

For Online Publication
**Appendix to “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*

January 2026

A Theory Appendix

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

A.A A Simple Love-of-Variety Framework

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. 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. 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.

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 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,¹ 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."² We derive the welfare comparison between

¹We 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.

²As 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.

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.

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:

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

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:

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

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.³ This number increases to 46 percent in the sample of phone applications.⁴

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 $\lambda^{\text{Counterfactual}} = 0.5 \times \lambda^F + 0.5 \times \lambda^M = 0.2735$.

³Note 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 is 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.

⁴To 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.

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.⁵ 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, which we now investigate formally in a quantitative growth model.⁶

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.

A.B Endogenous Growth Model

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

A.B.1 Derivations

In this section, we present derivations related to the model setup in Section IV.A.

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.

⁵We 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.

⁶We 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 innovators, who will have a disproportionate impact on female consumers through homophily.

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 frictions 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 frictions, 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 market 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 $r^* = \frac{1}{\varepsilon-1} \cdot \hat{\eta}_s N(t) \frac{L_{sM}}{N_s(t)}, \forall s$. 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, $\frac{N_1(t)}{N_2(t)} = \frac{\int_{\hat{\eta}_1}^{\infty} x f_1(x) dx}{\int_{\hat{\eta}_2}^{\infty} x f_2(x) dx}$. 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 $P_s = \left(\int_0^{N_s} p(x)^{1-\epsilon} dx \right)^{1/(1-\epsilon)} = N_s^{1/(1-\epsilon)}$, which we plug into each consumer group's price index, $P_g = P_1^{\alpha_g} P_2^{1-\alpha_g}$.

A.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 $\tilde{\phi}_i \sim U[0, 1]$ for all individuals. For agents in the minority group, we draw an access friction value from $\tilde{\tau}_i \sim U[0, 1]$. All individuals in the majority group (i.e., men or the top income quintile) are assumed to face no barriers. 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 $(\tau, \bar{\eta}, \phi)$, we convert the numerical draws described above into Bernoulli draws. Agents from the minority group are exposed if $\tilde{\tau}_i < \tau$, and we assign agents to the sector for which they have a taste preference if $\tilde{\phi}_i < \phi$. We also scale the research productivity draws by $\bar{\eta}$ to control the shape of the innovation productivity distribution, resulting in a distribution with CDF $F(x) = 1 - \left(\frac{\bar{\eta}}{x}\right)^\alpha$. 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 $(\tau, \bar{\eta}, \phi)$ 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 equal access” counterfactual, where all agents can enter innovation, or “targeted exposure” counterfactuals where we give access to all individuals in the top x percent of the productivity distribution (among those who previously faced

barriers).

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.

A.B.3 Evidence on Wedges

Here, we provide additional details related to the discussion in the main text on empirical evidence related to access wedges.

First, we study innovation productivity, comparing model predictions, depending on the wedge, to the patterns in the data. Panel A of Table C11 reports the probability of being a “top innovator” (top 10% or top 5% of the innovation productivity distribution) conditional on innovating, across models.⁷ The model with wedge 2 makes stark predictions: women should be highly likely to be top innovators conditional on innovating. Indeed, in this model only the most productive female innovators find it worthwhile to enter the innovation sector (given the cost $\tilde{\tau}_{gi}$ affecting their returns to innovation), while their productivity is unaffected. In contrast, the models with wedges of type 1 or 3 both predict that women should have lower probabilities of being a top innovator.⁸

Panel B of Table C11 presents the patterns in the data. We examine the probability of being in the top 10% or top 5% of innovators for phone applications (ranking by time usage

⁷Innovation productivity is governed by η_i for models with wedges of Type 1 and 2, and by $(1 - \hat{\tau}_{gi})\eta_i$ for the model with the Type 3 wedge.

⁸This under-representation at the top comes from the fact that, in equilibrium, the productivity cutoff for entry of innovator is slightly lower in the sector focusing on female customers (as noted above), where there are relatively more female innovator due to innovator-consumer homophily. This results from the interaction of demand and supply forces in equilibrium. On the demand side, customer demand is a bit lower for the female-focused sector because of the gender wage gap, which reduced the relative purchasing power of female customers. On the supply side, there is much lower supply of innovators in the female-focused sectors, because of the frictions female innovators face. To ensure that supply meets demand in equilibrium, the productivity cutoff must be lower and the price index is higher in the sector focusing on female customers. This mechanism applies to the three types of wedges and push for slightly lower productivity for women, who disproportionately enter the female-focused sector through homophily. With wedge 2, the selection effect (only the most productive female innovators decide to enter innovation) vastly offsets the heterogeneous cutoffs effect. With wedge 1 and 2, there is no offsetting selection effect, so female innovators are slightly less productive in equilibrium and, in particular, are a bit under-represented among top innovators. Finally, it is worth noting that the heterogeneous cutoffs mechanism is consistent with the finding that, in the consumer packaged goods data, female entrepreneurs start companies with higher unit prices.

of apps), consumer packaged goods (ranking by sales), and patents (ranking by citations). In all cases, we find that female innovators are *under-represented* at the top of the production. These empirical patterns reject the model with Type 2 wedge and favor models with Type 1 or Type 3 wedges.⁹

While a model with Type 1 wedges predicts that the distribution of *latent* innovator productivity should be similar for men and women, a model with Type 3 wedges predicts that women who becomes innovators come overwhelmingly from the top of the latent innovation productivity distribution. Although the latent innovation productivity distribution is unobserved, a simple test is to use math test scores in childhood.¹⁰ Focusing on patent inventors, Bell et al. (2019a) report that, compared to men, women and minorities remain much less likely to become inventors even when they have math test scores two standard deviations above the mean (see Panel A of Figure C5). Focusing on exceptional math students in the American Mathematics Competition (AMC), Calaway (2025) shows that even among these exceptional math students in middle school or high school, women and students from certain demographic groups remain less likely to attend selective universities and major in STEM fields. Calaway (2025) highlights that the gender gap is particularly striking: exceptional female students pursue STEM majors at a rate that is 19pp/20pp lower than male students (Panel B of Figure C5). These patterns are inconsistent with models with Type 3 wedges – where the female inventors with the highest productivity should always enter innovation, and do less well than men *ex-post* due to productivity barriers.

A.B.4 Gender Ratios among Top Inventors

In this appendix, we derive simple formulas for the gender ratio among the top X percent of inventors under different types of wedges. The formulas show that, under our calibrated parameters, the gender ratio among top inventors is almost identical under Type 1 and Type 3 wedges.

⁹As a complement to the simulation results in Table C11, we show analytically in Appendix A.B.4 that models with Type 1 and Type 3 wedges make very similar predictions for the probability of becoming a top inventor by gender.

¹⁰Several paper have shown that high mathematical ability in childhood strongly correlates with knowledge production in adulthood (e.g., Aghion et al., 2023; Akcigit et al., 2017; Bell et al., 2019a; Agarwal and Gaulé, 2020).

We work with the CDF of the basic Pareto distribution, $Pr(X > x) = \left(\frac{x_m}{x}\right)^\alpha$. When comparing the female fraction in the top X percent of inventors for the Type 3 wedge model, female inventors have their productivities multiplied by τ_{real} (a penalty), so the CDF for the productivity of female inventors becomes $Pr(\tau_{real}X > x) = \tau_{real}^\alpha \left(\frac{x_m}{x}\right)^\alpha$, while men keep the basic CDF, $Pr(X > x) = \left(\frac{x_m}{x}\right)^\alpha$. Taking ratios of CDFs, the fraction of female inventors above any cutoff x (greater than the entry cutoff in the male sector) is given by $\frac{\tau_{real}^\alpha}{1+\tau_{real}^\alpha}$. In our calibration, $\tau_{real} = 0.1683$, $\alpha = 1.26$, so $\tau_{real}^\alpha \approx 0.106$.

In the Type 1 wedge model, the probability of being above a certain innovator productivity cutoff shrinks uniformly for women, $Pr(X > x) = \tau_{exposure} \left(\frac{x_m}{x}\right)^\alpha$. Therefore, the gender ratio is $\frac{\tau_{exposure}}{1+\tau_{exposure}}$. In our calibration, $\tau_{exposure} = 0.111$, which is very close to the value of τ_{real}^α found above. Thus, the gender ratio is similar at each percentile cutoff in models with Type 1 and Type 3 wedges.

A.B.5 Endogenizing the Gender Wage Gap

Our baseline model features an exogenous gender earnings gap, governed by the exogenous parameter δ . This parameter scales in the same way production worker wages $w(t)$ and the returns to innovation $V_j(t)$ for each group g (men or women), so that occupational choice is left unaffected.¹¹ This parameter remains identical when we relax access barriers to the innovation sector for women. In this version of the model, for tractability we ignore differences in the share of women in entrepreneurship when computing earnings gap by gender such that, given our calibration of the earnings penalty parameter δ , men remain 20.5% richer than women regardless of the counterfactual.

In this appendix, we introduce a model with an endogenous gender earnings gap. The model is identical to the baseline model except that we now compute the difference in total earnings by adding up earnings in production work and earnings in entrepreneurship. In this version, differences in earnings between men and women account for the fact that being under-represented in innovation careers reduces average earnings for women, as these careers tend to be more lucrative. In this version of the model, the endogenous gender earnings gap

¹¹This assumption is consistent with the fact that certain factors that lead to lower earnings for the median female worker, such as the child penalty, also operate for entrepreneurs with similar magnitudes (e.g., Rutigliano (2024)).

is 1.568, compared to 1.205 in the baseline model. This increase in the gender earnings gap comes from the fact that innovators earn on average 4.93 more than production workers in the calibrated model.¹² In this version of the model, we solve for a fixed point equilibrium such that the market size of each sector is consistent with total earnings of each group.

Panel A of Table C17 shows that the two versions of the model deliver similar results for growth and cost-of-living counterfactuals. There is a fall in cost-of-living inequality of 16.3 percentage points in the model with endogenous wage gaps, compared to 18.20pp in the baseline model. The growth impacts of equal access to innovation careers is 1.38pp with endogenous wage gaps, compared to 1.44pp in the baseline model. The differences across the two models in terms of growth and cost-of-living inequality are small because the endogenous gender earnings gap affects innovation dynamics through market size, which only affects “marginal” inventors close to the productivity cutoff.

Panel B of Table C17 provides the calibrated parameters in both versions of the model, which explain why the cost-of-living and growth impacts are slightly smaller in the model with endogenous wages. The calibrated parameters are different in the model with endogenous gender wage gaps because of market size effects at the initial equilibrium. The relative earnings of men compared to women are larger in the endogenous gender wage gap model than in the baseline model because entrepreneurial income is taken into account. As a result, the market size of the sector targeting female customers, relative to the market size of the sector targeting male customers, is smaller in the endogenous gender wage gap model. This ends up changing several aspects of the baseline calibration. First, there is an impact on the calibrate homophily parameter, ϕ . As noted in the main text, we calibrate the model to target our empirical estimates of the average difference in female sales shares between male and female entrepreneurs, divided by the baseline rate of female consumption – namely, female entrepreneurs should sell about 25% more to female customers. Because the baseline sales share of female customers (relative to male customers) is smaller in the endogenous model, we no longer need as large an absolute difference in sales shares to male vs. female customers to match our regression coefficient. Consequently, the parameter ϕ in

¹²Production worker wages act as the numeraire, so the gap between innovators and production workers should be close to the mean of the Pareto distribution for inventor productivity. Indeed, with $\lambda = 1.26$, as in our calibration, the Pareto mean is $\lambda/(\lambda - 1) = 4.85$.

the model, which captures an absolute difference in the tendency to sell to male vs. female customers, is calibrated to be smaller in the endogenous wage model than in the baseline model. Therefore, the cost-of-living impacts of equal access to innovation by women turns out to be slightly smaller, at -16.3pp in the endogenous wage model compared to -18.2pp in the baseline model. Second, the wedge parameter τ is also affected. The smaller value of ϕ means that there is less sectoral specialization, i.e. that women are less shielded from competition from men (by focusing on a different sector) than they were in the baseline model. Consequently, matching the overall female inventor fraction requires a higher τ (lower barriers) in the endogenous wage model than in the baseline model. The counterfactual bringing τ to one (i.e., eliminating barriers) therefore constitutes a smaller change in the endogenous gender wage gap model, with slightly lower growth impacts, at 1.38pp instead of 1.44pp.

While the impacts on growth and cost-of-living inequality are limited, there is a large impact on relative earnings between men and women. The results are reported in Panel C of Table C17. In our baseline model with exogenous wage gap, the earnings ratio remain fixed at 1.205. In the model taking into account endogenous entrepreneurial income, the male to female ratio falls from 1.57 to 1.212 when relaxing barriers to innovation for women.¹³ Indeed, relaxing barriers leads to an increase in the share of women in the innovation sectors, which raises the average earnings of women as this sector offers higher compensation. The level of inequality in the counterfactual without access barriers equilibrium thus falls back to the level of inequality set by the penalty parameter δ .¹⁴ The fall in overall inequality is thus much stronger than in the baseline model.¹⁵

Because the expanded model takes twenty times longer to run than the baseline model

¹³The latter number would converge to 1.205 with an infinite sample; the reported result differs due to sampling error in our 100 bootstrap runs.

¹⁴The inequality patterns in our calibrate model are broadly consistent with external benchmarks. First, innovators earn 22.2 % of aggregate income in the model, similar to the share of income earned by the top 1 % in the United States (World Income Inequality Database). Second, the gender wage gap for the median worker is 20.5 % according to the U.S. Bureau of Labor Statistics, which we match exactly by setting δ . Third, when accounting for access barriers for innovation, the total gender earnings gap is such that men earn 57 % more than women, similar to the overall earnings disparities by gender documented by Andrew, Bandiera, Costa Dias, and Landais (2024).

¹⁵Panel C of Table C17 also reports the change in the earnings ratios of entrepreneurs to production workers with and without innovation barriers for women. In the counterfactual without barriers, the earnings ratio falls from 4.93 to 4.63. This change occurs because there are now relatively more female entrepreneurs among innovators and they are subject to the earnings penalty δ .

and that the gender wage gap is not our focus, it is not used in our baseline analysis, for tractability.

B Heterogeneity and Mechanisms

In this appendix, we investigate where homophily is strongest and how it might arise. Overall, we find that homophily is a broad-based phenomenon, which is not driven by a small number of products, industries, or consumer types — even though meaningful heterogeneity exists across categories. Figure 3 in the main text summarizes the main heterogeneity patterns. Panel A focuses on consumer packaged goods, for which we are able to analyze heterogeneity across the largest number of dimensions – by product categories, product features, and consumer characteristics. Panel B reports the results for phone applications. We discuss these results in the remainder of this section, along with evidence from our other datasets.¹⁶

Heterogeneity across product categories. We start by reporting innovator-consumer homophily estimates across the product space at a relatively granular level in the U.S. micro data (second to fourth rows in Panels A and B of Figure 3). For consumer packaged goods, we find that homophily is particularly strong for health and household products. Similarly, when we segment the product space into three broad categories for phone applications – “lifestyle, health, and leisure”, “technology and tools”, and “information and commerce” – homophily estimates are positive everywhere and appear larger for the first two groups. These results suggest that homophily operates broadly and that it may be of particular relevance in health- and household-related areas.

Heterogeneity by product characteristics. Next, we run three complementary analyses that characterize heterogeneity by product characteristics, considering in turn the roles of product popularity, vertical differentiation, and horizontal differentiation.

First, we analyze whether homophily is also present for the highest-selling products. Indeed, it could be that only “marginal” products, with lower commercial impact, might be subject to homophily. This is an important source of heterogeneity to uncover to discipline the model in the next section: the welfare impact of homophily would be limited if homophily

¹⁶Detailed estimates and standard errors are presented in Tables C18 to C29.

were relevant only for products with lower sales. To assess this, we repeat our analysis with sales weights for consumer packaged goods and time use weights for phone applications, reflecting the importance of the product. The results are reported in Figure 3 for consumer packaged goods and phone apps (fifth rows in Panels A and B): the estimates are very similar compared to the baseline results without weights.¹⁷ Similarly, using the Finnish data, 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 C22).

Second, we analyze whether vertical differentiation could help explain homophily. We use the price as a proxy for vertical differentiation – i.e., “quality” – within detailed product categories (e.g., Moshary et al., 2023). We conduct the analysis in the sample of consumer packaged goods, where we have detailed price information. To start, we show that female-founded teams tend to have significantly higher unit prices (Table C12), about a 24 percent difference between all-female and all-male teams within detailed product categories. Next, we estimate gender homophily while controlling for prices. As reported in the sixth row of Panel A in Figure 3 and detailed in Table C23, controlling non-parametrically for prices (with a fourth-order polynomial) leaves the homophily coefficient almost unaffected; if anything, it increases slightly, suggesting that homophily would be even stronger if unit prices were similar. Thus, homophily is not driven by vertical differentiation.

Third and finally, we examine whether product characteristics (i.e., observable horizontal product differentiation) can account for homophily. We first add three basic controls, which are available for many products: the size of the item, whether it is part of a “multipack”, and whether it is an organic product. These controls leave the homophily coefficient almost unchanged (Table C24). Next, we consider all product characteristics available in the NielsenIQ dataset, selecting the relevant variables using LASSO. The estimate in the seventh row of Panel A in Figure 3 shows that the homophily coefficient remains similar as we add characteristics under a LASSO specification.¹⁸ These results suggest that gender homophily

¹⁷Tables C19 and C20 provide detailed estimates. Although we focus on expenditure shares in the main text, as this specification can be linked to welfare (see Section IV), in Table C21 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 estimates are imprecise.

¹⁸For the full sample, we draw a random subset of 50% of all firms – given the very large number of

operates through characteristics that are not easily captured by traditional characteristics data.

Consumer characteristics. Next, we document whether homophily varies depending on consumer type. To do so, we compute female consumer shares within subgroups of consumers (e.g., families with college-educated heads of household). The results are reported in the last six rows of Panels A and B in Figure 3. For consumer packaged goods, higher education, higher income, and younger consumers appear to exhibit more homophily. However, these differences are difficult to estimate precisely given our limited statistical power. The results for phone applications are similar, with higher gender homophily for more educated, higher income, and younger consumers. Although statistical power is limited in some specifications, the point estimates remain positive in all categories.¹⁹

Heterogeneity across industries. We use the Finnish data to further characterize the patterns that drive our between-industry homophily estimates. Specifically, we estimate the strength of homophily by broad industry sectors (manufacturing vs. other sectors), and between or within 2-digit industries. Results are presented in Table C28. We observe prevalent homophily patterns by gender and income for patent inventors and entrepreneurs within both broad sectors and finer 2-digit industries, with homophily primarily occurring between rather than within 2-digit industries.

Innovator characteristics. The Finnish data offer rich information on innovator’s characteristics, so we proceed to study whether these characteristics relate to the degree of innovator-consumer homophily. Results are presented in Table C29. We do not detect any statistically significant differences in gender or income homophily by patent inventors’ education (college vs. non-college), age (below or above 40 years), or family background (high vs. low parental income). In the entrepreneur sample, with a substantially larger number

characteristics in the NielsenIQ data, we run into computational issues with a larger sample. When using control variables selected for the 50% sample, results are similar in the full sample, although less precise because the set of control variables is not optimized for this sample. Table C25 shows that the results are similar in the CPG startup sample.

¹⁹Tables C26 and C27 report all estimates and standard errors.

of observations, we find weaker gender homophily among highly-educated entrepreneurs in Finland. While the results suggest some nuanced differences in the degree of gender and income homophily by innovator type, nearly all subgroup-specific homophily coefficients turn out to be positive (21 out of 24), indicating the prevalence of gender and income homophily across innovator groups.

Team composition. Given the increased importance of teams for innovation (Jones, 2009b; Jaravel et al., 2018), it is instructive to investigate potential non-linear homophily patterns within teams. Figure C6 shows that there is a steady increase in team size for startup founders and patent inventors over time. While there is a long-term downward trend in the prevalence of all-male teams, such teams remain very common by the end of the study period (Figure C7).

We document two facts on team formation and non-linear effects. First, there is significant gender clustering in teams compared to random assignment (Figure C8). Second, we find that it is generally sufficient to have one female founder or inventor in the team to increase the share of sales to female consumers (Table C30 for the United States and Table C31 for Finland).²⁰ Statistical imprecision makes it challenging to assess the extent to which a larger number of female founders or inventors within the team further impacts homophily.

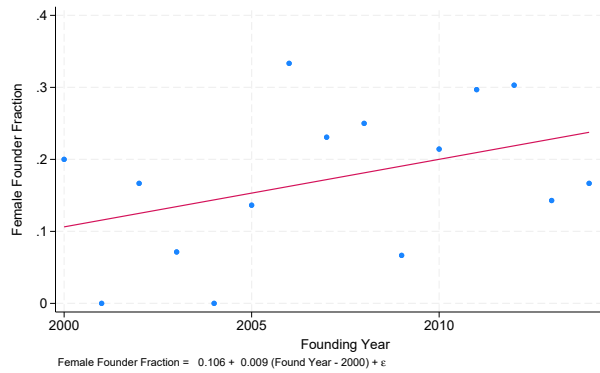
Social mechanisms independent of market size. In a companion paper (Einiö et al., 2025), we provide direct evidence that a mechanism for innovator-consumer homophily is that innovators’ social experiences have a causal impact on the direction of innovation, independent of financial incentives. Using a quasi-experimental study-peer design in Finnish education programs, we examine the impacts of the gender and socioeconomic composition of the peer group of a student who later becomes an entrepreneur on the types of consumers they cater to. The results provide direct evidence that social factors can affect the direction of innovation, which stands in contrast with mechanisms based on financial incentives, such as market size.

²⁰For income homophily, the relationship between the parental income composition of the team and the direction of innovation appears to be more linear (Table C31).

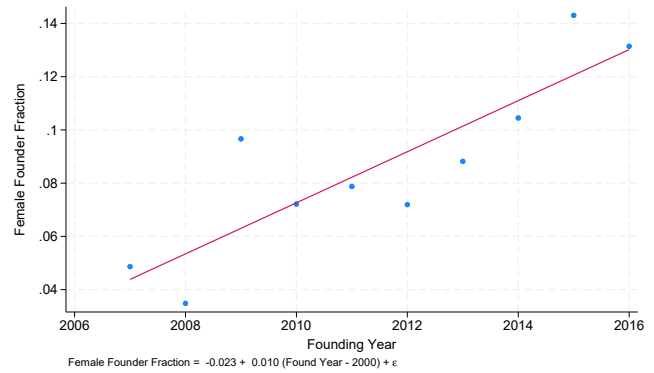
C Additional Figures and Tables

Figure C1: Gender Gaps in Entrepreneurship in the United States

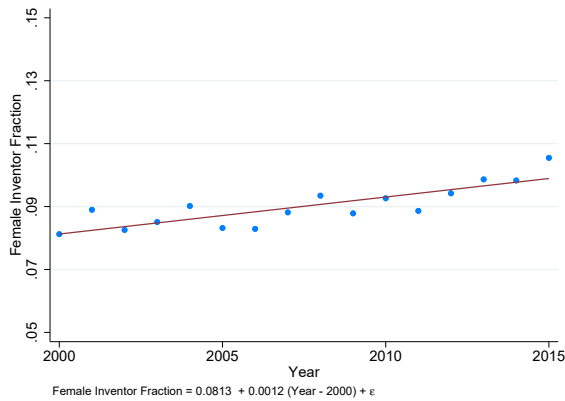
A. Fraction of Female Startup Founders within Consumer Packaged Goods



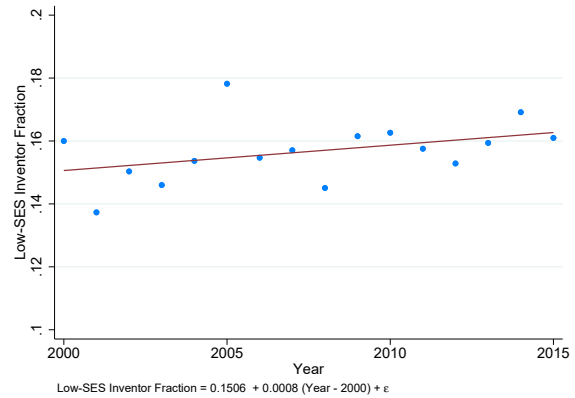
B. Fraction of Female Startup Founders within Phone Applications



C. Fraction of Female Patent Inventors

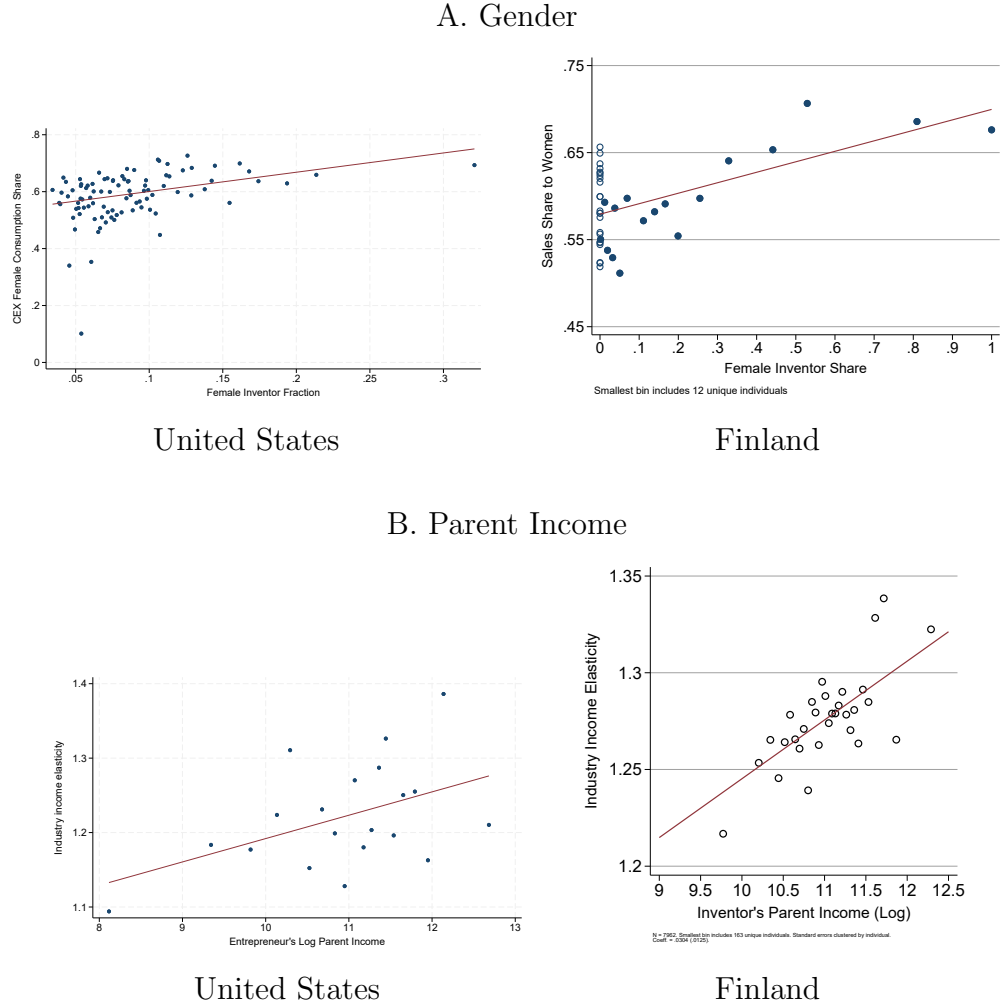


D. Fraction of Low-SES Patent Inventors



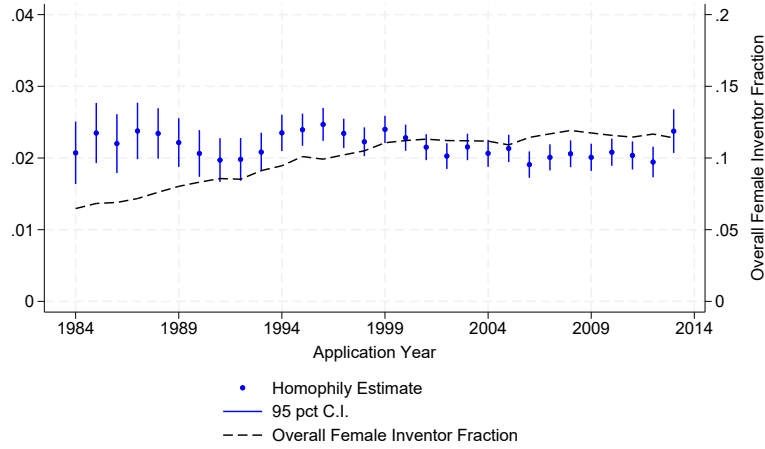
Notes: This figure presents the fraction of female startup founders in the United States for consumer packaged goods (Panel A) and phone applications (Panel B); the fraction of female patent inventors in Finland (Panel C); and the fraction of low-SES patent inventors, whose parental income is in the lowest quintile, in Finland (Panel D). The trend lines show that, at this rate, it will take 50 years (from the start of the sample) to reach gender parity among founders in the United States; it will take over 60 years to reach parity by gender and parental income among patent inventors in Finland.

Figure C2: Binned Scatter Plots for Innovator-Consumer Homophily across Industries



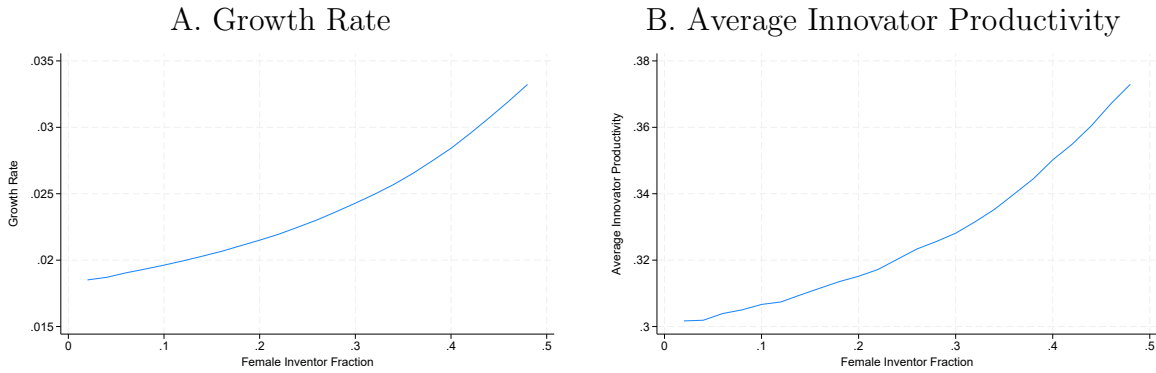
Notes: This figure presents binned scatter plots depicting innovator-consumer homophily by gender (Panel A) and social class (Panel B), for both the United States and Finland. The samples are described in Section II in the main text. Due to data limitations, the sample for the analysis by parent income in the United States only includes 300 observations. We start from the full PSID sample in 2017, which includes 26,445 individuals. We then link these individuals to their parents in earlier waves of the PSID: we can link 13,172 individuals. We only keep individuals between the ages of 21 and 69, reducing the sample to 5,257 individuals. In this sample, we have 325 entrepreneurs, defined as self-employed business owners. For the regression, we further exclude outliers – industries with the maximum or minimum industry income elasticities, and observations in the top 1% or bottom 1% of the parent income distribution.

Figure C3: Homophily Trends – U.S.-Based Inventors



Notes: The figure displays the estimates of the inventor-level homophily regression by filing year with industry-level consumption measures, based only on patents with all U.S.-based inventors instead of all U.S. patents (about 48 percent of the overall sample). The outcome is female consumer fraction, based on CEX data. We show 95 percent confidence intervals using robust standard errors, and also report the female inventor fraction in each year.

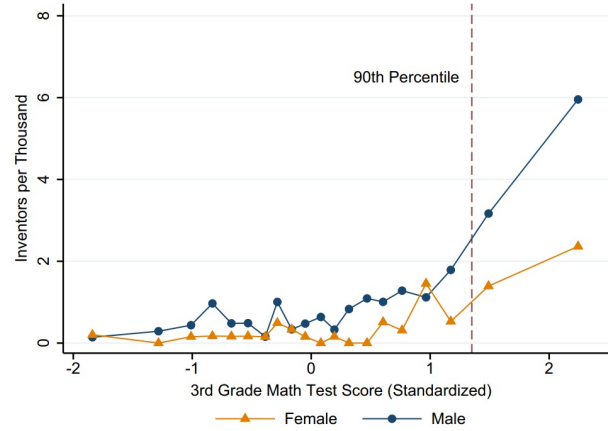
Figure C4: Change in Growth Rate and Innovator Productivity vs. Change in Female Inventor Fraction



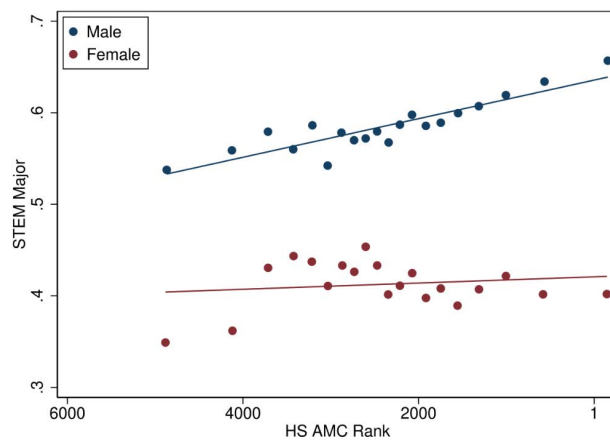
Notes: This figure reports the equilibrium change in growth rate (Panel A) and average innovator productivity (Panel B) against the equilibrium change in the fraction of women among innovators (x-axis), as we reduce access barriers in the model with Type 1 wedge.

Figure C5: Probability of Becoming an Innovator by Gender and Latent Ability

Panel A – Bell et al. (2019): Gender Gaps in Patent Rates
by 3rd Grade Math Test Scores

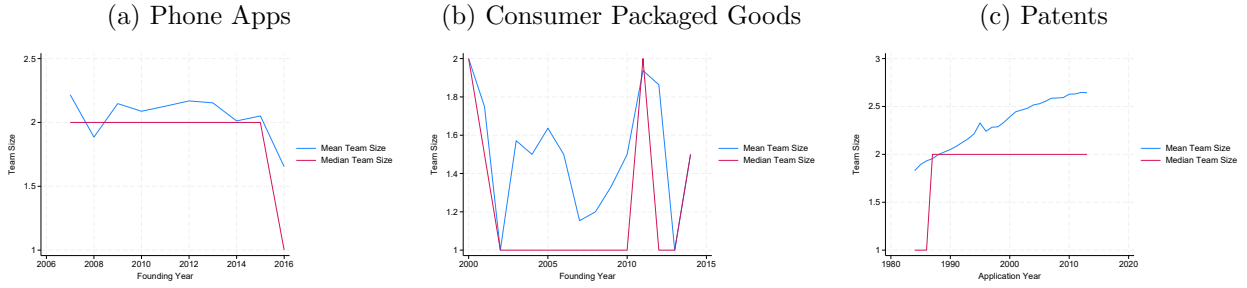


Panel B – Calaway (2025): Gender Gap in STEM
for Exceptional Math Students



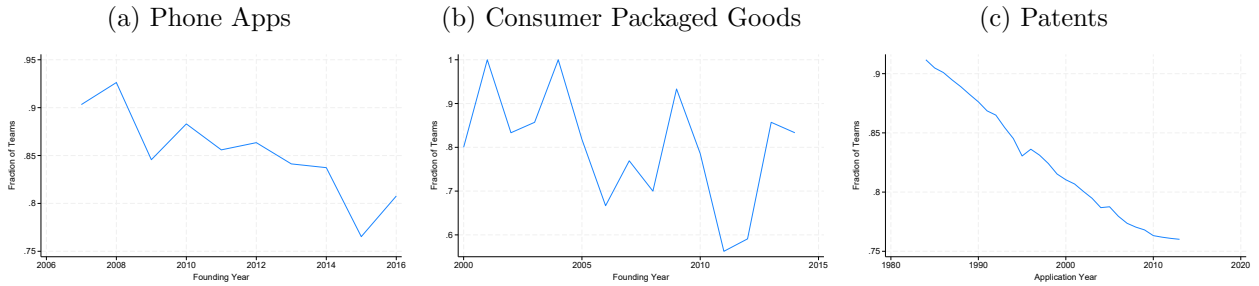
Notes: This figure reports evidence from Bell et al. (2019a) on Panel A and Calaway (2025) in Panel B, using math test scores as (noisy) proxies for latent innovation abilities. The panels show that gender gaps in becoming patent inventors and STEM majors persist even among exceptionally talented students.

Figure C6: Team Size over Time



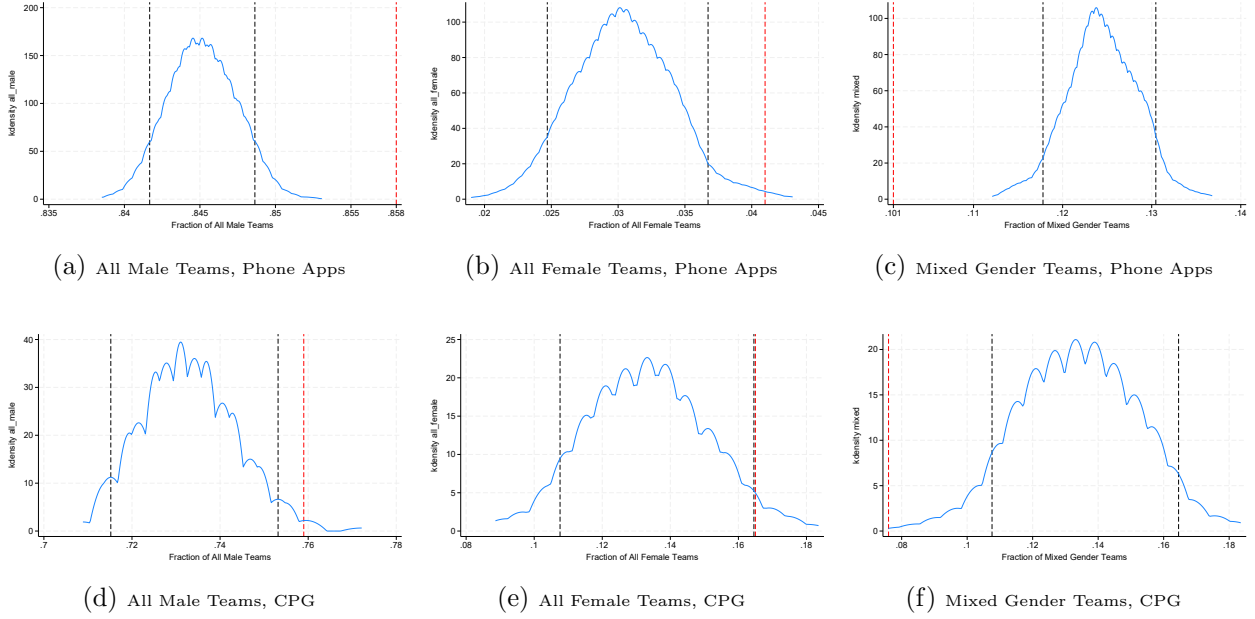
Notes: The panels of this figure report the mean and median time sizes over time for founder teams of phone apps startups (panel a) and of consumer packaged goods startups (panel b), as well as inventor teams in the U.S. patent data (panel c). The patterns for consumer packaged goods are noisy due to the limited sample size in each year.

Figure C7: Fraction of All-Male Innovator Teams over Time



Notes: The panels of this figure report the fraction of teams with only male founders in phone apps startups (panel a) and consumer packaged goods startups (panel b), and with only male inventors in the U.S. patent data (panel c). The patterns for consumer packaged goods are noisy due to the limited sample size in each year.

Figure C8: Gender Clustering in Teams



Notes: The panels of this figure report the distributions of the fraction of all-female, all-male, and mixed-gender teams under random assignment (histogram), their 5th and 95th percentile values (black lines), and means for these statistics in the data (red line). The distribution of the fraction of mix-gender teams under random assignments is obtained via bootstrap. The analysis is conducted separately on the phone apps and consumer packaged goods samples.

Table C1: Innovator-Consumer Geographic Homophily for Phone Applications

	User State Share	
	(1)	(2)
Founder State	0.088*** (0.007)	0.038*** (0.009)
Fixed Effects	None	Category-by-State
Sample Size	46,512	46,176

Notes: The sample used in this table includes all phone applications for VC-backed startups. The analysis is at the app by state level. The outcome variable is the fraction of time usage of a given app from a given state, which has a sample mean of 0.02. The independent variable is an indicator for whether the company is located in the state. With category-by-state fixed effects, the estimate is identified off whether apps in the same category have different shares in a given state based on where the startups are located. The sample size varies because singleton observations are dropped in the second column. Standard errors are clustered at the startup level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C2: 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		0.0184*** (0.0047)
Product Module F.E.	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 599,148$

Notes: This table is identical to Columns (1) and (2) of Table 3 in the main text, except that we measure consumer gender by weighting sales by the fraction of female members in the household, rather than the gender of the household head. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C3: 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* (0.0244)	0.0434*** (0.0155)
Fixed Effects	Age Group	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 NielsenIQ (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 C4: Gender Homophily – Revised Analysis for CPG Sample with Patents

	Share of Sales to Women										
	All UPCs			New UPCs			New Brands			New Manufacturers	
	Female HH	Female Wt		Female HH	Female Wt		Female HH	Female Wt		Female HH	Female Wt
Female Founder Fraction	0.0277*** (0.00585)	0.0184*** (0.00472)		0.0287*** (0.00619)	0.0167*** (0.00502)		0.0182*** (0.00860)	0.0238*** (0.00826)		0.00688 (0.0234)	0.0407* (0.0223)
Constant	0.268*** (0.00278)	0.543*** (0.00151)		0.269*** (0.00297)	0.546*** (0.00134)		0.271*** (0.00209)	0.544*** (0.00190)		0.273*** (0.00467)	0.536*** (0.00493)
Fixed Effects											
Unique firms	2,333	2,333		2,213	2,213		1,917	1,917		964	964
Sample Size	572,786	572,786		433,096	433,096		111,090	111,090		24,831	24,831

Notes: This table presents an alternative analysis of homophily, in the sample of firms that obtain a patent, within the consumer packaged goods sample. While the first two columns keep all barcodes, as in the main text, in the other columns we construct three new subsamples of products: 1) under new UPCs, 2) under new brands, and 3) by new manufacturers, entering the data in 2005 or later. The estimates remain broadly similar in these subsamples and suggesting that the results in the baseline sample are not driven by established products. The coefficients for the new manufacturers' subsample are large but imprecisely estimated. Keeping all barcodes, as in the first two rows, has the advantage of keeping patents affecting process innovations, rather than new products alone. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C5: Gender Homophily – Subsample with High Revenue Per Patent

	Share of Sales to Women							
	All UPCs		New UPCs		New Brands		New Manufacturers	
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0284*** (0.00596)	0.0184*** (0.00485)	0.0295*** (0.00631)	0.0169*** (0.00517)	0.0205** (0.00896)	0.0241*** (0.00879)	0.0140 (0.0277)	0.0479* (0.0268)
Constant	0.269*** (0.00289)	0.543*** (0.00157)	0.270*** (0.00309)	0.546*** (0.00139)	0.272*** (0.00220)	0.544*** (0.00203)	0.280*** (0.00542)	0.535*** (0.00595)
Fixed Effects								
Unique firms	1,170	1,170	1,152	1,152	1,040	1,040	270	270
Sample Size	549,568	549,568	415,197	415,197	103,908	103,908	19,218	19,218

Notes: This table presents an alternative analysis of homophily, in the sample of firms that obtain a patent, within the consumer packaged goods sample. The analysis is the same as in Table C4, except that we further reduce the sample to the subset of firms with an average revenue in the NielsenIQ data per patent above the median of the distribution. Indeed, some firms may have activities outside the domain of consumer packaged goods and may obtain patents for these activities rather than for the set of products we observe. Focusing on firms with a high NielsenIQ revenue per patent may mitigate this potential issue and identify more reliably a set of innovative firms within consumer packaged goods. The table shows that the results are similar to the full sample. The outcome is measured using either female household head ("Female HH") or weights within the households ("Female Wt") to compute the share of sales to female customers. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C6: 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
<i>N</i> industries	330	451	245	417	330	451
<i>N</i> 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 C7: 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*** (0.133)	0.1032*** (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 (0.105)	0.0486* (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
<i>N</i> 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 C8: Homophily Regression Coefficients – U.S.-Based Inventors

	Share of Industry Sales to Women	
	(1)	(2)
Female Patent Inventor	0.0231*** (0.000284)	0.0214*** (0.000293)
Weighting	Unweighted	Log patents
<i>N</i> industries	325	325
<i>N</i> individuals	1,109,280	1,109,280

Notes: The table displays the estimates of the industry-level homophily regression, including patents with only U.S.-based inventors instead of all U.S. patents.

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** (0.0034)			0.0053* (0.0027)	
Entrepreneur's Log Parent Income	0.00764** (0.00343)		0.0371*** (0.0008)	0.00543** (0.00251)		0.0247*** (0.0006)
Country	U.S	Finland	Finland	U.S	Finland	Finland
Mean	0.704	0.6997	0.6809	0.632	0.6241	0.6122
<i>N</i> industries	17	253	441	17	253	441
<i>N</i> 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 3.14% relative to the baseline rate when an entrepreneur comes from a family from the top income quintile instead of the bottom ($= 0.0313 \times \log(269,356/14,859)/0.7053$). 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 ($= 3.14/3.9$), we obtain that the overall income homophily coefficient for the United States is 17.6%. 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: Consumption Dissimilarity Indices

	Consumer Expenditure Survey	Consumer Packaged Goods	Phone Applications
	(1)	(2)	(3)
Male vs. female, single-gender households	0.135	0.320	0.242
Male vs. female, reference person gender	–	0.247	
Top vs. bottom income quintiles	0.336	0.335	0.354
Top vs. bottom income deciles	0.399	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 NielsenIQ dataset covering consumer packaged goods, we compute the dissimilarity index between the two consumer groups 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 C11: Probability of Becoming a Top Innovator by Gender, Model vs. Data

Panel A – Model Predictions				
	Probability of Top 10%		Probability of Top 5%	
	Men	Women	Men	Women
Wedge 1 (“Untapped Marie Curies”)	10.4%	7.4%	5.2%	3.7%
Wedge 2 (“Untapped Marginal Female Inventors”)	5.74%	39.0%	2.87%	19.5%
Wedge 3 (“Stifled Marie Curies”)	10.4%	7.4%	5.2%	3.7%

Panel B – Empirical Patterns				
	Probability of Top 10%		Probability of Top 5%	
	Men	Women	Men	Women
Phone Apps	10.5%	5.7%	5.3%	3.4%
NielsenIQ CPG	11.3%	8.7%	6.7%	2.2%
Citation Weighted Patents	10.3%	7.6%	5.2%	3.6%

Notes: Using models with the three types of wedges described in the text, Panel A presents the fractions of male and female innovators above the top 10% or top 5% productivity threshold, using model parameter η_i for models with Types 1 or 2 wedges and $(1 - \hat{\tau}_{gi})\eta_i$ for the model with the Type 3 wedge. Panel B reports empirical patterns, examining the probability of being in the top 10% or top 5% of innovators for phone applications (ranking by time usage of apps), consumer packaged goods (ranking by sales), and patents (ranking by citations). All ranks are computed within starting cohorts (founding year for phone apps, five-year groups for consumer packaged goods, and year of first filing for patent inventors).

Table C12: Price versus founding team composition

	Log Price	
	(1)	(2)
Female Founder Fraction	0.281** (0.129)	0.243* (0.144)
Constant	1.497*** (0.0336)	1.503*** (0.0233)
Fixed Effects	Group	Module
Sample Size	4044	4054

Notes: This table documents the relationship between founder gender and log unit prices within the consumer packaged goods sample. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C13: Productivity of New/Exiting Inventors with Changes in Access vs. Market Size

		Avg. Productivity	
		Removing Access Barriers	Equalizing Market Size
		(1)	(2)
Male Inventors, Male-Focused Sector	Enter	7.34	–
	Exit	–	7.578
Male Inventors, Female-Focused Sector	Enter	–	4.12
	Exit	5.43	–
Female Innovators, Male-Focused Sector	Enter	32.79	–
	Exit	–	7.56
Female Innovators, Female-Focused Sector	Enter	34.06	4.12
	Exit	5.43	–

Notes: This table presents additional statistics related to two of the counterfactuals discussed in the main text. For each counterfactual (removing access barriers in Column (1) and equalizing market size in Column (2)), we identify the individuals who enter and exit innovation in each sector between the status quo and the counterfactual scenario, and compute the average research productivity (parameter η_i) within each group. The model uses the Type 1 wedge.

Table C14: Features of the Baseline and Counterfactual Economies

	Baseline			Full Equal Access Counterfactual			Top 1% Equal Access 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)
<i>Fraction of innovators</i>									
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
<i>Average productivity of innovators</i>									
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 equal access” counterfactual, setting $\tau_i = 1$ for all women. Columns (7) to (9) describe the “top 1% equal access” scenario, with $\tau_i = 1$ for all women in the top 1% of the innovation productivity distribution.

Table C15: Baseline Parameters for the Analysis by Income Quintiles

<i>Panel A: Parameters calibrated outside of the model</i>			
Model Parameter	Parameter Definition	Source	Value
$ \alpha_{Q1} - \alpha_{Q5} $	Expenditure dissimilarity index by income quintiles	Consumer Expenditure Survey, NielsenIQ data, phone applications data (cf. Table C10, 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

<i>Panel B: Jointly estimated model parameters</i>			<i>Panel C: Targeted moments and model fit</i>		
Model Parameter	Parameter Definition	Value	Targeted Moment [Source]	Data	Model
τ	Barrier to enter innovation careers	0.121	Share of patent inventors in bottom vs. top parent income quintiles [Bell et al. (2019b)]	0.116	0.116
ϕ	Sectoral assignment	0.596	Income homophily regression coefficients [Table C9, column (1) and table notes]	0.176	0.176
$\bar{\eta}$	Pareto scale parameter of of innovators' productivity	0.011	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 C16: Model Predictions: Growth Impacts of Observed Changes in Women's Representation in Innovation Careers from 1985 to 2014

	1985	2014
Female inventor share (data)	0.050	0.128
τ	0.033	0.105
Growth	1.88pp	2.00pp
Cost-of-Living Inequality	22.20pp	18.70pp

Notes: This table presents results from the backwards looking counterfactual. 2014 quantities are based on the core calibration. 1985 equilibrium numbers are computed under a value of τ that generates a female innovator fraction of 0.05.

Table C17: Models with Endogenous vs. Exogenous Gender Wage Gap

Panel A: Growth and Cost-of-Living Counterfactuals

	Model with Exogenous Gender Wage Gap	Model with Endogenous Gender Wage Gap
Δ Cost-of-Living Inequality, Baseline vs. Equal Access	-18.20pp	-15.20pp
Δ Labor Productivity Growth, Baseline vs. Equal Access	+1.44pp	+1.39pp

Panel B: Calibration Parameters

	Model with Exogenous Gender Wage Gap	Model with Endogenous Gender Wage Gap
ϕ	0.7321	0.3957
τ	0.1050	0.1188
η	0.0114	0.0112

Panel C: Male vs. Female Earnings Ratios

	Model with Exogenous Gender Wage Gap	Model with Endogenous Gender Wage Gap
<u>Male-to-Female Earnings Ratio, Total Earnings</u>		
Baseline	1.205	1.568
Equal Access Counterfactual	1.205	1.212
<u>Share of Women in Innovation Sector</u>		
Baseline	12.5 %	12.5 %
Equal Access Counterfactual	50 %	50 %
<u>Earnings Ratios, Innovators vs. Production Workers</u>		
Baseline	N/A	4.93
Equal Access Counterfactual	N/A	4.63
<u>Male-to-Female Earnings Ratio for Production Workers</u>		
Baseline	1.205	1.205
Equal Access Counterfactual	1.205	1.205
<u>Male-to-Female Earnings Ratio for Innovators</u>		
Baseline	N/A	1.235
Equal Access Counterfactual	N/A	1.220

Notes: Equilibrium market size ratio and counterfactual growth and inequality results under a model that allows for endogenous market size. All results are medians over 100 simulations. For the model with endogenous wage gap, we apply a wage penalty (83%) to women who pursue entrepreneurship and production work.

Table C18: Homophily, Dissimilarity, and Consumption Shares

Panel A – NielsenIQ CPG

	Female HH			Female Wt			Share of startup rev	Fem. Innov. Frac.	#Firms
	β	Dissimilarity	Fem. Rev. Frac.	β	Dissimilarity	Fem. Rev. Frac.			
Food	0.048 (0.030)	0.21	0.24	0.036 (0.018)	0.08	0.50	0.84	0.14	85
Health and Household	0.147 (0.032)	0.31	0.27	0.132 (0.017)	0.12	0.52	0.13	0.08	52
General Merchandise	-0.007 (0.043)	0.33	0.25	-0.012 (0.042)	0.13	0.50	0.03	0.21	45

Panel B – Phone Apps

	β			Share of Startup Usage			#Firms
	Dissimilarity	Fem. Usage Frac.	Fem. Rev. Frac.	Dissimilarity	Fem. Usage Frac.	Fem. Rev. Frac.	
Lifestyle, Health and Leisure	0.097 (0.043)	0.31	0.54	0.45	0.12	0.795	795
Technology and Tools	0.112 (0.066)	0.18	0.56	0.42	0.066	0.532	532
Information and Commerce	0.053 (0.091)	0.25	0.55	0.13	0.067	0.352	352

Notes: These tables present gender homophily estimates in subsets of the product space, for consumer packaged goods (Panel A) and phone applications (Panel B). For each subset, we report the gender homophily coefficient (β), the consumption dissimilarity index between male and female consumers across all goods in the subset (“dissimilarity”), the fraction of revenue or usage from female consumers across both startup and other goods in the subset (“Fem. Rev. Frac.” and “Fem. Usage Frac.”), the share of startup revenue that the subset takes up, the female innovator fraction (“Fem. Innov. Frac.”) across all firms selling in that subset (not weighted by number of founders and firms can be in multiple areas), and number of firms. For consumer packaged goods, the outcome is measured using either female household head (“Female HH”) or weights within the households (“Female Wt”) to compute the share of sales to female customers.

Table C19: 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		0.0262*** (0.0058)	
Founder Age			0.127** (0.0532)
Product Module F.E.	Yes	Yes	Yes
Sample Size	Startups, $N = 4,058$	All manufacturers with patents, $N = 572,786$	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 0.273 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 C20: 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	0.0997** (0.0369)	0.0829** (0.0357)	0.0772* (0.0433)		
Female VC Fraction			0.152** (0.0761)		
Founder State				0.0875*** (0.0062)	0.0365*** (0.0079)
Fixed Effects	None	Subcategory	None	None	Subcategory
Sample Size	3211	3211	1391	43680	43392

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 C21: Innovator-Consumer Homophily for Phone Applications, Raw Usage (Logarithm of Time Use)

	Log Female Usage	
	(1)	(2)
Female Founder Fraction	0.0903 (0.715)	0.0267 (0.741)
Fixed Effects	None	Subcategory
Sample Size	$N = 3,211$	

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 $\log(1+\text{female usage})$, rather than usage shares. 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 C22: 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*** (0.000203)	0.0478*** (0.0042)							
Female Entrepreneur			0.0306*** (0.0003)						
Patent Inventor's Log Parent Income				0.0300*** (0.0145)					
Entrepreneur's Log Parent Income					0.0302*** (0.0110)	0.1418*** (0.0034)			
Patent Inventor Age							0.00439*** (0.000832)	0.0579*** (0.0057)	
Entrepreneur Age									0.0160*** (0.0010)
Country	U.S.	Finland	Finland	Finland	U.S.	Finland	U.S.	Finland	Finland
<i>N</i> industries	325	342	476	253	17	441	323	342	476
<i>N</i> individuals	2,217,864	9,643	344,698	3,812	300	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 C23: Gender Homophily with Price Controls

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0485** (0.0209)	0.0379** (0.0154)	0.0530*** (0.0202)	0.0416*** (0.0155)	0.0513** (0.0204)	0.0401*** (0.0154)
Unit Price	-0.000770 (0.000597)	-0.000862*** (0.000315)				
Log Price			-0.0233*** (0.00825)	-0.0204*** (0.00459)	-0.0266 (0.0224)	-0.0189 (0.0122)
Log Price ²					0.00931 (0.00865)	0.00814** (0.00406)
Log Price ³					-0.00310 (0.00226)	-0.00302*** (0.000891)
Log Price ⁴					0.000253 (0.000474)	0.000180 (0.000218)
Constant	0.257*** (0.00656)	0.541*** (0.00410)	0.285*** (0.0125)	0.564*** (0.00697)	0.279*** (0.0172)	0.556*** (0.00902)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	4054	4054	4054	4054	4054	4054

Notes: This table presents homophily estimates with flexible price controls, using the consumer packaged goods sample. The outcome is measured using either female household head (“Female HH”) or weights within the households (“Female Wt”) to compute the share of sales to female customers. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C24: NielsenIQ – Adding Basic Controls

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0485** (0.0212)	0.0376** (0.0157)	0.0540*** (0.0195)	0.0437*** (0.0154)	0.0615 (0.0398)	0.0802*** (0.0233)
Constant	0.247*** (0.00367)	0.532*** (0.00297)	0.286*** (0.0123)	0.564*** (0.00689)	0.289*** (0.0193)	0.571*** (0.00989)
Controls						
Module FEs	x	x	x	x	x	x
Log(Price), Size, Multipack, Organic, etc.			x	x	x	x
Indicators for other characteristics					x	x
Fixed Effects			Module			
Sample Size	4015	4015	4015	4015	4015	4015

Notes: This table documents homophily after controlling for specific sets of product characteristics, to assess the role of horizontal product differentiation. The second set of controls (“Log(Price), Size...”) includes log unit price, multipack indicator, product size, USDA organic indicator, and male/female targeted variables based on NielsenIQ Homescan data and the brand text approach in Moshary et al. (2023). We then include indicators for different values of the following variables in NielsenIQ’s extra product attributes file: flavor code, form code, formula code, container code, salt content code, style code, type code, product code, variety code, strength code, scent code, and target skin condition code. The outcome is measured using either female household head (“Female HH”) or weights within the households (“Female Wt”) to compute the share of sales to female customers. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C25: NielsenIQ – Adding Controls with LASSO

Panel A – Startup Sample

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Female Founder Fraction	0.0455** (0.0213)	0.0399*** (0.0144)	0.0510** (0.0214)	0.0434*** (0.0146)	0.0391 (0.0429)	0.0684** (0.0273)
<i>Available Controls (all with LASSO)</i>						
Module FEs	x	x	x	x	x	x
Log(Price), Size, Organic, etc.			x	x	x	x
Indicators for other characteristics with LASSO					x	x
Sample Size	4015	4015	4015	4015	4015	4015

Panel B – Broader Sample (50% of all firms)

	Share of Sales to Women					
	Female HH	Female Wt	Female HH	Female Wt	Female HH	Female Wt
Startup	-0.0409*** (0.00567)	-0.00818** (0.00380)	-0.0370*** (0.00568)	-0.00756** (0.00381)	-0.0289*** (0.00789)	-0.00917* (0.00510)
Startup × Female Founder Fraction	0.0573*** (0.0173)	0.0339*** (0.0121)	0.0580*** (0.0173)	0.0332*** (0.0120)	0.0573** (0.0233)	0.0421*** (0.0141)
<i>Available Controls</i>						
Module FEs	x	x	x	x	x	x
Log(Price), Size, Organic, etc.			x	x	x	x
Indicators for other characteristics					x	x
Sample Size	250367	250367	250361	250361	250361	250361

Notes: This table documents homophily after controlling for specific sets of product characteristics, under a penalized regression (LASSO) approach. We provide results from our core startup sample and from adding a random sample of 50% of non-startup manufacturers in order to better populate the characteristics space. The second set of controls (“Log(Price), Size...”) includes log unit price, multipack indicator, product size, USDA organic indicator, and male/female targeted variables based on NielsenIQ Homescan data and the brand text approach in Moshary et al. (2023). We then include indicators for different values of the following variables in NielsenIQ’s extra product attributes file: flavor code, form code, formula code, container code, salt content code, style code, type code, product code, variety code, strength code, scent code, and target skin condition code. The outcome is measured using either female household head (“Female HH”) or weights within the households (“Female Wt”) to compute the share of sales to female customers. In Panel B, “Startup” represents the difference between all-male founded startups and the overall module average and the interaction term represents the difference between female startups and male startups. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C26: Gender homophily within consumer subgroups, consumer packaged goods

Panel A – Female Weighted

	Share of Sales to Women					
	Education		Income		Age	
	College	Non-College	>= 60k	< 60k	45+	<45
Female Founder Fraction	0.0373** (0.0146)	0.0173 (0.0252)	0.0434*** (0.0133)	0.0166 (0.0205)	0.00641 (0.0191)	0.0557*** (0.0166)
Constant	0.526*** (0.00248)	0.553*** (0.00368)	0.502*** (0.00228)	0.593*** (0.00406)	0.545*** (0.00349)	0.530*** (0.00267)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	3867	3469	3797	3558	3870	3369

Panel B – Single Households

	Share of Sales to Women					
	Education		Income		Age	
	College	Non-College	>= 60k	< 60k	45+	<45
Female Founder Fraction	0.0787** (0.0370)	0.0317 (0.0721)	0.0555 (0.0506)	0.0547 (0.0354)	0.0232 (0.0522)	0.103** (0.0422)
Constant	0.673*** (0.00577)	0.718*** (0.00801)	0.638*** (0.00724)	0.725*** (0.00626)	0.686*** (0.00725)	0.720*** (0.00670)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	2666	2388	2217	2772	2844	2014

Notes: This table investigates heterogeneity in gender homophily depending on characteristics of the customers, in the consumer packaged goods sample. Panel A uses weights within the households to compute female consumer fraction. Panel B focuses households with one member. The column labels note the demographic variable and the subgroup of consumers within which we compute the female consumer fraction (e.g., within the set of college-educated households). The number of data points varies across subgroups because some products do not have any purchasing by households from that group. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C27: Gender homophily within consumer subgroups, phone applications

	Female Usage Share					
	Education		Income		Age	
	College	Non-College	>= 75k	< 75k	45+	<45
Female Founder Fraction	0.144*** (0.0420)	0.0320 (0.0374)	0.128*** (0.0402)	0.0453 (0.0406)	0.0437 (0.0425)	0.0981*** (0.0365)
Constant	0.504*** (0.0112)	0.562*** (0.0122)	0.482*** (0.00947)	0.568*** (0.0129)	0.546*** (0.0114)	0.526*** (0.0118)
Fixed Effects	Subcategory	Subcategory	Subcategory	Subcategory	Subcategory	Subcategory
Sample Size	2,364	2,812	2,141	2,921	2,443	2,770

Notes: This table investigates heterogeneity in gender homophily depending on characteristics of the customers, in the sample of phone applications. The column labels note the demographic variable and the subgroup of consumers within which we compute the female consumer fraction (e.g., within the set of college-educated households). The number of data points varies across subgroups because some products do not have any purchasing by households from that group. Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C28: Innovator-Consumer Homophily across Industries by Sector

	Patent Inventors			Entrepreneurs				
	(1) Manuf.	(2) Manuf.	(3) Other	(4) Other	(5) Manuf.	(6) Manuf.	(7) Other	(8) Other
A. Outcome: Female Consumption Share								
Female	0.0332*** (0.0039)	0.0035* (0.0019)	0.0540*** (0.0045)	0.0076*** (0.0023)	0.0680*** (0.0017)	0.0127*** (0.0015)	0.0254*** (0.0003)	0.0071*** (0.0002)
Constant	0.5444*** (0.0011)	0.5465*** (0.0007)	0.6284*** (0.0017)	0.6332*** (0.0008)	0.5505*** (0.0008)	0.5640*** (0.0006)	0.6328*** (0.0002)	0.6394*** (0.0001)
2-digit industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
N individuals	6945	6945	3082	3082	28147	28147	322613	322613
N industries	21	21	41	41	22	22	42	42
B. Outcome: Income Elasticity								
Log Parent Income	0.0289** (0.0132)	0.0168* (0.0099)	0.0302 (0.0286)	0.0030 (0.0059)	0.0193** (0.0089)	0.0087* (0.0050)	0.1487*** (0.0035)	0.0062*** (0.0012)
Constant	0.9496*** (0.1471)	1.0839*** (0.1101)	0.9634*** (0.3186)	1.2643*** (0.0669)	0.9186*** (0.0961)	1.0324*** (0.0541)	-0.4779*** (0.0384)	1.0526*** (0.0134)
2-digit industry FEs	No	Yes	No	Yes	No	Yes	No	Yes
N individuals	2769	2769	1194	1194	6675	6675	93829	93829
N industries	21	21	39	39	20	20	42	42

Notes: This table presents homophily estimates for different sectors in Finland. We present the results separately for patent inventors (Columns 1-4) and entrepreneurs (Columns 5-8) in manufacturing and non-manufacturing sectors. Each even column includes fixed effects for 2-digit industries. For gender homophily (Panel A), we observe consistently positive and statistically significant coefficients, indicating that female patent inventors and entrepreneurs cater more to female consumers within sectors and industries. The fixed effects estimates are considerably smaller but statistically significant, meaning that while gender homophily is prevalent also within 2-digit industries, variation across industries accounts for a significant fraction of it. For income homophily (Panel B), all coefficients are positive, with six out of ten being statistically significant at the 10 percent or lower risk level. The estimate for patent inventors in the non-manufacturing sector in Column 3 has low precision due to limited sample size (this sample includes only patent inventors in non-manufacturing sector whose parental income is observed). However, the point estimate of 0.0302 is almost equivalent to the baseline estimate of 0.0304 in the full sample in the main text. Standard errors are clustered at the individual level.

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

Table C29: Innovator-Consumer Homophily across Industries by Innovator Type

	Patent Inventors			Entrepreneurs		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Outcome: Female Consumption Share						
Female	0.0447*** (0.0083)	0.0510*** (0.0109)	0.0456*** (0.0044)	0.0312*** (0.0004)	0.0262*** (0.0010)	0.0305*** (0.0006)
Female × College	0.0049 (0.0092)			-0.0039*** (0.0010)		
Female × Age > 40		0.0084 (0.0068)			-0.0006 (0.0007)	
Female × High-income			-0.0075 (0.0120)			0.0106*** (0.0013)
Sample Size	9643	9643	3812	344698	344698	99189
B. Outcome: Income Elasticity						
Log Parent Income	0.0213 (0.0248)	0.0289** (0.0129)	0.0278** (0.0130)	0.1396*** (0.0037)	0.1596*** (0.0043)	0.1363*** (0.0034)
Log Parent Income × College	0.0057 (0.0284)			0.0160* (0.0089)		
Log Parent Income × Age > 40		0.0214 (0.0351)			0.0532*** (0.0090)	
Log Parent Income × Female			0.0142 (0.0463)			-0.0452*** (0.0067)
Sample Size	3812	3812	3812	99189	99189	99189

Notes: This table estimates heterogeneity in homophily by innovator type in Finland. We present the results separately for patent inventors (Columns 1-3) and entrepreneurs (Columns 4-6). Panel A investigates heterogeneity in gender homophily by whether the patent inventor/entrepreneur holds a college/university degree (Columns 1 and 4), is more than 40 years old (Columns 2 and 5), and has parental income above the median (Columns 3 and 6). Panel B investigates heterogeneity in income homophily by the same characteristics, except in Columns 3 and 6 the interaction is with a binary indicator for female. Standard errors are clustered at the individual level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C30: Gender Homophily by Founding Team Composition

	Share of Sales to Women					
	All Startups			Multi-founder Startups		
	CPG (HH) (1)	CPG (Wt) (2)	Phone Apps (3)	CPG (HH) (4)	CPG (Wt) (5)	Phone Apps (6)
1 Female Founder	0.0283** (0.0142)	0.0260*** (0.00946)	0.0870*** (0.0184)	0.00515 (0.0271)	0.0115 (0.0176)	0.0788*** (0.0217)
2+ Female Founders	0.0168 (0.0553)	-0.0113 (0.0369)	-0.0155 (0.0516)	0.0367 (0.0701)	-0.0137 (0.0454)	-0.0231 (0.0504)
Constant	0.251*** (0.00453)	0.533*** (0.00303)	0.528*** (0.0112)	0.242*** (0.00824)	0.530*** (0.00533)	0.533*** (0.0122)
Fixed Effects	Module	Module	Module	Module	Module	Module
Sample Size	4054	4054	3211	1595	1595	2057

Notes: This table investigates heterogeneity in homophily depending on team composition, using the consumer packaged goods on phone apps datasets. To examine whether one woman in the team is sufficient, we regress the market share of female customers (“HH” refers to measurement based on female head of household and “Wt” refers to measurement based on overall gender composition of the household) on dummies for the number of female founders, specifically whether the firm has one female founder or more than one. Columns (1) to (3) consider all startups, while Columns (4) to (6) focus on the subset of startups with more than one founder. For phone applications, Columns (3) and (6) show that it is sufficient to have at least one female founder for the application to be more tailored towards female customers. The other columns show that, for consumer packaged goods, the patterns appear to be driven by single founders. Indeed, in columns (4) and (5) the point estimates become much smaller in magnitude (although we cannot rule out substantial homophily given the size of the standard error). Standard errors are clustered at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C31: Homophily by Team Composition, Finland

Outcome:	Patent Inventors		Entrepreneurs	
	Female Share	Income Elasticity	Female Share	Income Elasticity
1 Female Inventor	0.03423*** (0.0090)		0.0346*** (0.0105)	
2+ Female Inventors	0.0458*** (0.0114)		0.0488*** (0.0114)	
1 Low-Income Inventor		-0.0157 (0.0307)		-0.0221 (0.0338)
2+ Low-Income Inventors		-0.0579* (0.0313)		-0.0560* (0.0313)
Constant	0.5396*** (0.0029)	1.2504*** (0.0164)	0.5366*** (0.0035)	1.2670*** (0.0134)
N firms	2172	1060	1442	867

Notes: This table examines whether homophily varies with team composition in Finland. Inventor teams are defined based on firm-year cells. We find that, in line with the the U.S. results for founder teams, inventor teams with at least one female member cater more female-intensive consumer markets, on average. We find similar results for entrepreneurs. Moreover, unlike for U.S. founder teams, the degree of homophily is not tied to having at least one female inventor or entrepreneur in the team, as having additional female patent inventors or entrepreneurs in the team further strengthens it. We recover similar pattern for entrepreneurs from low-income backgrounds, with homophily increasing as we move from teams with a single low-income member to teams with multiple low-income members. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.