

Social Exposure, Innovator-Consumer Homophily, and Inequality: Evidence from College Peers

October 2025

Abstract

We investigate the importance of social networks for the direction of entrepreneurship and innovation. Using quasi-experimental variation in the gender and socioeconomic composition among college peers in Finland, we show that exposure to lower-income peers increases the likelihood that an entrepreneur founds a firm in a necessities industry without affecting entrepreneurial income. Likewise, increased exposure to female peers increases entrepreneurial activities targeting female consumers. These effects are largest among groups over-represented in innovation: men and individuals from high-income backgrounds. We assess the macro implications of this heterogeneity in an endogenous growth model and find that differences in college peer composition explain around a quarter of the observed inventor-consumer homophily by gender, inducing a significant cost-of-living disadvantage for women. These findings show that innovators' social experiences have a causal impact on the direction of innovation, independent of financial incentives.

Economic theories of entrepreneurship emphasize personal abilities, skills, and financial incentives as the main drivers of entrepreneurial activities (Lucas, 1978; Jovanovic, 1982; Holmes and Schmitz, 1990; Acemoglu, 2002). However, recent work has documented that entrepreneurs and inventors have a strong tendency to focus on consumers from similar socio-demographic backgrounds as themselves (Koning et al., 2021; Truffa and Wong, 2025; Einiö et al., 2023). Moreover, certain groups are strongly underrepresented in the innovation system, such as women and individuals from less advantaged backgrounds (e.g., Bell et al., 2019). The equilibrium implications of these two patterns, particularly for cost-of-living inequality, depend on the extent to which innovators are responding to financial and non-financial incentives when deciding on the direction of their activities.¹

In this article, we document one potentially important non-financial channel for determining an innovators' direction of innovation: social exposure. Our analysis proceeds in two steps. First, we provide direct causal evidence that innovators' social experiences have a causal impact on the direction of innovation. Using a quasi-experimental study-peer design in Finnish colleges, we examine whether variation in the gender and socioeconomic composition of an individual's peer group has an impact on the direction of entrepreneurship. 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 activities targeting female consumers. These results provide direct evidence that social factors independent of financial incentives affect

¹As long as individuals are purely responding to financial incentives, then there will be minimal cost-of-living inequality in equilibrium, because innovators from the overrepresented group (e.g., men) will also target consumers from groups underrepresented in innovation (e.g., women). However, if non-financial incentives to pursue opportunities benefiting consumers from similar backgrounds are large, then, in the resulting equilibrium, groups underrepresented in innovation benefit less from innovation and have to pay more to attain a given level of utility compared to overrepresented groups.

the direction of innovation.

Second, we apply the causal estimates to the model in Einiö et al. (2023) to assess the degree of economic inequality induced by variation in peer exposure. We first calculate the level of homophily one would expect to see due to the identified peer effects and observed differences in peer composition between people with different gender or different family backgrounds. This exercise can be thought of as quantifying the consequences of the fact that people with different demographic or socio-economic characteristics tend to have systematically different peers. Our causal estimates suggest that differences in the average fraction of female peers between male and female college students can explain around 26.6% of the overall innovator-consumer homophily. Reducing homophily by the same amount – corresponding to a counterfactual where peer composition across groups is equalized – cost-of-living inequality, arising from a disproportionate share of business activities targeting male consumers, declines by around 6.5 pp. as higher exposure to female peers leads the majority male inventor group to invent more for female consumers.

This article primarily contributes to the literature on the social mechanisms underlying inventive and entrepreneurial activities. Research shows that the decision to found a business is affected by social exposure to entrepreneurship (e.g., Nanda and Sørensen, 2010; Lerner and Malmendier, 2013; Guiso et al., 2021) and by entrepreneurial contacts, with substantial homophily within networks by industry, gender, and socio-economic status (Fluegge and Bailey, 2024). Howell and Nanda (2023) show that male startup entrepreneurs benefit from exposure to male venture capitalists, whereas female startup entrepreneurs do not. Truffa and Wong (2025) show that exposure to female study peers increases the likelihood that individuals publish scientific research focused on female subjects and gender differences. Our key contribution to this literature is to show how peers' gender and socio-economic background

affect the choice of customer segments individuals choose to serve as entrepreneurs.

Our work also adds to the literature on the direction of innovation. Much of the existing economics literature has focused on market size effects (Schmookler, 1966; Acemoglu, 2002, 2007), likely due to the measurability and availability of demand shocks. We contribute to this literature by using administrative data to isolate variation in social peers and then track long-run impacts of social exposure on the direction chosen by innovators. One advantage of our setting is that we are able to measure both the direction of innovation and entrepreneurial earnings, which helps us disentangle the importance of financial versus non-financial incentives.

Finally, we contribute to the macroeconomics literature on growth, inequality, and interactions. Our analysis makes use of the canonical endogenous growth frameworks (e.g., Romer, 1990), and also addresses the issue of cost-of-living inequality studied in Foellmi and Zweimüller (2006) and Einiö et al. (2023). Our work is also related to Lucas and Moll (2014) and Akcigit et al. (2018), which study the role played by interactions between innovators in determining growth. We contribute to this literature in two ways. First, we focus on analyzing the role played by interactions between innovators and social peers. Second, our causal estimates allow us to quantify, within an endogenous growth model framework, the macro-economic implications of the fact that individuals from different socio-demographic backgrounds are exposed to different social peer groups.

The paper is organized as follows. We present the research design in Section II, the falsification tests in Section III, the main results in Section IV, and robustness checks in Section V.

I Setting and Data

We estimate social exposure effects among vocational school and university students in Finland. We refer to them as college students. 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. We draw data covering all college students in Finland from the student register maintained by Statistics Finland. Our sample covers the 1999–2013 period and includes individual-level information on the educational institution (also referred to as school), study program, and program start year for college students.² We identify college peers as individuals observed in the same freshman class, which comprise students who start to study in the same study program and school in the same year. Study programs are based on 6-digit education codes, which define the level and field of study at a detailed level. For individuals who are observed to start in several programs, we define peers as co-students in the last freshman class the student enters. However, we use all co-students in the freshman class observed in the full student population data when constructing measures of peer composition in order not to induce measurement error in them.

We link to students in the student register the Finnish employer-employee population panel (FOLK) compiled by Statistics Finland. The FOLK is based on administrative registers and provides individual-level information on income, occupation, entrepreneurship, and industry of employment. The data set also includes information on family links, which allows merging parent income to children.

A key variable in FOLK is entrepreneurship status, which is based on pension contribution and income tax records. We use the status for the last week of the year which allows for

²Our 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.

temporal consistency across variables.³ An individual is defined as an entrepreneur if she/he has received only entrepreneurial income, and no employee salary income, during the year and is associated with a private business in the entrepreneur pension insurance system in the last week of the year. She/he is also identified as an entrepreneur in the last week of the year if she/he has made entrepreneur pension contributions in that week. If an individual has both entrepreneur and employee pension contributions in the last week of the year, she/he will be defined as an entrepreneur if the entrepreneurial income associated with the contributions is larger.

A second key variable is the unique company identifier that provides information on the company a worker or entrepreneur worked in. This information 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.

We link industry income elasticities and female consumption shares to the population panel by the industry code of the company an individual is associated with in the last week of the year. The match rate for the sample of individuals with industry codes in the population panel is 80%.⁴ Our final linked sample includes 602,658 individuals who start studies in college in the years 1999 to 2013, whose long-term outcomes are observed (from age 28 onwards), and for whom parental income is observed. Of these individuals 51,186 become an entrepreneur.

Summary statistics for key variables are displayed in Table 1. There is significant variation in the industry-level outcome measures, entrepreneurial earnings, and peer group measures within the entrepreneurs sample. About 30 percent of entrepreneurs enter luxury goods

³The data includes also codes for the company which is associated with the longest employment spell during the year.

⁴Industry codes are based on employment or entrepreneurship and hence not available for unemployed individuals. In our extensive margin analysis, we use the full data including also unemployed individuals.

industries and about 40 percent enter industries with sales share to women greater than 0.5.

Table 1: Summary statistics

	Entrepreneurs		All	
	Mean	SD	Mean	SD
A. Outcomes				
Entrepreneur	1	0	0.085	0.278
Industry expenditure elasticity	1.158	0.520		
Industry sales share to women	0.635	0.103		
Entrepreneur in industry with expenditure elasticity over 1			0.026	0.159
Entrepreneur in industry with sales share to women over 0.5			0.035	0.183
Income	27.32	19.62	27.98	15.42
Share of sales to rich (above 100k over below 30k)	0.689	0.137		
Share of sales to rich (above 60k over below 60k)	0.618	0.103		
B. Own and peer group characteristics				
Female	0.40	0.49	0.51	0.50
Fraction female peer	0.48	0.34	0.52	0.33
Parental income	49.4	29.2	53.9	29.61
Peer parental income	51.1	12.3	55.1	12.83
Observations	51,186		602,658	

Notes: The table displays summary statistics for the baseline estimation sample consisting of 51,186 individuals who become entrepreneurs (Columns 1 and 2) and for the full student sample (Columns 3 and 4). In Panel A, outcome variables are means from age 28 onward, except the binary indicator for being an entrepreneur.

II Research Design

Following Hoxby (2000) and the subsequent literature, we exploit idiosyncratic variation in the composition of peers across classes within the same program and school to assess whether peers have a causal effect on the direction of innovation.

We examine in turn the effect of: (i) peer gender; and (ii) peer parent income composition on the direction of innovation. We estimate both extensive and intensive margin effects. At the extensive margin, we estimate the impacts of social exposure on the likelihood of becoming an entrepreneur. Intensive margin effects are estimated for the sample of students who

become entrepreneurs. In addition, we examine entrepreneurs' income levels and variance to further assess the role of financial incentives and entrepreneurial risk. We can thus investigate both the extent to which the innovator-consumer homophily arises from differences in social exposure and the importance of financial incentives.

Our baseline model is a standard linear-in-means peer regression for individual i who starts in a study 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} :

$$Y_i = \beta \bar{X}_{(i)jks} + \gamma_1 X_i + \gamma_2 W_i + \alpha_{jk} + \lambda_{ks} + \varepsilon_{ijks}. \quad (1)$$

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.⁵ The main regressor of interest is the peer mean, $\bar{X}_{(i)jks}$, 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, $\bar{X}_{(i)jks}$ is the leave-one-out fraction of female peers or the leave-one-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).⁶ 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 (excluding X_i), 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,

⁵We focus on outcomes from age 28 onwards, and average them across years for entrepreneurs who are active over several years.

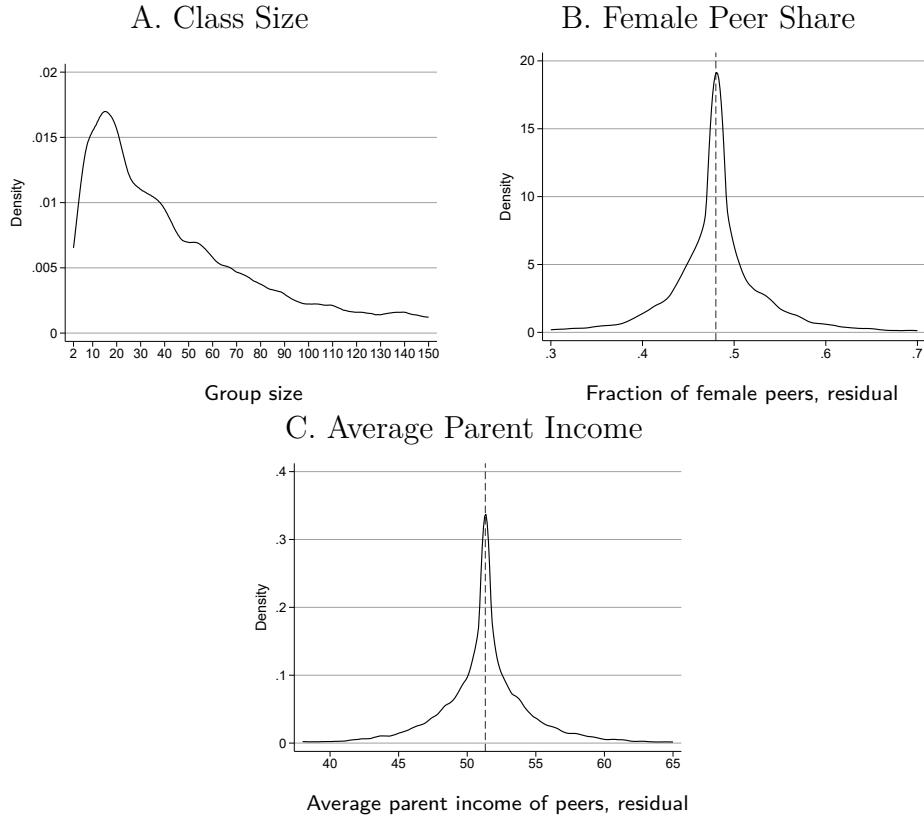
⁶Controlling for own characteristic is a standard practice to eliminate the mechanical correlation between this variable and the peer mean, which may arise in a peer regression where an individual is allowed to be both the subject of peer effects and the peer (Angrist, 2014).

the level of variation in our design.

The parameter of interest in Equation (1) is β , the coefficient on the average characteristic of college peers. Conditioning on school-by-program fixed effects means that the peer effect is identified from idiosyncratic variation in peer composition across classes within the same program and school.⁷ 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).

⁷We note that the estimation of this model is facilitated by our panel data that includes several starting classes by school-by-program cell.

Figure 1: Class Size Distribution and Identifying Variation in Peer Attributes



Notes: Panel A depicts the variation in class size in our baseline sample, using data on 21,009 freshman classes in which at least one individual becomes an entrepreneur. Classes 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 classes 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 8, 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.

The key identifying assumption is that, while there can be selection into schools and programs, variation in peer gender or parent income composition across classes in the same study program and school is uncorrelated with other determinants of the direction of entrepreneurship. 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 an individual's subsequent 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.

Our sample includes 556 schools and 21,009 peer groups. Panel A in Figure 1 displays the size distribution of peer 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 1 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 classes 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 ($1/15 \approx 0.07$).

III Falsification tests

To examine the plausibility of the identification assumption underlying Equation (1), we run an analogous specification to assess whether peer characteristics are predictive of pre-

determined characteristics of individual i :

$$\tilde{Y}_i = \tilde{\beta} \bar{X}_{(i)jks} + \tilde{\gamma}_1 X_i + \tilde{\alpha}_{jk} + \tilde{\lambda}_{ks} + \tilde{\varepsilon}_{ijks}. \quad (2)$$

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. As the outcome, we use predicted outcome based on the control variables W_i in Equation 1 (excluding X_i from the prediction model). The advantage of this approach is that it assigns larger weights to the characteristics that best predict future outcomes. Moreover, it provides a single test for an outcome and the coefficient can be compared to the main estimate graphically (see e.g., Carrell et al., 2018).⁸ For completeness, we also provide results for each predetermined outcome separately.

Table 2 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 classes 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.⁹ In Panels A and B, all coefficients for both measures of predicted direction are close to zero and statistically insignificant. This

⁸The 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. 2) 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.

⁹Coefficients on the set of pre-determined characteristics in the prediction models are presented in Appendix Table A1.

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. For completeness, Appendix Table A2 reports results separately for each predetermined characteristic and both peer variables (30 separate regressions). After adjusting p-values for multiple hypotheses testing, 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.

Table 2: 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 (2) 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 8. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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.

IV Results

Main estimates. We start our discussion of the results with the extensive margin effects on the direction of innovation in Table 3, estimating peer effects on the likelihood of entering any sector, female-intensive sector (female share above 0.5), or high-income-intensive sector (income elasticity above 1).¹⁰

We find that peer gender composition has minimal impacts on the likelihood of becoming an entrepreneur in any sector or in a female-intensive sector for male students. However, we do find that exposure to female peers reduces the likelihood of catering to female-intensive sectors among female students. This is because exposure to female peers has relatively strong negative effect on the likelihood of becoming an entrepreneur and female entrepreneurs are more likely to cater to female-intensive sectors at the baseline. Hence, the negative effect on entrepreneurship among female students disproportionately reduces the fraction of entrepreneurs in female-intensive sectors among female students. Finally, we find no extensive margin effects of peer parent income on the likelihood of becoming an entrepreneur or catering to higher-income consumers.

¹⁰We note that since the consumer type is only observed conditional on becoming an entrepreneur (e.g., the share of sales to women), we cannot run regressions on continuous consumer type variables in the full sample, as the outcome is missing for all individuals who do not become entrepreneurs.

Table 3: Study Peers, Entrepreneurship, and the Direction of Innovation, Extensive Margin Effects

Dependent variable	Fraction female among study peers			Average parent income of study peers		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Entrepreneur	-0.0172*** (0.0053)	-0.0205*** (0.0067)	-0.0116 (0.0086)	-0.000039 (0.000072)	-0.000049 (0.000112)	-0.000038 (0.000093)
Dependent mean	0.085	0.069	0.101	0.085	0.092	0.078
B. In industry with sales share to women above 0.5	-0.0101*** (0.0034)	-0.0087** (0.0041)	-0.0056 (0.0056)			
Dependent mean	0.035	0.026	0.043			
C. In industry with income elasticity above 1				0.000063 (0.000043)	0.000086 (0.000066)	0.000040 (0.000056)
Dependent mean				0.026	0.028	0.024
Sample	All	Women	Men	All	Own parent income below median	Own parent income above median
Students	602658	308376	294282	602658	276395	326263
Study groups	46999	35680	36261	46999	42543	40637
Schools	585	565	569	585	582	576

Notes: The table displays the estimates of the impact of study peers on the dependent variable, indicated by the row label, for the full sample of students, not conditioning on entrepreneurship. 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 (1). 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 8. 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 585 schools and covers 15 start years.
 $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

Next, we 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 4 report the estimated effects 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.¹¹ 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.¹²

Columns (4) to (6) in Row B of Table 4 report the estimated impacts 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 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 backgrounds. The point estimate is not statistically significant when considering entrepreneurs with parent income below median in Column (5). The point estimate for entrepreneurs with parent

¹¹We base our assessment of magnitudes on sample standard deviation rather than residual standard deviation reported for Panel B of Figure 1, 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).

¹²The 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 4.

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 A3. We find no peer effects on the variance of entrepreneur's income, suggesting that exposure to female or high-income peers does not lead to systematically more or less risky entrepreneurial careers.

As an alternative strategy, we estimate peer effect for a sample including patent inventors and entrepreneurs in 4-digit industries that generate patents.¹³ We find very similar results as for the entrepreneur sample. We interpret this as suggestive evidence that the estimated peer effects are relevant for innovative firms.

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.¹⁴ 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

¹³The sample of students who become patent inventors is too small to be used alone for peer effect analysis (includes 443 individuals).

¹⁴Note 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.

low-income backgrounds.¹⁵

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 find untapped market opportunities, and therefore earn higher entrepreneurial income. Alternatively, it could be that exposure to peers shifts an entrepreneur's intrinsic motivation or ideas to target specific consumers, independently of financial incentives. If the intrinsic motivation is large enough, it could even lead to lower entrepreneurial income.

To assess the role of financial incentives, in row C of Table 4, 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 A5, 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.

¹⁵A 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 4: Impacts of Study Peers on the Direction of Innovation and Entrepreneur's Income, Intensive Margin Effects

Dependent variable	Fraction female among study peers			Average parent income of study peers		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Sales share to women	0.0013 (0.0056)	-0.0080 (0.0110)	0.0187** (0.0082)			
B. Industry income elasticity				0.00080** (0.00040)	-0.00020 (0.00071)	0.00141** (0.00055)
C. Income	-2.165* (1.117)	-0.735 (1.847)	-2.873* (1.710)	0.0016 (0.017)	-0.0171 (0.0277)	0.0198 (0.0284)
D. Income variance	-41.45* (23.54)	-31.88 (34.97)	-52.55 (38.14)	0.504 (0.404)	0.204 (0.700)	0.481 (0.665)
E. Industry sales	-0.037 (0.076)	-0.030 (0.131)	-0.113 (0.113)	0.00115 (0.00114)	0.00016 (0.00186)	0.00143 (0.00183)
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 (1). 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 8. All control variables are measured one year before entering the study program. Industry sales are in logs and measured by 3-digit industry. Standard errors clustered at the school-by-start-year level are in parenthesis. The sample includes 556 schools and covers 15 start years. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

For peer parent income, there is no evidence of any response of income. In columns (4) to (6) of row C in Table 4, exposure to peers from different parts of the parent income distribution leads to no significant impact on average income. Appendix Table A5 confirms this finding, with no statistically significant responses across the income distribution. We also find no evidence of peer composition affecting the size of the market an entrepreneur caters, which we measure as log sales by 3-digit industry (Row E of Table 4).

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 4 to the descriptive homophily estimates from Einiö et al. (2023). The comparisons are reported in Table 5, 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 Einiö et al. (2023). 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 0.80,¹⁶ 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.

Table 5: Study Peer Effects vs. Homophily Estimates

	Share of sales to women (1)	Industry income elasticity (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 4 to our homophily estimates from the main text. The calculation steps are provided in the text.

¹⁶We multiply the female study peer coefficient for male students, equal to 1.87pp according to Row A of Column (3) of Table 4, 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.

Column (2) of Table 5 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%.¹⁷ 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 0.0091,¹⁸ i.e., 10% ($=0.0091/0.091$) of the overall difference. This sizable effect is plausible: while college peers constitute a subset of all social interactions of an individual, they can be expected to be particularly important for the direction of innovation.¹⁹

Macro implications. To assess the macro implications of our results, we employ the endogenous growth model in Einiö et al. (2023) allowing for unequal access to innovation, heterogeneity in tastes, and differences in the direction of innovation stemming from social factors rather than financial incentives. Appendix B summarizes the model and describes the counterfactual in greater detail. We study in turn a “no homophily” scenario 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.

Figure 2 reports the results for both counterfactuals, compared to the baseline calibrated economy. Effects on cost-of-living inequality are large, with limited impacts on growth.

¹⁷We use the homophily regression coefficient of 0.1416 from Einiö et al. (2023). 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 $0.14 \times \log(108,680/24,325) = 0.091$.

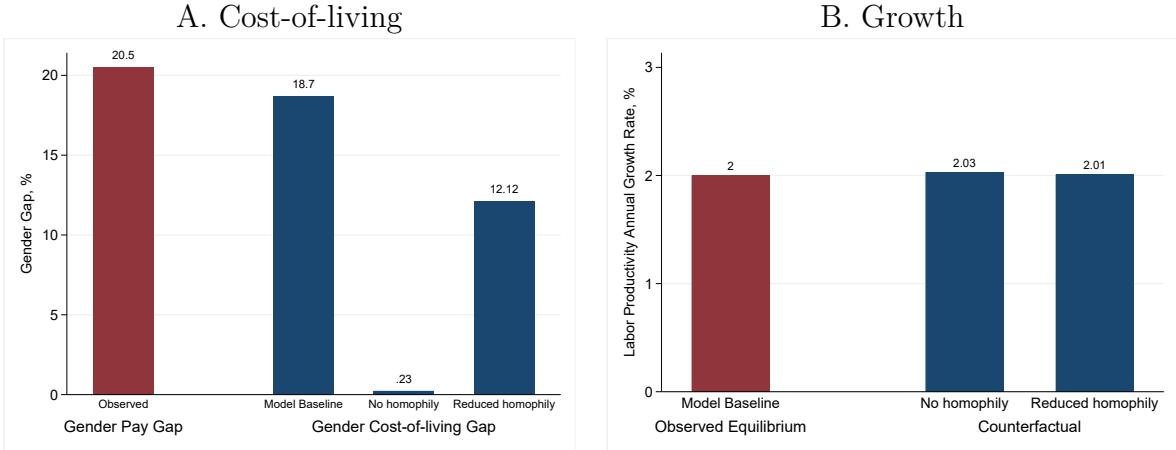
¹⁸We multiply the peer effect coefficient capturing the causal effect of the parent income of study peers (in thousands of euros) on the industry income elasticity, equal to 0.00080 in Row B of Column (4) of Table 4, by the difference in the average peer parent income for entrepreneurs across the parent 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.

¹⁹We obtain similar results when studying the share of sales to high- vs. low-income consumer groups (Table A6).

Panel A reports significant reductions in cost-of-living inequality, because more male innovators end up innovating in the female-preferred sector. We find that the gender cost-of-living gap falls by 18.47pp when homophily is removed completely. By reducing gender homophily by 26.6%, i.e., by the fraction of homophily implied by the difference in the average fraction of female study peers between male and female students and our causal estimate, we obtain a meaningful fall in inequality (-6.58pp). 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.

Panel B shows that neither counterfactual change affects growth rates very much, because very few productive inventors enter as a result.

Figure 2: Main Counterfactual Estimates



Notes: this figure 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 5).

V Robustness

We now discuss several robustness checks. First, Figure A1 presents binned scatter plots of the main estimates from Tables 2 and 4. The figure shows that the falsification tests

and main results are not driven by outliers and are appropriately summarized by linear regressions. Second, Table A7 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 4 without controls, as well as when we use weighted regressions and study groups with 25 students or less.

VI Conclusion

In this article, we have shown that social peers have a quantitatively significant impact on an individual’s subsequent direction of innovation. Using administrative data from Finland, we are able to isolate quasi-random variation in the gender and socioeconomic composition of college peers. We then show that exposure to high-SES or female peers increases the likelihood that an individual starts a business in an industry that serves consumers from that background.

The key implication of our work is that factors beyond financial incentives can play a significant role in determining the types of innovation an individual pursues. This result is consistent with the view that innovators use ideas and preferences that are shared by their social peers as inputs to their work. Because innovators have significantly different backgrounds to those of consumers, on average, our results indicate that the current innovation system generates significant cost-of-living inequality – too little innovation and missing goods for consumer groups underrepresented among innovators.

Our results also raise several additional issues that could be addressed in future work. First, we find evidence of asymmetric peer effects on the direction of innovation, which suggests some variation in the baseline exposure to specific groups (e.g., due to media). Second, we find that entrepreneurs end up earning similar income when they start a business

in different industry due to exposure to peers from different backgrounds. This likely reflects the importance of general entrepreneurial ability rather than sector-specific ability. Finally, our work also speaks to the management of innovation. Some firms create interactions between their R&D personnel and target customers. Our results suggest that such practices could influence the types of new goods created by those firms.

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.
- 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**, “Who becomes an inventor in America? The importance of exposure to innovation,” *The Quarterly Journal of Economics*, 2019, 134 (2), 647–713.
- Carrell, Scott E., Mark Hoekstra, and Elira Kuka**, “The long-run effects of disruptive peers,” *American Economic Review*, 2018, 108 (11), 3377–3415.
- 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.
- Einiö, Elias, Josh Feng, and Xavier Jaravel**, “Social push and the direction of innovation,” *SSRN Working Paper 3383703*, 2023.
- Fluegge, Robert B. and Michael Bailey**, “Where Do Friends Matter for Founders? The Role of Social Connections in U.S. Entrepreneurship,” 2024. Working Paper.
- 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.
- Guiso, Luigi, Luigi Pistaferri, and Fabiano Schivardi**, “Learning Entrepreneurship from Other Entrepreneurs?,” *Journal of Labor Economics*, 2021, 39 (1), 135–191.
- 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.
- Holmes, Thomas J. and James A. Schmitz**, “A Theory of Entrepreneurship and Its Application to the Study of Business Transfers,” *Journal of Political Economy*, 1990, 98 (2), 265–294.
- 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.

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

Jovanovic, Boyan, “Selection and the Evolution of Industry,” *Econometrica*, 1982, 50 (3), 649–670.

Koning, Rembrand, Sampsia 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.

Lerner, Josh and Ulrike Malmendier, “With a Little Help from My (Random) Friends: Success and Failure in Post-Business School Entrepreneurship,” *Review of Financial Studies*, 2013, 26 (10), 2411–2452.

Lucas, Robert E., “On the Size Distribution of Business Firms,” *Bell Journal of Economics*, 1978, 9 (2), 508–523.

— and **Benjamin Moll**, “Knowledge growth and the allocation of time,” *Journal of Political Economy*, 2014, 122 (1), 1–51.

Nanda, Ramana and Jesper B. Sørensen, “Workplace Peers and Entrepreneurship,” *Management Science*, 2010, 56 (7), 1116–1126.

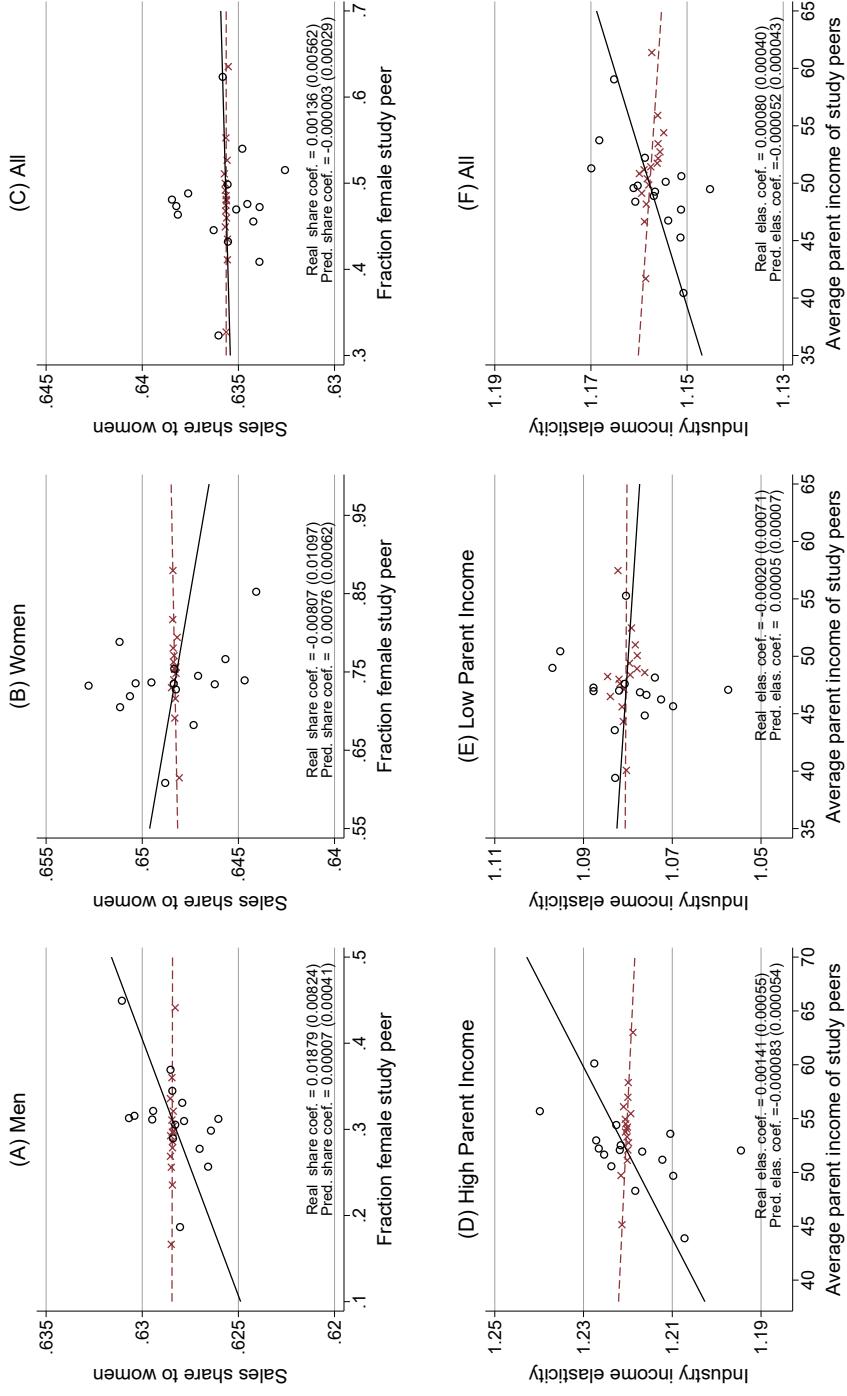
Romano, Joseph P. and Michael Wolf, “Stepwise multiple testing as formalized data snooping,” *Econometrica*, 2005, 73 (4), 1237–1282.

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.

Truffa, Francesca and Ashley Wong, “Undergraduate Gender Diversity and the Direction of Scientific Research,” *American Economic Review*, July 2025, 115 (7), 2414–48.

Figure A1: 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 8. 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.

Table A1: Coefficients for Predicted Outcomes

Outcome:	For peer parent income regressions (exclude own parent income)			For peer gender composition regressions (exclude female dummy)		
	Income Elasticity	Female Share	Income	Income Elasticity	Female Share	Income
Female (%)	-0.040991*** (0.005602)	0.010954*** (0.001200)	-5.686135*** (0.233779)			
Own parent income				0.000800*** (0.000091)	-0.000058*** (0.000020)	0.042400*** (0.003832)
Labor earnings	0.001019*** (0.000208)	0.000118*** (0.000045)	0.254387*** (0.008697)	0.001131*** (0.000207)	0.000082* (0.000045)	0.272559*** (0.008710)
Years of schooling	-0.003687*** (0.001254)	-0.000981*** (0.000269)	0.408430*** (0.052348)	-0.004486*** (0.001249)	-0.000766*** (0.000268)	0.296639*** (0.052442)
Employed (%)	0.000054 (0.000058)	-0.000020 (0.000012)	-0.003763 (0.002405)	0.000033 (0.000058)	-0.000016 (0.000012)	-0.006009** (0.002418)
Married (%)	-0.000037 (0.000064)	-0.000013 (0.000014)	-0.023959*** (0.002683)	-0.000014 (0.000064)	-0.000019 (0.000014)	-0.020810*** (0.002695)
Foreign (%)	0.000243 (0.000254)	0.000004 (0.000054)	-0.017000 (0.010613)	0.000329 (0.000254)	-0.000006 (0.000055)	-0.010791 (0.010678)
Primary language Finnish (%)	0.000284** (0.000130)	0.000024 (0.000028)	0.013643** (0.005407)	0.000259** (0.000130)	0.000027 (0.000028)	0.011715** (0.005438)
Unemployment benefits	0.004567*** (0.001010)	-0.000177 (0.000216)	-0.180990*** (0.042157)	0.004663*** (0.001010)	-0.000193 (0.000217)	-0.171980*** (0.042398)
General housing allowance	0.012350*** (0.002323)	0.001155** (0.000498)	0.114008 (0.096967)	0.011893*** (0.002322)	0.001256** (0.000498)	0.059757 (0.097504)
Parental years of education	0.001534*** (0.000406)	-0.000179** (0.000087)	0.039674** (0.016943)	-0.000254 (0.000460)	-0.000072 (0.000099)	-0.045258** (0.019312)
Parental employed (%)	0.000642*** (0.000063)	-0.000001 (0.000013)	0.005131** (0.002613)	0.000536*** (0.000064)	0.000008 (0.000014)	-0.001021 (0.002671)
Parental pension income	0.000569*** (0.000216)	-0.000054 (0.000046)	0.002724 (0.009005)	0.000494** (0.000216)	-0.000048 (0.000046)	-0.001585 (0.009063)
Parental unempl. benefits	0.003810*** (0.000581)	0.000304** (0.000124)	-0.027671 (0.024232)	0.004349*** (0.000585)	0.000283** (0.000125)	-0.007219 (0.024553)
Parental housing allowance	0.002880* (0.001571)	-0.000148 (0.000337)	0.258390*** (0.065567)	0.004536*** (0.001581)	-0.000277 (0.000339)	0.349985*** (0.066392)

Continues on next page

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

Table A1: Coefficients for Predicted Outcomes, continued

Outcome:	For peer parent income regressions (exclude own parent income)			For peer gender composition regressions (exclude female dummy)		
	Income Elasticity	Female Share	Income	Income Elasticity	Female Share	Income
<i>Coefficients on age dummies:</i>						
16	-0.003175 (0.013688)	0.003947 (0.002933)	2.441475*** (0.571246)	0.000341 (0.013673)	0.002920 (0.002933)	2.967502*** (0.574078)
17	-0.014154 (0.017057)	0.004100 (0.003654)	1.907683*** (0.711835)	-0.010422 (0.017042)	0.003030 (0.003656)	2.457172*** (0.715534)
18	-0.021610 (0.019586)	0.001679 (0.004196)	1.451423* (0.817407)	-0.018421 (0.019579)	0.001030 (0.004200)	1.804117** (0.822020)
19	-0.031404*** (0.011029)	-0.002877 (0.002363)	1.364798*** (0.460269)	-0.033940*** (0.011019)	-0.002159 (0.002364)	0.995045** (0.462643)
20	<i>Reference category</i>					
21	0.003135 (0.010081)	-0.003617* (0.002160)	-0.455956 (0.420724)	0.005478 (0.010066)	-0.004541** (0.002159)	0.000407 (0.422617)
22	0.008678 (0.011081)	0.001509 (0.002374)	-1.904537*** (0.462446)	0.010718 (0.011070)	0.000757 (0.002375)	-1.530372*** (0.464786)
23	-0.003421 (0.011918)	-0.004829* (0.002553)	-1.776738*** (0.497381)	-0.002569 (0.011912)	-0.005315** (0.002555)	-1.544878*** (0.500119)
24	0.005256 (0.012464)	-0.000385 (0.002670)	-2.145816*** (0.520154)	0.006365 (0.012457)	-0.000912 (0.002672)	-1.890682*** (0.523004)
25	-0.002326 (0.012998)	-0.004976* (0.002785)	-3.196341*** (0.542471)	-0.000464 (0.012993)	-0.005435* (0.002787)	-2.955270*** (0.545501)
26	-0.002761 (0.013592)	-0.005412* (0.002912)	-3.060139*** (0.567230)	-0.001637 (0.013586)	-0.005816** (0.002914)	-2.858738*** (0.570411)
27	0.028568** (0.014247)	0.001454 (0.003052)	-3.399858*** (0.594578)	0.030002** (0.014241)	0.001048 (0.003055)	-3.190533*** (0.597923)
28	-0.000907 (0.014747)	-0.002925 (0.003160)	-3.688504*** (0.615464)	-0.000922 (0.014743)	-0.003189 (0.003162)	-3.572949*** (0.618981)
29	-0.006579 (0.015059)	-0.001358 (0.003226)	-3.600703*** (0.628466)	-0.006451 (0.015054)	-0.001548 (0.003229)	-3.514194*** (0.632061)
30	0.031363** (0.015411)	-0.001440 (0.003302)	-3.520309*** (0.643144)	0.031671** (0.015405)	-0.001752 (0.003305)	-3.376130*** (0.646802)
31	0.015320 (0.015862)	-0.001527 (0.003398)	-3.575188*** (0.661990)	0.014949 (0.015858)	-0.001527 (0.003402)	-3.582788*** (0.665792)
32	0.011622 (0.016293)	-0.004812 (0.003491)	-2.935035*** (0.679973)	0.011203 (0.016289)	-0.004901 (0.003494)	-2.904941*** (0.683882)
33	0.011838 (0.016840)	-0.002765 (0.003608)	-2.443325*** (0.702785)	0.010930 (0.016836)	-0.002849 (0.003611)	-2.422553*** (0.706848)
34	0.010071 (0.017153)	-0.000505 (0.003675)	-3.557613*** (0.715871)	0.010276 (0.017148)	-0.000687 (0.003678)	-3.473131*** (0.719968)
35	0.033881* (0.017853)	0.002498 (0.003825)	-3.081283*** (0.745076)	0.033375* (0.017848)	0.002341 (0.003829)	-3.022827*** (0.749363)
36	0.008385 (0.018499)	-0.000445 (0.003963)	-1.405797* (0.772037)	0.008615 (0.018494)	-0.000566 (0.003967)	-1.347599* (0.776461)
37	-0.015230 (0.019723)	-0.001876 (0.004226)	-2.031575** (0.823115)	-0.014726 (0.019717)	-0.002125 (0.004229)	-1.911147** (0.827818)
38	0.019320 (0.019819)	-0.002202 (0.004246)	-1.867591** (0.827115)	0.018078 (0.019814)	-0.002182 (0.004250)	-1.902608** (0.831893)
39	-0.043888** (0.021103)	-0.003594 (0.004521)	-3.120462*** (0.880721)	-0.044092** (0.021097)	-0.003755 (0.004526)	-3.054158*** (0.885774)
40	-0.006690 (0.022698)	0.000186 ²⁸ (0.004863)	-2.221937** (0.947287)	-0.004855 (0.022691)	-0.000264 (0.004867)	-1.985169** (0.952676)

Continues on next page

Table A1: Coefficients for Predicted Outcomes, continued

Outcome:	For peer parent income regressions (exclude own parent income)			For peer gender composition regressions (exclude female dummy)		
	Income Elasticity	Female Share	Income	Income Elasticity	Female Share	Income
41	0.005638 (0.024187)	0.008639* (0.005182)	-3.926921*** (1.009429)	0.004810 (0.024181)	0.008607* (0.005187)	-3.930452*** (1.015235)
42	0.014357 (0.026112)	-0.004238 (0.005594)	-3.577288*** (1.089736)	0.014854 (0.026104)	-0.004500 (0.005599)	-3.451679*** (1.095972)
43	-0.021955 (0.029473)	-0.005603 (0.006315)	-2.360152* (1.230026)	-0.019037 (0.029463)	-0.006295 (0.006320)	-1.993799 (1.237003)
44	-0.027851 (0.032239)	-0.007773 (0.006907)	-2.528817* (1.345465)	-0.026316 (0.032229)	-0.008278 (0.006913)	-2.274483* (1.353136)
45	-0.006135 (0.035509)	0.002574 (0.007608)	-2.283914 (1.481934)	-0.003390 (0.035499)	0.002105 (0.007615)	-2.019874 (1.490428)
46	-0.048869 (0.043791)	0.003230 (0.009382)	-1.439035 (1.827562)	-0.045894 (0.043776)	0.002432 (0.009390)	-1.025151 (1.837966)
47	0.003652 (0.050726)	-0.015696 (0.010868)	-4.624421** (2.117012)	0.008191 (0.050708)	-0.016928 (0.010877)	-3.986250* (2.128991)
48	-0.039363 (0.061117)	-0.000373 (0.013094)	-4.114260 (2.550647)	-0.035969 (0.061097)	-0.001335 (0.013106)	-3.619229 (2.565194)
49	0.011190 (0.079562)	0.008070 (0.017046)	-9.604751*** (3.320437)	0.020798 (0.079533)	0.005894 (0.017061)	-8.444677** (3.339216)
50	0.116394 (0.116198)	0.024350 (0.024896)	-2.111243 (4.849378)	0.120910 (0.116163)	0.023307 (0.024918)	-1.556926 (4.877154)
51	-0.095809 (0.155865)	0.040909 (0.033394)	-1.36e+01** (6.504835)	-0.092872 (0.155820)	0.040420 (0.033425)	-1.33e+01** (6.542152)
52	-0.755767*** (0.199376)	-0.026450 (0.042717)	-1.32e+01 (8.320724)	-0.764583*** (0.199322)	-0.026251 (0.042757)	-1.34e+01 (8.368617)
52+ not shown due to small number of obs.						
Constant	1.072728*** (0.023278)	0.646511*** (0.004987)	23.303633*** (0.971501)	1.073632*** (0.023260)	0.648824*** (0.004990)	22.304253*** (0.976582)

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

Table A2: Separate Falsification Tests for Each Background Characteristics, Study Peer Design

	Dependent variable	Fraction female among study peers			Average parent income of study peers		
		Coeff.	s.e.	Adjusted p-value	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 (2) 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 (5) 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 A3: Impacts of Study Peers on the Direction of Innovation, Share of Sales across Consumer Groups

Dependent variable	Average parent income of study peers		
	(1)	(2)	(3)
A. Share of sales to rich (above 100k over below 30k)	0.00018* (0.00010)	-0.00005 (0.00019)	0.00031** (0.00014)
B. Share of sales to rich (above 60k over below 60k)	0.00013 (0.00008)	-0.00007 (0.00014)	0.00024** (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 (1). 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 8. 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 A4: Impacts of Study Peers on the Direction of Innovation and Income, Entrepreneurs and Inventors in Patenting Industries

Dependent variable	Fraction female among study peers			Average parent income of study peers		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Sales share to women	0.00115 (0.00619)	-0.00365 (0.01273)	0.02014** (0.00889)			
B. Industry income elasticity				0.00074* (0.00044)	-0.00023 (0.00079)	0.00160*** (0.00058)
C. Income	-2.49233** (1.23158)	-0.64292 (2.09294)	-3.57579** (1.80196)	0.01812 (0.01953)	0.00275 (0.03093)	0.03566 (0.03107)
Sample	All	Women	Men	All	Own parent income below median	Own parent income above median
Students	46247	18382	27865	46247	21598	24649
Study groups	19882	10358	13149	19882	12624	13540
Schools	555	509	517	555	533	523

Notes: Sample consisting of 46,247 individuals who become inventors or entrepreneurs in 4-digit industries that generate patents. 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 (1). 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 8. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A5: Impacts of Study Peers on Top Incomes

Dependent variable	Fraction female among study peers			Average parent income of study peers		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Income above 50th pc.	-0.0130 (0.0232)	0.0311 (0.0458)	0.0071 (0.0319)	-0.00019 (0.00036)	-0.00056 (0.00061)	0.00054 (0.00053)
B. Income above 90th pc.	-0.0473** (0.0186)	-0.0258 (0.0301)	-0.0592** (0.0288)	0.00021 (0.00027)	0.00006 (0.00042)	0.00035 (0.00043)
C. Income above 99th pc.	-0.0068 (0.0094)	-0.0116 (0.0166)	-0.0079 (0.0143)	-0.00009 (0.00016)	-0.00010 (0.00022)	-0.00011 (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 (1). 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 8. 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 A6: 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 A3 to our homophily estimates from the main text. Specifically, using the homophily estimates from the main text, 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 A3 (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 A7: Additional Estimates for the Study Peer Design

Specification:	Fraction female study peers			Average parent income of study peers		
	Share of sales to women			Industry income elasticity		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Baseline	0.0013 (0.0056)	-0.0080 (0.0110)	0.0187** (0.0082)	0.00080** (0.00040)	-0.00020 (0.00071)	0.00141** (0.00055)
N	51,186	20,714	30,472	51,186	23,889	27,297
B. No additional controls	0.0014 (0.0056)	-0.0074 (0.0109)	0.0187** (0.0082)	0.00074* (0.00040)	-0.00010 (0.00071)	0.00136** (0.00055)
N	51,186	20,714	30,472	51,186	23,889	27,297
C. Weighted	0.0034 (0.0057)	0.0009 (0.0104)	0.01543* (0.0082)	0.00095** (0.00045)	-0.000142 (0.00075)	0.00195*** (0.00061)
N	51,186	20,714	30,472	51,186	23,889	27,297
D. Peer group size ≤ 25	0.01231 (0.0084)	0.0187 (0.0196)	0.0283* (0.0155)	0.0010* (0.0005)	-0.00011 (0.00115)	0.00197** (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 (1). 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 8. 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Model and Counterfactual Results

We briefly summarize the model in Einiö et al. (2023), and then describe the counterfactual exercises involving homophily and social interactions.

B.A Model Summary

The model is an extension of the canonical Romer (1990) framework. To account for the observation homophily documented in Einiö et al. (2023), we allow for heterogeneity in

consumer tastes and differences in the direction of innovation pursued by innovators from different backgrounds:

- The economy has two equally-size groups indexed by $g \in \{M, W\}$
- There are two sectors. Each group has different preferences across the sectors:
 - Agents in the economy maximize lifetime discounted utility $\int_0^\infty e^{-\rho t} \log(C_i(t)) dt$, where $C_i(t) = C_{1i}(t)^{\alpha_{g(i)}} \cdot C_{2i}(t)^{1-\alpha_{g(i)}}$.
 - Preference parameters α_g are specific to each group, and determine spending shares allocated to each sector.
 - $C_{ji}(t) = \left(\int_0^{N_j(t)} c_{ji}(\nu, t)^{(\varepsilon-1)/\varepsilon} d\nu \right)^{\varepsilon/(\varepsilon-1)}$, 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.
- 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.
 - Agent $i \in g$ is assigned 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
 - Agents can decide whether to innovate only in the specific sector they were assigned to. 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}
- Finally, we account for unequal access by creating a binary wedge τ_{gi} that determines whether the agent is able to enter the innovation sector, regardless of ability. For

simplicity, we assume that the wedge only affects individuals in the W group.

- Cost-of-living inequality is determined by the ratio of price indices across sectors: $\frac{P_{Wt}}{P_{Mt}} = \left(\frac{N_1(t)}{N_2(t)}\right)^{\frac{\alpha_M - \alpha_W}{\varepsilon - 1}}$
 - Agents who prefer sectors with fewer goods will face higher implied prices (more spending is allocated to fewer varieties)

To solve for the equilibrium, we can guess cutoff productivities in each sector. We then check against the first order conditions to find cutoffs that meet equilibrium conditions (intertemporal and across sectors).

B.B Counterfactual exercises

Einiö et al. (2023) calibrates the model parameters using observed data on female innovator fraction, growth rates, and estimated homophily coefficients. The paper then focuses on counterfactuals where barriers to female innovator participation (τ) are relaxed.

Here, we use the calibrated parameters from that paper ($\tau = 0.111, \phi = 0.725, \bar{\eta} = 0.011$), but assess counterfactuals related to the sector assignment parameter ϕ . First, we run a counterfactual economy with no homophily ($\phi = 0$), keep all other parameters the same. In addition, we also compute a counterfactual where ϕ is reduced, reflecting the peer effects estimates ($\phi = 0.661$).