Supplementary Materials for

The value of complementary co-workers

Frank M. H. Neffke*

*Corresponding author. Email: fneffke@gmail.com

Published 18 December 2019, Sci. Adv. 5, eaax3370 (2019)
DOI: 10.1126/sciadv.aax3370

This PDF file includes:

Section SA. Data
Section SB. Definition of measures
Section SC. Co-worker synergy and substitutability by industry and occupation (Fig. 2)
Section SD. Wages
Section SE. Career paths (Fig. 3)
Section SF. Wage premiums
Fig. S1. Part-time work in Sweden.
Fig. S2. Scatter plot of educational synergy against educational substitutability.
Fig. S3. Educational synergies network.
Fig. S4. Educational substitutability network.
Fig. S5. Selected synergy communities in the synergy and substitutability graphs.
Fig. S6. Nested community structure.
Fig. S7. Levels and contents of education in synergy and substitutability communities.
Fig. S8. Co-worker variables and establishment size.
Fig. S9. Average co-worker synergy and substitutability by industry and occupation.
Fig. S10. SD synergy (observed versus theoretical).
Fig. S11. Estimated effect of co-worker synergy by error-variance bin.
Fig. S12. Estimated effect of co-worker synergy by error-variance bin (homoscedastic case).
Fig. S13. Co-worker complementarity and work experience.
Table S1. Sample sizes.
Table S2. Educational pairs: Top synergy.
Table S3. Educational pairs: Top substitutability.
Table S4. Educational pairs: Top synergy, controlling for substitutability.
Table S5. Descriptive statistics main variables.
Table S6. Survival analysis.
Table S7. Co-worker educational fit and wages.
Table S8. Wage regressions, fixed effects models.
Table S9. IV regression, first differences.
Table S10. Wage effects of within-education local supply shifts.
Table S11. Fixed effects models, within transformation versus first differences.
Table S12. Errors-in-variables estimates.
Table S13. Wage regressions: Summary.
Table S14. Returns to education.
Table S15. Large-plant premium.
Table S16. UWP, conditional on co-worker fit.
References (35–40)
Section SA. Data

All analyses are based on employer-employee linked data derived from Sweden’s official registries as provided by Statistics Sweden (SCB). The data are described in detail in (22). Access to these data can be requested at https://www.scb.se/en/services/guidance-for-researchers-and-universities. The dataset contains yearly individual-level observations for the entire population of Sweden. It records a number of sociodemographic characteristics, such as gender, age, and municipality of residence, together with work-related variables, such as the individual’s main establishment of work and annual wage-income. Apart from individual-level characteristics, the data also contain detailed (5-digit) industry and municipality codes for all work establishments in the Swedish economy.

Unfortunately, there is no information on total hours worked. Consequently, it is impossible to calculate hourly wage rates. Variation in annual wages may therefore reflect differences in hourly pay or differences in hours worked. The complications that arise from this, as well as the solutions chosen to deal with them are explained below.

Occupations are coded into one of 335 4-digit occupation classes of the Swedish occupational classification system, Standard för svensk yrkesklassificering 1996, which derives from the United Nations’ ISCO-88 classification. Examples include “1231: Personnel management” and “7232: Aircraft mechanics.” Occupational data are available from 2001 on for about 90% of the working population. For this reason, the analysis concentrates on the period 2001-2010. However, the data before 2001 are still used to construct work experience variables. Finally, and, from the paper’s perspective, most importantly, from 1990 to 2010, the dataset contains detailed information on each individual’s highest absolved education. These data are provided to SCB on a yearly basis by the country’s educational institutions, or, in the case of immigrants, collected by SCB through surveys.

A.1 Samples

The data are split into two separate samples to avoid mechanical relations between measurement and effect estimates. The first data set, the measurement sample, consists of a 75% random sample of the Swedish working population. To characterize the full set of interactions among coworkers, only minimal restrictions are imposed on this sample. First, workers who are self-employed are dropped, given that these workers do not have coworkers. Second, to avoid counting irregular work, workers are dropped if they earn below the subsistence income as defined by Statistics Sweden. Third the sample is limited to working age adults, aged 18-65. Fourth, workers for whom the establishment (and therefore their coworkers) is unknown are dropped from the data set. This happens, for instance, when workers are employed through employment agencies or when workers are not assigned to an establishment because they have no fixed work location (such as itinerant repairmen).

The remaining 25% of the data set, the estimation sample, is used to estimate the effects of coworker synergy and coworker substitutability on careers and wages. Because government wages tend to be more regulated than private-sector wages, only employees of private-sector firms are retained in this sample. Furthermore, workers with missing educational or establishment information, workers with annual wages below the subsistence income, workers with rare education codes and workers employed through employment agencies are dropped from this sample. Rare educational tracks have fewer than 500 individuals a year in the measurement sample. Dropping workers with such education ensures that the coworker fit measures are based on a reasonably large number of individuals. Finally, to remove outliers with extreme wages, workers below the 0.5th and over the 99.5th wage percentile are excluded from the estimation data set as well.
The lack of information on how many hours an employee works in a given year causes particular concern when part-time work is common. Fig. S1 provides information on labor force participation and part-time work in Sweden, taken from the Swedish Labor Surveys as provided by Statistics Sweden. It shows that, as individuals exit their early twenties, labor force participation rises quickly to above 90% for men. However, it stays well below this mark for women, leading to a gender gap in participation rates of about 7 percentage points. This may raise sample-selection issues, because workers will self-select into the labor force. More concerning in light of the annual nature of the wage data is that those women who are gainfully employed tend to work shorter hours than men.

Among prime-age workers in active employment, men work on average 37 hours per week (Fig. S1b), which is close to the regular 40-hours work week as defined in Sweden’s Working Hours Act. In contrast, prime-age employed women work, on average, only 29 hours a week. This difference seems to be related to raising children: Fig. S1c shows that among parents with children of up to eighteen years old, 93% of men are full-time employed, against just over half (56%) of women.

The large prevalence of part-time employment among women makes it hard to accurately describe their career progression using the SCB data. For instance, any productivity premium that workers may obtain from complementarity-rich work environments may not be reflected in higher annual wages, but in the opportunity to negotiate flexible work arrangements. This will be particularly attractive to workers who assume child care responsibilities. Such choices will lead to a misrepresentation of the progression of a person’s productivity in the available annual wage data. Fig. S1 suggests that the risk of such distortions are much more pronounced among women than among men and among workers close to the retirement age, regardless of gender. To minimize the statistical problems caused by part-time employment and self-selection into the labor force, the estimation sample is restricted to the subsample of prime-age (i.e., 20-60 years old) male workers among whom part-time employment and non-participation are less prevalent. This leaves a data set of about 365,000 employees, observed for up to ten years, in which the paper’s hypotheses about career and wage dynamics are tested.

Table S1 provides the sizes of subsamples of workers with different levels of education in the estimation sample. For completeness, it also reports the number of workers with only primary or secondary education, although this part of the sample will not be used in the main analysis. Because 17% of workers have a college degree or higher, whereas workers with only primary school degrees represent 3% of the sample, moving from a primary school to a college education amounts to an 80-percentiles rise in educational attainment. Consequently, the shift from a variable’s 10th to its 90th percentile is comparable to the shift from primary to college education. This provides a useful anchor to interpret effect sizes: where convenient, parameter estimates are described in terms of the implied effect of moving from the 10th to the 90th percentile in a variable’s distribution.
# observations | share
---|---
primary | 83,390 | 3.2%
sec. | 348,609 | 13.5%
upper sec. | 1,522,026 | 59.1%
post-sec. | 184,033 | 7.1%
college | 426,827 | 16.6%
post-grad. | 12,079 | 0.5%

Table S1. Sample sizes. Sizes of educational-level subsamples of the estimation sample. Observations represent worker-year combinations.

## A.2 Educational information

The educational information in the data consists of two components. The first component is an alphanumeric code that divides educations into 351 different fields ("content" or "expertise" areas), such as “344z: Accounting and taxation,” “214a: Fashion design” or “725f: Radiology nursing program.” The second component is a 3-digit code that distinguishes 49 different levels of education, such as “337: Vocationally oriented program, three years” or “640: Doctoral program.” However, because at the 2-digit and 3-digit levels of aggregation, educational levels cannot be unambiguously ranked in increasing order of length or complexity, this study only uses the first digit of this code. This first digit distinguishes among six levels of education: primary school, secondary school, upper secondary school, post-secondary education, tertiary (henceforth referred to as “college”) education and post-graduate education. Next, 1-digit educational level information and 4-digit educational field identifiers are combined to create 491 educational tracks.

## Section SB. Definition of measures

### B.1 Educational synergy

To measure synergies between educational tracks, the paper relies on the assumption that, if workers with two different types of expertise are often found to be working together, it is likely that these types of expertise are synergistic. Ideally, we would have information on which workers collaborate as teams within an establishment. However, most establishments in the economy are small. Therefore, the workforce of an establishment can be approximately thought of as a team of coworkers.

To avoid conflating educations that are just very ubiquitous in the economy with highly synergistic educations, we need to define what it means that an educational track is substantially (i.e., more than expected given its overall size) present in an establishment. Let $E_{ep}$ be the number of workers with education $e$ in establishment $p$. $e$'s presence in $p$ is defined as:

$$P_{pe} = \begin{cases} 
1 & \text{if } \frac{E_{pe}}{E_p} > 1 \\
0 & \text{elsewhere} 
\end{cases}$$

(S1)

where omitted subscripts indicate summations over the corresponding dimension. Eq. (S1) says that $e$ is present in $p$, denoted by
$P_{ep} = 1$, if the share of workers with education $e$ in establishment $p$ is at least as large as their share in the economy as a whole. The quantity $\frac{E_{pe}/E_p}{E_e/E_e}$ is widely used in trade (where it is known as revealed comparative advantage), economic geography (known as location quotient) and computer science (lift) as a measure of overrepresentation. The definition of what constitutes a presence based on this quantity follows (14) as well as the economic complexity literature (2,23), which assesses in an analogous way which products are significantly present in the export basket of a country.

Synergies between educations are now inferred from co-occurrence counts: in how many establishments do both educations have a significant presence? However, the number of co-occurrences that an establishment generates rises in proportion to the square of the number of different educations it hosts. As a consequence, co-occurrence counts will be dominated by large establishments. To avoid this, co-occurrences are normalized such that each establishment contributes a total of one co-occurrence (a synergy variable created without this step is highly correlated ($\rho = 0.977$, $N = 102,655$) with the one used in this study):

$$N_{ee'} = \sum_p \frac{P_{pe} P_{pe'}}{\sum_a \sum_b P_{pa} P_{pb}}$$

In a final step, the co-occurrence count is compared to a random benchmark to account for the fact that some educations are present in more establishments than others. Under the assumption that co-occurrences form at random, the expected value of $N_{ee'}$ is:

$$\hat{N}_{ee'} = N_e \frac{N_{e'}}{N}$$

Normalizing observed with expected co-occurrences yields:

$$\tilde{c}_{ee'} = \frac{N_{ee'}}{N_{ee'}}$$

which is essentially the same quantity as what was used in eq. (S1) to define presences.

The ratio expressed by $\tilde{c}_{ee'}$ has a strongly skewed distribution: for overrepresented educational combinations, $\tilde{c}_{ee'}$ ranges from 1 to infinity, whereas for underrepresented combinations, $\tilde{c}_{ee'}$ lies between 0 and 1. To address this, the following monotonous transformation of $\tilde{c}_{ee'}$ is applied:

$$c_{ee'} = \frac{\tilde{c}_{ee'}}{\tilde{c}_{ee'} + 1}$$

which maps $\tilde{c}_{ee'}$ symmetrically around 0.5 onto the interval $[0,1)$. This is eq. (2) in the main text. Henceforth, $c_{ee'}$ will be referred to as the synergy between educations $e$ and $e'$. Furthermore, to mimic the educational substitutability defined below, educations are assumed to be fully synergistic with themselves: $c_{ee} := 1$.

### B.2 Educational substitutability

As pointed out in the main text, many workers who work together end up doing so, not because they complement one another, but because they can substitute for one another. The educational synergy defined above may therefore result from either circumstance.
However, having complementary coworkers may arguably have markedly different consequences than having coworkers that can easily substitute the worker. Therefore, it is important to be able to assess the extent to which the estimated synergy to coworkers actually reflects a substitutability by coworkers.

From an employer’s perspective, two educational tracks are substitutes if workers with either track can carry out the same tasks. Because occupations are essentially bundles of tasks, the substitutability of two educational tracks should be reflected in the occupational opportunities they give access to. To quantify the similarity between such occupational opportunities, let $E_{oe}$ represent the number of workers with education $e$ who work in occupation $o$. Using the Pearson correlation as a similarity metric, the degree to which education $e$ can substitute for education $e'$ is measured as (see also eq. (3) in the main text):

$$s_{ee'} = \text{corr}(E_{oe}, E_{oe'})$$  (S3)

where correlations are taken over all 355 occupations in the economy. Note that, according to this definition, an educational track is a perfect substitute to itself: $s_{ee} = 1$.

In principle, educational synergies could be quantified in a similar way. To see this, let matrix $P$ collect elements $P_{pe}$. This matrix will have tens of thousands of rows, one for each establishment. Because workers are spread across many establishments, most rows will contain only a handful of the 491 educational tracks. Matrix $P$ is therefore very sparse, especially when compared to matrix $E$, which collects elements $E_{oe}$. This latter matrix has just 355 rows, one for each occupation and is therewith much less sparse than $P$.

The sparsity of $P$ suggests that there is little information in the fact that two educations are simultaneously absent from an establishment. This is why educational synergy, $c_{ee'}$, is based on co-occurrences, which emphasizes co-presences over co-absences. If, instead, we were to calculate the correlations between the columns of $P$, we would weight the information on co-presences and co-absences of educations in establishments equally. By contrast, looking only at the occupational opportunities that a worker has, while ignoring the ones she does not have would waste valuable information on how similar the occupational opportunities are that are associated with two different educational tracks. This is why $s_{ee'}$ is based on the correlation between occupational vectors, not on the number of occupations in which two educations are significantly overrepresented (i.e., in which they co-occur).

### B.3 Comparing educational synergy and educational substitutability

Tables S2 and S3 show which educational pairs have the highest synergy and substitutability. High synergy values often coincide with high levels of substitutability. Apparently, educations that often co-occur in establishments also tend to give access to similar occupations. However, the overlap is far from perfect. For instance, Table S2 shows that, although workers with a background in agricultural management often work with workers with degrees in agricultural science, they cannot easily substitute for one another.

To further illustrate the difference between synergy and substitutability, Table S4 shows the educational tracks with the highest synergy, controlling for how substitutable these tracks are. To be precise, it ranks educational pairs by the residual of a non-parametric regression of $c_{ee'}$ on $s_{ee'}$. The table shows that some educations with high synergy may be poor substitutes. This non-parametric regression is also shown in the scatter plot of Fig. S2.
Table S2. Educational pairs: Top synergy. Pairs of educational tracks with the strongest synergies that have at least 250 employees a year in the estimation sample. For each of nine broad content areas, the top two pairs are shown. Numbers preceding an educational track’s name represent educational levels: 1: primary school; 2: secondary school; 3: upper secondary school; 4: post-secondary education; 5: college; 6: post-graduate. Column “syn.” reports the synergy of the two educational tracks as measured in eq. (2), column “subst.” their substitutability as measured in eq. (3) of the main text.

| Rank | Edu. (1) | Edu. (2) | Syn. | Subst. |
|------|----------|----------|------|--------|
| 1    | General, Other | General, Other | 0.773 | 0.866 |
| 2    | General, Other | Humanities, Other | 0.757 | 0.797 |
| 1    | Teaching Methods | Other Psychology | 0.971 | 0.954 |
| 2    | Teaching Methods | Sociology | 0.976 | 0.989 |
| 1    | Pastoral | Divinity | 0.976 | 0.286 |
| 2    | Humanities, Other | Art & Media, Other | 0.979 | 0.967 |
| 1    | Economics | Political Science | 0.985 | 0.840 |
| 2    | Political Science | Sociology | 0.985 | 0.975 |
| 1    | Biochemistry | Pharmacists | 0.971 | 0.283 |
| 2    | Biochemistry | Other Pharmacy | 0.971 | 0.522 |
| 1    | Other Construction | Other Architecture | 0.967 | 0.918 |
| 2    | Construction Engineering | Construction Engineering | 0.970 | 0.920 |
| 1    | Veterinary Science | Other Animal Health | 0.984 | 0.571 |
| 2    | Msc Forestry | Forest Management | 0.981 | 0.576 |
| 1    | Pharmacists | Dispensers | 0.992 | 0.175 |
| 2    | Other Pharmacy | Dispensers | 0.991 | 0.846 |
| 1    | Water Transport | Water Transport | 0.972 | 0.681 |
| 2    | Staff Officers | Higher Officers | 0.980 | 0.998 |

The aforementioned scatter plot reveals another interesting fact about educational synergy and substitutability: the relation is asymmetric. In particular, whereas a high substitutability almost certainly means that two educations are synergistic, high levels of synergy can exist between educations with widely varying levels of substitutability. That is, whereas workers who have very similar skills often work together, working together does not imply that workers have very similar skills. This asymmetry is explored in greater detail in the next subsection.

B.4 Networks (Fig. 1)

Figs S3 and S4 show enlarged and labeled versions of the networks in Fig. 1 of the main text. The edges in Fig. S3 reflect the estimated synergy between two educational tracks as defined in eq. (2) of the main text. That is, they reflect the degree to which educational pairs are overrepresented in establishment co-occurrences. The edges in Fig. S4 reflect the substitutability of two educational tracks as defined in eq. (3) of the main text. That is, they show the extent to which two educations give access to jobs in the same occupations.

The synergy network of S3 only shows edges that are statistically significantly overrepresented at the 1% level (i.e., where $c_{ee'}$ significantly exceeds 0.5). Confidence levels are based on eq. (S31) in section D.5. Because testing for overrepresentation takes
Table S3. Educational pairs: Top substitutability. Idem Table S2, showing the most substitutable educational pairs.

| Rank | Edu. (1)                | Edu. (2)                | Syn  | Subst |
|------|-------------------------|-------------------------|------|-------|
| 1    | 3: General, General     | 3: Other Nursing, General | 0.689| 0.976 |
| 2    | 3: General, General     | 3: Other Therapy        | 0.654| 0.970 |
|      | **General education**   |                         |      |       |
| 1    | 5: Teacher, Lower, Swedish | 5: Teacher, Lower, Science | 0.903| 0.999 |
| 2    | 5: Teacher, After-School| 5: Teacher, Pre-School  | 0.833| 0.999 |
|      | **Teaching methods and teacher education** |          |      |       |
| 1    | 6: Other Languages      | 6: Art & Media, Other   | 0.971| 0.992 |
| 2    | 6: History              | 6: Art & Media, Other   | 0.975| 0.991 |
|      | **Humanities and art**  |                         |      |       |
| 1    | 6: Other Languages      | 6: Art & Media, Other   | 0.971| 0.992 |
| 2    | 6: History              | 6: Art & Media, Other   | 0.975| 0.991 |
|      | **Social sciences, law, commerce, administration** | |      |       |
| 1    | 3: Medical Secretary    | 5: Medical Secretary    | 0.910| 1.000 |
| 2    | 5: Psychology           | 5: Psychology, Incl. Internship | 0.955| 1.000 |
|      | **Natural sciences, mathematics and computing** | |      |       |
| 1    | 5: Software Engineering | 4: Computing, General   | 0.829| 0.991 |
| 2    | 4: Computing, General   | 5: Computing, General   | 0.841| 0.993 |
|      | **Engineering and manufacturing** | |      |       |
| 1    | 5: Electronics          | 5: Other Electronics    | 0.850| 0.982 |
| 2    | 6: Other Engineering    | 6: Physics              | 0.964| 0.985 |
|      | **Agriculture and forestry, animal health** | |      |       |
| 1    | 6: Other Agriculture    | 6: History              | 0.852| 0.924 |
| 2    | 3: Horticulture         | 3: Other Horticulture   | 0.896| 0.924 |
|      | **Health care and nursing, social care** | |      |       |
| 1    | 5: Internal Medicine    | 5: Surgery              | 0.970| 1.000 |
| 2    | 5: Surgery              | 5: General Medicine     | 0.937| 1.000 |
|      | **Services**            |                         |      |       |
| 1    | 5: Tactical Military    | 5: Professional Officers| 0.875| 0.999 |
| 2    | 5: Staff Officers       | 5: Tactical Military    | 0.940| 0.999 |

Fig. S2. Scatter plot of educational synergy against educational substitutability. Scatter plot of synergy against substitutability edge weights. The triangular shape means that strong substitutability implies synergy, but not vice versa: synergy exists at any degree of substitutability.
### Educational Pairs: Top Synergy, Controlling for Substitutability

| Rank | Edu. (1)                          | Edu. (2)                          | Syn. | Subst. |
|------|-----------------------------------|-----------------------------------|------|--------|
| 1    | 3: General, General               | 5: Nursing, Geriatric             | 0.615| 0.028  |
| 2    | 3: General, Natural Science       | 5: Veterinary Science             | 0.579| -0.017 |
|      | **Teaching methods and teacher education** |                                |      |        |
| 1    | 5: Teacher, Commerce              | 5: Vocational Guidance            | 0.917| 0.030  |
| 2    | 5: Teacher, Nat. Resources        | 5: Veterinary Science             | 0.914| 0.007  |
|      | **Humanities and art**            |                                   |      |        |
| 1    | 5: Divinity                       | 5: Music                          | 0.960| -0.000 |
| 2    | 5: Pastoral                       | 5: Music                          | 0.945| 0.022  |
|      | **Social sciences, law, commerce, administration** |                          |      |        |
| 1    | 5: Medical Secretary              | 5: General Medicine               | 0.951| -0.003 |
| 2    | 5: Psychology                     | 5: Psychiatry                     | 0.955| 0.001  |
|      | **Natural sciences, mathematics and computing** |                          |      |        |
| 1    | 6: Biology                        | 5: Biomedical                      | 0.900| 0.021  |
| 2    | 5: Biochemistry                   | 5: Paediatrics                    | 0.881| 0.023  |
|      | **Engineering and manufacturing** |                                   |      |        |
| 1    | 3: Construction Engineering      | 6: Other Architecture             | 0.905| 0.037  |
| 2    | 5: Architecture                   | 5: Environmental Protection       | 0.883| 0.029  |
|      | **Agriculture and forestry, animal health** |                          |      |        |
| 1    | 3: Animal Care                    | 5: Veterinary Science             | 0.965| 0.003  |
| 2    | 3: Animal Husbandry               | 5: Veterinary Science             | 0.941| -0.003 |
|      | **Health care and nursing, social care** |                          |      |        |
| 1    | 5: Nursing, Radiology             | 5: Radiology                      | 0.990| -0.003 |
| 2    | 5: Dental Surgery                 | 5: Dental Hygiene                 | 0.990| 0.000  |
|      | **Services**                      |                                   |      |        |
| 1    | 4: Occupational Safety            | 5: General Medicine               | 0.841| -0.008 |
| 2    | 3: Fire-Fighting                  | 5: Structural Engineering         | 0.918| 0.052  |

Table S4. Educational pairs: Top synergy, controlling for substitutability. Idem Table S2, but now showing the strongest synergies, controlling for substitutability.
place in hundreds of thousands of edges, significance levels are Bonferroni adjusted. To keep the graph fully connected, all edges that are on the maximum spanning tree are retained regardless of their statistical significance. This procedure yields a network with 3,283 edges. As shown in section D.5, noise causes less concern in the substitutability network. However, to maximize comparability between the two networks, also for this network only the strongest 3,283 edges are retained.

Both networks exhibit clear communities, often organized around common educational contents. However, the communities in the substitutability network are smaller and tighter than the ones in the synergy network. What is more, substitutability communities are often to be nested in synergy communities. To illustrate this, Fig. S5a highlights the nodes that belong to the health care community of the synergy network, once in the synergy and once in the substitutability network. Educations of this synergy community clearly form a number of separate communities in the substitutability network, such as college-level educations in medicine, specialized nursing, therapeutics and psychiatric care (pink) and upper secondary and college degrees in health-care administration (orange). Although the educations from these various substitutability communities are arguably synergistic, they are sufficiently distinct to be poor substitutes for one another. Figs S5b and S5c show that the same holds for the synergy community around teaching (light-blue), which contains large sections of four different substitutability communities and for the synergy community around electronics and physics educations, which contains nodes from two separate substitutability communities.

B.4.1 Communities

To explore this nesting of substitutability communities within synergy communities systematically, Fig. S6 uses the infomap community detection algorithm (39). The algorithm analyzes a network's community structure through the lens of information theory. It calculates the average length of the binary code needed to describe the path of a random walker on the graph. If graphs have tight communities, the code length can be reduced by leveraging this community structure, because the random walker will make most steps within communities rather than between them. The infomap algorithm aims to partition nodes into communities in a way that minimizes average code length.

In the synergy network, the community structure allows reducing the average code length from 8.2 to 6.7 bits. In the substitutability network, code lengths are reduced from 8.1 to 5.7 bits. This confirms that the substitutability graph exhibits more tightly defined communities than the synergy graph.

The fact that substitutability communities are often largely nested in synergy communities can be seen in Fig. S6. It shows the histograms of the shares of nodes that belong to a synergy community’s modal substitutability community and vice versa. The green line represents a random benchmark. This benchmark is created by reshuffling the substitutability communities to which a node belongs 1,000 times, while keeping the node’s synergy community as is. This operation breaks the connection between substitutability and synergy communities, but maintains the size distributions of both types of communities. The Kolmogorov-Smirnov test rejects the hypothesis that the histograms are drawn from the same distribution.

The histograms show that often (in about a third of all communities) all of a substitutability community’s nodes belong to the exact same synergy community. Furthermore, in 21 out of 54 substitutability communities, 80% of the nodes belong to the same synergy community. This is highly statistically significant: on average, this happens in fewer than one of the simulated communities of the
Fig. S4. Educational substitutability network.
Fig. S5. Selected synergy communities highlighted in the synergy and substitutability graphs.

Fig. S6. Nested community structure. Histogram of the maximum number of nodes of a synergy community that belong to the same substitutability community (left) and vice versa (right). Dotted green lines show a random benchmark created by randomly reshuffling substitutability community membership 1,000 times. N: number of synergy communities, p80: number of synergy communities for which at least 80% of nodes also belong to the modal substitutability community, ΔCL: improvement in infomap code length. KS: Kolmogorov-Smirnov statistic for whether or not the observed histogram follows the same distribution as the simulated histogram. p: p-value of KS test.
benchmark. In the reverse case, for only 4 out of 39 synergy communities do 80% of the nodes belong to the same substitutability community. This shows that, whereas substitutability communities are often largely nested in synergy communities, synergy communities tend to consist of nodes that belong to multiple substitutability communities. This finding was already foreshadowed by the scatter plot of educational synergy against educational substitutability (Fig S2). The triangular shape of this plot means that highly substitutable educational pairs are also highly synergistic, but pairs that are synergistic may not necessarily be substitutable.

B.4.2 Educational levels and content areas per community

As mentioned in the main text, synergy communities also tend to contain a wider range of educational levels than substitutability communities. In fact, in 35% of all substitutability communities, all members have the same educational level. In contrast, not a single synergy community is completely homogeneous in terms of educational levels, although 15% have educational tracks in just one of the eleven high-level content areas used to color Figs S3 and S4.

Fig. S7 analyzes the concentration of educational levels and content areas in greater detail. It depicts the effective number of educational levels or content areas in synergy and substitutability communities. This quantity is calculated as follows:

$$N_{\kappa,\lambda}^* = \frac{1}{\sum_{\lambda} \left( \frac{|C_{\kappa,\lambda}|}{|C_{\kappa}|} \right)^2}$$

where $|C_{\kappa,\lambda}|$ is the number of elements in the set of education tracks of the level or content area $\lambda$ in community $\kappa$. Content areas are here aggregated into eleven general expertise areas. In other words, $N_{\kappa,\lambda}^*$ is the inverse of the Herfindahl-Hirschman Index (HHI) of the distribution of educational tracks across levels or across content areas within a community. Note that when a community’s educational tracks are equally distributed across $n$ categories (i.e., when each category contains the same number of tracks), the inverse of the HHI, $\frac{1}{HHI}$, equals $n$. Therefore, $N_{\kappa,\lambda}^*$ can be regarded as the effective number of educational categories in a community, i.e., the number of categories that, if shares had been distributed equally, would yield the same HHI.

Figs S7a and S7b show histograms of $N_{\kappa,\lambda}^*$ for the communities of the synergy and of the substitutability network. Synergy communities typically host a larger effective number of educational levels than substitutability communities. In contrast, the differences between synergy and substitutability communities are less pronounced when focusing on the effective number of educational content areas.

To assess how significant these differences are, Figs S7c and S7d normalize $N_{\kappa,\lambda}^*$ by subtracting the mean and dividing by the standard deviation of a simulated random benchmark. These normalized values are shown as vertical lines in the plots. The distribution of the 1,000 draws of the simulated benchmark is plotted using kernel densities. The benchmark was created by reshuffling educational levels and educational content areas. This operation keeps the number of elements by level (content area) and by community fixed, but breaks the relation between levels (contents) and communities.

The average effective number of educational levels in synergy communities is well below the random benchmark. However, expressed in standard deviations, this difference is twice as large in substitutability communities. In other words, whereas both types of communities are to some extent homogeneous in their educational levels, substitutability communities are more homogeneous in
Homogeneity in level (a) and content (b) of educations within communities. Effective number of educational levels and content areas by synergy and substitutability community. Simulated benchmarks (centered on zero) in (c) and (d) represent averages across 1,000 runs, where educational level and content area labels are randomly shuffled across communities. The vertical lines show the average observed effective number of educational levels or content areas in a community in units of standard deviations of the simulated benchmark.

levels than synergy communities.

Homogeneity in content areas is even more pronounced: the observed substitutability and synergy communities host educational tracks from far fewer content areas than the simulated ones. However, the differences between the two types of communities are now relatively small: synergy and substitutability communities are more or less equally homogeneous in educational content. In sum, Fig. S7 shows that, whereas belonging to a community of educational tracks that can substitute one another requires that tracks have the same educational level and educational content, belonging to a community of synergistic tracks requires that these tracks have the same content, but not necessarily the same level.

B.5 Worker-establishment aggregation

The networks in sec. B.4 represent $T_e \times T_e$ matrices, with $T_e$ the number of different educational tracks in the economy. These matrices describe the relations between educational tracks. However, what is needed is a quantification of the synergy and substitutability between a worker and her team of coworkers. One way to arrive at such a quantification is counting how many coworkers have high synergy or substitutability to w’s education. Using a logarithmic transformation, we can define a worker’s coworker synergy and coworker substitutability as:

\[
C_{wpwt}^\# = \log_{10} \left( \sum_e E_{epwt} 1(e_{ew} > \zeta_c) \right) \tag{S4}
\]

\[
S_{wpwt}^\# = \log_{10} \left( \sum_e E_{epwt} 1(s_{ew} > \zeta_s) \right) \tag{S5}
\]

where $e_{uw}$ is w’s education and $p_{uw}$ w’s work establishment in year $t$. $\zeta_c$ and $\zeta_s$ are thresholds above which two educations are said to be highly synergic or substitutable. These thresholds were chosen such that, if two workers are matched at random, there is a 1% probability that they are highly synergic or highly substitutable. Note that, because a worker is a perfect substitute and has maximal synergy to herself, the arguments of the logarithms in $S_{wpwt}^\#$ and $C_{wpwt}^\#$ are strictly greater than zero, ensuring that eqs (S4) and (S5) are always well-defined. In the main paper, the quantities $S_{wpwt}^\#$ and $C_{wpwt}^\#$ are used to construct Figs 5c and 5f.
The disadvantage of these measures is that they are highly correlated with an establishment’s size. This makes it hard to disentangle effects related to the size of an establishment from effects related to coworker fit. An alternative that does not suffer from this is \( w \)’s employment-weighted average synergy to coworkers:

\[
C_{wpw} = \sum_e \frac{E_{epw} - \delta_{ew} \lambda}{\sum_e' E_{e'p} - 1} \epsilon_{ew}
\] (S6)

Similarly, \( w \)’s substitutability by coworkers can be defined as:

\[
S_{wpw} = \sum_e \frac{E_{epw} - \delta_{ew} \lambda}{\sum_e' E_{e'p} - 1} \epsilon_{ew}
\] (S7)

where \( \delta_{ew} \) is a Kronecker delta function: \( \delta_{ew} = 1 \) if \( e = e_w \) and 0 otherwise. These quantities are identical to eqs (4) and (5) in the main text, but now written as weighted averages across educations, not unweighted averages across workers. This different notation will be useful when turning to the errors-in-variables corrections of sec. D.5.1. \( C_{wpw} \) and \( S_{wpw} \) are all but uncorrelated with establishment size, as shown in Fig. S8. The correlation between the logarithm of establishment size and weighted average coworker synergy (substitutability) is negative at -0.11 (-0.14, \( N = 2,144,965 \)). It is thus unlikely that findings based on these measures will be driven by establishment-size effects. The main paper therefore uses the latter measures and coworker synergy and coworker substitutability will henceforth refer to these weighted-average based measures.

The main text argues that one reason why educations co-occur in the same establishment is that they are similar to one another. That is, synergistic educations may not just enhance one another but also substitute for one another. To disentangle these two relations, eq. (6) in the main text proposes a notion of coworker complementarity, measured as the residual of a regression of coworker synergy on coworker substitutability:

\[
C_{wpw} = \alpha_{lw} + \beta_{lw} S_{wpw} + m_{wpw}
\] (S8)

---

Fig. S8. Co-worker variables and establishment size. Binned weighted-average synergy (blue squares) and substitutability (red triangles) against average establishment-size in a bin, 90% confidence intervals overlaid in brighter colors.

---
### Table S5. Descriptive statistics main variables.

Descriptive statistics for workers with at least upper secondary education in the estimation sample. *Cow. syn.* and *cow. subst.* are weighted average synergy and substitutability to coworkers, as in eqs (S6) and (S7); *# syn. cow.* and *# subst. cow.* count the number of coworkers with synergy and substitutability relations over thresholds $\zeta_c$ and $\zeta_s$ in eqs (S4) and (S5); *establishment size* in number of workers, *sh(own edu.):* share of coworkers with the focal worker’s education; *age* in years.

| Variable          | mean  | median | st. dev. | p90 - p10 |
|-------------------|-------|--------|----------|-----------|
| cow. syn.         | 0.592 | 0.579  | 0.109    | 0.264     |
| cow. subst.       | 0.393 | 0.386  | 0.185    | 0.481     |
| # syn. cow.       | 29.7  | 3.0    | 116.0    | 50.0      |
| # subst. cow.     | 28.0  | 4.0    | 99.8     | 51.0      |
| establishment size| 345.0 | 40.0   | 1008.0   | 759.0     |
| sh(own edu.)      | 0.081 | 0.019  | 0.146    | 0.250     |
| age               | 38.8  | 38.0   | 10.4     | 29.0      |

Note that, in a regression analysis, the effect of complementarity can be assessed in two ways. First, we can regress a dependent variable, say wages, on $m_{w_p,t}$. Second, we can regress wages on coworker synergy and coworker substitutability simultaneously. The resulting partial regression coefficient of coworker synergy is now, according to the Frisch-Waugh-Lovell theorem, equivalent to the regression coefficient of the residual of a regression of coworker synergy on coworker substitutability, i.e., to $\hat{m}_{w_p,t}$ (for this equivalence to be exact, we would need $\alpha_{c,t} = \alpha$ and $\beta_{c,t} = \beta^c$ in eq. (S8) and any further control variables would have to be added to the equation as well). Therefore, we can interpret the partial regression coefficient of coworker synergy as the effect of coworker complementarity, as long as the regression also controls for coworker substitutability.

Table S5 provides summary statistics of the main variables of interest. Apart from means and standard deviations, it also lists the difference between 90th and 10th percentile of a variable’s distribution.

In the main text, we analyze the effects of coworker complementarity and substitutability on wages of individual workers. If complementarities indeed improve the productivity of coworker teams, we would expect these productivity effects to be manifest also at the aggregate level of an economic establishment. Unfortunately, the data do not contain any information on the economic output of establishments. However, we can recover the longevity of such establishments. Table S6 shows the statistical associations between the team composition of an establishment and the likelihood that the establishment survives for at least five years. To be precise, it estimates Linear Probability Models (LPMs) with a dependent variable that assumes a value of one if an establishment survives for five years or longer. The independent variables are the average coworker synergy and substitutability across all coworkers, as well as the size of the establishment at birth. To avoid left- or right-censored observations, we focus on establishments founded between 2001 and 2005.

Establishments with a high level of synergies among coworkers indeed on average have a higher likelihood that they will survive...
Table S6. Survival analysis. ***: p<.01; **: p<.05, *: p<.1. Linear probability model for 5-year survival rates of establishments founded between 2001 and 2005. Cow. syn. and cow. subst. are the synergy and substitutability to coworkers, as in eqs. (S6) and (S7), averaged across all coworkers in the establishment and measured in the first year of the establishment’s existence. Robust standard errors in parentheses.

for five year or longer. In contrast, establishments with homogeneous workforces (i.e., with high substitutabilities among their workers) tend to have lower survival rates. These results are suggestive of the hypothesized productivity effects at the team level. However, it is important to note that the analysis in Table S6 is purely correlational and may not reflect causal relations. For instance, it is theoretically possible that the relation between longevity and complementarity runs in the opposite direction: successfully run establishments may survive longer and therefore attract complementary teams.

**Section SC. Co-worker synergy and substitutability by industry and occupation (Fig. 2)**

Which industries offer highly synergistic work environments? And which occupations offer workers a good coworker fit? In the main text, this is explored in Fig. 2. The graph plots the average coworker synergy against the average coworker substitutability in an industry. Industries are classified at the 3-digit level of the Swedish SNI 2002 classification, which corresponds to the European NACE Revision 1.1 classification system. Occupations are classified at the 4-digit level of the Swedish SSYK classification, which is based on the international ISCO 88 classification.

The plot displays the following quantities $\bar{C}_s$ and $\bar{S}_s$:

$$\bar{C}_s = \frac{1}{|W_s|} \sum_{(w,t) \in W_s} C_{wp,t} \quad (S9)$$

$$\bar{S}_s = \frac{1}{|W_s|} \sum_{(w,t) \in W_s} S_{wp,t} \quad (S10)$$

where $W_s$ represents the set of worker-year observations in occupation (Fig. 2a) or industry (Fig. 2b) s.

As a further illustration, Fig. S9 shows the same graph, but now for occupations within the category of managers and for the 4-digit industries that belong to the business services sector. The graph also displays a linear regression fit of $\bar{C}_s$ against $\bar{S}_s$. Note that the underlying regression was run on all occupations (industries), not just the subsets that are depicted here. The distance to this regression line can be interpreted as the average coworker complementarity of managers (Fig. S9a) or among workers employed in
Fig. S9. Average co-worker synergy and substitutability by industry and occupation. (a) Average $C_{wp_{wt}}$ against average $S_{wp_{wt}}$ in the estimation sample as defined in eqs (S6) and (S7) across all workers in managerial occupations. The gray line represents a linear fit based on the full set of occupations in the economy. (b) Idem, but showing the average $C_{wp_{wt}}$ against the average $S_{wp_{wt}}$ in the estimation sample for workers employed in business service industries.

Business services (Fig. S9b).

Fig. S9a shows that there is quite some variation in the average level of synergy and substitutability between a manager and her coworkers. A combination of high levels of synergy and relatively low levels of substitutability (i.e., high levels of complementarity) seems to be typical for managers that are responsible for processes related to a firm’s technological know-how, such as the management of R&D and of production and operations. Low levels are more often found in the management of support services, such as financial administration, human resources, IT and marketing and sales.

Similarly, business services can display a wide range of synergy and substitutability within their workforces (Fig. S9b). In particular, human-capital-intensive business services, like legal services, R&D, software publishing, hardware consultancy and technical testing, tend to hire teams of mostly synergistic workers. In contrast, lower-skill activities, such as cleaning, security and rental agencies, tend to employ workforces that are more homogeneous: their workers can often substitute one another but create relatively few synergies.

**Section SD. Wages**

**D.1 Correlational analyses (Table 1)**

To study the relation between the coworker variables and wages, consider the following regression model:

$$\log_{10}(wage_{wt}) = X_{wt}\beta_x + Q_{p_{wt}}\beta_p + \beta_x C_{wp_{wt}} + \beta_s S_{wp_{wt}} + \epsilon_{wt}$$  \hspace{1cm} (S11)

where $X_{wt}$ is a vector of worker characteristics and $Q_{p_{wt}}$ a vector of establishment characteristics. The main variables of interest are $C_{wp_{wt}}$ and $S_{wp_{wt}}$. Table S7 shows results for different model specifications, where columns (1), (2), (3) and (5) repeat the specifications reported in Table 1 of the main text.

Across all models, an increase in coworker synergy is associated with an increase in wages, whereas an increase in coworker
Table S7. Co-worker educational fit and wages. ***, p<.01; **, p<.05, *, p<.1. OLS regressions with \( \log_{10}(\text{wage}) \) as a dependent variable. The variable edu.-occ. match is defined in eq. (S13); diversity is the logarithm (base 10) of the number of distinct educational tracks in the establishment, including the worker’s own. Standard errors, clustered at the worker level, in parentheses.
substitutability is associated with a drop in wages. The estimated effects are substantial. In model (3), which estimates the effects of both variables simultaneously, but contains no other control variables, an increase from the 10th to the 90th percentile in coworker synergy translates into 25.3% higher wages. Note that, as explained in sec. B.5, the partial regression coefficient of coworker synergy can be interpreted as the effect of coworker complementarity. In contrast, moving from the 10th to the 90th percentile of coworker substitutability is associated with a wage reduction of 19.8%.

Effects of coworker complementarity drop when adding control variables. Column (4) adds worker-level characteristics such as age and educational level. This drastically reduces the negative association of coworker substitutability and, to a lesser extent, the positive association of coworker complementarity with wages. The drop in parameter estimates can be attributed to the fact that coworker complementarity rises with a worker’s work experience – and therefore her age – and is positively correlated with educational attainment. These phenomena are discussed in greater detail in sec. E and F.1.2.

Model (5) reports the results of the main text’s preferred specification. It controls for establishment-size and important worker characteristics, without adding variables that use information on the coworker environment. Adding such variables is in principle undesirable, because it may partial out aspects of coworker fit that should be considered part of coworker synergy and substitutability. In this model, an 80-percentiles rise in coworker synergy is associated with a wage increase of 18.1%, whereas a similar increase in substitutability is associated with a wage drop of 4.8%.

In spite of the aforementioned reservations, columns (6) to (9) add control variables that contain information about a worker’s coworker environment. Column (6) adds the share of coworkers with the same education as the worker herself. This leaves effect estimates for coworker complementarity and coworker substitutability all but unchanged. Column (7) controls for the shares of coworkers in each of six educational levels, which reduces the coworker-complementarity coefficient and increases the substitutability coefficient somewhat, but not overwhelmingly so. Column (8) controls for how well a worker’s education fits her occupation. It does so by adding an occupational match variable, \( M_{eo} \):

\[
M_{eo} = \frac{E_{oe}/E_o}{E_e/E_o} \quad \text{(S12)}
\]

\( M_{eo} \) is calculated in the measurement sample and quantifies the degree to which education \( e \)’s employment share in occupation \( o \) exceeds its employment share in the overall economy. According to this definition, an education matches an occupation if the education is typical for the occupation. For instance, if workers with a college degree in “462z: Statistics” are overrepresented in occupation “2413: Market analysts” statisticians are deemed well-suited for jobs as market analysts. To reduce skew, \( M_{eo} \) is mapped onto the interval \([0, 1)\), using the same transformation as in eq. (S2):

\[
\mu_{eo} = \frac{M_{eo}}{M_{eo} + 1} \quad \text{(S13)}
\]

Controlling for this occupational match affects results more than controlling for any of the other control variables. It reduces the coworker-complementarity parameter by about a third and the substitutability parameter by one fifth vis-à-vis the preferred specification. Note that, whereas \( M_{eo} \) addresses the match of a worker’s human capital to her job in a traditional, individual-centric, manner, coworker synergy and substitutability describe this match in terms of her relation to coworkers. The fact that the
complementarity effect estimate drops when occupational match is added as a control variable means that the occupational fit and the coworker fit explain overlapping parts of a worker’s wage. That is, workers whose high wages the model attributes to being well-suited to do a specific job also tend to have skills that complement the establishment’s wider workforce and vice versa. Indeed, the main paper argues that workers are suitable for a specific job, because they complement their coworkers. From that perspective, the fact that the bulk of the association of coworker synergy and substitutability with wages survives when controlling for the individual-centric part of the job match, $M_{eo}$, supports the claim that human capital has an inherently social orientation.

Finally, column (9) controls for workforce diversity by adding the logarithm of the number of different educational tracks in an establishment. Adding this control variable strengthens the estimated wage associations of the coworker variables slightly. Overall, however, none of the variables in columns (6) to (9) can account for the lion’s share of the statistical associations we observe between wages and coworker complementarity or substitutability.

### D.2 Fixed effects models

One concern about the analysis in Table S7 is that the estimated wage associations reflect that more capable workers are better positioned to sort themselves into work environments that fit their qualifications. Note that such sorting would not necessarily invalidate the conclusion that a good coworker-fit is important. After all, the fact that workers sort themselves into well-fitting work environments suggests that such environments are attractive. However, if this attractiveness has other than monetary reasons, the correlation between coworker synergy and substitutability on the one hand and wages on the other may derive from the fact that productive workers prefer complementarity-rich environments. Similarly, the most productive (and high-wage) employers may hire teams of complementary workers. Either case would result in a positive correlation between wages and coworker synergy and a negative correlation between wages and coworker substitutability, even if the skills of a worker’s coworkers are of no consequence to her wage. It is therefore unclear how much of the observed wage effects can be attributed to good coworker fit as opposed to a higher intrinsic quality of workers typically found in well-matched coworker environments.

To control for such sorting effects, Table S8 adds worker, establishment, and worker-establishment fixed effects. As a result, these models estimate wage effects by comparing the same worker across (possibly different) establishments in different years, different workers in the same establishment and year and the same worker in the same establishment but with different coworkers in different years.

Adding either worker (column 2) or establishment-year (column 3) fixed effects reduces the estimated effect of coworker complementarity by roughly 40%, while leaving the effect of substitutability unchanged. Column (4) controls for both effects using the high-dimensional fixed effects estimator of (36) and column (5) interacts worker and establishment fixed effects.

Across all specifications a higher coworker synergy is significantly associated with higher wages and a higher coworker substitutability with lower wages at p-values below 1%. However, controlling for worker and establishment fixed effects reduces the size of the estimated coworker effects substantially, which suggests that ability-based sorting might be at play. Note however that, if ability-based sorting were a problem, this problem should have been largely resolved in column (2), which bases effect estimates on changes in coworkers for the same individual across her career. Yet, most of the drop in effect estimates occurs in the final specification, where
Table S8. Wage regressions, fixed effects models. 

| dep. var.: log(wage) | (1) | (2) | (3) | (4) | (5) |
|----------------------|-----|-----|-----|-----|-----|
| cow. syn.            | 0.274*** 0.165*** 0.160*** 0.103*** 0.057*** |
|                      | (0.0030) (0.0054) (0.0068) (0.0058) (0.0075) |
| cow. subst.          | -0.044*** -0.046*** -0.048*** -0.027*** -0.016*** |
|                      | (0.0018) (0.0029) (0.0030) (0.0030) (0.0039) |
| log(est. size)       | 0.044*** 0.028*** 0.044*** 0.051*** |
|                      | (0.0003) (0.0004) (0.0015) (0.0020) |
| 4th polyn. age?      | yes yes yes yes yes |
| edu level dum?       | yes yes yes yes yes |
| fixed effects?       | yr yr, w. yr×est. yr, w., est. yr, w.×est. |
| R²                   | 0.300 0.810 0.648 0.834 0.886 |
| # obs.               | 2,144,965 2,144,965 2,144,965 2,065,237 2,144,965 |
| # clust.             | 364,642 364,642 144,371 144,371 |

Dep. var. = log(wage) (1) (2) (3) (4) (5)

Coworker synergies (cow. syn.) and substitutes (cow. subst.) are measured using log-transformed wage, log-transformed estimated establishment size, and establishment + year fixed effects. The models consecutively add fixed effects for years (yr); years and workers (yr, w.); establishment-year combinations (yr×est.); years, workers and establishments (yr, w., est.); and for years and worker-establishment combinations (yr, w.×est.). Standard errors (in parentheses) are clustered at the worker level in columns (1) and (2), at the worker and the establishment level in column (3) and at the establishment level in columns (4) and (5).

In column (4) and (5) standard errors are clustered at the establishment level. The worker fixed effects are identified solely from variation in the coworker variables that are due to coworkers’ entering or exiting an establishment, while the worker herself does not change establishments. The fact that the wage effects of the coworker variables remain significant even in this strictest regression reassures us that sorting cannot adequately explain coworker effects. Sec. D.4 will argue that there is an alternative explanation for why effect sizes drop in the fixed effects regressions: measurement error. Because fixed effects control for parts of the signal in a variable, fixed effects exacerbate measurement-error induced attenuation bias (26). Fixed effects models may therefore substantially underestimate the true coworker effects.

**D.3 Instrumental variables estimation**

In spite of the fact that worker fixed-effects models would correct for ability-based sorting, one may still wonder whether there are other concerns that prevent us from interpreting estimated effects as causal. First, although worker fixed effects can eliminate biases related to ability-based sorting, they do not necessarily do so. For instance, worker fixed effects will not absorb time-varying aspects of ability, nor can they adequately correct for workers’ intrinsic ability if this ability is only revealed with time (see (38)). A second concern is that workers may be assigned to jobs where they earn the highest wages as in the Roy model (21). In this case, observed wages can be thought of as a (nonrandomly) selected sample of a population of potential wage offers.

This section makes an additional attempt to estimate the causal effect of coworker synergies on wages using instrumental variables (IV) estimation. This approach should correct simultaneously for various sources of endogeneity, as well as for measurement error. Note that, given that workers do not randomly choose jobs and firms do not randomly choose workers, it is hard to imagine any source of exogenous variation in the level of synergy between coworkers. A more promising place to look for such variation is in changes in synergy and in particular for situations in which the focal worker does not actively try to improve her coworker fit. These situations can be found when workers do not change establishments or education (note that the latter, changes in educational attainment itself,
are exceedingly rare). Consider therefore wage equation (S11) in first differences:

\[
\Delta \log_{10}(wage_{wt}) = \Delta X_{wt}\beta_x + \Delta Q_{pw_{t}}\beta_p + \Delta C_{wp_{t}}\beta_c + \Delta S_{wp_{t}}\beta_s + \Delta \epsilon_{wt}
\]  

(S14)

where \(\Delta\) denotes the first-differencing operator. The main concern is that \(\Delta C_{wp_{t}}\) and \(\Delta S_{wp_{t}}\) are correlated with \(\Delta \epsilon_{wt}\) because of sample-selection biases, learning-about-ability or measurement error. Therefore, we must identify sources of exogenous variation in changes in \(C_{wp_{t}}\) and \(S_{wp_{t}}\). One potential source of such exogenous variation can be found in shifts in the local supply of coworkers with a particular education. The underlying idea is that if there were an exogenous increase in the local supply of synergistic workers, this would lower the price of these workers, and therefore increase the likelihood of them being hired by the worker’s employer.

The supply shift we will exploit here is based on the number of graduates in a region. However, because students’ educational choices may reflect future employment prospects, observed graduation rates may not be exogenous. Therefore, this section will predict these graduation rates from information that should be exogenous to wage dynamics. In particular, it predicts the number of local graduates in each educational track by combining information on the local educational structure between 1990 and 1995 with the national growth rate of the educational track in question.

To illustrate this identification strategy, assume that, between 1990 and 1995, Gothenburg trained 20% of all Swedish automotive engineers. Furthermore, assume that, in 2001, a total of 800 students graduated in automotive engineering in Sweden but outside Gothenburg. If Gothenburg’s educational share of automotive engineers in Sweden didn’t change, Gothenburg should produce 200 graduates in this field in 2001. This suggests predicting the number of graduates (\(G_{emt}\)) for educational track \(e\) in municipality \(m\) and year \(t\) as:

\[
\hat{G}_{emt} = \frac{q_{em}}{1 - q_{em}} \sum_{m \neq m'} G_{em't}
\]  

(S15)

where \(q_{em} = \frac{\sum_{1995}^{1990} G_{emt}}{\sum_{t=1995} G_{emt}}\) represents municipality \(m\)’s historical share of graduates in educational track \(e\). The number of local graduates that are highly synergistic to education \(e\) is based on the metric defined in eq. (S4). That is, it counts the number of workers with educations whose synergy to education \(e\) exceeds \(\zeta_c\). Next, the predicted number of such synergistic graduates is divided by the number of synergistic workers living (the instrument thus uses information on the municipality of residence, not of work) in the region at the beginning of the period. This number represents an exogenous shift in the local availability of workers synergistic to education \(e\). The instrument is constructed for two different thresholds \(\zeta_c\) (the first (education-specific) threshold is chosen such that, for each education, the expected likelihood that a random pair of workers is synergistic is 0.1%, whereas the second sets this expected likelihood at 1%) and at two spatial scales, namely, at the level of municipalities and at the level of labor market areas. The worker’s own municipality is omitted from a labor market area to ensure that the two instruments reflect different supply shifts. Finally, region-industry-year fixed effects correct for changes in local market conditions.

One complication of the specification in first-differences is that the wage data refer to a worker’s annual wage, i.e., to total wages earned in a given year. Because coworkers may enter or leave an establishment at any time during the year, the change in coworker synergy between years \(t\) and \(t+1\) may already have affected wages in year \(t\). Therefore, wage changes are measured as the annualized
changes in wage over a two-year period, from \( t - 1 \) to \( t + 1 \). Furthermore, the sample is restricted to workers who work continuously at the same establishment between the years \( t - 2 \) and \( t + 1 \) to ensure that wages are earned in only one establishment.

Due to their collinearity, it is infeasible to instrument coworker synergy and substitutability simultaneously using this identification strategy: When instrumenting both variables simultaneously, the Kleibergen-Paap F-statistic drops to 1.9 when using four instruments and 4.7 when using only one. Table S9 therefore only shows two-stage least squares (2SLS) estimates for the effect of coworker synergy.

The preferred specification uses all four instruments and controls for industry-municipality-year fixed effects. Model (1) shows the 2SLS, reduced form and first-stage estimates for this specification. The point estimate of 0.637 implies that a move from the 10th to the 90th percentile in coworker synergy translates into a 47% increase in wages. This is over twice as high as the original OLS effect (column (5) of Table S7). However, given the large standard errors, the OLS estimate is still well within a 95% confidence interval around this point estimate.

To explore the robustness of these findings, model (1) is re-estimated with only year (column 2), industry-year (column 3) and region-year fixed effects (column 4). Although point estimates in these models exceed the ones in model (1), differences are well within the margin of error.

The lower Kleibergen-Paap statistics of models (2) and (3) suggests that these increases may reflect weak-instrument problems. Moreover, the large, imprecisely estimated effects in model (1) still raise concerns over weak identification, even though its Kleibergen-Paap F-statistic of 14.1 indicates that instruments are reasonably strong. This possibility is explored in two additional models: model (5), which uses a Limited Information Maximum Likelihood (LIML) estimator and model (6), which estimates a just-identified model using only the strongest instrument. Both models produce similar results to model (1), suggesting that the high point-estimates are not due to weak instruments. Moreover, the Kleibergen-Paap statistic in model (6) rejects the possibility that the preferred instrument is weak.

The Hansen J-statistic on over-identifying restrictions does not differ significantly from zero in any of the models with multiple instruments. That is, it does not raise concerns about the instruments’ exclusion restriction. In spite of this failure to reject our identification strategy, a potential concern is that shifts in the local supply of synergistic workers may directly impact wages by improving a worker’s outside options and, therewith, her bargaining position in her current establishment. If such bargaining effects were to occur, they should also appear when the labor supply in a worker’s own education increases. Such an expansion in competition from local workers outside the own establishment should deteriorate a worker’s bargaining position, i.e., it should have a negative effect on wages. This hypothesis is investigated by regressing a worker’s wage on the predicted number of own-education graduates in her location. Table S10 shows that there is no statistically significant evidence for such a bargaining effect.

To further explore the validity of the identification strategy, the IV analyses of Table S9 are rerun on a sample of static establishments, i.e., establishments that do not change their workforce. In other words, only workers in establishments that neither hire nor fire anyone in a given year are retained. In such establishments, the local supply shift can only affect workers’ wages directly, not indirectly through a change in a worker’s coworkers. The combined direct and indirect effects of an instrument are measured in the reduced-form equation: in this case, the regression of a worker’s change in wages on the local supply shift. In the full sample,
Table S9. Instrumental variable regression in first differences. First-differenced regression as specified in eq. (S14) using workers who do not change establishments or educations between the years $t - 2$ and $t + 1$. The dependent variable is the base-10 logarithm of the annualized wage growth between years $t - 1$ and $t + 1$. Model (1) shows results from a 2SLS regression (together with the first stage - fs - and reduced form - rf - estimates), using four instruments: the number of highly synergistic (threshold of 0.1%) and of synergistic (threshold of 1%) graduates in the establishment’s municipality and analogous variables at the level of labor market areas. Models (1), (5) and (6) control for year-municipality-industry (yr×m×i) fixed effects. Models (2), (3) and (4) show analyses with only year, year-municipality and year-industry fixed effects. Model (5) uses the LIML instead of the 2SLS estimator. Model (6) uses only the preferred instrument (threshold of 0.1% at the labor market area level).

| Model          | 2SLS | fs | rf | 2SLS | 2SLS | 2SLS | liml | pref |
|----------------|------|----|----|------|------|------|------|------|
| \( \text{cow. syn.} \) | 0.637** | 1.099*** | 0.948*** | 0.956*** | 0.677** | 0.788*** |      |      |
| (0.265)        | (0.286) | (0.274) | (0.283) | (0.283) | (0.301) |      |      |
| \( \text{log(est. size)} \) | 0.027*** | -0.017*** | 0.016*** | 0.040*** | 0.035*** | 0.033*** | 0.027*** | 0.029*** |
| (0.005)        | (0.001) | (0.001) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) | (0.005) |
| \( 4^{th} \text{polyn. of age?} \) | yes | yes | yes | yes | yes | yes | yes | yes |
| IV - synergistic grad. (0.1%, m) | 0.0013*** | 0.0013* |      |      |      |      |      |      |
| (0.0004) | (0.0007) |      |      |      |      |      |      |
| IV - synergistic grad. (1%, m) | 0.0014*** | 0.0006 |      |      |      |      |      |      |
| (0.0005) | (0.0009) |      |      |      |      |      |      |
| IV - synergistic grad. (0.1%, r) | 0.0004 | 0.0029* |      |      |      |      |      |      |
| (0.0007) | (0.0015) |      |      |      |      |      |      |
| IV - synergistic grad. (1%, r) | 0.0022** | -0.0015 |      |      |      |      |      |      |
| (0.0009) | (0.0018) |      |      |      |      |      |      |
| Kleibergen-Paap | 14.1 | 14.1 | 14.1 | 11.7 | 11.7 | 11.9 | 14.1 | 44.4 |
| P-val. Hansen J | 0.220 | 0.457 | 0.363 | 0.694 | 0.224 |      |      |      |
| # obs. | 727,982 | 727,982 | 727,982 | 771,684 | 771,684 | 771,470 | 727,982 | 727,982 |
| # clust. | 115,802 | 115,802 | 115,802 | 123,820 | 123,820 | 123,791 | 115,802 | 115,802 |
| Instruments | all | all | all | all | all | all | all | all |

***: p<.01; **: p<.05; *: p<.1, standard errors clustered at yr-mun-edu track level.
Table S10. Wage effects of within-education local supply shifts. Effect of the predicted shift in local supply of same-education graduates in the sample of workers who do not change establishments. Model (1) controls for year-municipality-industry \((yr \times mun)\) fixed effects. Models (2), (3) and (4) show analyses with only year, year-municipality and year-industry fixed effects.

|                | (1)      | (2)      | (3)      | (4)      |
|----------------|----------|----------|----------|----------|
| \(\log(\text{pred. grad. edu} \times \text{mun})\) | -0.0003  | 0.0002   | -0.0006  | -0.0001  |
|                | (0.0004) | (0.0005) | (0.0004) | (0.0004) |
| \(\log(\text{est. size})\) | -0.0021*** | -0.0019*** | -0.0020*** | -0.0018*** |
|                | (0.0003) | (0.0003) | (0.0003) | (0.0003) |
| \(4^{\text{th}}\) polyn. age? | yes      | yes      | yes      | yes      |
| edu level FE?  | yes      | yes      | yes      | yes      |
| yr yrXedu yrXmun yrXeduXmun |          |          |          |          |
| # clust.       | edu, mun | edu, mun | edu, mun | edu, mun |
| # obs.         | 1,474,639| 1,474,639| 1,474,639| 1,474,639|

***: \(p<.01\); **: \(p<.05\); *: \(p<.1\), standard errors clustered at mun. and edu. track level.

these reduced-form effects are highly significant. For instance, in model (6) of Table S9 the preferred instrument’s reduced-form effect on wages has a \(t\)-value of 2.83. In contrast, in the static-establishment sample, the corresponding \(t\)-value is -0.09. This corroborates the claim that there is no direct effect of the instrument on wages, i.e., there is no evidence for an effect that is not channeled through the hiring of synergistic coworkers.

D.4 Measurement error

As pointed out at the end of sec. D.2, the fact that coworker effects weaken in fixed-effects specifications is consistent with the presence of unobserved ability-based sorting. However, the lion’s share of such ability-based sorting effects should arguably be controlled for with worker fixed effects. It is, therefore, surprising that the most precipitous drop in point estimates results when controlling for worker-establishment, not worker fixed-effects. An alternative explanation for why point estimates are lower in fixed-effects models is that the variables of interest are mismeasured. In particular, if a variable’s actual values are more strongly autocorrelated than its measurement error, fixed effects will absorb large parts of the variable’s signal, but little of its noise. In fact, the addition of control variables that are correlated with the variable’s signal, but not with the noise has similar consequences in the models of Table S7. The consequent deterioration in signal-to-noise ratio will exacerbate existing attenuation biases. This subsection will first provide evidence that coworker synergy is estimated with substantial error and then present two strategies to quantify the resulting errors-in-variables bias.

D.4.1 Symptoms of measurement error

A first symptom of noise in the educational synergy variable is that the estimated synergies in a given educational pair vary markedly from one year to the next, with a correlation between two consecutive years of, on average, 0.88 \((N = 120,295)\). Assuming that there is no abrupt change in the intrinsic synergy between two educations, a correlation below one would indicate measurement error (note that the implied measurement-error variance-share of \(1 - 0.88^2 = 0.23\) underestimates true measurement error because teams typically last for several years, which artificially inflates year-on-year correlations). To the extent that measurement errors
Table S11. Fixed effects models, within transformation versus first differences. Standard errors (in parentheses) are clustered at the worker level in model (1) and robust in columns (2) to (5).

|                  | (1) | (2) | (3) | (4) | (5) |
|------------------|-----|-----|-----|-----|-----|
| dep. var.        | log10(wage) | OLS          | within | 1st dif.          | within | 1st dif.          |
|                  | cow. syn.  | 0.274***  | 0.165***  | 0.076***  | 0.057***  | 0.045***  |
|                  |         | (0.0030) | (0.0049) | (0.0031) | (0.0058) | (0.0042) |
|                  | cow. subst. | -0.044***  | -0.046***  | -0.020***  | -0.016***  | -0.013***  |
|                  |         | (0.0018) | (0.0026) | (0.0017) | (0.0032) | (0.0024) |
|                  | log(plant size) | 0.044***  | 0.028***  | 0.015***  | 0.051***  | 0.048***  |
|                  |         | (0.0003) | (0.0004) | (0.0003) | (0.0009) | (0.0008) |
| 4th polyn.of age? | yes  | yes  | yes  | yes  | yes  | yes  |
| edu. level dum.? | yes  | yes  | yes  | yes  | yes  | yes  |
| fixed effects?   | yr    | yr   | yr   | yr   | yr   | yr   |
|                  | yr, w.  | yr, w.  | yr, w.  | yr, w.×est. | yr, w.×est. | yr, w.×est. |
| R²               | 0.300  | 0.277  | 0.242  | 0.242  | 0.242  | 0.242  |
| N                | 2,144,965 | 2,144,965  | 1,697,584  | 2,144,965  | 1,474,639  | 1,474,639  |

***: p<.01; **: p<.05; *: p<.1.

carry over to the worker-establishment level, the estimated effects in Table S7 will be biased towards zero. Compared to synergy, educational substitutability seems much less noisy. In fact, the estimated substitutability in a given educational pair is highly stable, with a correlation of 0.98 (N = 120, 295) between consecutive years. A plausible reason for this is that, whereas synergy estimates are based on co-occurrences in establishments, substitutability is measured as a relation between the occupation and education of individuals. The precision of synergy estimates depends therefore on the total number of establishments in which two educations are present, whereas the precision with which educational substitutability is estimated depends on how many individuals absorbed these educations. Because the latter is much larger than the former, substitutability will be measured more accurately than synergy.

If the reduction in point estimates in the fixed-effect models is indeed due to measurement error, it matters how fixed effects are corrected for. Doing so by means of a within-transformation typically yields less-attenuated parameter estimates than first-differencing the data (26) (first-differencing will particularly inflate the error-to-signal ratio vis-à-vis a within-transformation if the number of observations per individual is large and if the signal is strongly and positively autocorrelated, whereas the noise is not).

Table S11 shows that this is indeed the case. Model (1) repeats the preferred OLS specification of Table S7. Models (2) and (3) include worker fixed-effects, models (4) and (5) worker-establishment fixed-effects. The within-transformation is used in models (2) and (4), whereas fixed-effects are eliminated through first-differencing in models (3) and (5).

The difference between the two panel-data approaches is striking. Although the within-transformed models in columns (2) and (4) should, theoretically, give the same results as the first-differenced models in columns (3) and (5), estimated effects in the latter are much lower than in the former, supporting the notion that (some of) the drop in point estimates in fixed effects models can be attributed to measurement error.

D.4.2 Error-variances in educational synergy

This section derives theoretical expressions for the error-variance in the coworker synergy variable (see also (24)) used to truncate edges in the networks of B.4. The main idea behind this derivation is that measurement error arises when an education participates in few co-occurrences, such that the effective sample in which co-occurrences for this education are counted is small.
Let the estimated synergy in year \( t \) be composed of two unobserved components: the true synergy and measurement error. That is:

\[
\hat{c}_{ee'}t = c_{ee'} + \nu_{ee'}t \tag{S16}
\]

where \( \hat{c}_{ee'}t \) is the observed synergy between educational tracks \( e \) and \( e' \), \( c_{ee'} \) the underlying, actual synergy and \( \nu_{ee'}t \) a measurement error that is uncorrelated with \( c_{ee'} \). Let \( N_{ee'} \) be the number of co-occurrences of education \( e \) with \( e' \). Furthermore, omitted subscripts indicate a summation over the corresponding dimension, i.e.: \( N_e = \sum_e N_{ee'} \), \( N_{ee'} = \sum_e N_{ee'} \) and \( N = \sum_e \sum_{ee'} N_{ee'} \). Now assume that \( N_{ee'} \) is drawn from a Binomial distribution, \( N_{ee'} \sim BIN(\Pi_{ee'}, N) \), and, therefore has a variance of:

\[
V[N_{ee'}] = N \Pi_{ee'} (1 - \Pi_{ee'}) \tag{S17}
\]

\( \Pi_{ee'} \), the unknown probability of a co-occurrence between educations \( e \) and \( e' \) can be estimated as the observed relative frequency of \( N_{ee'} \). Denoting observed or estimated quantities by a hat (“^”), we get: \( \hat{N}_{ee'} = \frac{\hat{N}_{ee'}}{N} \). Consequently, (S17) can be written as:

\[
\hat{V}[N_{ee'}] = N \frac{\hat{N}_{ee'}}{N} \left( 1 - \frac{\hat{N}_{ee'}}{N} \right) \tag{S18}
\]

One problem is that, for the vast majority of educational pairs, \( \hat{N}_{ee'} \) equals zero. Therefore, eq. (S18) leads to the implausible conclusion that synergies are measured without error in these pairs. The reason why this happens is that the uncertainty in \( \hat{N}_{ee'} \) is not taken into account. To address this, \( \hat{\Pi}_{ee'} \) is re-estimated in a Bayesian framework. First, expectations and variances for \( N_{ee'} \) are derived under the assumption that \( N_{ee'} \) is drawn from a Hypergeometric distribution that takes the total number of co-occurrences in which educations \( e \) and \( e' \) participate as given and equal to the observed quantities \( \hat{N}_e \) and \( \hat{N}_{e'} \). These variances and expectations suggest a prior distribution for \( \Pi_{ee'} \), which can then be updated using information on the actually observed number of co-occurrences, \( \hat{N}_{ee'} \). The posterior expectation of \( \Pi_{ee'} \), \( \hat{\Pi}_{ee'}^{post} \), that results from this exercise is always strictly greater than zero. As a consequence, the estimated variance that follows from eq. (S17) when using \( \hat{\Pi}_{ee'}^{post} \) is always nonzero.

In particular, consider the following Binomial model for the probability of observing \( n_{ee'} \) co-occurrences between education \( e \) and \( e' \):

\[
Pr[N_{ee'} = n_{ee'} | N = n, \Pi_{ee'} = \pi_{ee'}] = \binom{n}{n_{ee'}} \pi_{ee'}^{n_{ee'}} (1 - \pi_{ee'})^{n - n_{ee'}} \tag{S19}
\]

Using Bayes’ law, we get:

\[
Pr[\Pi_{ee'} = \pi_{ee'} | N = n, N_{ee'} = n_{ee'}] = \frac{Pr[N_{ee'} = n_{ee'} | N = n, \Pi_{ee'} = \pi_{ee'}] Pr[\Pi_{ee'} = \pi_{ee'} | N = n]}{\int_0^1 Pr[N_{ee'} = n_{ee'} | N = n, \Pi_{ee'} = \pi_{ee'}] Pr[\Pi_{ee'} = \pi_{ee'} | N = n] d\pi_{ee'}}
\]
Choosing the $BETA[\alpha_{ee'}, \beta_{ee'}]$ distribution (the Binomial’s conjugate prior), as a prior for $\Pi_{ee'}$, this becomes:

$$Pr [\Pi_{ee'} = \pi_{ee'} \mid N = n, N_{ee'} = n_{ee'}] = \frac{Pr [N_{ee'} = n_{ee'} \mid N = n, \Pi_{ee'} = \pi_{ee'}] \pi_{ee'}^{n_{ee'} - 1} (1 - \pi_{ee'})^\beta_{ee'} - 1 \frac{\Gamma(\alpha_{ee'} + \beta_{ee'})}{\Gamma(\alpha_{ee'}) \Gamma(\beta_{ee'})}}{\int_0^1 Pr [N_{ee'} = n_{ee'} \mid N = n, \Pi_{ee'} = q_{ee'}] q_{ee'}^{n_{ee'} - 1} (1 - q_{ee'})^\beta_{ee'} - 1 \frac{\Gamma(\alpha_{ee'} + \beta_{ee'})}{\Gamma(\alpha_{ee'}) \Gamma(\beta_{ee'})} dq_{ee'}} \tag{S20}$$

where $\Gamma$ represents the gamma function. Using the expression for the binomial probability density function (pdf), (S20) simplifies to:

$$Pr [\Pi_{ee'} = \pi_{ee'} \mid N = n, N_{ee'} = n_{ee'}] = \frac{\pi_{ee'}^{n_{ee'} + \alpha_{ee'} - 1} (1 - \pi_{ee'})^{n_{ee'} - n_{ee'} + \beta_{ee'} - 1}}{\int_0^1 q_{ee'}^{n_{ee'} + \alpha_{ee'} - 1} (1 - q_{ee'})^{n_{ee'} - n_{ee'} + \beta_{ee'} - 1} dq_{ee'}} \tag{S21}$$

Because the denominator of (S20) has to integrate to one (it is an integral of a pdf over its domain), we get:

$$\int_0^1 t^{n_{ee'} + \alpha_{ee'} - 1} (1 - t)^{n_{ee'} - n_{ee'} + \beta_{ee'} - 1} dt = \frac{\Gamma(n_{ee'} + \alpha_{ee'}) \Gamma(n - n_{ee'} + \beta_{ee'})}{\Gamma(n + \alpha_{ee'} + \beta_{ee'})} \tag{S22}$$

Substituting (S22) into (S21), yields:

$$Pr [\Pi_{ee'} = \pi_{ee'} \mid N = n, N_{ee'} = n_{ee'}] = \pi_{ee'}^{n_{ee'} + \alpha_{ee'} - 1} (1 - \pi_{ee'})^{n_{ee'} - n_{ee'} + \beta_{ee'} - 1} \frac{\Gamma(n + \alpha_{ee'} + \beta_{ee'})}{\Gamma(n_{ee'} + \alpha_{ee'}) \Gamma(n - n_{ee'} + \beta_{ee'})} \tag{S23}$$

which describes a $BETA[n_{ee'} + \alpha_{ee'}, n - n_{ee'} + \beta_{ee'}]$ distribution. In other words, the posterior distribution of $\Pi_{ee'}$ is:

$$\Pi_{ee'} \mid N = n, N_{ee'} = n_{ee'} \sim BETA[n_{ee'} + \alpha_{ee'}, n - n_{ee'} + \beta_{ee'}] \tag{S24}$$

This leaves the task of choosing parameters $\alpha_{ee'}$ and $\beta_{ee'}$ such that they reflect plausible priors for the mean and variance of $\Pi_{ee'}$. To arrive at such priors, assume that the total number of co-occurrences in which educations $e$ and $e'$ participate are given. In other words, think of co-occurrences as arising from a process in which each time an education $e$ is present in an establishment, it draws a random second education from the pool of educational presences. Moreover, because the total number of co-occurrences in the economy $N \ggg N_{ee'}$ is large for any $(e, e')$, $N$ can be considered fixed. Consequently, co-occurrences follow a Hypergeometric distribution, with the following prior means and variances for $\Pi_{ee'}$:

$$E[\Pi_{ee'}] = E\left[\frac{N_{ee'}}{N}\right] = \frac{1}{N} E\left[N_{ee'}\right] \approx \frac{1}{N} N_e \frac{N_{ee'}}{N} \tag{S25}$$

$$V[\Pi_{ee'}] = \frac{1}{N^2} V\left[N_{ee'}\right] \approx \frac{1}{N^2} N_e N_{ee'} (N - N_e) (N - N_{ee'}) \tag{S26}$$

where $\approx$ indicates an equality by assumption of the Hypergeometric data generating process. The $BETA[\alpha_{ee'}, \beta_{ee'}]$ distribution implies:

$$E[\pi_{ee'}] = \mu_{ee'} = \frac{\alpha_{ee'}}{\alpha_{ee'} + \beta_{ee'}} \tag{S27}$$
\[ V[\pi_{ee'}] = \sigma^2_{ee'} = \frac{\alpha_{ee'}\beta_{ee'}}{(\alpha_{ee'} + \beta_{ee'})^2 (\alpha_{ee'} + \beta_{ee'} + 1)} \]  

(S28)

Solving for \( \alpha_{ee'} \) and \( \beta_{ee'} \), yields:

\[
\alpha_{ee'} = \frac{\mu^2_{ee'} (1 - \mu_{ee'}) - \mu_{ee'}}{\sigma^2_{ee'}}
\]  

(S29)

\[
\beta_{ee'} = \mu_{ee'} \left( \frac{(1 - \mu_{ee'})^2}{\sigma^2_{ee'}} + 1 \right) - 1
\]  

(S30)

Eqs (S23), (S25), (S26), (S29) and (S30) now define a posterior expectation, \( \hat{\Pi}_{ee'}^{\text{post}} \), of \( \Pi_{ee'} \) for each educational pair. Now recall equation (2) in the main text, which defines the synergy in educational pair \((e, e')\), \(c_{ee'}\), as follows:

\[
c_{ee'} = \frac{\kappa_{ee'} \hat{N}_{ee'}}{\kappa_{ee'} \hat{N}_{ee'} + 1}
\]

where \( \kappa_{ee'} = \frac{N_e}{N_e N_{e'}} \). Its variance is given by:

\[
V[c_{ee'}] = V \left[ \frac{\kappa_{ee'} \hat{N}_{ee'}}{\kappa_{ee'} \hat{N}_{ee'} + 1} \right]
\]  

(S31)

Approximating this variance using the delta method, we get:

\[
V[c_{ee'}] \approx V \left( \frac{\kappa_{ee'} \hat{N}_{ee'}}{\kappa_{ee'} \hat{N}_{ee'} + 1} \right)^2
dc_{ee'} \frac{d\kappa_{ee'}}{dN_{ee'}} = -\hat{N} \frac{N_e + N_{e'}}{N_e N_{e'}} \approx k \frac{N_e + N_{e'}}{N_e N_{e'}},
\]

where the approximation uses the fact that \( \hat{N} \left( \frac{N_e + N_{e'}}{N_e N_{e'}} \right) \gg N_e N_{e'} \). Using the expectation for \( N_{ee'} \), \( \hat{N}_{ee'} \), and its variance, \( \hat{N}_{ee'} \hat{N}_{ee'} \left( 1 - \hat{N}_{ee'}^{\text{post}} \hat{N}_{ee'} \right) \), results in the following (non-zero) error-variance for the synergy in educational pair \((e, e')\):

\[
V[c_{ee'}] \approx \hat{N} \hat{N}_{ee'} \left( 1 - \hat{N}_{ee'}^{\text{post}} \frac{N_e + N_{e'}}{N_e N_{e'}} \right)^2
\]  

(S32)

D.4.3 Empirical estimate of the magnitude of measurement error in educational synergy

How well does eq. (S32) perform? To answer this, we need to calculate an empirical counterpart to the theoretical error-variance. One way to do so is exploiting the timeseries dimension of the data: because we observe the synergy between two educations in multiple years, we can estimate the variance of the estimated synergy within each educational pair across time.

The theoretical standard deviation that follows from eq. (S32) correlates surprisingly well with its empirical counterpart: the estimated rank correlation between the two quantities is 0.826 \((N = 113, 409)\). Fig. S10 shows a log-log scatterplot of empirical against theoretical standard deviations, where each dot represents an educational pair. At 1.06, the estimated slope is close to 1 and, with an \( R^2 \) of 0.60, the regression has a good fit. Overall, therefore, there is substantial empirical support for the model-based eq. (S32) as a prediction of the heteroscedasticity in \( c_{ee'} \)'s measurement error.
D.5 Correcting for measurement error in wage regressions

D.5.1 Measurement error at the worker-establishment level

Worker-establishment level coworker synergy, $C_{wpw_t}$, is calculated from synergies in educational pairs, $c_{ee'}$, by taking the weighted average of the latter quantity across a worker’s coworkers (see eq. (S6)). Assuming that measurement errors are uncorrelated across educational pairs and with an establishment’s employment composition, the error-variance of $C_{wpw_t}$ becomes:

$$V[C_{wpw_t}] = \sum_e \left( \frac{E_{epw_t} - \delta_{ewt}}{E_{pwt} - 1} \right)^2 V[c_{ewt}]$$  \hspace{1cm} (S33)

The smaller $V[C_{wpw_t}]$ is, the more accurately measured $C_{wpw_t}$ will be. If $V[C_{wpw_t}]$ were known for each observation in the data, we could try to avoid attenuation biases by restricting the sample to observations with minimal measurement error. Following this logic, worker-year observations are divided into ten error-variance bins with equal numbers of observations. Next, a separate regression analysis is run in each bin, $B_b$, with $b \in \{1, 2, ..., 10\}$:

$$\log_{10}(wagew_{wt}) = X_{wt}\beta_x + Q_{pw_t}\beta_p + \gamma^h_c C_{wpw_t} + \gamma^b_s S_{wpw_t} + \epsilon_{wt}$$  \hspace{1cm} (S34)

where $(w, t) \in B_b$ denotes the set of observations in decile $b$ and $\epsilon_{wt}$ is an error term. The control variables collected in $X_{wt}$ and $Q_{pwt}$ are a 4th order polynomial of worker age, year fixed-effects, educational-track fixed-effects, coworker-shares by educational level, and the logarithm of establishment size. Controlling for these characteristics ensures that error-variances and wages are approximately conditionally independent.

The error-variance in coworker-synergy, $V[C_{wpw_t}]$, can be derived from eq. (S32) for each observation in the data. Dropping subscripts for notational clarity, collecting all regressors other than $C_{wpw_t}$ in vector $Z$ and letting $y$ denote $\log_{10}(wagew_{wt})$, (S34) can be written as:

$$y = \gamma^h_c C + Z\Theta_b + \epsilon$$  \hspace{1cm} (S35)
In this equation, $C$ is a mismeasured version of the underlying quantity $\hat{C}$:

$$C = \hat{C} + \eta$$

where $\eta$ is a measurement error, which is assumed to be uncorrelated with the true coworker synergy, $\hat{C}$, the regressors in $Z$ and the disturbance term $\epsilon$. Furthermore, let’s assume that the actual relation between $y$ and $C$ is given by:

$$y = \gamma_c \hat{C} + Z \Theta + \tilde{\epsilon}$$ (S36)

That is, the real effect of coworker synergy is constant across error bins. Although this will not be imposed in the empirical analyses, for expositional convenience, it is also assumed that other parameters are constant across bins, i.e., $\Theta_b = \Theta$ for all $b$. The estimate of $\gamma_b$ in (S35) will be biased and the size of the bias will depend on $V[\eta]$ in bin $b$. Given the Frisch-Waugh-Lovell theorem, the OLS estimate of $\gamma^b_c$ can be written as:

$$\hat{\gamma}^b_c = \frac{\text{Cov}[\hat{C}, y]}{\text{Var}[\hat{C}]}$$ (S37)

where $\hat{C}$ is the residual of a regression of $C$ on $Z$:

$$C = Z \Theta + \hat{C}$$ (S38)

Similarly, $\hat{\hat{C}}$ represents the residual of a regression of $\hat{C}$ on $Z$:

$$\hat{\hat{C}} = Z \Theta + \hat{\hat{C}}$$ (S39)

Because $\eta$ is uncorrelated with the columns of $Z$, (S38) and (S39) must have the same parameters, $\Theta$. Consequently:

$$\hat{\hat{C}} = \hat{C} + \eta$$ (S40)

Using (S40) and (S36), the numerator of (S37) can be written as:

$$\text{Cov}[\hat{\hat{C}}, y] = \text{Cov}[\hat{C} + \eta, \gamma_c \hat{C} + Z \Theta + \tilde{\epsilon}]$$

$$\text{Cov}[\hat{\hat{C}}, y] = \gamma_c \text{Cov}[\hat{\hat{C}}, \hat{C}] + \sum_k \Theta_k \text{Cov}[\hat{\hat{C}}, Z_k] + \text{Cov}[\hat{\hat{C}}, \epsilon] + \gamma_c \text{Cov}[\eta, \hat{C}] + \sum_k \Theta_k \text{Cov}[\eta, Z_k] + \text{Cov}[\eta, \epsilon]$$ (S41)

where $k$ indexes columns of matrix $Z$. Except for the first term, all terms in (S41) are equal to zero. The second term equals zero, because $\hat{\hat{C}}$ is the residual of a regression on $Z$ and therefore orthogonal to each column of $Z$. The third term equals zero because $\hat{\hat{C}} = \hat{C} + \eta$ and $\hat{C} \perp \epsilon$ because $\hat{C}$ is a linear combination of the regressors in (S35) and $\eta \perp \epsilon$ by assumption. The fourth to sixth
terms equal zero by the assumption that measurement errors are uncorrelated with regressors, the true, underlying coworker synergy, and with the residual in (S35). We therefore get:

$$\text{Cov}[\hat{C}, y] = \gamma_c \text{Cov}[\hat{C}, \hat{C}]$$

which, using (S39), can be written as:

$$\text{Cov}[\hat{C}, y] = \gamma_c \text{Cov}[\hat{C}, \hat{C} + Z\Theta]$$

Given that $\hat{C}$ is the residual of a regression on $Z$, this becomes:

$$\text{Cov}[\hat{C}, y] = \gamma_c V[\hat{C}]$$  \hspace{1cm} (S42)

Using (S42) in (S37), the following expression for $\hat{\gamma}_b^c$ results:

$$\hat{\gamma}_b^c = \gamma_c \left(1 - \frac{V[\eta]}{V[\hat{C}] + V[\eta]} \right).$$  \hspace{1cm} (S43)

**D.5.2 Errors-in-variables correction: extrapolation**

Eq. (S43) relates $\hat{\gamma}_b^c$, the estimated partial regression coefficient of coworker synergy in bin $b$, to the variable’s real (partial) effect, $\gamma_c$, and the ratio of $V[\eta]$, the error-variance, to $V[\hat{C}]$, the variance of $C_{wp,t}$ conditional on the other regressors of (S34), calculated as the variance of the residual of a regression of $C_{wp,t}$ on $X_{wt}$, $Q_{p,t}$ and $S_{wp,t}$ in bin $b$). It thus tells us for each bin how much bias we should expect in the estimated coworker-synergy effect.

Fig. S11a plots the estimated $\hat{\gamma}_b^c$’s against $\frac{V[\eta]}{V[\hat{C}]}$, where $V[\eta]$ is estimated using (S33). The relation between the estimated coworker-synergy effect and the error-variance share is striking: the lower the error-variance, the higher point-estimates become. Extrapolating the linear trend implied in (S43) suggests that the unbiased effect of coworker synergy (i.e., the value at which the trend line crosses the vertical axis) lies roughly between 0.4 and 0.55. This trend line downweights the two outliers with extreme error-variances. It is plausible that measurement errors are overestimated in these bins, because the educational track dummies in (S34) also absorb structural measurement-error components that are specific to educational tracks. Moreover, the delta method approximates standard deviations only imperfectly. As a consequence, the residual measurement-error variance may be smaller than what is plotted along the horizontal axis.

Fig. S11b shows that, whereas $\hat{\gamma}_c^c$ rises as $\frac{V[\eta]}{V[\hat{C}]}$ becomes smaller, $\hat{\gamma}_b^c$, the estimated effect of $S_{wp,t}$, drops. This is indeed what
one would expect if $\frac{V[\eta]}{V[C]}$ quantified the mismeasurement of $C_{wp_t}$. In that case, the strong positive correlation between coworker synergy and substitutability would mean that the downward bias in the estimated effect of the former induces an upward bias in the estimated effect of the latter.

Interestingly, disregarding the heteroscedasticity in the measurement errors of educational synergy yields very similar results, at least for the synergy effect. In this homoscedastic case, where $V[c_{ew}] = \sigma^2_\eta$ for all $(e, w, e)$, eq. (S33) simplifies to:

$$V[C_{wp_t}] = \sigma^2_\eta \sum_e \left( \frac{E_{wp_t} - \delta_{ew}}{E_{wp_t} - 1} \right)^2$$  \hspace{1cm} (S44)

In other words, the difference in error-variance across worker-establishment observations is driven by the sum of squared educational employment shares in the establishment. Although this approach ignores heterogeneity in the precision with which $c_{ew}$ is estimated, it still allows sorting observations (albeit imperfect) by their coworker-synergy error-variance. Fig. S12 shows the results when bins are created using (S44) instead of (S33). Note that (S44) yields error-variances up to a scaling factor $\sigma^2_\eta$. As a consequence, $\frac{V[\eta]}{V[C]}$ need not lie between 0 and 1. However, the scale of the horizontal axis is inconsequential for the extrapolation to a measurement-error-free world. The implied unbiased coworker-synergy effects are all but indistinguishable from those based on heteroscedastic measurement errors in Fig. S11.

D.5.3 Errors-in-variables correction: 2SLS

A different way to correct for measurement errors in coworker synergy is to instrument this variable with an alternative proxy for how well a worker fits her work environment (15). A plausible candidate for such a proxy is the match between a worker’s education and her occupation, $\mu_{eo}$, as defined in (S13). Assuming that measurement errors in coworker-synergy and education-occupation match are uncorrelated, using the latter as an instrument for the former should remove the attenuation bias. Table S12 compares the results
Fig. S12. Estimated effect of co-worker synergy by error-variance bin (homoscedastic case). Idem Fig. S11, but assuming homoscedastic measurement errors in educational synergy.

| dep. var.: | (1) | (2) |
|-----------|-----|-----|
| log(wage) | OLS | 2SLS |
| cow. syn. | 0.274*** | 0.547*** |
|           | (0.0030) | (0.0071) |
| cow. subst. | -0.044*** | -0.163*** |
|           | (0.0018) | (0.0033) |
| log(est. size) | 0.044*** | 0.043*** |
|           | (0.0003) | (0.0003) |
| 4th polyn. of age? | yes | yes |
| edu. level dum.? | yes | yes |
| fixed effects? | yr | yr |
| # obs. | 2,144,965 | 1,640,144 |
| # clust. | 364,642 | 323,400 |

First Stage

| match edu.-occ. | 0.164*** |
| t-stat. | 281.7 |

***: p<.01; **: p<.05; *: p<.1, standard errors clustered at worker level.

Table S12. Errors-in-variables estimates.

of this approach (column 2) to the preferred OLS specification of Table S7 in column (1).

With $\mu_{eo}$ as an instrument for $C_{wp_{t}}$, the 2SLS estimate of the wage effect of coworker synergy exceeds its OLS counterpart by a substantial margin. Interestingly, once again the negative effect of substitutability strengthens considerably as well. In fact, the effects of coworker synergy (i.e., of coworker complementarity) and substitutability reported in column (2) are – given their confidence intervals – indistinguishable from the ones implied by Fig. S11.

Table S13 concludes this section by collecting various estimates of the associations of wages with coworker synergy and substitutability. Compared to a simple OLS estimation, fixed effects models reduce the estimated effects of coworker synergy (complementarity) and substitutability. However, the causal effects of column (4), as well as the error-in-variables corrected models of column (5) and (6) suggest that the true effects in reality exceed the OLS estimates by a wide margin. Taking into consideration estimates’ standard errors, the last three columns report surprisingly similar wage effects. This suggests that the OLS-based effects reported in
the main text are conservative: they are likely to substantially underestimate the true coworker effects.

**Section SE. Career paths (Fig. 3)**

This section focuses on the evolution of coworker fit along workers’ career paths. To summarize coworker fit in a single variable it uses *coworker complementarity*, i.e., the component of coworker synergy that is orthogonal to coworker substitutability, $\hat{m}_{w_{wp},t}$, as defined in eq. (6).

Panel a of Fig. 3 in the main text shows that coworker complementarity increases with work experience along a curve that is remarkably similar to the Mincerian wage curve. The Mincerian wage curve expresses the relation between a worker’s wage and her work experience and was first proposed in (25). In Fig. 3, this curve is created by plotting the residual of the following wage regression against work experience:

$$\log_{10}(wage_{wt}) = \alpha_{ct} + \omega_{wt}$$  \hspace{1cm} (S45)

where $\alpha_{ct}$ is an interaction of year dummies with educational-level dummies.

Fig. S13 repeats the plot of the estimated residual, $\hat{\omega}_{wt}$, as well as of complementarity, $\hat{m}_{w_{wp},t}$ (the residual of coworker synergy after controlling for coworker substitutability), against work experience, while adding similar plots for four different subsamples. The subsamples are restricted to workers with upper secondary, post-secondary, college or post-graduate degrees.

As workers progress in their careers, coworker complementarity rises along a concave curve in all of the depicted samples. Note that, although wages closely track complementarity in all samples, the wage increase over a worker’s career is an order of magnitude larger than what the increase in complementarity could account for, given the estimated effects on wages reported in section D. This suggests that a worker’s productivity improves at a diminishing rate for a number of other reasons besides coworker fit.

Another striking fact is that educational choices leave a surprisingly long-lasting imprint on workers’ careers: the complementarity to coworkers (which is based on the relation to a worker’s own education) keeps rising for up to 20 years after a worker has started her career (and typically, finished her education). In the sample of workers with upper secondary degrees, the complementarity curves are somewhat steeper and flatten sooner than in the samples of higher educated workers. This suggests that the higher a worker’s...
level of education, the longer the opportunities for improvements in co-worker complementarity will persist.

Section SF. Wage premiums

The analyses on wages have shown that a high level of co-worker complementarity is associated with a significant wage premium. How does this wage premium relate to other well-known wage premiums? This section answers this question for the returns to schooling, the large-plant premium and the urban wage premium. It will also show that co-worker complementarities do not just explain substantial parts of these premiums, but also strongly enhance them. These findings corroborate the existence of a social component of human capital, explaining how well-known premiums associated with work environments — as opposed to with individual characteristics — reflect how well the skill sets of coworkers relate to one another.

F.1 Returns to schooling

F.1.1 Comparison to United States

To put Swedish returns to schooling into perspective, Table S14 compares returns to schooling in Sweden to those in the United States. All analyses contain a 4th order polynomial of a worker’s age and year dummies as control variables. Col. (1) regresses log_{10}(wage) on a worker’s educational level in Sweden. Column (2) repeats this analysis using a worker’s percentile rank in the overall wage distribution as a dependent variable. Cols (3) and (4) show analogous regressions, using US census data for the years 2001 to 2010 retrieved from (40) and cleaned following the procedures outlined in (37).

US workers are divided into educational categories that are meant to mimic the Swedish categories as closely as possible. In the US models, the omitted category is workers with at most middle school, “sec.” refers to workers who completed 11th grade, “upper sec.” to workers who completed grade 12 or have a high school degree, “post-sec.” to workers with associate degrees or some college education, “college” to workers with a bachelor’s, master’s or professional degree and “post-graduate” to workers with a doctoral degree.

The sample is constructed using similar criteria as those used for the Swedish data: apart from using the same age restrictions, self-employed, female and government employed workers as well as workers in employment agencies are excluded. Moreover, workers
Table S14. Returns to education. ***: p<.01; **: p<.05, *: p<.1. Standard errors (in parentheses) are clustered at the individual level when using Swedish data and robust in the US data sample.

below the poverty threshold in real 2010 USD are excluded as well, as are workers in the top and bottom 0.5th wage-percentile. Wages are annual wages and the regressions are weighted by workers’ sample weights. Col. (5) repeats the preferred specification of Table S7, using a worker’s percentile rank in the wage distribution as a dependent variable to show the effect of coworker synergy and substitutability on a worker’s wage rank. Note that this regression uses only workers with at least upper secondary education.

At 183%, the absolute wage premium of college-over-primary education in the U.S. is over three times Sweden’s premium of 54%. However, in relative terms, the difference is less pronounced. College-educated workers in Sweden are found on average 26.4 wage percentiles above workers with only primary education, whereas in the U.S., this premium is 42.2 percentiles. Again, the returns to coworker synergy reported in column (5) are substantial. The point estimate implies that moving from the 10th to the 90th synergy percentile is associated with a 10.3 percentiles rise in the wage distribution, whereas the drop in wage-rank associated with a similar increase in substitutability is 2.9 percentiles.

F.1.2 Educational wage premiums by coworker complementarity quintile (Fig. 4)

Fig. 4, panel a in the main text shows that not just wages, but also the returns to schooling a worker receives may depend on whether or not she works with complementary coworkers. The graphs in this figure are created as follows. First, all worker-year observations are divided into five equally-sized bins based on their coworker complementarity, $\hat{m}_{wp,t}$. Next, educational level dummies are
interacted with these complementarity quintiles in the following regression equation:

$$\log_{10}(\text{wage}) = X_{wt}\beta_x + Q_{pwt}\beta_p + \sum_{b=1}^{5} \sum_{l=1}^{6} \beta_{b\times l} B_b L_l + \epsilon_{wt}$$  \hfill (S46)

$b_b$ is a dummy for complementarity quintile $b$ and $L_l$ a dummy for educational level $l$. $X_{wt}$ and $Q_{pwt}$ follow the preferred specification and control for establishment size, year dummies and a 4th order polynomial of age.

Panel b of Fig. 4 repeats this analysis, controlling for worker fixed effects. Given that very few workers in the estimation sample change their educational attainment, the variance that allows estimating the $\beta_{b\times l}$ parameters comes mostly from workers’ switching complementarity quintiles. Note that the estimated returns to education are substantially lower in this specification. This is in large part due to the fact that the average effect of having a higher level of education is mostly absorbed in the worker fixed effects. This happens because of the aforementioned paucity of workers who upgrade their education. The plotted coefficients should therefore be interpreted as increases over these average effects. Taking this into account, the dependence of educational wage premiums on complementarity remains strikingly strong for higher educated workers.

F.2 Large-plant premium (Fig. 5, panel a, b and c)

The main text also studies the large-plant premium (LPP), the premium associated with working in large establishments. The conjecture is that the LPP reflects that large establishments have more complementary workforces. The force behind this explanation is that larger establishments allow for a deeper division of labor, which leads to greater interdependencies among workers with different, yet complementary skills.

If this were the case, the LPP should disappear once coworker fit is controlled for. Contradicting this, Table S15 (depicted graphically as panel c of Fig. 5 in the main text) shows that, controlling for weighted-average coworker-synergy and substitutability as defined in eqs (4) and (5) doesn’t affect the wage elasticity with respect to establishment size, regardless of a worker’s level of education. However, given that the definition based on weighted averages was chosen to ensure that the coworker-fit variables would be close to uncorrelated with establishment size, this is not surprising. When instead controlling for the number of synergistic coworkers and of close substitutes among one’s coworkers (i.e., using the indices of eqs. (S4) and (S5)), the LPP is reduced by 53% in the full sample. Moreover, the LPP disappears completely for workers with over upper secondary education. Given that, depending on the educational level, synergistic coworkers make up on average only between 6.4% and 15.6% of coworkers and close substitutes just between 6.8% and 11.4% of coworkers, it is remarkable that controlling for the presence of these relatively small sets of coworkers can account for the entire LPP.

Panels a and b of Fig. 5 show that the benefits of complementarity rise with establishment size. These panels were constructed using an analogous regression to eq. (S46) above:

$$\log_{10}(\text{wage}) = X_{wt}\beta_x + \sum_{b=1}^{5} B_{b\cdot \text{est. size}} \hat{m}_{wpwt} + \epsilon_{wt}$$  \hfill (S47)

where $B_{b\cdot \text{est. size}}$ represents dummy variables for the establishment size quintiles, $\hat{m}_{wpwt}$ coworker complementarity and the remaining
Table S15. Large-plant premium. Wage elasticity with respect to establishment size. First row: analyses with year dummies, a 4th polynomial of age, and, in the last column, educational level dummies as controls. Second row: controls added for average coworker synergy and substitutability as defined in eqs (S6) and (S7). Third row: controls instead for the number of synergistic coworkers and the number of coworkers that are close substitutes to the focal worker as defined in eqs (S4) and (S5). Standard errors in parentheses.

Variables are defined as in eq. (S46).

F.3 Urban wage premium (Fig. 5, panel d, e and f)

The final premium studied in the main text is the urban wage premium (UWP): the positive elasticity of wages with respect to city size. For instance, (35) find that, for every doubling of a region’s population, wages rise with about 5%. The UWP is often attributed to better learning opportunities in larger cities. Here, an alternative explanation is put forward: large cities help workers find complementary coworkers.

If the UWP indeed reflects better coworker matching opportunities in large cities, accounting for coworker fit should reduce the estimated wage elasticity with respect to region size. Table S16 (visualized in panel f of Fig. 5 of the main text) shows that this is the case for workers with high levels of education. The table summarizes how the UWP changes when controlling for coworker synergy and coworker substitutability. The reported results are derived from regressions of \(\log_{10}(wage)\) on \(\log_{10}(region\ size)\), where region size is the number of employees in each of Sweden’s 110 labor market areas. These regressions are run without any worker-level control-variables, such that they capture the raw (unconditional) UWP. Adding worker-level control-variables reduces the overall UWP, but does not substantially alter the relative reductions in UWP when adding coworker-fit controls. The first row of Table S16 shows UWPs that were estimated without controlling for coworker variables. The second row adds average coworker synergy and substitutability as defined in eqs (S6) and (S7) as control variables. The third row uses as control variables the count of highly synergistic coworkers and coworkers who are close substitutes to the focal worker, as defined in eqs (S4) and (S5).

Controlling for average coworker synergy and substitutability does not change the point estimates of the UWP for workers with low levels of education. However, it does substantially change the observed UWP of workers with college or post-graduate degrees, by, respectively 30% and 50%. Controlling for the count-based coworker variables further reduces the UWP by 21% for workers with post-secondary degrees, by 34% for workers with a college degree and by 74% for workers with a post-graduate degree. The UWP for workers with upper secondary school degrees is left unchanged.

Panel c of Fig. 5 in the main text plots UWP-estimates for workers in different complementarity quintiles. These estimates were derived as interaction effects in the following wage regression:
| controls | upper sec. OLS | post-sec. OLS | tertiary OLS | PhD OLS | all OLS |
|----------|----------------|---------------|--------------|---------|---------|
| no controls | 0.024*** | 0.037*** | 0.050*** | 0.023*** | 0.047*** |
|           | (0.0026) | (0.0044) | (0.0046) | (0.0060) | (0.0035) |
| $C_{wp,t} + S_{wp,t}$ cntrls | 0.027*** | 0.040*** | 0.035*** | 0.011** | 0.051*** |
|           | (0.0029) | (0.0060) | (0.0047) | (0.0056) | (0.0045) |
| $C_{wp,t} + S_{#wp,t}$ cntrls | 0.025*** | 0.029*** | 0.033*** | 0.006 | 0.043*** |
|           | (0.0029) | (0.0044) | (0.0048) | (0.0052) | (0.0034) |

***: p<.01; **: p<.05; *: p<.1, standard errors clustered at the labor market area level.

Table S16. UWP, conditional on co-worker fit. Wage elasticity with respect to the size of a labor market area (total employment in the area). Standard errors in parentheses.

\[
\log_{10}(wage) = \sum_{b=1}^{5} \beta_{b \times P \times S} B_{b}^{comp} \log_{10}(region \ size) + \epsilon_{wt}
\]  

where $B_{b}^{comp}$ represents dummy variables for quintiles of complementarity, $\hat{m}_{wp,t}$, and $\log_{10}(region \ size)$ is the (base-10) logarithm of the population size of Swedish labor market areas. Standard errors are clustered at the level of labor market areas.

The UWP is much more pronounced for workers who participate in highly complementary work environments than for those who don’t, rising about five-fold between the bottom and the top complementarity quintiles. The analysis is repeated with worker fixed effects added in panel d of Fig. 5. This specification controls for ability-based sorting of workers into large cities. Although the average UWP drops, suggesting that the UWP may in part reflect spatial sorting of workers, the interaction with coworker complementarity remains clearly visible.