the majority group (i.e., men or the top income quintile) are assumed to be exposed. We thus obtain draws from uniform distributions once at the beginning of the simulation, avoiding us to redrawing random variables under each new parameter guess.

Second, for a given parameter vector guess $(\tau, \overline{\eta}, \phi)$, we convert the numerical draws described above into Bernoulli draws. Agents from the minority group are exposed if $\widetilde{\tau}_i < \tau$, and we assign agents to the sector for which they have a taste preference if $\widetilde{\phi}_i < \phi$. We also scale the research productivity draws by $\overline{\eta}$ to control the shape of the innovation productivity distribution, resulting in a distribution with CDF $F(x) = 1 - \left(\frac{\overline{\eta}}{x}\right)^{\alpha}$. This allows us to then compute the empirical distributions $f_1(x), f_2(x)$, corresponding to the distribution of productivity of individuals who are exposed and assigned to sector 1 or sector 2.

Finally, we numerically search for the cutoffs $\hat{\eta}_1$ and $\hat{\eta}_2$ that satisfy the two first-order conditions, FOC1 and FOC2, derived in the previous subsection. Given these cutoffs, we then compute the growth rate, the propensity to pursue an innovation career for each group, and homophily. We perform a derivative-free search ("fminsearch" in MATLAB) over the set of parameters $(\tau, \overline{\eta}, \phi)$ to match the observed moments, which we are able to match exactly.

Once we have calibrated the baseline economy, we run counterfactual analyses. Specifically, we vary the parameter values for τ, and recompute the BGP growth rate and cost-of-living inequality. We either run a "full exposure" counterfactual, where all agents receive exposure to innovation, or "targeted exposure" counterfactuals where we expose all "unexposed" individuals in the top x percent of the productivity distribution among "unexposed" individuals.

To check that the results are not sensitive to the specific draws of the random variables, we consider an economy with a large number of agents, 8 million. We thus have 4 million agents from each group, where groups refer to gender or top/bottom income quintiles. Furthermore, we repeat the analysis one hundred times and report values for median draws.

B.B.3 Roy Rational Sorting Models

We also simulate two Roy rational sorting models.

First, we consider a version of the model where all agents receive exposure to innovation, but the productivity of agents in the minority group is scaled down by $(1 - \tau_{Roy})$. In this model, individuals at the top of the innovation productivity distribution always enter the innovation sector, but they are less productive than they would be in the absence of this friction. To solve the model, we solve for productivity cutoffs as described previously. As we reduce τ_{Roy} towards zero in counterfactuals, existing minority inventors in the system become more productive and new minority innovators enter the innovation system.

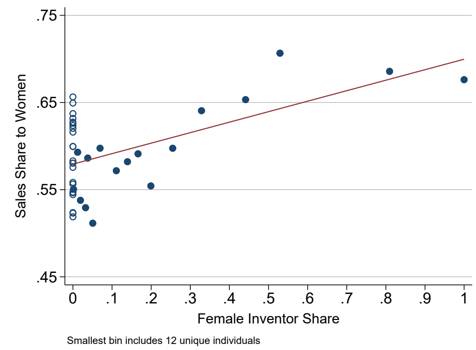
Second, we consider a Roy model akin to Hsieh et al. (2019). We set both majority and minority agents to have the same research productivity distribution and full exposure. We then exclude minority inventors from the innovation system if they are assigned to sector s and their research productivity is $\eta_i < \frac{\hat{\eta}_s}{(1-\tau_{Roy'})}$. The parameter $\tau_{Roy'}$ acts as a barrier to entry into the innovation sector for minority groups. However, once they enter, they are equally effective at innovating, in contrast with the first Roy model described above. In this case, as we reduce $\tau_{Roy'}$ towards zero, the most productive minority inventors are already in the system and do not benefit, so most of the gains come from new, less-productive minority innovators entering the innovation system. Therefore, the impacts of reducing frictions – in terms of both growth and cost-of-living inequality – are smaller than under the alternative versions of the model. Intuitively, in this model the most productive innovators always pursue innovation careers. We note that empirical evidence on innovators' backgrounds supports the alternative versions of the model, i.e. there are many "Lost Einsteins" – people who would have had very impactful innovations in the absence of frictions. See Bell et al. (2019b) and Bell et al. (2019a) for a discussion of the relevant evidence.

C Additional Figures and Tables

Figure C1: Binned Scatter Plots for Innovator-Consumer Homophily across Industries

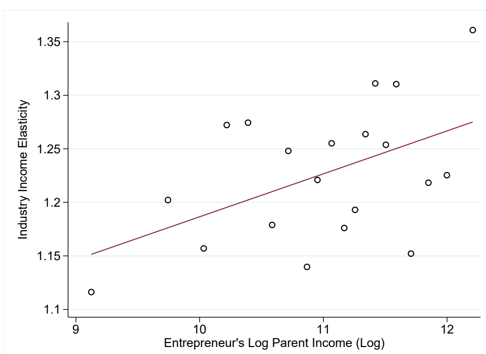
A. Gender

United States



Finland

B. Parent Income

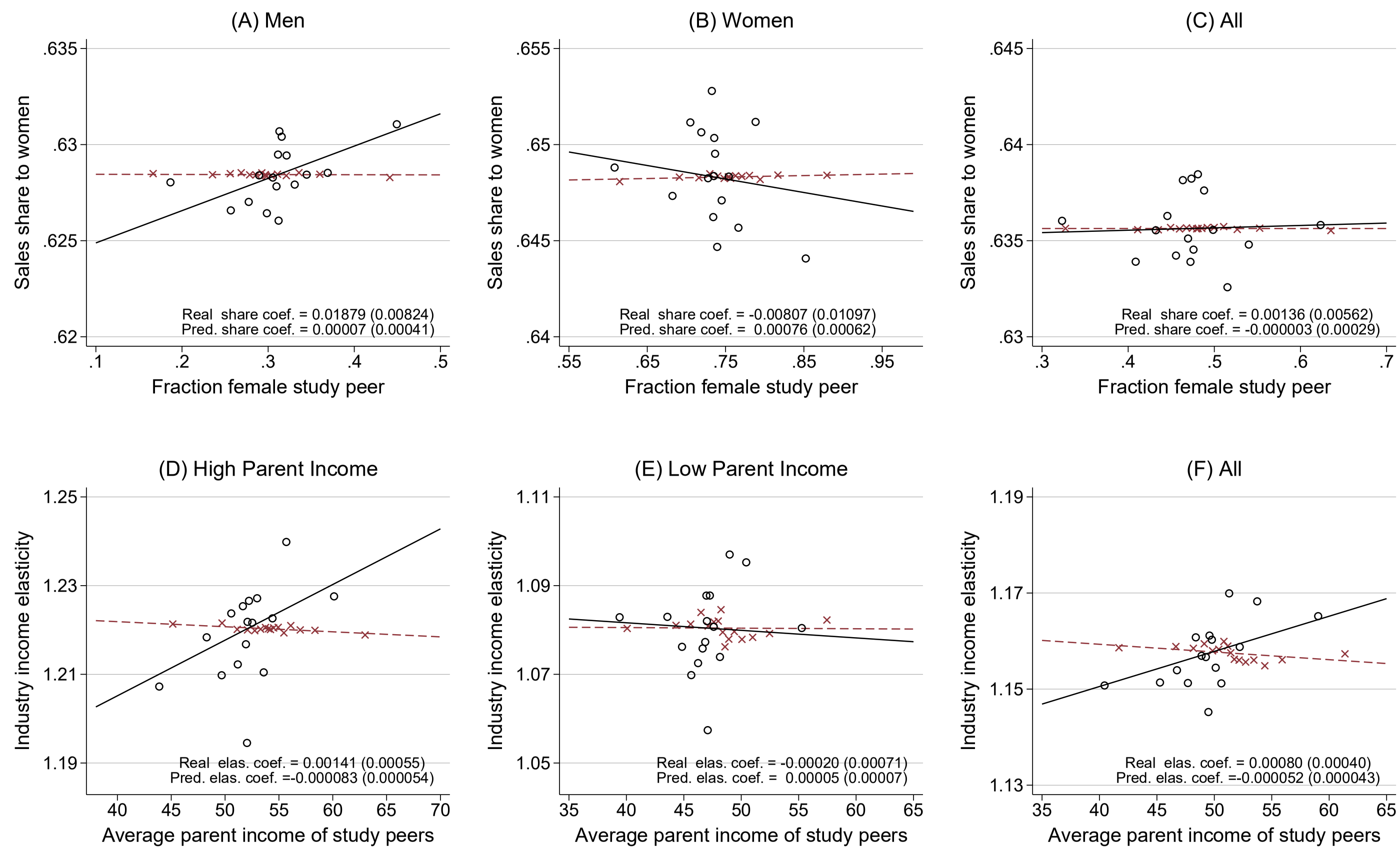


United States

Finland

Notes: This figure presents binned scatter plots depicting innovator-consumer homophily by gender and social class. The samples are the same as in Table 4 in the main text.

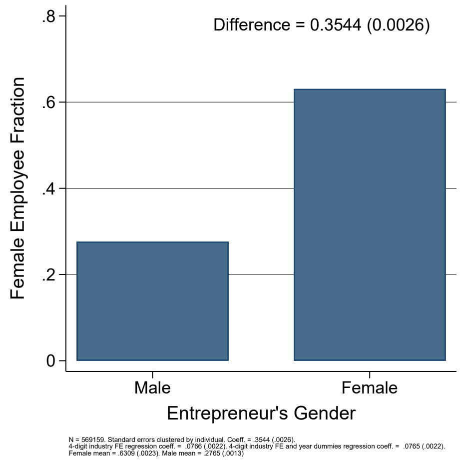
Figure C2: Effect of Study Peers on the Direction of Innovation



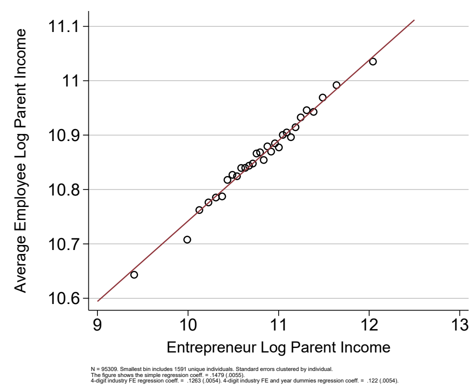
Notes: Panels A to C display the effect of the fraction of female study peers on the realized and predicted industry sales shares to women. Panels D to F display the effect of average parent income of study peers on the realized and predicted industry income elasticity. Predicted outcomes are constructed as the best linear prediction based on the pre-determined characteristics listed in footnote 16. Outcomes are means from age 28 onward. For realized outcomes, the figure plots the residuals from separate regressions of the x- and y-axis variables on pre-determined characteristics, dummies for calendar year, and school-by-start-year and study program fixed effects. For predicted outcomes, the figure plots residuals from otherwise similar regressions which do not control for pre-determined characteristics. The fitted regression lines pass through co-ordinates corresponding to the sample means of the variables on the horizontal and vertical axes. Income is expressed in thousands of euros. Standard errors clustered at the school-by-start-year level are in parentheses.

Figure C3: Entrepreneur-Employee Homophily

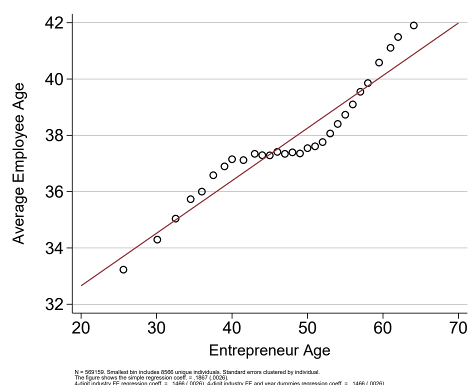
A. Gender



B. Parent Income



C. Age



Notes: This figure depicts the relationship between entrepreneur and employee gender, parent income, and age, in the Finnish administrative data set.

Table C1: Innovator-Consumer Homophily for Phone Applications, Weighted Regressions (Logarithm of Time Use)

	Female User Share			Founder State User Share	
	(1)	(2)	(3)	(4)	(5)
Female Founder Fraction	0.0781**	0.0675**	0.0778*		
	(0.0349)	(0.0327)	(0.0429)		
Female VC Fraction			0.163**		
			(0.0738)		
Founder State				0.0999***	0.0506***
				(0.00773)	(0.0104)
Fixed Effects	None	Subcategory	None	None	Subcategory
Sample Size			$N = 3,380$		

Notes: The sample used in this table includes all phone applications for VC-backed startups. This table is identical to Table 2 in the main text, except that all applications are now weighted by the logarithm of time use. The specification thus gives more weight to applications that are more widely used. In columns (1) to (3), the outcome variable is the fraction of time usage of an app accounted for by female users. In columns (4) and (5), the outcome variable is the fraction of time usage of the app by users located in the same U.S. state as the founder of the app. The coefficients are very similar to the main text, showing that the results are not driven by "marginal" innovations. The level of observation is an app. Standard errors are clustered at the startup level. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: Innovator-Consumer Homophily for Phone Applications, Raw Usage (Logarithm of Time Use)

	Female Usage	
	(1)	(2)
Female Founder Fraction	0.0569	0.0466
	(0.673)	(0.694)
Fixed Effects	None	Subcategory
Sample Size	$N = 3,380$	

Notes: The sample used in this table includes all phone applications for VC-backed startups. This table is identical to Table 2 in the main text, except that the outcome is log(1+female usage). The level of observation is an app. Standard errors are clustered at the startup level. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: Innovator-Consumer Homophily for Phone Applications

	Founder State User Share	
	(1)	(2)
Founder State	0.086***	0.037***
	(0.007)	(0.008)
Fixed Effects	None	Category-by-State
Sample Size	$N = 3,380$	

Notes: The sample used in this table includes all phone applications for VC-backed startups. The outcome variable is the fraction of time usage by users located in the same U.S. state as the founder of the app, with a sample mean of 0.02. The level of observation is an app. Standard errors are clustered at the startup level. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C4: Innovator-Consumer Homophily for Consumer Packaged Goods, Alternative Consumer Gender Measure

	Share of Sales to Women	
	(1)	(2)
Female Founder Fraction	0.0367**	
	(0.0159)	
Female Patent Inventor Fraction		0.0361***
		(0.0128)
Product Module F.E.	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 1,094,229$

Notes: This table is identical to Columns (1) and (2) of Table 3 in the main text, except that we measure consumer gender by using the fraction of female members of the households, rather than the gender of the household head. The coefficients are very similar to the main text, showing that our results are robust to alternative measures of consumer gender. Standard errors are clustered at the firm level. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: Innovator-Consumer Homophily for Consumer Packaged Goods, Weighted Regressions (Logarithm of Sales)

	Share of Sales to Women		Average Consumer Age, Sales-weighted
	(1)	(2)	(3)
Female Founder Fraction	0.0408*		
	(0.0212)		
Female Patent Inventor Fraction		0.0272*	
		(0.0153)	
Founder Age			0.127**
			(0.0532)
Product Module F.E.	Yes	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 1,094,229$	Startups, $N = 4,058$

Notes: This table is identical to Table 3 in the main text, except that all products are now weighted by the logarithm of sales. In columns (1) and (2), the outcome variable is the fraction of sales to households with a female head. The sample means are 0.256 in column (1) and 0.265 in column (2). In column (3), the outcome variable is the average age of consumers, where the average is obtained using sales weights. The level of observation is a product. The coefficients are very similar to the main text, showing that the results are not driven by "marginal" product innovations. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C6: Innovator-Consumer Homophily for Consumer Packaged Good, Controlling for Prices

	Share of Sales to Women		Average Consumer Age, Sales-weighted
	(1)	(2)	(3)
Female Founder	0.0485**		
	(0.0209)		
Female Patent Inventor Fraction		0.0259*	
		(0.0148)	
Founder Age			0.135**
			(0.0528)
Unit Price	-0.0008	-0.0009***	-0.000157
	(0.0006)	(0.0001)	(0.0152)
Product Module F.E.	Yes	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 1,094,229$	Startups, $N = 4,058$

Notes: This table is identical to Table in the main text, except that we now control for unit prices in all regressions. In columns (1) and (2), the outcome variable is the fraction of sales to households with a female head. The sample means are 0.256 in column (1) and 0.265 in column (2). In column (3), the outcome variable is the average age of consumers, using sales weights, with a sample mean of 47.2. The level of observation is a product. The estimated homophily coefficients remain very similar, demonstrating that homophily is not driven by prices. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C7: Innovator-Consumer Homophily by Age Group, Alternative Consumer Age Measure

	Share of Sales to Age Group	
	(1)	(2)
Founder of Same Age Group	0.0433*	0.0434***
	(0.0244)	(0.0155)
Fixed Effects	None	Module-by-Age Group
Sample size	$N = 4,058$	

Notes: This table presents an alternative analysis of age-based homophily. We create five age group bins based on quintiles of the consumer age distribution in Nielsen (the cutoffs are 42.5, 52.5, 60, and 65). We then compute the share of sales to each age group for each product. Finally, we create a binary indicator, "Founder of Same Age Group," equal to one if the average age of the founding team at founding matches the given age group. The baseline probability of sales is mechanically 20% across all groups. The results show that having a founder in the same age group increases the probability of sales by about 21.5% relative to baseline. Standard errors are clustered at the firm level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C8: Innovator-Consumer Homophily across Industries in Finland, Excluding Agriculture

	Share of Industry Sales to Women		Industry Income Elasticity		Average Consumer Age, Sales-weighted	
	(1)	(2)	(4)	(5)	(8)	(9)
Female Patent Inventor	0.0488***					
	(0.0037)					
Female Entrepreneur		0.0401***				
		(0.0005)				
Patent Inventor's Log Parent Income			0.0269**			
			(0.0125)			
Entrepreneur's Log Parent Income				0.0119***		
				(0.0027)		
Patent Inventor Age					0.0537***	
					(0.0051)	
Entrepreneur Age						0.0253***
						(0.0013)
Country	Finland	Finland	Finland	Finland	Finland	Finland
Mean	0.5688	0.6369	1.279	1.300	0.5688	51.81
N industries	330	451	245	417	330	451
N individuals	9,592	274,785	3800	83,316	9,592	274,785

Notes: All regressions are run at the level of an individual innovator, with outcomes measured at the industry level. Standard errors are clustered at the individual level. This table is identical to Table 4 in the main text, except that we exclude agriculture. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C9: Innovator-Consumer Homophily across Industries, by Income Group Expenditure Shares

	Above 100k vs. below 30k			Above 60k vs. below 60k		
	(1)	(2)	(3)	(4)	(5)	(6)
Patent Inventor's Log Parent Income		0.0082**			0.0053*	
		(0.0034)			(0.0027)	
Entrepreneur's Log Parent Income	0.0101**		0.0371***	0.0076**		0.0247***
	(0.00428)		(0.0008)	(0.00299)		(0.0006)
Country	U.S	Finland	Finland	U.S	Finland	Finland
Mean	0.7053	0.6997	0.6809	0.6328	0.6241	0.6122
N industries	17	253	441	17	253	441
N individuals	275	3,812	99,189	275	3,812	99,189

Notes: This table reports industry-level homophily estimates for the United States and Finland. To compute the share of sales to households earning above \$100k ("high-income") or below \$30k ("low-income") depending on an entrepreneur's own family income background in the U.S., we proceed as follows. We take the regression coefficient in column (1) and use average parent income across the U.S. income distribution, equal to \$14,859 for entrepreneurs from the bottom 20%, and \$269,356 for the top 20% in 2021, according to the U.S. Census Bureau Historical Income Tables (see here, Table H-3). We obtain that the share of sales to high-income households increases by 4.15% relative to the baseline rate when an entrepreneur comes from a family from the top income quintile instead of the bottom ($= 0.0101 \times \log(269,356/14,859)/0.7053$). We can use this between-industry homophily estimate to extrapolate and estimate the overall homophily coefficient. Specifically, we assume that the relative magnitudes of gender homophily and income homophily are the same within and between industries; therefore by rescaling our overall gender homophily estimate (= 21.8) by the ratio of between-industry income and gender homophily estimates (= 4.15/3.9), we obtain that the overall income homophily coefficient for the United States is 23.8%. For Finland, we take the regression coefficient in column (3) and use average income across the Finnish disposable income distribution, equal to \$16,581 for the bottom 20% and \$84,547 for the top 20%, according to official statistics in 2013. Thus, the share of sales to high-income households increases by 3.85% relative to baseline when an entrepreneur comes from the top income quintile instead of the bottom ($= 0.0371 \times \log(84,547/16,581)/0.6809$). $*p < 0.1$, $**$ $p < 0.05$, $***$ $p < 0.01$.

Table C10: Innovator-Consumer Homophily across Industries, Weighted Regressions

	Share of Industry Sales to Women			Industry Income Elasticity			Average Consumer Age, Sales-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Female Patent Inventor	0.0209***	0.0478***							
	(0.000203)	(0.0042)							
Female Entrepreneur			0.0306***						
			(0.0003)						
Patent Inventor's Log Parent Income				0.0300**					
				(0.0145)					
Entrepreneur's Log Parent Income					0.0385**	0.1418***			
					(0.013013)	(0.0034)			
Patent Inventor Age							0.00313***	0.0579***	
							(0.000814)	(0.0057)	
Entrepreneur Age									0.0160***
									(0.0010)
Country	U.S.	Finland	Finland	Finland	U.S.	Finland	U.S.	Finland	Finland
N industries	325	342	476	253	17	441	323	342	476
N individuals	2,219,193	9,643	344,698	3,812	275	99,189	48,156	9,643	344,698

Notes: All regressions are run at the level of an individual innovator, with outcomes measured at the industry level. Regressions are weighted as follows: columns studying patent inventors (col. (1), (2), (4), (7) and (8)) use $\log(1 + \text{patents}_i)$ as weights; columns studying entrepreneurship (cl. (3), (5), (6) and (9)) use $\log(1 + \text{income}_i)$ as weights. The coefficients are similar to those reported in the main text, indicating that homophily between industries is not driven by marginal innovators. Standard errors are clustered at the individual level. $^*p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C11: Industry-Level Regressions for Innovator-Consumer Homophily, with Industry-Level Independent Variables

	Share of Industry Sales to Women			Industry Income Elasticity		Average Consumer Age, Sales-weighted		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Female Patent Inventor Fraction	0.677***	0.1032***						
	(0.133)	(0.0163)						
Female Entrepreneur Fraction			0.120***					
			(0.021)					
Patent Inventor's Log Parent Income				0.0812**				
				(0.0393)				
Entrepreneur's Log Parent Income					0.1377**			
					(0.0581)			
Patent Inventor Age						0.121	0.0486*	
						(0.105)	(0.0278)	
Entrepreneur Age								0.1097***
								(0.0357)
Country	U.S.	Finland	Finland	Finland	Finland	U.S.	Finland	Finland
Mean	0.593	0.5843	0.592	1.1478	1.1267	48.97	49.62	49.4101
N industries	325	476	476	253	441	323	342	476

Notes: This table reports industry-level homophily estimates for the United States and Finland. The independent variables are observed at the individual level but averaged to the industry level. This table is thus identical to Table 4 in the main text, except that the regressions are implemented at the industry level instead of the individual level. Since most of the variation in innovator covariates occurs within industries, where the outcome does not vary, the point estimates in this table tend to be larger than in Table 4. Standard errors are clustered at the industry level. $^*p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table C12: Separate Falsification Tests for Each Background Characteristics, Study Peer Design

	Dependent variable	Fraction female among study peers: Coeff.	s.e.	Adjusted p-value	Average parent income of study peers: Coeff.	s.e.	Adjusted p-value
Pre-determined individual characteristics	A. Labor earnings	-1.522*	(0.873)	0.54	-0.0182	(0.0148)	0.91
	B. Employed (%)	-3.474	(2.443)	0.87	-0.0018	(0.0381)	0.99
	C. Years of schooling	-0.099	(0.137)	0.99	0.0006	(0.0023)	0.99
	D. Married	-3.280	(2.596)	0.97	-0.0080	(0.0391)	0.99
	E. Foreign	-0.320	(0.522)	0.99	-0.0016	(0.0077)	0.99
	F. Primary language Finnish	-1.285	(1.130)	0.94	0.0430**	(0.0190)	0.15
	G. Unemployment benefits	0.306**	(0.154)	0.32	-0.0029	(0.0022)	0.90
	H. General housing allowance	0.059	(0.055)	0.95	-0.0006	(0.0008)	0.99
	I. Age	0.250	(0.348)	0.99	-0.0092*	(0.0049)	0.48
	J. Female	-	-	-	0.0000	(0.0003)	0.99
Pre-determined parent characteristics	K. Number of parents employed	-2.583	(2.595)	0.95	-0.0417	(0.0371)	0.94
	L. Parent years of schooling	-0.484	(0.368)	0.90	0.0086*	(0.0047)	0.52
	M. Parent pension income	0.592	(0.871)	0.99	-0.0233	(0.0164)	0.87
	N. Parent unemployment benefits	0.397*	(0.222)	0.53	0.0042	(0.0031)	0.89
	O. Parent general housing allowance	0.161**	(0.075)	0.21	-0.0007	(0.0011)	0.99
	P. Parent income	0.425	(1.772)	0.99	-	-	-

Notes: The baseline estimation sample consists of 51,186 individuals who become entrepreneurs. Each cell presents a coefficient from a separate regression, following Equation (3) and using the pre-determined characteristics indicated by the row label as the outcome. Pre-determined characteristics are measured one year before the first study year. All specifications include program-by-school and school-by-start-year fixed effects. Income, earnings, benefits, allowances, and pensions are in thousands of euros. Columns (3) and (6) report the stepdown p-values (Romano and Wolf, 2005) adjusted for multiple hypothesis testing of 15 coefficients on the two peer characteristics. Standard errors in Columns (2) and (4) are clustered at the school-by-start-year level and are not corrected for multiple hypothesis testing. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C13: Impacts of Study Peers on the Direction of Innovation, Share of Sales across Consumer Groups

	Average parent income of study peers		
Dependent variable	(1)	(2)	(3)
A. Share of sales to rich (above 100k over below 30k)	0.00018*	-0.00005	0.00031**
	(0.00010)	(0.00019)	(0.00014)
B. Share of sales to rich (above 60k over below 60k)	0.00013	-0.00007	0.00024**
	(0.00008)	(0.00014)	(0.00011)
Sample	All	Own parent income below median	Own parent income above median
Students	51,186	23,889	27,297
Study groups	21,009	13,485	14,468
Schools	556	539	526

Notes: The table displays the estimates of the impact of study peers on the dependent variable indicated by the row label. The baseline estimation sample consists of 51,186 individuals who become entrepreneurs. Each cell presents a coefficient from a separate regression, following Equation (2). Outcomes are means from age 28 onward. All specifications include program-by-school and school-by-start-year fixed effects and control for dummies for the year of outcome measurement and pre-determined characteristics listed in footnote 16. Parent income, income, and earnings are in thousands of euros. Standard errors robust for clustering at the school-by-start-year level are in parentheses. $*p < 0.1$, $**\ p < 0.05$, $***\ p < 0.01$.

Table C14: Impacts of Study Peers on Top Incomes

	Fraction female among study peers			Average parent income of study peers		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
A. Income above 50th pc.	-0.0130	0.0311	0.0071	-0.00019	-0.00056	0.00054
	(0.0232)	(0.0458)	(0.0319)	(0.00036)	(0.00061)	(0.00053)
B. Income above 90th pc.	-0.0473**	-0.0258	-0.0592**	0.00021	0.00006	0.00035
	(0.0186)	(0.0301)	(0.0288)	(0.00027)	(0.00042)	(0.00043)
C. Income above 99th pc.	-0.0068	-0.0116	-0.0079	-0.00009	-0.00010	-0.00011
	(0.0094)	(0.0166)	(0.0143)	(0.00016)	(0.00022)	(0.00026)
Sample	All	Women	Men	All	Own parent income below median	Own parent income above median
Students	51,186	20,714	30,472	51,186	23,889	27,297
Study groups	21,009	11,212	13,884	21,009	13,485	14,468
Schools	556	516	518	556	539	526

Notes: The table displays the estimates of the impact of study peers on the dependent variable indicated by the row label. We consider two sets of study peer characteristics: gender (columns (1)-(3)) and parent income (columns (4)-(6)). The baseline estimation sample consists of 51,186 individuals who become entrepreneurs. Each cell presents a coefficient from a separate regression, following Equation (2). Outcomes are measured from age 28 onward. All specifications include program-by-school and school-by-start-year fixed effects and control for dummies for the year of outcome measurement and pre-determined characteristics listed in footnote 16. Income percentiles are calculated from the sample including all students; parent income is measured in thousands of euros. Standard errors are clustered at the school-by-start-year level. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C15: Study Peer Effects vs. Homophily Estimates

	Share of Sales across Consumer Groups
A. Difference b/w top and bottom quintiles of parent income	2.40pp
B. Effect of difference in average study peer parent income b/w top and bottom quintiles of own parent income	0.21pp
⇒ Ratio A/B	8.8%

Notes: This table compares the magnitudes of study peer effects from Table C13 to our homophily estimates from Table C9. Specifically, using the homophily estimates from Table C9, we estimate that the industry sales share to households making above \$100k, relative to those below \$30k, is 2.4pp larger for entrepreneurs from a family in the top 20% of the income distribution, compared with those from the bottom 20%. According to the study peer estimates from Table C13 (Column (1), Row B), the change in average peer parent income across these groups leads to an increased in this sales share of 0.21pp, accounting for 8.8% of the overall difference.

Table C16: Impacts of Study Peers on the the Probability of Becoming an Entrepreneur

	Fraction female among study peers			Average parent income of study peers		
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
Entrepreneur	-0.0172***	-0.0205***	-0.0116	-0.000039	-0.000049	-0.000038
	(0.0053)	(0.0067)	(0.0086)	(0.000072)	(0.000112)	(0.000093)
Sample	All	Women	Men	All	Own parent income below median	Own parent income above median
Observations	602,658	308,376	294,282	602,658	276,395	326,263

Notes: The table displays the estimates of the impact study peers on the probability of becoming an entrepreneur. We consider two sets of study peer characteristics, gender (columns (1)-(3)) and parent income (columns (4)-(6)). The estimation sample consists of all students in the panel. The average probability of becoming an entrepreneur in this sample is 7%. Each cell presents a coefficient from a separate regression, following Equation (2). All specifications include program-by-school and school-by-start-year fixed effects and control for dummies for the year of outcome measurement and pre-determined characteristics listed in footnote 16. Standard errors robust for clustering by school and program start year are in parentheses. $^{*}p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table C17: Additional Estimates for the Study Peer Design

	Fraction female study peers Share of sales to women			Average parent income of study peers Industry income elasticity		
Specification:	(1)	(2)	(3)	(4)	(5)	(6)
A. Baseline	0.0013	-0.0080	0.0187**	0.00080**	-0.00020	0.00141**
	(0.0056)	(0.0110)	(0.0082)	(0.00040)	(0.00071)	(0.00055)
N	51,186	20,714	30,472	51,186	23,889	27,297
B. No additional controls	0.0014	-0.0074	0.0187**	0.00074*	-0.00010	0.00136**
	(0.0056)	(0.0109)	(0.0082)	(0.00040)	(0.00071)	(0.00055)
N	51,186	20,714	30,472	51,186	23,889	27,297
C. Weighted	0.0034	0.0009	0.01543*	0.00095**	-0.000142	0.00195***
	(0.0057)	(0.0104)	(0.0082)	(0.00045)	(0.00075)	(0.00061)
N	51,186	20,714	30,472	51,186	23,889	27,297
D. Peer group size $\leq$ 25	0.01231	0.0187	0.0283*	0.0010*	-0.00011	0.00197**
	(0.0084)	(0.0196)	(0.0155)	(0.0005)	(0.00115)	(0.00081)
N	16,974	7,473	9,501	16,974	7,939	9,035
Sample	All	Women	Men	All	Own parent income below median	Own parent income above median

Notes: The table displays the estimates of the impact study peers on the dependent variable indicated by the column panel title. We consider two sets of study peer characteristics, gender (columns (1)-(3)) and parent income (columns (4)-(6)). The baseline estimation sample consists of 51,186 individuals who become entrepreneurs. Each cell presents a coefficient from a separate regression, following Equation (2). Outcomes are means observed at age 28 onward. All specifications include program-by-school and school-by-start-year fixed effects and control for dummies for the year of outcome measurement and pre-determined characteristics listed in footnote 16. Row panel C weights regressions by the inverse of the number of observations available for each entrepreneur in the panel from age 28 onwards. Parent income, income, and earnings are in thousands of euros. Standard errors robust for clustering at the school-by-start-year level are in parentheses. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C18: Consumption Dissimilarity Indices

	Consumer Expenditure Survey	Consumer Packaged Goods	Phone Applications
	(1)	(2)	(3)
Male vs. female, single-gender households	0.14	0.320	0.242
Male vs. female, reference person gender	0.067	0.247	
Top vs. bottom income quintiles	0.34	0.335	0.354
Top vs. bottom income deciles	0.38	0.472	0.441

Notes: The table displays the estimates of dissimilarity indices between male and female consumers, as well as across the income distribution, for three datasets. For the consumer expenditure survey and the Nielsen dataset covering consumer packaged goods, we compute the dissimilarity index between the two consumer group using expenditure shares and the most detailed products available in each dataset. For gender, we conduct the analysis either by focusing on single-gender households or by using the reference person gender for all households. For phone applications, we compute the dissimilarity index by using time shares, rather than expenditures. Since the use of phone applications is directly observed at the individual level, we do not draw a distinction between single-gender households and reference person gender in Column (3). Column (1) only takes into account differences in expenditure shares that arise between product categories, while Columns (2) and (3) account for the differences within categories. The between-category estimate in Column (1) ignores some of the relevant variation, while the within-category dissimilarity indices in Columns (2) and (3) may not be representative of other categories. For the analysis by gender, we take 0.24 as our baseline dissimilarity index, which we view as conservative since the dissimilarity index is much higher in the first row of Column (2). For income groups, the dissimilarity indices are very similar for the top and bottom income quintiles; we take 0.34 as our baseline value.

Table C19: Features of the Baseline and Counterfactual Economies

	Baseline			Full Exposure Counterfactual			Top 1% Exposure Counterfactual		
	All	Sector 1	Sector 2	All	Sector 1	Sector 2	All	Sector 1	Sector 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fraction of innovators									
All	0.059	0.028	0.031	0.081	0.040	0.041	0.053	0.032	0.021
Men	0.102	0.054	0.048	0.080	0.058	0.022	0.087	0.059	0.028
Women	0.016	0.002	0.014	0.082	0.022	0.060	0.019	0.005	0.015
Average productivity of innovators									
All	0.329	0.470	0.202	0.404	0.426	0.384	0.523	0.488	0.574
Men	0.345	0.474	0.201	0.424	0.446	0.366	0.396	0.439	0.305
Women	0.227	0.374	0.205	0.385	0.371	0.390	1.100	1.106	1.098

Notes: The table displays the fraction of innovators and their average productivity in different scenarios. Columns (1) to (3) report the results for the baseline equilibrium. Columns (4) to (6) focus on the "full exposure" counterfactual, setting $\tau_i = 1$ for all women. Columns (7) to (9) describe the "top 1% exposure" scenario, with $\tau_i = 1$ for all women in the top 1% of the innovation productivity distribution.

Table C20: Baseline Parameters for the Analysis by Income Quintiles

Panel A: Parameters calibrated outside of the model

Model Parameter	Parameter Definition	Source	Value
$\lvert\alpha_{Q1} - \alpha_{Q5}\rvert$	Expenditure dissimilarity index by income quintiles	Consumer Expenditure Survey, Nielsen data, phone applications data (cf. Appendix Table C18, third row)	0.34
ε	Elasticity of substitution between varieties	DellaVigna and Gentzkow (2019)	1.9
ρ	Discount rate, annual	Kaplan et al. (2018)	0.051
λ	Pareto parameter of innovators' productivity	Bell et al. (2019a)	1.26
$\frac{1-\delta}{\delta}$	Ratio of disposable income (after taxes and transfers) between top and bottom income quintiles	Congressional Budget Office	6.46

Panel B: Jointly estimated model parameters

Model Parameter	Parameter Definition	Value
τ	Exposure to innovation careers	0.121
ϕ	Sectoral exposure	0.596
$\bar{\eta}$	Pareto scale parameter of of innovators' productivity	0.011

Panel C: Targeted moments and model fit

Targeted Moment [Source]	Data	Model
Share of patent inventors in bottom vs. top parent income quintiles [Bell et al. (2019b)]	0.116	0.116
Income homophily regression coefficients [Table C9, column (1) and table notes]	0.238	0.238
Annual growth rate of labor productivity, 1990-2020 [Saint Louis Fed]	0.02	0.02

Notes: This table presents the baseline parameters of the growth model for the analysis by income quintiles. In Panel A, the model parameters are set directly to match the value observed in data or taken from the literature. In Panel B, the three parameters are estimated jointly to match the moments from the model with moments observed in the data, displayed in Panel C.

Table C21: Clean Patents and Inventor Gender

	Clean Patent	
	(1)	(2)
Female Inventor Fraction	0.326***	
	(0.015)	
Any Female Inventor		0.212***
		(0.007)
N	1403	1401

Notes: This table reports the propensity to create "clean patents" by inventor gender, following Aghion et al. (2016) to classify clean vs. other energy patents. The level of observation is a patent, in the sample of energy patents. *$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.