Automation Experiments and Inequality

Seth Gordon Benzell
Chapman, MIT, & Stanford

Kyle R. Myers
Harvard & NBER

August 14, 2026

Many experiments study the productivity effects of automation technologies such as generative algorithms. A central question in these experiments relates to inequality: does the technology increase output more for high- or low-skill workers performing the same job? However, the theoretical content of this empirical test has been unclear. To better understand this test, we develop a partial-equilibrium model of task-based production in which workers differ in task-specific skills and perform each task themselves or delegate it to technology. Importantly, we show that the inequality effect of automation depends on the interaction of two factors: (*i*) the correlation in task-level skills across workers, and (*ii*) workers' skills relative to the technology's capability. This interaction can produce opposite inequality effects in seemingly similar experiments and reversals within the same setting; as technology improves, inequality may first fall and then rise. Still, the model yields clear predictions when key factors are observable. We verify one of these predictions in a reanalysis of an experiment involving radiologists and a diagnostic technology. Looking forward, the model and some descriptive statistics of skill correlations suggest that the diversity of automation technologies will shape the evolution of inequality.

Benzell: benzell@chapman.edu. Myers: kmyers@hbs.edu. This project received support from the Harvard Business School. There are no financial relationships or other potential conflicts of interest that apply to any of the authors. We appreciate helpful feedback from Rem Koning and Philip Trammell. Thanks also to Alex Moehring for assistance with radiology experiment data.

1. Introduction

Recent advances in the capabilities of algorithms have led to a surge of investigations into the economics of automation technologies. An important, open question is the effect of these technologies on inequality. Often, the question is: does the technological advance help low-skilled workers more (reducing inequality) or does it help high-skilled workers more (increasing inequality)?

Researchers have approached this question using both macroeconomic and microeconomic methods. In this paper, we are interested in what can and cannot be learned from some of the microeconomics.[1] Specifically, we focus on the microeconomics of automation experiments: studies that use exogenous variation in the availability of a technology to estimate productivity effects. These experiments are valuable because they yield internally valid estimates (typically, in partial equilibrium) and provide managers and policymakers information about technologies at a relatively high frequency.

In order to answer questions about inequality, experimenters often test whether a technology differentially increases the output of lower- or higher-skilled workers with the same job.[2] In June of 2025, *The Economist* summarized the inequality effects from automation experiments to date: "*Although early studies suggested that lower performers could benefit ... newer studies look at more complex tasks... In these contexts, high performers benefit far more than their lower-performing peers,*" which implies that the skill *level* of the workforce is an important feature and suggests a future of skill-biased technical change, as implied by the article's title "*How AI will Divide the Best from the Rest*" (The Economist 2025).

But what exactly do we learn if we see an automation experiment lead to more, or less, inequality? What features of the workforce are relevant to the inequality effect? If a technology increases inequality now, will it continue to increase inequality as it becomes more capable?

To answer this group of questions, we introduce a model in the task-based tradition, with output increasing in task-level inputs. However, unlike many macroeconomic models of task-based production, we explicitly model jobs as production functions at the *individual level*. Individual workers are defined by their endowments of skill in each task, which motivates our focus on a feature of the workforce that is rarely highlighted: the correlation of workers' skills across tasks. This correlation describes, for example, whether the most analytical physicians

[1] The macroeconomic work typically asks: after general equilibrium effects play out, will the new distribution of jobs and earnings be less or more unequal? The answer often appears to be that technologies that reduce demand for certain types of labor in the short run may increase demand for that labor in the long run, and vice-versa. For overviews covering historical and current waves of automation, see: Katz and Autor (1999); Acemoglu and Autor (2011); Acemoglu and Restrepo (2019).

[2] For examples of automation trials that report inequality results, see: Noy and Zhang (2023); Chen and Chan (2024); Choi, Monahan, and Schwarcz (2024); Kreitmeir and Raschky (2024); Cui et al. (2025); Brynjolfsson, Li, and Raymond (2025); Dell'Acqua et al. (2025); Kanazawa et al. (2025); Kim et al. (2025); Otis et al. (2025); Roldán-Monés (2025).

are also the most empathetic, or whether the best fielder on the baseball team is also the best batter.

We are focused on the context of short-term automation experiments, in which the sample population and job responsibilities are fixed. The model allows for either positively or negatively correlated skills among workers. Automation technologies can perform one (or more) of the tasks in production, and a technology's capability may be below or above a given worker's skill. Workers are assumed to have rational expectations over whether using a technology will improve their performance.[3]

The model provides clear predictions about the change in inequality (i.e., the difference in output between worker types) given a marginal increase in an automation technology's capability. Even in simple settings with only two tasks, perfect skill correlations, and single-task automation, we obtain results that may seem counter-intuitive at first, but become very clear through the lens of the model. Changes in inequality do not depend merely on whether workers are generally high- or low-skilled, as the quote from *The Economist* implies or as intuition might suggest. It is the interaction of skill correlation and technological capability that determines the inequality effect. Furthermore, the inequality effect need not be monotonic in the automation technology's capability. Ceteris paribus, as a technology's capabilities improve, inequality can decrease and then increase, and vice versa. In several extensions to the model, allowing for imperfect-foresight in AI adoption, differential ability in AI use, and other modes of human-AI interaction, we show these key results hold.

Allowing for more than one task to be automated yields additional results. One is that a sure path to equality is when automation improvements are balanced across tasks. Intuitively, this is because when all human skills are surpassed by AI technologies, the gap in with-AI productivity between high- and low-skilled workers must be reduced. This motivates the potential importance of a diversity of technological advances when it comes to the pursuit of equality. However, the interpretation of this scenario as positive relies on two important assumptions. First is that the job still exists after human skill at each individual task is surpassed. This may be because there is some intrinsic need or desire for humans to be involved in the role. Second is that humans are homogeneous in providing this residual non-automatable function.

While several dynamics are possible, ours is not an "anything goes" model. When information is available about workers' task-level skills and the relative capability of the automation technology, our model can be used for prediction or extrapolation. It is designed to be useful for researchers, organizations, managers and policymakers as they design and interpret automation experiments.

[3]This assumption is not innocuous, as some automation experiments do find that workers perform worse than under placebo conditions when given access to generative algorithms (Dell'Acqua et al. 2025; Otis et al. 2025), which suggests the presence of biased beliefs.

We test one major prediction of the model in a re-analysis of the radiology experiment previously studied by Yu et al. (2024); Moehring et al. (2025); Agarwal et al. (2025). This setting is ideal for testing our model, because task-level AI quality is observed and the data allow us to estimate task-level skill correlations across radiologists. Yu et al. (2024) examine whether unaided performance predicts gains from AI assistance. Our analyses involve new investigations into the skill correlations in the sample and statistical tests of inequality effects motivated by the structure of our model. We find strong evidence of positive skill correlations, which our model predicts will lead to a negative inequality effect given the technology's capabilities. Consistent with some of the results in Yu et al. (2024), even after adjusting for regression to the mean, we find that access to automation reduced inequality in overall radiologist productivity, just as the model predicts.

When exact task-level technology capabilities or task-level skill correlations are not observed, our model can still be useful in predicting into which settings experiments can be generalized. This is because existing theory and data provide some guidance on when and where we should expect skill correlations to be positive or negative. We discuss this in the latter portion of the paper drawing on the literature on cognitive testing and on occupation and country-level skill correlations (based on SAT scores) from the US National Longitudinal Survey of Youth (BLS 2024). While cognitive skills are consistently positively correlated, cognitive and physical skills are often negatively correlated. We also apply Kremer's (1993) O-ring model of production to show why skills may be negatively correlated within firms even if they are positively correlated in the population. These discussions also prompt interesting questions about the endogenous evolution of workers' skills.

While we focus on partial equilibrium and microeconomics, our model suggests new insights into the macroeconomics of automation. In particular, the model's focus on skill correlations emphasizes the importance of understanding whether the "experiment" is occurring at the job-, organization-, sectoral-, or societal-level. We discuss a range of normative implications; for example, we show how a manager's optimal investment decision when choosing between improvements to human skill or technological capabilities depends on the features our model highlights.

After a brief literature review below, the paper is organized as follows: Section 2 lays out the model in detail; Section 3 derives results from the model; Section 4 tests a major prediction of the model in the context of radiology; Section 5 suggests normative implications of our model and connects our theoretical results to evidence on skill correlations across the economy; and Section 6 concludes with a discussion of our model's implications, limitations, and possible extensions.

Related Literature

Our paper builds on research on the productivity impact of emerging technologies, most notably, generative algorithms, theories that model jobs as multi-task production functions, and the broader economics of automation.

Generative Algorithms and Inequality. Table 1 summarizes recent automation experiments that have estimated inequality effects due to generative algorithms. As the table shows, findings are decidedly mixed. Our theoretical framework helps rationalize these seemingly inconsistent findings. Given the rapid pace of algorithmic progress, experimental results at one point in time may not predict effects as technologies continue to advance. Moving forward, experiments that measure not only productivity effects but also task-level skill distributions in the sample, the types of workers who use the technology most intensively, and the technology's capability relative to worker skills will provide more robust guidance for anticipating distributional consequences.

TABLE 1
Inequality Results in Recent Automation Experiments

	job	*methodology*	*inequality effect*
Brynjolfsson, Li, and Raymond (2025)	Customer support	Natural Experiment	↓
Chen and Chan (2024)	Ad copywriting	RCT	↓
Choi, Monahan, and Schwarcz (2024)	Legal work	RCT	↓
Cui et al. (2025)	Coding	RCT	↓
Dell'Acqua et al. (2025)	Consulting	RCT	↓
Kanazawa et al. (2025)	Taxi driving	Natural Experiment	↓
Maršál and Perkowski (2025)	Banking	RCT	↓
Noy and Zhang (2023)	Professional writing	RCT	↓
Kim et al. (2025)	Investment analysis	RCT	↑
Otis et al. (2025)	Entrepreneurship	RCT	↑
Roldán-Monés (2025)	Debating	RCT	↑

Notes: Summarizes recent automation experiments. Studies are sorted based on whether they find automation to decrease (↓) or increase (↑) inequality, whether they use a formal randomized controlled trial (RCT) or a natural experiment, and then alphabetically.

Jobs as Production Functions. We are not the first to treat jobs as production functions. One close example is Autor and Handel (2013; see their Eq. 2), which models job-level output as an exponential function of workers' task-level skills and job-task-specific elasticities. There, the focus is not on automation, but rather on predicting job-level sorting. Deming (2017) measures the relationship of workers' math and social skills to wages, finding that they are complementary in production. But he does not focus on the implications of different correlations in these skills across workers. Dessein and Santos (2006) also consider jobs as bundles of tasks, although the focal question is how organizations endogenously decide which tasks are bundled together. Our model's highlighting of task-level correlation in skills should prove useful in future work using worker-level production functions, specifically, and the division of labor more generally (e.g., Becker and Murphy 1992; Autor, Levy, and Murnane 2003). We review empirical studies of skill correlations later in Section 5.3.

Automation in General. Related work in the broader economics of automation includes Athey, Bryan, and Gans (2020), who develop a framework for optimal delegation between humans and algorithms, and Xu et al. (2025), who study how generative algorithms affect organizational structure and the allocation of tasks within firms. Trammell (2025) focuses on the automation of tasks when there is learning-by-doing in a sequence of tasks that constitute the workflow for a job. There is also an emerging stream focused on the design of human-technology collaborations (e.g., Kleinberg et al. 2018; Mullainathan and Obermeyer 2022; Agarwal et al. 2025; Agarwal, Moehring, and Wolitzky 2025). Regarding the macroeconomics of automation and inequality, Katz and Autor (1999), Acemoglu and Autor (2011), and Acemoglu and Restrepo (2019) provide comprehensive overviews of how technological change affects the wage structure through general equilibrium channels. More recently, Autor and Thompson (2025) examine how generative algorithms affect the returns to expertise, while Garicano and Rayo (2025) study implications for training and apprenticeship.

Our partial equilibrium approach abstracts from general equilibrium effects in order to isolate the direct productivity channel, which yields predictions about which workers benefit from automation in the short run before labor markets adjust. This is the relevant experimental context and the focus of this paper. However, a natural extension of our model is to embed it into leading general equilibrium models of automation (e.g., Zeira 1998; Acemoglu and Restrepo 2018) to predict longer-term and general equilibrium effects.

2. Model

Here, we outline our model of jobs as production functions and the automation of tasks. The tasks of the production functions and the skills of workers are exogenous and fixed. This short-run, partial-equilibrium view reflects the format of most automation experiments, given their pursuit of internal validity.[4] We begin with a graphical preview of the model that delivers the intuition behind two key results. Throughout, we model jobs as worker-level

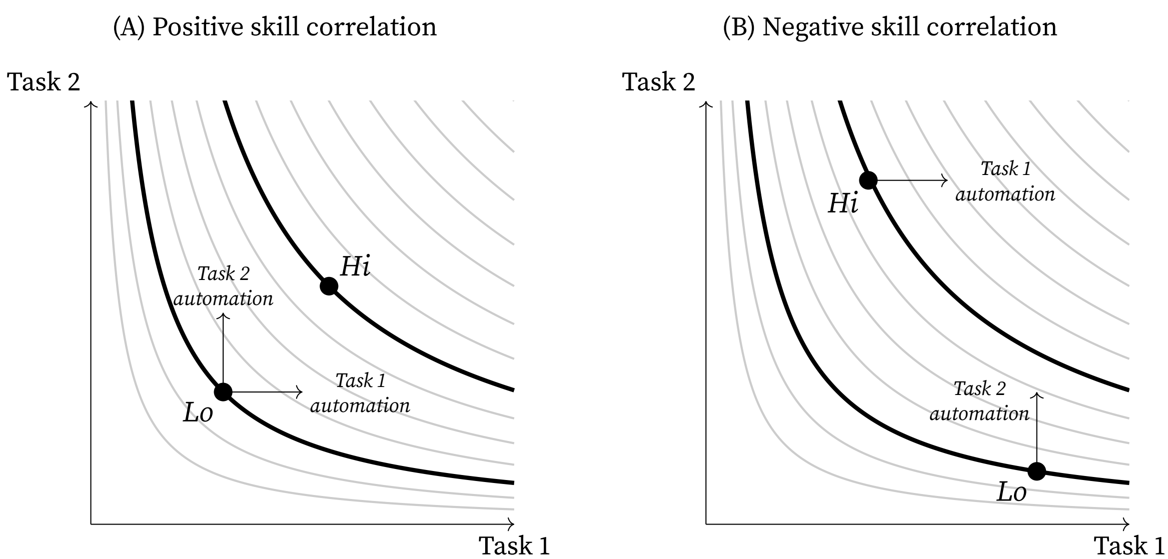
[4]See the literature review above for some work on the long-run, general-equilibrium effects of automation.

production functions that aggregate task-level output (per workers' skills) to worker-level output.

2.1. Mechanism Preview

Consider two samples of workers with identical constant elasticity of substitution (CES) production functions over a pair of tasks. Further, assume both samples are identical in their overall productivity distribution: one-half of workers are high-productivity, and one-half are low-productivity. However, in one sample, skills are positively correlated across workers, while in the other they are negatively correlated. As Figure 1 shows, an increase in the capability of a technology that automates one task can *decrease* inequality in one sample while *increasing* inequality in the other.

FIGURE 1
Initial Changes in Output after Automation



Note: Plots isoquants of production possibilities, highlighting the isoquant for a worker with higher output (*Hi*) and a worker with lower output (*Lo*) in two scenarios, where task-level skills are positively correlated (Panel A) or negatively correlated (Panel B). Workers' pre-automation endowments are shown by the black circles. If a task-specific automation technology is made available and the technology's capability exceeds the workers' skill endowment, they will use the technology, which will shift them to a new production frontier as shown by the arrows. For illustrative purposes, the first worker to benefit from automation of a given task is highlighted.

In Figure 1, Panel A, skills are positively correlated and in Panel B skills are negatively correlated. From an aggregate perspective, the cases are very similar. Still, as the figure illustrates, a marginal increase in automation technology will have very different effects on inequality across the pair of cases.

In the positive skill correlation case, as automation technology improves from an initial capability of zero, the first user will be the low type. However, in the negative correlation case either type may use the technology first – in this example, the high-type benefits first from task 1 technology improvements. Thus, the low type will benefit first and inequality will decrease in the Panel A case. But the high type will benefit first and inequality will increase in the Panel B case.

A second key result of the model is that the inequality effect need not be monotonic in the automation technology's capability. Continuing with Figure 1 Panel A, let's again focus on Task 1 automation. As just noted, the low type will be the initial users, which decreases inequality initially. However, once the technology's capability surpasses the high type's skill and the high type begins using it, output differences will be governed by differences in Task 2 skills. Now, since the high type is more skilled at Task 2 (and tasks are complementary), inequality will *increase* as the Task 1 automation technology improves.

2.2. Modeling Worker Productivity Without AI

We now outline the complete model, first without any AI available. There is a single job in which each worker i produces a non-negative quantity of output, Y_i. The job consists of a set of tasks $t \in \{1, 2, \ldots, T\}$. Each worker has an exogenous, task-specific skill level $S_{it} \geq 0$. We describe worker i's skill endowment across tasks by the vector $\mathbf{S}_i \equiv (S_{i1}, \ldots, S_{iT})$.

Workers may also have access to a task-specific automation technology with capability $A_t \geq 0$. For each task, a worker can either perform the task themselves, producing S_{it} units of task-level output, or delegate the task to the technology, producing A_t units. Because using the technology is costless, the effective task-level output for task t is therefore: $\max(A_t, S_{it})$.[5]

Worker-level output is produced according to a general CES production function:

$$Y_i(\mathbf{A}) = \left(\sum_t \left(\alpha_t \cdot (\max\{A_t, S_{it}\})^\rho \right) \right)^{1/\rho}, \qquad \rho = \frac{\sigma - 1}{\sigma} \leq 1, \quad \sum_t \alpha_t = 1, \tag{1}$$

where σ is the job-specific elasticity of substitution across tasks and α_t is the input share of task t.

While most jobs contain a vast array of tasks, automation experiments often study a technology that can perform one particular task or group of tasks. It is therefore useful to partition the tasks in a job into two groups: a set of tasks that may be delegated to the

[5]Implicitly, our model assumes each worker inelastically supplies a unit of effort towards each task, and in return produces $\max\{A_t, S_{it}\}$ of task-level output. In other words, in the baseline model, output is a function of the worker's skill along each dimension, not any endogenous effort decisions. Still, a stable allocation of time across tasks can arise endogenously in a Cobb-Douglas setting if technology is interpreted as multiplying rather than substituting for worker skill. We thank Phil Trammell for this last point. Below, we discuss work in our Appendix that relaxes this and other assumptions.

technology, denoted by t, and a set that may not be delegated, denoted by t^-. So throughout the paper, we focus on the simple case of two tasks that have equal input shares (i.e., $\alpha_t = \alpha_{t^-} = 1/2$). The production function then becomes:

$$Y_i(A_t) = \left(\frac{1}{2} \cdot \left((\max\{A_t, S_{it}\})^\rho + S_{it^-}^\rho \right) \right)^{1/\rho}, \tag{2}$$

where we also consider scenarios where both of the two tasks may be automated.[6]

For simpler results, we sometimes focus on the limiting case in which $\rho \to 0$. This yields the familiar Cobb-Douglas production function:

$$Y_i(A_t) = (\max\{A_t, S_{it}\} \cdot S_{it^-})^{1/2}. \tag{3}$$

2.3. Automation Experiments

We model an automation experiment as an exogenous increase in the capability of the technology, A_t. The introduction of a new technology can equivalently be understood as an increase from no useful technology, $A_t = 0$, to some positive capability level.

The technology may be worse than, equal to, or better than a particular worker at the automated task (i.e., $A_t \lesseqgtr S_{it}$). We assume that workers know the capability of the technology and use it whenever it produces more task-level output than they would produce themselves. Thus, worker i delegates task t if $A_t > S_{it}$ and performs the task themselves if $A_t \leq S_{it}$.

This setup deliberately abstracts from differences in workers' costs of using the technology, their skill in using it, and their beliefs about its capability. It also abstracts from changes in effort and the reallocation of effort across tasks. We consider these possibilities in the Appendix. Our baseline model instead isolates the direct effect of technological capability interacting with workers' existing task-specific skills.

2.4. Skill Correlations

We consider two limiting cases for the correlation of workers' skills across tasks. In both cases, there are two groups of workers: those with initially high output ($i = Hi$) and those with initially low output ($i = Lo$).[7]

Table 2 summarizes the positive- and negative-correlation cases. We use B and C to describe workers' absolute and comparative advantages, where $1 < C < B$.

[6]When both tasks may be automated, output is: $Y_i(A_1, A_2) = (\max\{A_1, S_{i1}\} \cdot \max\{A_2, S_{i2}\})^{1/2}$.

[7]This binary categorization is simple, and it reflects the common empirical exercise of proxying for workers' skills using their pre-experiment output levels to divide them into two groups (e.g., Noy and Zhang 2023; Chen and Chan 2024; Choi, Monahan, and Schwarcz 2024; Cui et al. 2025; Brynjolfsson, Li, and Raymond 2025; Dell'Acqua et al. 2025; Kanazawa et al. 2025; Otis et al. 2025; Roldán-Monés 2025).

TABLE 2
Skill Correlation Scenarios — Worker (i) and Job Task (t)

	positive corr.		*negative corr.*	
	$t = 1$	$t = 2$	$t = 1$	$t = 2$
$i = Hi$	B	$B \cdot C$	B	1
$i = Lo$	1	C	1	C

Notes: Reports workers' task-specific skill levels in the positive- and negative-correlation cases, where $1 < C < B$. In the negative-correlation case, skill is more dispersed across workers in task 1.

In the positive-correlation case, the high-output worker is more skilled at both tasks. Skills therefore have the same ranking across tasks. In the negative-correlation case, each worker has a comparative advantage in a different task. The high-output worker is more skilled at task 1, while the low-output worker is more skilled at task 2. Because $B > C$, the worker labeled *Hi* has higher initial output despite having lower skill at task 2.

We view the positive-correlation case as a simple representation of settings in which common factors, such as education or general human capital, raise workers' skills across many tasks. The negative-correlation case captures settings in which workers specialize in different tasks within the same job. We return to the empirical relevance of these two skill structures below.

2.5. Inequality Measures and Effects

There are many ways to measure inequality. In practice, worker-level output inequality is often evaluated by comparing changes in output between workers with different initial output levels. If initially high-output workers benefit more from automation than initially low-output workers, inequality increases. If initially low-output workers benefit more, inequality decreases.

We therefore define the output gap as:

$$\Delta \equiv \Delta(A_t) = Y_{Hi}(A_t) - Y_{Lo}(A_t) \ . \tag{4}$$

Our main effect of interest is:

$$\frac{\partial \Delta}{\partial A_t} \ , \tag{5}$$

which describes how the output gap changes as the capability of the automation technology increases. A positive value indicates that automation increases inequality, while a negative value indicates that automation decreases it.

As discussed in the introduction, automation experiments generally identify short-run, partial-equilibrium effects. Extrapolating from these estimates to long-run inequality is difficult without a structural model. The general-equilibrium effect of a technology depends not only on its direct effect on worker output, but also on substitution across workers and outputs, changes in worker sorting across jobs, and changes in the task composition of jobs.

A more modest extrapolation is to hold the nature of the job (i.e., the production function) fixed and consider how continued improvement in technological capabilities will affect inequality. Specifically, this relates to the second derivative of inequality with respect to the capability of the technology:

$$\frac{\partial^2 \Delta(A_t)}{\partial A_t^2} \,. \tag{6}$$

This derivative describes how the marginal effect of technology capability on inequality changes as the technology improves.

We also examine how the inequality effect depends on two features of the workplace. First, we consider how it depends on the size of the initial skill differences between workers:

$$\frac{\partial^2 \Delta}{\partial A_t \cdot \partial (B/C)} \,. \tag{7}$$

Second, we consider how it depends on the substitutability of tasks in the production function:

$$\frac{\partial^2 \Delta}{\partial A_t \cdot \partial \rho} \,. \tag{8}$$

In some cases, a sufficiently capable technology causes the initially low-output worker to become more productive than the initially high-output worker. We'll see that this rank reversal can occur under negatively correlated skills when the technology automates the task at which the initially high-output worker has an advantage. Once the ranking reverses, $\Delta(A_t)$ becomes negative even though the absolute difference in output may remain large.

We therefore also consider the absolute output gap:

$$|\Delta| \equiv |\Delta(A_t)| = |Y_{Hi}(A_t) - Y_{Lo}(A_t)| \,. \tag{9}$$

This measure captures the size of the output difference without depending on which worker is more productive.

Other common inequality measures include the coefficient of variation and the Gini coefficient. In our two-worker setting, these measures are simple transformations of the absolute output gap. In particular, the coefficient of variation is:

$$\frac{|Y_{Hi} - Y_{Lo}|}{Y_{Hi} + Y_{Lo}} \,,$$

and the Gini coefficient is one-half of this value. These scaled measures are useful for comparisons across settings with different average output levels. Because our theoretical comparisons generally hold the setting fixed, we focus on the simpler measures Δ and $|\Delta|$.

2.6. Effort, Rationality, and Other Alternative Assumptions

The baseline model makes several simplifying assumptions. It assumes that workers supply effort inelastically, correctly observe the capability of the technology, face the same cost of using it, and can substitute technology for their own task input one-for-one. Each of these assumptions may fail in practice. Effort may depend on worker skill or technology capability, workers may have incorrect beliefs about the technology, workers may differ in their ability or cost of using it, and technology and worker input may be imperfect substitutes within a task (e.g., Dell'Acqua et al. 2025; Otis et al. 2025; Fügener, Walzner, and Gupta 2026).

In Appendix A, we relax each of these assumptions. These extensions introduce additional heterogeneity and complicate the specific expressions for worker output. However, they do not overturn the main result of the baseline model: the direction of the inequality effect depends on the correlation of workers' skills across tasks and can be non-monotonic in technology capability. The purpose of the baseline assumptions is not to rule out these additional channels. Rather, it is to highlight the relationship between automation technologies' capabilities, the correlation of workers' skills across tasks, and inequality.

3. Model Results

3.1. Single-task Automation — Cobb-Douglas Case

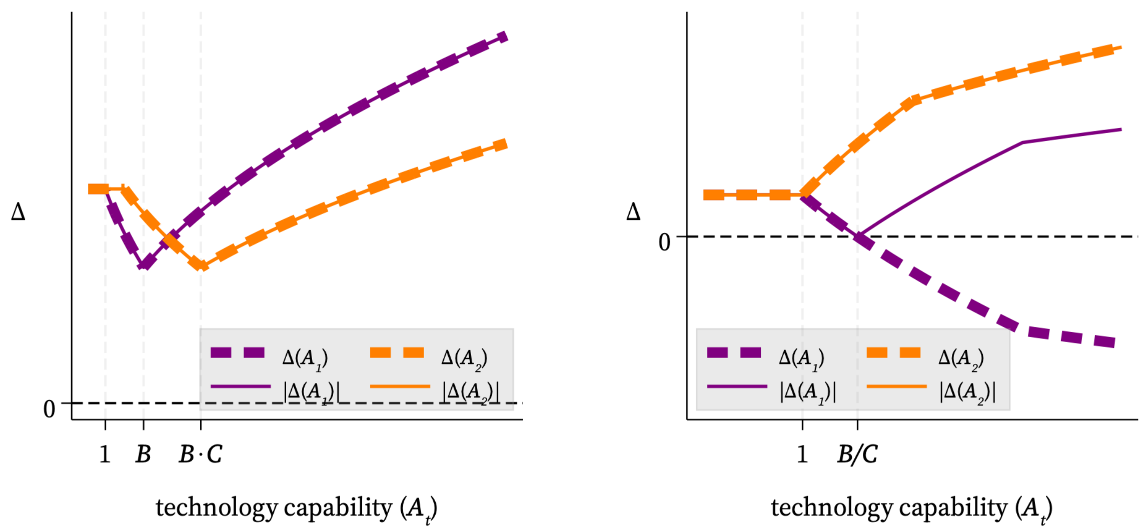
We begin with a focus on the simple case of Cobb-Douglas production and single-task automation. Figure 2 reports the results visually. It visualizes inequality as a function of the correlation in workers' skills (per panel), which task is automated (per line color), and the capability of the automation technology (per the x-axis).

A first result the figure illustrates is that when workers' skills are positively correlated, Δ must be weakly greater than zero. Equivalently, Δ and $|\Delta|$ align. In other words, when skills are positively correlated, there is no possibility for technology to "boost" the low-types above the high types. This intuitive result holds in the general CES case.

On the other hand, Figure 2 Panel A also shows that, regardless of whether task 1 or 2 is automated, there will be a non-monotonic relationship between inequality and the automation technology's capability. First, inequality declines, then it increases. Initially, improvements in automation help the low type "catch up" to the high-output type. This is because the technology's capability is greater than the low type (so, the low type uses the tool), but lower than the high type (so, the high type relies on their own skill). But eventually,

FIGURE 2
Inequality and Single-task Automation — Cobb-Douglas Production

(A) Positive skill correlation (B) Negative skill correlation



Note: Plots the level of inequality per first-difference inequality (Δ; solid line) or absolute inequality (|Δ|; dashed line) depending on skill correlations, the task being automated, and the capability of the technology.

the technology is better than both types, and the high-type's absolute advantage in the non-automated task becomes the key determinant of output differences and inequality grows due to the complementarity in the tasks. To summarize, when skills are positively correlated, low-capability automation helps the low-types to catch up; but very high-capability automation allows the high-types to shine.

In the case of negatively correlated skills, visualized in Panel B, the results are strikingly different. Here, the effect of single-task automation *is* monotonic in *first-difference* inequality for both tasks. When the task that high-skilled workers are better at is automated, the measure decreases. But when the task that low-skilled workers are better at is automated, high-skilled workers pull farther away.

In the special case of negatively correlated skills where, additionally, the skill the high-type is better at is being automated, the reduction in Δ is so extreme that the measure can turn negative. In this case, the effect on absolute inequality is non-monotonic: beyond a certain point, better automation of the skill that the low-types are worse at increases their lead over the ex-ante high-types (see Figure 2B). Just as in the positive correlation case, when automation at a single task becomes highly advanced, this increases inequality by magnifying differences in ability in the remaining non-automated task. However, unlike in the positive correlation case, this magnification may differentially benefit the group that performed worse initially. This divergence of first-difference and absolute inequality highlights the importance of tracking workers over time. Cross-sectional analyses of automation experiments may reveal no change in inequality while masking a major re-ordering of workers.

The specific points at which these non-monotonic switches in inequality occur will be context-specific. But overall, the simple Cobb–Douglas case clearly illustrates how improvements in automation may increase or decrease output inequality depending on the workers' skill correlations and the capability of the technology relative to each worker's skills. Furthermore, as we show in Appendix A.1, this non-monotonicity persists in a model where task-level labor supply is endogenous.

3.2. Single-task Automation — CES Case

TABLE 3
Absolute Inequality and Automation

technology	*tech. capability*	*sign* $\frac{\partial \lvert\Delta\rvert}{\partial A}$	*sign* $\frac{\partial^2 \lvert\Delta\rvert}{\partial A^2}$	*sign* $\frac{\partial^2 \lvert\Delta\rvert}{\partial A \cdot \partial B/C}$	*sign* $\frac{\partial^2 \lvert\Delta\rvert}{\partial A \cdot \partial \rho}$
Panel (A): Negative skill correlation					
A_1	$1 < A_1 < A^*$	$(-)$	$(+)$	0	$(+)$
A_1	$A^* < A_1$	$(+)$	$(-)$	0	$(-)$
A_2	$1 < A_2 < C$	$(+)$	$(-)$	$(+)$	$(-)$
A_2	$C < A_2$	$(+)$	$(-)$	0	$(-)$
Panel (B): Positive skill correlation					
A_1	$1 < A_1 < B$	$(-)$	$(+)$	0	$(\pm)^a$
A_1	$B < A_1$	$(+)$	$(-)$	$(+)$	$(-)$
A_2	$C < A_2 < B \cdot C$	$(-)$	$(+)$	0	$(-)$
A_2	$B \cdot C < A_2$	$(+)$	$(-)$	$(+)$	$(-)$

Notes: Reports the sign of the first and second derivatives of the absolute inequality effect ($\partial|\Delta|$) for the general CES production function with automation depending on whether workers' skills are negatively correlated (Panel A) or positively correlated (Panel B). $A^* = (B^\rho + 1 - C^\rho)^{1/\rho}$. $(\pm)^a$: sign is positive if $A < C$ and negative if $C < A < B$.

Table 3 generalizes these results to the CES production function. The table reports the signs of first and second derivatives of absolute inequality with respect to automation capability, providing a more complete picture of when and why automation's inequality effects may reverse.

In nearly all cases, the sign of this second derivative with respect to the technology's capability ($\partial^2|\Delta|/\partial A^2$) is opposite to the sign of the first derivative, which stems from the concavity of the production function .[8] This matters for practice: a firm piloting automation today may observe inequality changing, and even before the dramatic non-monotonicity occurs, the rate of change in inequality will decline.

The cross-derivatives with respect to the high-type's skill advantage ($\partial^2|\Delta|/\partial A \cdot \partial B/C$) reveal a few instances where initial skill gaps can amplify automation's inequality effects. In the positive skill correlation case, the skill gap (B/C) is relevant once the technology's capability exceeds both workers' skills — a larger pre-existing skill gap means automation will increase inequality more sharply. For organizations considering automation, this implies that the same technology will have different distributional consequences depending on how heterogeneous the workforce is to begin with. A narrow skill distribution may mute inequality changes; a wide distribution may amplify them.

The cross-derivatives with respect to task complementarity ($\partial^2|\Delta|/\partial A\partial\rho$) indicate whether automation's effect on inequality is amplified or not in jobs with more substitutable tasks ($\rho \to 1$). In the case of negative skill correlation, task substitutability behaves just as the second derivative; more substitutability mutes the inequality effect. With positively correlated skills, the cross derivative is negative for most technological capabilities, which amplifies inequality effects at low capability levels but depresses them at higher capability levels.

Practically, these comparative statics tell us that predicting automation's inequality effects requires knowing not just current worker skills and technology capabilities, but also the job's production structure and the trajectory of technological improvement. Perhaps most strikingly, across all cases the comparative statics of improving technology's capabilities for a single task at a time reveal that the effect of sufficiently super-human ability at a given task (i.e., $A_t \to \infty$) *must* be to increase inequality.

3.3. Full Automation

Returning to the Cobb-Douglas formulation for simplicity, Figure 3 extends the analysis to consider automation of both tasks, revealing how the task *composition* of automation technologies shapes inequality outcomes. Each panel plots absolute inequality compared to the pre-automation period as a function of both technologies' capabilities. The contrast between panels highlights how skill correlations determine not just whether inequality rises or falls, but which combinations of technologies prove most equalizing.

The figure delineates five different conditions. In the orange range, the automation technology is not advanced enough to be useful to any group. Therefore, in this range, inequality is identical to its starting level. In the gray range, inequality is reduced below the initial level. In the black range, inequality is higher than the initial level. In the white

[8] Exceptions are the Cobb-Douglas case of $\rho = 1$ and at kinks (i.e. the quality level where a technology first becomes worth using for a group) where the second derivative is undefined.

range, labeled 'full automation,' the technology has surpassed both worker types in both tasks. Finally, in the range bounded by the blue box, which overlaps the above ranges, at least one worker is fully surpassed by the AI in both skills.

These results should be interpreted while recalling that our two-task formulation is a low-dimensional illustration of the general model, in which a job consists of T tasks. In practice, a job may comprise dozens, hundreds, or even thousands of tasks, depending on how finely tasks are defined. Thus, the figure is best understood as illustrating the endpoints and logic of a potentially long path: technologies may surpass workers on some tasks long before they do so across every task in the job.

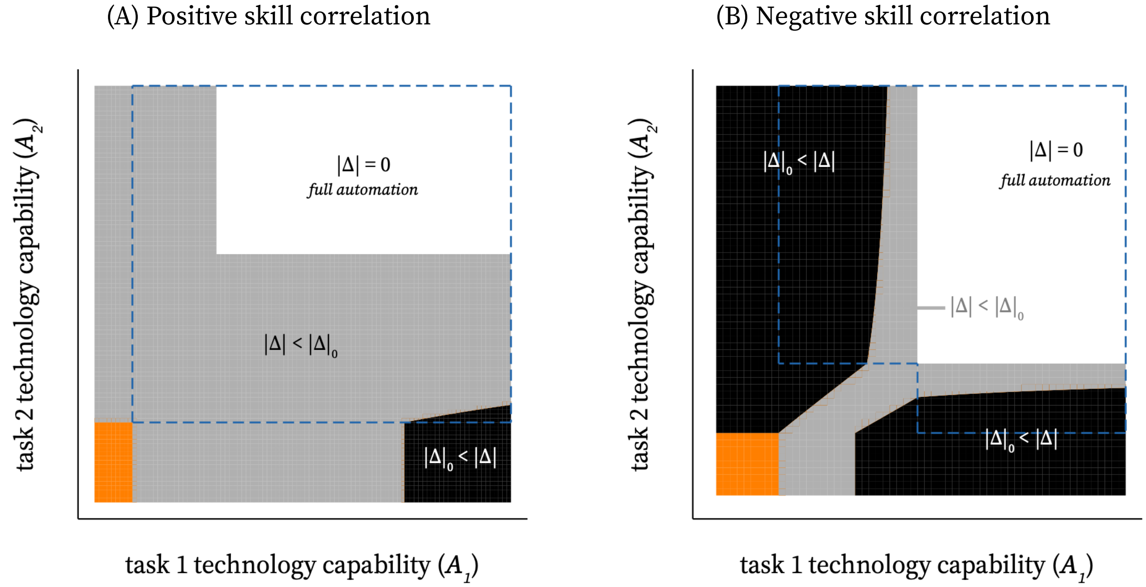
An important subtlety in interpreting these results is how to evaluate the blue-boxed and white regions, where one or both workers are fully surpassed by the automation technology in both skills. There are two possibilities. One possibility is that in these ranges, the surpassed worker retains their job, but their overall productivity is fully determined by the technology level. This may be because there is some regulatory or intrinsic need or desire for a human to be involved in performing the job. If so, we assume workers provide this residual, human-specific component homogeneously. The other possibility is that full automation will lead to the surpassed worker being fired, in which case, productivity inequality would actually be undefined. In this paper, we are focused on automation experiments and short-term extrapolations. In experiments that measure differential productivity effects, necessarily, workers do not get fired or re-sort across jobs (otherwise, you would not be able to measure their ex-post productivity). Therefore, we focus on the first interpretation of these cases, while noting that these cases, which are often good for the low-type worker, may be less positive for them in the long run after re-sorting or job reorganization.

With that caveat in mind, Figure 3 reveals the asymmetric paths to equality. Sufficiently large single-task automation eventually magnifies skill differences in the remaining human task, whereas sufficiently large full automation of both tasks can eliminate output inequality. This illustrates how the balance of technological capabilities is relevant to inequality. As more tasks are automated, the dimensionality along which workers differ shrinks, and inequality increasingly reflects skill differences in the remaining non-automated tasks.

With positive correlation (see Figure 3A), much of the capability space involves decreases in inequality compared to pre-automation periods. Effectively, technologies are always helping the low type catch up. Only if the task 2 technology has very low capability and the task 1 technology has very high capability can inequality exceed its pre-automation level.

With negative correlation (see Figure 3B), there is a much greater opportunity for increases in inequality. The intuition follows from the single-task results: when skills are negatively correlated, different workers are disadvantaged in different tasks, so equalizing outcomes requires helping both types of workers. A technology suite that is very capable in task 1 helps low-types, leaving low-types' comparative advantage in task 2 intact, which remains a

FIGURE 3
Absolute Inequality and Multi-task Automation — Additional Detail



Note: Plots the level of absolute inequality ($|\Delta|$) compared to the initial, pre-automation level of absolute inequality (defined here as: $|\Delta|_0$) depending on skill correlations and the capability of both automation technologies. The results here assume Cobb-Douglas production, and the orange areas indicate regions where neither technology has been adopted (i.e., $|\Delta| = |\Delta|_0$). The boxed area highlights the region of technological capabilities where both skills of at least one worker are fully surpassed by the automation technology.

(potential) source of inequality.

3.4. Alternative Modeling Assumptions and Robustness

The baseline model makes several simplifying assumptions: workers supply effort inelastically, correctly assess the technology's capability, face the same costs of using it, and treat technology and worker skill as perfect substitutes within a task. In Appendix A, we relax each of these assumptions. In Appendix A.1, we allow for endogenous effort supply decisions by the worker (i.e., allocating more effort towards tasks where they are more skilled). In Appendix A.2, we allow workers to have biased beliefs about the relative productivity of the technology compared to their skills; furthermore, those biases are allowed to be correlated with skill levels. Likewise, in Appendix A.3, we examine scenarios where workers must pay (potentially correlated) costs of using technologies. Lastly, motivated by Fügener, Walzner, and Gupta (2026), Appendix A.4 considers different conceptualizations of "augmentation" as opposed to the perfect, task-level substitution (or "automation") that technologies perform in our baseline model.

These extensions make the inequality paths more complicated, but leave the main result intact. Endogenous effort changes the levels and turning points, while biased expectations and heterogeneous use costs introduce additional adoption thresholds and can make inequality rise and fall several times as technology improves. Modeling augmentation introduces another worker- and task-specific source of heterogeneity. Across these cases, an inequality effect observed at one level of technology capability generally remains an *unreliable* guide to the effect of a more capable technology. Relaxing the baseline assumptions therefore reinforces our central result: automation's effect on inequality depends on workers' task-level skills and can be non-monotonic in the technology's capability.

3.5. Limits and the Long Run

Automation technologies are anticipated to improve considerably, so a natural case to consider is what happens in the limit as these technologies become arbitrarily powerful. As specified above, our model has clear implications for this limit: if only a portion of a job can be automated, then as the automation becomes more powerful inequality will increase among workers doing that same job. Alternatively, if all parts are automated, inequality will be eliminated.

However, in our model (and the vast majority of models of production functions) there are no bounds on the returns to inputs. That is, output is strictly increasing in task-level inputs for infinite levels of inputs.[9] However, in many practical settings of interest, it may be the case that effective skill at one component of a job may be capped. In practice, the nature of the job may be such that, beyond some maximum input level, output is no longer increasing in one or both inputs.[10] This would overturn our results that single-task automation is ultimately inequality-increasing (as shown in Figure 2 and Table 3). For instance, consider a Cobb-Douglas job with two tasks where technology is automating task 2, but there is a cap on the returns to that task:

$$Y_i = \left(S_{i1} \cdot \min\left\{\max\{S_{i2}, A_2\}, \overline{S}_2\right\}\right)^{1/2} ,$$

where $\overline{S}_2$ indicates the point beyond which additional task-2 skill (from the worker) or capability (from the technology) yields no additional output. In this case, the ultimate sign of the inequality effect as technological capabilities improve is undetermined since it depends on the relative magnitudes of $\overline{S}_2$ and A_2 (compared to workers' skill levels and correlations). For example, in our case of positive skill correlation, if $C < \overline{S}_2 < B \cdot C$, then the high type has already hit the cap on returns to task-2 input levels, so improvements in A_2 will only ever reduce inequality. In this example, when $C < A_2 < \overline{S}_2$, improvements in the technology raise

[9] To be clear, this is the norm in theoretical analyses as it prevents kinks in the production function.

[10] For example, if the worker is a plumber that repairs homes, there may be a practical limit as to how the plumber's physical skills translate into value for a homeowner — one plumber may be stronger than another, but if a bolt connecting two pipes can only be tightened so much, then additional strength yields no additional value.

the low type's output but leave the high type's output unchanged (inequality is reduced), but then once $\overline{S}_2 \leq A_2$, technology improvements have no effect on inequality.

Another case to consider: it may be that a job's productivity is unbounded in its arguments, but the technological progress necessary for automation to progress may take some time.[11] However, it is important to point out that we take no stance on *how long* said improvements may take. The long run in our model could be months, years, decades, or centuries – which makes a focus on the non-asymptotics of our model equally or more empirically relevant. Still, our framework emphasizes that it is not the absolute capability of the technology that is relevant, rather it is the technology's capability *relative* to the skill levels of different workers. It is when technologies cross certain thresholds based on workers' skill levels that the sign of inequality effects from automation changes (e.g., the B, $B \times C$ and B/C points flagged in Figure 2).

4. Empirical Case Study

While the model permits several different inequality effects, these effects depend on factors that can, in principle, be measured: the correlation of workers' task-level skills, and the capability of the technology relative to those skills. Here, we use the model to investigate an experiment granting radiologists access to diagnostic artificial intelligence.

4.1. Setting: Radiology & AI

We use data from the radiology experiment and Collab-CXR dataset described by Yu et al. (2024) and Moehring et al. (2025). In the experiment, professional radiologists evaluated historical chest X-rays under different, randomized AI-based treatments. All radiologists observed the X-ray. Depending on the experimental arm, they also received clinical-history information, predictions from a diagnostic AI system, both, or neither. Experimental treatments were randomized both across and within radiologists (across cases and pathologies).

The radiologists reported the probability that each X-ray exhibited a set of possible pathologies. The AI system reported a probability for each pathology using the X-ray alone. Diagnostic performance was evaluated against assessments from five top radiologists from the premier Mount Sinai Hospital with at least 10 years of experience and chest radiology as a subspecialty, each of whom independently evaluated all of the cases. Following Yu et al. (2024), we use the average assessment of these radiologists as the diagnostic standard.

The setting maps naturally into our model. A worker i is a radiologist, and a task t is assessing whether a particular pathology is present. The radiologist's unaided diagnostic

[11] To continue the above example, perhaps sufficiently advanced robotic plumbers truly could be "unboundedly" valuable to a homeowner.

performance on pathology t measures S_{it}, while the AI's diagnostic performance measures A_t. Overall radiology productivity is an aggregate of the radiologist's accuracy across pathologies.

Agarwal et al. (2025) show that the AI performs within the distribution of radiologist performance, rather than dominating every radiologist at every task (see their Figure 1). The technology is therefore in the intermediate-capability region highlighted by our model. A number of papers have explored this dataset and identified many different margins of heterogeneous effects (e.g., Yu et al. 2024; Agarwal et al. 2025). Yu et al. (2024) develop a measure of overall radiologist skill and find evidence that the higher-skilled radiologists benefit relatively less from the technology (i.e., see their Figure 3a and 3c), but do not find this relationship to be significant. However, they do not investigate skill correlations across radiologists, nor do they employ a CES production function to measure worker-level skill as we do. Our analysis includes 227 radiologists evaluating 324 cases. We restrict the sample to the 14 pathologies for which the data contain predictions from the radiologist, the AI, and the diagnostic-standard panel. We exclude the catch-all "any abnormality". The resulting sample contains 293,440 radiologist–case–pathology observations.

4.2. Measuring Task-Level and Overall Performance

We measure task-level diagnostic performance using Brier loss. For radiologist i, pathology t, and information treatment $x \in \{\text{no-AI}, \text{AI}\}$, define:

$$\ell_{it}^{x} = \frac{1}{|\mathcal{C}_{it}^{x}|} \cdot \sum_{c \in \mathcal{C}_{it}^{x}} \left(p_{itc}^{x} - g_{tc}\right)^{2} \,, \tag{10}$$

where $\mathcal{C}_{it}^{x}$ is the set of cases that radiologist i receives with pathology t and information treatment x, p_{itc}^{x} is the probability that case c has pathology t reported by radiologist i, and g_{tc} is the diagnostic-standard probability for that case c and pathology t. A lower Brier loss indicates better diagnostic performance.

We calculate one average Brier loss for each radiologist–pathology–treatment cell, retaining cells with at least five evaluations in both treatments. To measure overall performance with and without AI, we take the Cobb–Douglas aggregate of the pathology-level losses:

$$y_{i}^{x} = -\left(\prod_{t=1}^{T} \ell_{it}^{x}\right)^{1/T} \,, \tag{11}$$

where the negative rescaling yields the intuitive result that radiologists with larger values of y_i^x perform better as diagnosticians.[12] The Cobb–Douglas aggregation provides a direct

[12]We impose a floor of $\ell_{it}^{x} \geq 10^{-4}$ because some radiologist–pathology cells have losses at or near zero. There are 23 exactly zero no-AI cells and 352 no-AI cells below the floor, out of 3,178 retained cells.

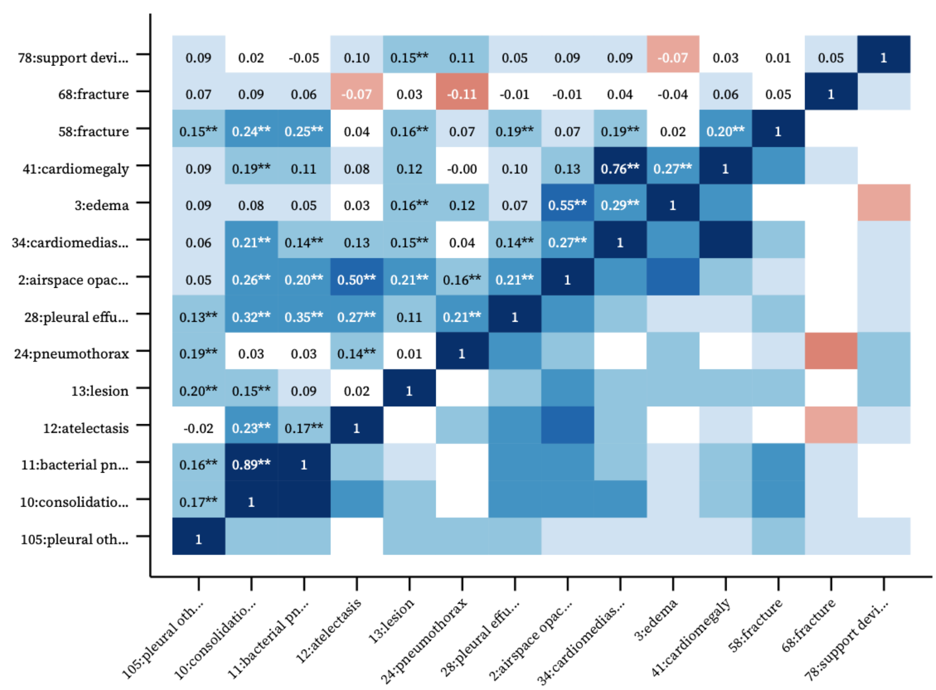
connection to the model while allowing performance to vary across more than two tasks.

4.3. Task-Level Skill Correlations

The model's prediction depends first on the correlation of workers' skills across tasks. We therefore calculate the pairwise correlation in unaided Brier losses across radiologists for each pair of pathologies. A positive correlation means that radiologists who perform relatively well on one pathology also tend to perform relatively well on the other (in the absence of AI).

Figure 4 reports the resulting correlation matrix. The task-level correlation structure is overwhelmingly positive. Across the 91 distinct pathology pairs, the average pairwise correlation is 0.14 , and 90.1 % of the pairwise correlations are positive. Thirty-nine correlations are significantly positive at the 5% level, while zero negative correlations are significant at the 5% level. We conclude that radiologists who are relatively accurate at diagnosing one pathology are relatively accurate at diagnosing other pathologies.

FIGURE 4
Pairwise Correlations in Condition-level Performance



Note: Shows the pair-wise correlations of Brier loss scores within radiologists for each possible pair of the 14 pathologies; ** indicates $p < 0.05$. The numbers leading each pathology correspond to unique identifiers in the data.

Through the lens of our model, this positive correlation structure generates a clear prediction. Because the AI diagnostic technology lies *within* the distribution of radiologist performance, access to AI should disproportionately benefit radiologists with worse initial performance. This is the scenario visualized in Figure 1A. We should therefore observe a compression of the overall performance distribution.

4.4. AI Access and Performance Inequality

Now, we can examine the relationship between a radiologist's baseline loss and their improvement under AI assistance. Define the AI-induced improvement in diagnostic performance as:

$$g_i = y_i^{\text{AI}} - y_i^{\text{no-AI}}. \tag{12}$$

To identify the inequality effect of this automation experiment, we want to estimate a regression of the form:

$$g_i = \alpha + \beta \cdot y_i^{\text{no-AI}} + \varepsilon_i, \tag{13}$$

but we face the problem of *regression to the mean*. To see this, note that baseline performance ($y_i^{\text{no-AI}}$) appears directly in the construction of the outcome (g_i). Thus, a radiologist who happens to perform poorly in the no-AI observations will therefore appear both worse at baseline and more improved in a different sample, even if their underlying ability is unchanged, purely due to chance.

We address this problem using a split-sample procedure. For each radiologist, we randomly divide the no-AI evaluations into two independent halves. The first half is used to construct the measure of baseline performance, while the second half is used to construct the no-AI component of the gain measure. We repeat this procedure across 50 random splits. We then standardize both variables (g_i and $y_i^{\text{no-AI}}$) so that the estimated relationship can be interpreted in units of standard deviations.[13]

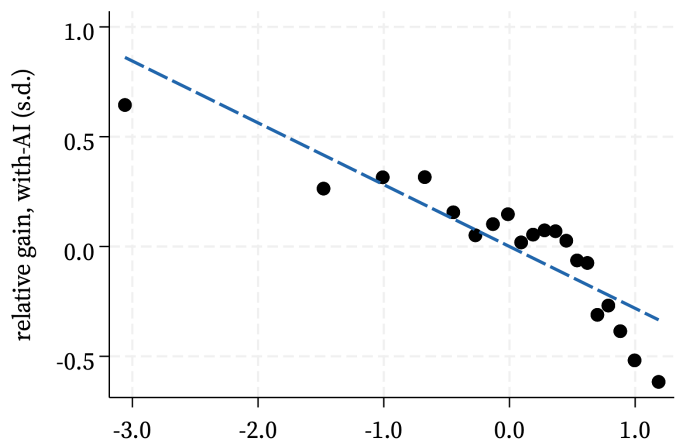
Figure 5 reports the relationship between baseline performance and gains based on this split-sample procedure designed to avoid regression to the mean. The slope is −0.282, and the radiologist-clustered standard error is 0.080 (radiologist-clustered bootstrap 95% C.I. = [−0.441,−0.136], based on 1,000 draws). As predicted by the model, gains from AI decline with baseline performance.[14]

The distributional results lead to the same conclusion. The cross-radiologist standard deviation of composite Brier loss falls from 0.0062 without AI to 0.0044 with AI. This represents a 29 % compression of the performance distribution.

[13]This split sample procedure mechanically introduces some classical measurement error into our baseline performance measure, which biases us towards the null and (by itself) suggests that our estimate will be conservative in absolute value.

[14]Using a different definition of overall radiologist performance, Yu et al. (2024) also estimate that AI access reduces performance inequality, though this is not highlighted as a finding of the paper (Figure 3, panels a and c) because by their measure the difference is not significant. They do not investigate correlation in unassisted radiologist performance across tasks.

FIGURE 5
Change in Radiologists' Performance per Baseline Performance



Note: Shows an equally binned scatterplot of radiologists' baseline performance (x-axis) and their change in performance with AI (y-axis); based on data from Moehring et al. (2025). The slope is –0.282, and the radiologist-clustered standard error is 0.080 (radiologist-clustered bootstrap 95% C.I. = [–0.441,–0.136], based on 1,000 draws).

Overall, the evidence follows the pattern predicted by the model. Radiologists' unaided skills are positively correlated across diagnostic tasks, the AI's capability lies within the distribution of those skills, and initially lower-performing radiologists experience larger gains from AI access. As a result, access to AI compresses the distribution of diagnostic performance. This example illustrates that the model can be used to predict which workers benefit most from new or improved automation technologies.

5. Towards Normative Implications

A wide range of managers' and policymakers' choices depend on understanding the inequality effects of automation: salary negotiations, tax policies, union and overtime rules, office culture, etc. In this section, we provide some guidance on how to make use of the predictions of our model.

In some settings, the manager has clear information on key objects: the distribution of workers' task-level productivities; automation's capability levels; costs and benefits of investments. In these settings, our model can be used to generate specific analytic conditions under which one investment strategy dominates another. We first focus on this case in Section 5.1.

In other settings, both automation capability levels and workers' detailed skill distributions are unknown. Even in these cases, our model can help make predictions about how technology deployment strategies are likely to impact inequality. In Section 5.2, we discuss such scenarios of limited information and, for example, when delaying technology deployment might be desirable.

Given the importance of understanding skill correlations, we close this section by reviewing the empirical studies that have estimated skill correlations (Section 5.3), presenting new evidence on the distribution of skill correlations across sectors (Section 5.4), and arguing theoretically that settings with positive assortative matching will tend to create negative skill correlations (Section 5.5).

5.1. Investing in Training versus Engineering

In Appendix C, we use the model to study a manager who can invest in either worker "training" to improve their skills (S), or "engineering" to improve the automation technology's capabilities (A) and who values both total output and equality. We focus on scenarios where neither investment produces superhuman performance, and again we find that the correlation of workers' skills is a key object, here, as an important determinant of investment.

With positively correlated skills, greater capability both raises total output and reduces inequality, so every manager chooses the largest improvement attainable by either training or engineering. With negatively correlated skills, investment depends on the strength of the manager's preferences over output versus equality. A manager placing relatively more weight on equality invests only until the workers have equal output, while a manager placing relatively more weight on total output chooses the largest attainable improvement. Conditional on this desired capability, the choice between training and engineering depends on their relative effectiveness and the technology's initial capability. Thus, the model does not imply a general preference for training or engineering. Instead, workers' skill correlation determines the manager's desired capability, while the relative costs of the two investments determine how the manager reaches it.

5.2. Investment with Limited Visibility

Now consider a decision-maker who cannot observe the full parameter vector that governs the optimal investment choice (e.g., as described in the preceding subsection). All else equal, the policy-maker wishes to delay frontier technology adoption when it increases inequality and accelerate adoption when it decreases inequality. In the baseline model, the effect of technology adoption corresponds to the sign of $\partial|\Delta|/\partial A$ reported in Table 3. Whether it does so depends on two key parameters, each of which may often be usefully estimated, even if they are not perfectly observable: the sign of the workers' skill correlation, and where the technology's capability sits relative to the distribution of worker skill at the automated task.

It turns out that if the policy-maker can guess at even one of these two parameters, they can gather useful information about the likely impact of the technology.

First, consider the role of the level of automation quality A. The two relevant cases are when the technology is better than some, but not all, workers, and when technology is superhuman at a task (i.e., it exceeds every worker's skill). Understanding whether the technology has achieved superhuman levels at a task may be relatively straightforward in comparison to measuring its exact position in the human skill distribution.

If it is clear that the technology has achieved superhuman quality at a task, even in the absence of any other information, the guidance to the policymaker is clear. In every row of Table 3, once capability exceeds the most skilled worker at the automated task, $\partial|\Delta|/\partial A > 0$: the technology has equalized the automated task completely, so each further increment of capability only magnifies the differences that remain in the task workers still perform themselves. Inequality weakly increases.[15] For a decision-maker who values equality, the prescription for advanced single-task automation is therefore uniform — slow its development — and it requires no other knowledge of parameters.

If the technology is not yet superhuman at a given task, the policy-maker will need to estimate or guess at the correlation structure of ability. When skills are positively correlated, the workers least skilled at the automated task are also the lowest-output workers. They adopt first and catch up, so capability growth compresses inequality. When skills are negatively correlated, the correlation's sign alone is not enough. At low capability, the identity of the group that is weaker at the automated task determines whether inequality increases when the high-type group benefits from the basic technology or decreases when the low-type group adopts first. Therefore, when an inequality- and risk-averse policy-maker is considering introducing a non-superhuman automation technology, they should be more pro-technology when they suspect skills are positively correlated, and more cautious when they are negatively correlated.

These conclusions lead naturally to the following question: In what settings would we expect relevant skills to be positively correlated, and in which would we expect them to be negatively correlated? In the following sections, we review the empirical and theoretical evidence.

Positive correlations are the norm when the deployment population resembles the general population or an entire occupation — all radiologists, say — and when the job's tasks are uniformly cognitive or social. Others' estimates of these correlations are reviewed in Section 5.3 and our own evidence in Section 5.4. In these settings, basic automation is relatively likely to reduce inequality.

We hypothesize that negative correlations are more likely to occur in two other circum-

[15] If the technology is already superhuman at the other task, then increasing its quality on this dimension has no impact on inequality.

stances. The first is when the job pairs mental with physical tasks, for which population-level correlation estimates turn negative. The second is when the deployment population has been selected for similar overall performance — a single firm, team, or business unit. Section 5.5 shows that positive assortative matching can generate negative within-group skill correlations even when the population-level correlation is positive. In settings that are highly selected in this way, or which require both mental and physical abilities, the model is ambiguous in its conclusions in the absence of additional information. In these settings, policy-makers who wish to minimize inequality should be more cautious about accelerating the adoption of automation.

5.3. Existing Estimates of Skill Correlations

At least since Willis and Rosen (1979), it has been appreciated that a single index of workers' skill will abstract away from important variation in the labor supply. However, empirical estimates of workers' skills on specific types of tasks remain a challenging pursuit. Approaches to date typically involve accessing confidential data, intense measurement, or the estimation of structural models of labor supply.

Most studies examining multi-dimensional skills have focused on aggregate categories: cognitive skills, social skills, and manual skills. While these categories may be much broader than the connotation of the "tasks" in our model, the logic is the same. Furthermore, these studies generally focus on large populations of workers (e.g., national statistics).

Broadly speaking, among these three categories of skills, cognitive and social skills show the most positive correlation. Estimates of the correlation in these skills span from approximately 0.1 up to 0.7 (Mayer, Roberts, and Barsade 2008; Baker et al. 2014; Deming 2017; Guvenen et al. 2020; Lise and Postel-Vinay 2020; Girsberger, Koomen, and Krapf 2022; Bárány and Holzheu 2025). Cognitive and manual skills exhibit negative to weakly positive correlations, roughly in the range of –0.4 to 0.1 (Lindenlaub 2017; Lise and Postel-Vinay 2020; Bárány and Holzheu 2025). Social and manual skills consistently show the most negative correlations of these categories, on the scale of –0.4 to –0.6 (Lise and Postel-Vinay 2020; Girsberger, Koomen, and Krapf 2022; Bárány and Holzheu 2025). Within the category of cognitive skills, data persistently reveal a strong positive correlation, for example, of approximately 0.7 across mathematical skills and literacy (or verbal) skills (Hampf, Wiederhold, and Woessmann 2017; Guvenen et al. 2020; Woessmann 2024).

5.4. SAT Score Correlations across the Economy

As noted above, prior work has generally found mathematical and verbal skills to be positively correlated per standardized tests. A common source of data used is the National Longitudinal Survey of Youth (NLSY), which reports participants' SAT scores for verbal and math components separately (BLS 2024). In Appendix B, we also make use of this data, but report the math-verbal skill correlation separately for alternative industries and occupations.

Appendix Figure B1 highlights that the strength of the math-verbal score correlation varies significantly across the economy. In some industries, math-verbal score correlations approach 0.8, while in other industries the correlation is as low as 0.2. Interpreted through the lens of the model, this heterogeneity suggests a common automation technology deployed in these different sectors could yield inequality effects of significantly different magnitudes.

5.5. An O-Ring Model with Two Tasks

The evidence in the prior section describes skill correlations at the population and occupational levels. However, many, if not most, automation experiments narrowly focus on a single occupation at a single firm, or possibly even more narrowly defined business units and jobs. In this section, we make the point that the skill correlations within these sample populations may be very different from those in the overall population.

In order to forecast how worker sorting could lead aggregate skill correlations to diverge from those within an experimental sample, we extend Kremer's (1993) O-ring model. Specifically, we consider the O-ring model when production involves two tasks, unlike the original, single-task model. The model setup is the same as described in Section 2, augmented with additional assumptions to incorporate worker sorting and firm-level production.

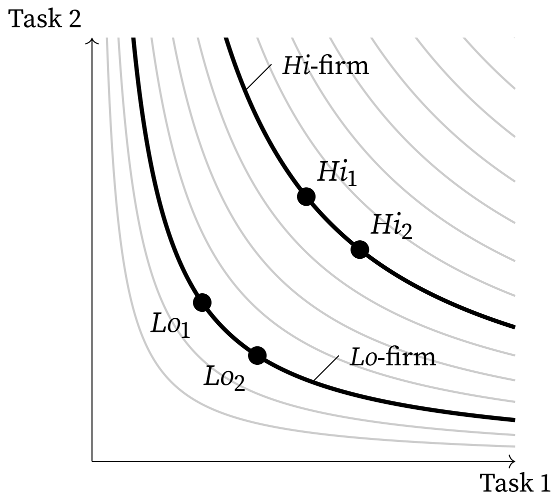
The main result of this model is that skills must be negatively correlated within a firm or team no matter the population correlation in these skills. The intuition is straightforward. The O-Ring model assumes that output is complementary in the different inputs to a team, firm, or country. This leads to perfect assortative matching in ability across teams or firms: the best teams can't afford any weak links, the least productive firms can't afford higher quality workers. Firms in the middle hire mediocre workers. In our extension, workers' quality is a function of their ability in multiple skills. Because the only way for a worker to be mediocre overall is to be good at one task but bad at another, or to be mediocre at both, this yields our main finding.

We visualize this intuition in Figure 6. Each productivity isoquant represents both a worker productivity level and a firm type (because all workers at a given firm must have the same productivity level). We plot a pair of high-type workers working at a high-type firm, and a pair of low-type workers working at a low-type firm. In the given example, if we assume each node has the same population of workers, the negative correlation in skills within firms occurs despite the fact that there is a positive correlation in skills overall.

Generally, the fact that isoquants are weakly downward-sloping requires weakly negative skill correlations in the presence of positive assortative matching.[16]

[16]We say weakly because, in the case of Leontief production, isoquants have undefined or zero slopes.

FIGURE 6
Skill Correlations and Assortative Matching



Note: Plots isoquants of production possibilities, highlighting the isoquant for workers with higher (*Hi*) or lower output (*Lo*). Workers are chosen to illustrate a scenario where skills are positively correlated in the population (including all four workers). Since there are only two types of workers (and each type exists on the same isoquant), positive assortative matching in the context of two-worker firms will generate a negative within-firm skill correlation.

In the remainder of this section and Appendix D, we formalize this logic.

Consider an economy with a continuum of workers and firms. Each firm employs n workers, and each worker performs two tasks.[17] Worker i has task-specific skill levels S_{i1}, S_{i2}. Worker i's production function is again the simple Cobb-Douglas form: $Y_i = \sqrt{S_{i1} \cdot S_{i2}}$.[18] Following Kremer (1993), firm-level output (Z) is the simple product of worker-level output:

$$Z = \prod_{i=1}^{n} Y_i \,, \tag{14}$$

where we have abstracted away from capital used in production for simplicity, and because we are not interested in metrics related to, for example, labor shares.

Skills are jointly log-normally distributed in the aggregate population. Since our focus is on the correlation in skills, we'll focus on the simple case where the means (μ) and standard

[17] For simplicity, assume n is large such that the realized within-firm skill distribution converges to the conditional population distribution.

[18] We focus on Cobb-Douglas production in this analysis because the general CES form leads to non-linear constraints that do not have straightforward analytical solutions.

deviations (σ) of both skills are equal:

$$\begin{pmatrix} \ln S_{i1} \\ \ln S_{i2} \end{pmatrix} \sim \mathcal{N}\begin{pmatrix} \mu & \sigma^2 & corr_{\text{pop}} \cdot \sigma^2 \\ \mu & corr_{\text{pop}} \cdot \sigma^2 & \sigma^2 \end{pmatrix}, \tag{15}$$

where $corr_{\text{pop}} \in (-1, 1)$ is the population correlation of logged skills.

The multiplicative firm-level structure ensures supermodularity in worker-level output: $\partial^2 Z/(\partial Y_i \cdot \partial Y_j) > 0$ for $i \neq j$. As in Kremer (1993), equilibrium features positive assortative matching on the scalar index Y_i.

Now, we can derive the relationship between the population-level skill correlation ($corr_{\text{pop}}$) and the within-firm skill correlation, which we'll denote with $corr_{\text{firm}}$, after workers sort into firms. In Appendix D we solve for the equilibrium, within-firm correlation in skills and obtain:

$$corr_{\text{firm}} = -\exp\left(-\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right) < 0\,. \tag{16}$$

In this example, the within-firm skill correlation is always negative regardless of the population-level correlation. The sharpness of this result stems from the assumptions of (multiplicative) complementarity among all workers within the firm and perfect positive assortative matching. As we highlighted above with Figure 6, this leads to firms where all workers produce on the same isoquant, which, by definition, yields a negative skill correlation. This is a stark example of "Berkson's paradox" (Berkson 1946).

Certainly, the presence of substitutable workers within firms or imperfect positive assortative matching across firms would attenuate this effect.[19] Thus, in practice, within-firm skill correlations need not always be negative.

Still, this model highlights the importance of understanding how workers sort into different levels of aggregation when predicting skill correlations. Under common scenarios, the model indicates that as narrower groups of workers are selected for inclusion in an automation experiment, their skill correlations will become more negative. This, in turn, implies that the choice of scope for an automation experiment can have an important effect on the skill correlation of the workers in the experiment and, in turn, the sign of the inequality effect.

6. Discussion

Our model of automation's short-term effect on productivity inequality is focused on the following insight: the inequality effects of automation depend on the interaction between

[19] For example, with random matching, the within-firm and population-level skill correlations will become equal.

skill correlations and the technology's capability relative to worker skills, rather than on the workforce's absolute skill level.

We highlight three results that are novel relative to the task-based automation literature. First, the model explains how the same technology can have opposite-signed inequality effects in two workforces that are indistinguishable by job production functions and productivity distributions. Moreover, because of the non-monotonic effects of automation improvements, an experiment showing an inequality reduction from today's technology may not imply an inequality reduction from tomorrow's. Frameworks that rank workers on a single index, or that impose a monotone comparative-advantage ordering across tasks (e.g., Autor, Levy, and Murnane 2003; Acemoglu and Autor 2011), treat these two workforces identically and therefore do not allow for such an outcome. Similarly, aggregate task-based models generate non-monotonicity in the effect of AI on inequality through the creation of new tasks or general-equilibrium reallocation (Acemoglu and Restrepo 2019); ours arises within a fixed job production function and group of employees.

Thus, the mix of seemingly inconsistent results documented in automation experiments thus far (Table 1) could plausibly be attributed to differences in fundamentals without the introduction of any behavioral biases.[20] A pair of seemingly similar automation experiments will vary in their economic impact because of (*i*) the task-level skill correlation of workers, (*ii*) which task is (or tasks are) being automated, and (*iii*) the capability of the automation technology relative to the workers' skill levels. While our model is flexible enough to match the heterogeneity in the observed data, it does so by attributing the different outcomes to parameters that are fundamentally measurable. We put one such prediction — that intermediate automation will reduce inequality in the presence of positive skill correlations — to the test in the case of AI-assisted radiology in Section 4, finding the anticipated effect.

Second, the model predicts reversals in the effect of automation on inequality at observable thresholds. As a single-task technology improves, dispersion often first compresses and then widens, with the turning point located at the most skilled worker's level in the automated task (or, under negative correlation, at the overtaking point A^*). In our extensions, which allow for bias or errors in rational adoption, this threshold is fuzzed or shifted, but generally preserved. Testing to see whether these reversals exist where they are predicted is an excellent direction for future research.

Our third main novelty is our managerial and policy implications. When parameters are observable, our model can be combined with preferences over inequality and aggregate productivity to help firms plan technology investments. More generally, even if the details of skill distributions or automation quality are unknown, our model suggests settings where automation technology is more or less likely to increase productivity inequality. Restricting

[20] Of course, that is not to say such biases are irrelevant.

attention to single-task automation, superhuman automation tends to increase inequality, while the effect of more basic automation depends on the skill correlation. Multidimensional skills appear in prior work on sorting and wages (Autor and Handel 2013; Deming 2017), but no automation framework has made the within-sample correlation of task skills a determinant of the technology's distributional effect. We provide suggestive evidence that population and occupation-level settings for jobs that combine mental tasks will tend to have positive correlations, making inequality reductions from automation more likely. On the other hand, we connect to a literature suggesting negative skill correlations are present in jobs combining mental and physical skills and in settings with assortative matching, making the effect of automation on inequality in these settings more ambiguous.

For researchers planning future automation experiments, our framework suggests several priorities. A checklist of considerations motivated by this model includes:

Document sorting into the experimental population. Our O-ring analysis highlights the effect that assortative matching can have on skill correlations. Researchers should report how workers sorted into the population being sampled for the experiment.

Measurement of skill correlations. Skill correlations across workers play a key role in determining inequality effects. While measuring task-specific skills has historically been a challenging exercise, making progress on this front will give empiricists an improved view of the workforce and facilitate better interpretations of the inequality effects obtained.

Multiple technology capabilities. The model highlights how technological progress leads to non-monotonicities. Thus, experimenters could begin to forecast trends by testing for effects using multiple technologies with different capability levels.

Task "automatability" measurement. Since non-automated tasks drive inequality in our model, especially in the long run, documenting which tasks are more or less likely to be automated can improve forecasts beyond the duration of experiments.

Our simple model of automation experiments highlights a unique feature of the workforce: the correlation of workers' skills across tasks. While many discussions of emerging automation technologies focus on how workers with different *levels* of skill are affected, we emphasize here that it is the *correlation* in skills across tasks that is key for understanding how automation affects inequality.

The empirical evidence suggests a negative correlation between physical and mental skills in the population; however, it remains unclear to what extent this is exogenous or an endogenous result of individuals' choices about which skills to acquire and when to acquire them. Digital technologies in the arena of "artificial intelligence" have clearly made a great leap in the automation of mental skills. Thus, for jobs requiring both physical and mental skills, our model suggests initial advances in automation will decrease inequality so long as the task-composition of those jobs remains fixed. Our investigation of SAT math and verbal scores reveals substantial heterogeneity in cognitive skill correlations across industries and occupations; however, all correlations are positive. Therefore, for knowledge workers in the population as a whole, and when technology is not (yet) superhuman, our model predicts automation will disproportionately help lower-skill workers, decreasing inequality. On the other hand, those with jobs like these often find themselves in teams which are highly selected for a particular skill level, for reasons discussed in our O-Ring model extension. In these settings it is possible for the effect of automation on inequality to flip, making extrapolations from the general population to selected teams (or the other direction) problematic. Notably, whatever the skill correlation, there is a clear prediction of what happens as technology becomes increasingly superhuman in a subset of tasks — inequality will increase as the remaining human differences in ability become increasingly magnified.

While formalizing these long-run general equilibrium dynamics is beyond the scope of this paper, our framework provides qualitative intuition for how shifts in skill correlations would alter the trajectory of inequality. If technology adoption prompts firms to redesign jobs around specialization (e.g., by having some workers focus entirely on technical oversight while others focus on interpersonal tasks), a historically positive skill correlation could weaken or turn negative. Through the lens of our model, this shift would dilute the initial "catch-up" effect of intermediate technological capabilities, because lower-output workers would no longer systematically be the weakest at the automated task. If the correlation turns fully negative, the distributional impact of further automation would depend entirely on which specific comparative advantage is automated next. Conversely, if technological capabilities disproportionately raise the returns to general cognitive traits, workers might invest in ways that strengthen positive skill correlations. In this scenario, our framework implies that the late-stage magnification of inequality (once the technology becomes superhuman) would become much more severe, as differences in the remaining un-automated tasks would grow even wider.

The base model's simplicity is both a strength and a limitation. By abstracting from general-equilibrium adjustments, we isolate the direct productivity channel that experiments measure. This modeling choice makes our predictions directly comparable to experimental results, but it leaves the model silent on longer-run inequality effects in general equilibrium. Incorporating skill correlations into macroeconomic models of automation is a natural next step to explore these broader consequences.

An intentional, but major, limitation of our model is the short-term partial equilibrium assumption. Throughout, we have treated workers' skills, and therefore their skill correlations, as fixed. This is a fitting abstraction for the window of an experiment but not for predicting long-term consequences. Natural extensions to our model would allow for workers to invest in changing their skills, for workers to re-sort themselves into new jobs and firms, and for firms to reorganize and invent new jobs. This is an exciting direction for future research. For now, we will only speculate that these additional, time-delayed, dimensions of adjustment may serve to make more common the phenomenon of non-monotonic effects of automation technology on inequality.

More broadly, our framework suggests that the diversity of automation technologies will play a key role in the evolution of inequality. Technologies that excel at a narrow set of tasks may amplify inequality by primarily benefiting workers who specialize in those tasks or in complementary skills. In contrast, technologies with moderate capabilities across many tasks may compress inequality by helping workers overcome their weakest skills. This observation motivates further theoretical and empirical work on how the portfolio of available automation tools shapes the returns to various skill combinations.

Finally, our focus on task-level production and skill correlations connects to broader questions about the organization of work. When skills are negatively correlated across workers, firms have an incentive to exploit comparative advantage by assigning each worker to the tasks where their relative skill is highest. When skills are positively correlated, such specialization opportunities are limited, since the same workers tend to outperform others in all tasks. The interaction between skill correlations, automation, and the endogenous organization of production is a promising direction for future research.

References

Acemoglu, Daron, and David Autor. 2011. "Skills, tasks and technologies: Implications for employment and earnings." In *Handbook of Labor Economics*, vol. 4, 1043–1171: Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The race between man and machine: Implications of technology for growth, factor shares, and employment." *American economic review* 108 (6): 1488–1542.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Automation and new tasks: How technology displaces and reinstates labor." *Journal of Economic Perspectives* 33 (2): 3–30.

Agarwal, Nikhil, Alex Moehring, Pranav Rajpurkar, and Tobias Salz. 2025. "Combining human expertise with artificial intelligence: Experimental evidence from radiology." *Mimeo*.

Agarwal, Nikhil, Alex Moehring, and Alexander Wolitzky. 2025. "Designing Human-AI Collaboration: A Sufficient-Statistic Approach." *Mimeo*.

Athey, Susan C, Kevin A Bryan, and Joshua S Gans. 2020. "The allocation of decision authority to human and artificial intelligence." In *AEA Papers and Proceedings*, vol. 110: 80–84, American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Autor, David H, and Michael J Handel. 2013. "Putting tasks to the test: Human capital, job tasks, and wages." *Journal of labor Economics* 31 (S1): S59–S96.

Autor, David H, Frank Levy, and Richard J Murnane. 2003. "The skill content of recent technological change: An empirical exploration." *The Quarterly Journal of Economics* 118 (4): 1279–1333.

Autor, David, and Neil Thompson. 2025. "Expertise." *Journal of the European Economic Association*: jvaf023.

Baker, Crystal A, Eric Peterson, Steven Pulos, and Rena A Kirkland. 2014. "Eyes and IQ: A meta-analysis of the relationship between intelligence and "Reading the Mind in the Eyes"." *Intelligence* 44: 78–92.

Bárány, Zsofiá L, and Kerstin Holzheu. 2025. "The Two Faces of Worker Specialization." *Mimeo*.

Becker, Gary S, and Kevin M Murphy. 1992. "The division of labor, coordination costs, and knowledge." *The Quarterly Journal of Economics* 107 (4): 1137–1160.

Berkson, Joseph. 1946. "Limitations of the application of fourfold table analysis to hospital data." *Biometrics Bulletin* 2 (3): 47–53.

BLS, (Bureau of Labor Statistics; U.S. Department of Labor). 2024. "National Longitudinal Survey of Youth 1997 Cohort.", Center for Human Resource Research (CHRR), The Ohio State University.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*: qjae044.

Chen, Zenan, and Jason Chan. 2024. "Large language model in creative work: The role of collaboration modality and user expertise." *Management Science* 70 (12): 9101–9117.

Choi, Jonathan H, Amy B Monahan, and Daniel Schwarcz. 2024. "Lawyering in the age of artificial intelligence." *Minnesota Law Review* 109: 147.

Cui, Zheyuan Kevin, Mert Demirer, Sonia Jaffe, Leon Musolff, Sida Peng, and Tobias Salz. 2025. "The effects of generative AI on high skilled work: Evidence from three field experiments with software developers." *Mimeo*.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani. 2025. "Knowledge Worker Productivity and Quality in Generative AI's Jagged Technological Frontier: Field Experimental

Evidence." *Mimeo*.

Deming, David J. 2017. "The growing importance of social skills in the labor market." *The Quarterly Journal of Economics* 132 (4): 1593–1640.

Dessein, Wouter, and Tano Santos. 2006. "Adaptive organizations." *Journal of Political Economy* 114 (5): 956–995.

Fügener, Andreas, Dominik D Walzner, and Alok Gupta. 2026. "Roles of artificial intelligence in collaboration with humans: Automation, augmentation, and the future of work." *Management Science* 72 (1): 538–557.

Garicano, Luis, and Luis Rayo. 2025. "Training in the Age of AI: A Theory of Apprenticeship Viability." *Mimeo*.

Girsberger, Esther Mirjam, Miriam Koomen, and Matthias Krapf. 2022. "Interpersonal, cognitive, and manual skills: How do they shape employment and wages?" *Labour Economics* 78: 102235.

Guvenen, Fatih, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer. 2020. "Multidimensional skill mismatch." *American Economic Journal: Macroeconomics* 12 (1): 210–244.

Hampf, Franziska, Simon Wiederhold, and Ludger Woessmann. 2017. "Skills, earnings, and employment: Exploring causality in the estimation of returns to skills." *Large-scale Assessments in Education* 5 (1): 12.

Kanazawa, Kyogo, Daiji Kawaguchi, Hitoshi Shigeoka, and Yasutora Watanabe. 2025. "AI, skill, and productivity: The case of taxi drivers." *Management Science*.

Katz, Lawrence F, and David H Autor. 1999. "Changes in the wage structure and earnings inequality." In *Handbook of Labor Economics,* vol. 3, chap. 26, 1463–1555: Elsevier.

Kim, Alex, David S Kim, Maximilian Muhn, Valeri V Nikolaev, and Eric C So. 2025. "AI, Investment Decisions, and Inequality." *Mimeo*.

Kleinberg, Jon, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. 2018. "Human decisions and machine predictions." *The Quarterly Journal of Economics* 133 (1): 237–293.

Kreitmeir, David, and Paul A. Raschky. 2024. "The Heterogeneous Productivity Effects of Generative AI." *Mimeo*.

Kremer, Michael. 1993. "The O-ring theory of economic development." *The Quarterly Journal of Economics* 108 (3): 551–575.

Lindenlaub, Ilse. 2017. "Sorting multidimensional types: Theory and application." *The Review of Economic Studies* 84 (2): 718–789.

Lise, Jeremy, and Fabien Postel-Vinay. 2020. "Multidimensional skills, sorting, and human capital accumulation." *American Economic Review* 110 (8): 2328–2376.

Maršál, Ales, and Patryk Perkowski. 2025. "A Task-Based Approach to Generative AI: Evidence from a Field Experiment in Central Banking." *Mimeo*.

Mayer, John D, Richard D Roberts, and Sigal G Barsade. 2008. "Human abilities: Emotional intelligence." *Annu. Rev. Psychol.* 59 (1): 507–536.

Moehring, Alex, Manasi Kutwal, Ray Huang, Oishi Banerjee, Adam Jacobi, Corey Eber, Dexter Mendoza, Mike Chung, Etan Dayan, Yogesh Gupta et al. 2025. "A Dataset for Understanding Radiologist-Artificial Intelligence Collaboration." *Scientific data* 12 (1): 739.

Mullainathan, Sendhil, and Ziad Obermeyer. 2022. "Diagnosing physician error: A machine learning approach to low-value health care." *The Quarterly Journal of Economics* 137 (2): 679–727.

Noy, Shakked, and Whitney Zhang. 2023. "Experimental evidence on the productivity effects of generative artificial intelligence." *Science* 381 (6654): 187–192.

Otis, Nicholas, Rowan Clarke, Solene Delecourt, David Holtz, and Rembrand Koning. 2025. "The uneven impact of generative AI on entrepreneurial performance." *Mimeo*.

Roldán-Monés, A. 2025. "When GenAI increases inequality: Evidence from a university debating competition." *Mimeo*.

Steyvers, Mark, Heliodoro Tejeda, Aakriti Kumar, Catarina Belem, Sheer Karny, Xinyue Hu, Lukas W Mayer, and Padhraic Smyth. 2025. "What large language models know and what people think they know." *Nature Machine Intelligence* 7 (2): 221–231.

The Economist. 2025. "How AI will divide the best from the rest."

Trammell, Philip. 2025. "Workflows and Automation." *Mimeo*.

Willis, Robert J, and Sherwin Rosen. 1979. "Education and self-selection." *Journal of Political Economy* 87 (5, Part 2): S7–S36.

Woessmann, Ludger. 2024. "Skills and earnings: A multidimensional perspective on human capital." *Annual Review of Economics* 17.

Xu, Fasheng, Jing Hou, Wei Chen, and Karen Xie. 2025. "Generative AI and Organizational Structure in the Knowledge Economy." *Mimeo*.

Yu, Feiyang, Alex Moehring, Oishi Banerjee, Tobias Salz, Nikhil Agarwal, and Pranav Rajpurkar. 2024. "Heterogeneity and predictors of the effects of AI assistance on radiologists." *Nature Medicine* 30 (3): 837–849.

Zeira, Joseph. 1998. "Workers, machines, and economic growth." *The Quarterly Journal of Economics* 113 (4): 1091–1117.

Supplementary Appendix

Appendix A. Additional Results and Model Modifications

A.1. Endogenous Task-level Labor Supply

In the main text, we assume labor is inelastically supplied by all workers at the same cost. Thus, variation in output (and in turn utility) is governed only by the differences in skill endowments. Here, the objective function of the workers in the case of Cobb-Douglas production involves no optimization and is simply given by:

$$\left(\max\{A_t, \bar{l}_{it} \cdot S_{it}\} \cdot \bar{l}_{it-} \cdot S_{it-}\right)^{1/2} - 0 \quad \text{where} \quad \bar{l}_{it} = \bar{l}_{it-} = 1\,\forall i\,, \tag{A1}$$

where, as in the main text, we use the notation t to indicate the task being automated and $t-$ to indicate the non-automated task.

Of course, in reality, workers supply effort at some cost and choose how to allocate that effort across tasks in accordance with their expectations. One objective function that more closely captures that reality is:

$$\max_{l_{it}, l_{it-}} \left(\max\{A_t, l_{it} \cdot S_{it}\} \cdot l_{it-} \cdot S_{it-}\right)^{1/2} - (l_{it} + l_{it-})^2\,, \tag{A2}$$

which introduces a convex cost of (the sum of) effort and yields closed-form solutions for workers' optimal choices (l^*_{it}, l^*_{it-}) that will depend on their skills and the technology's capabilities.

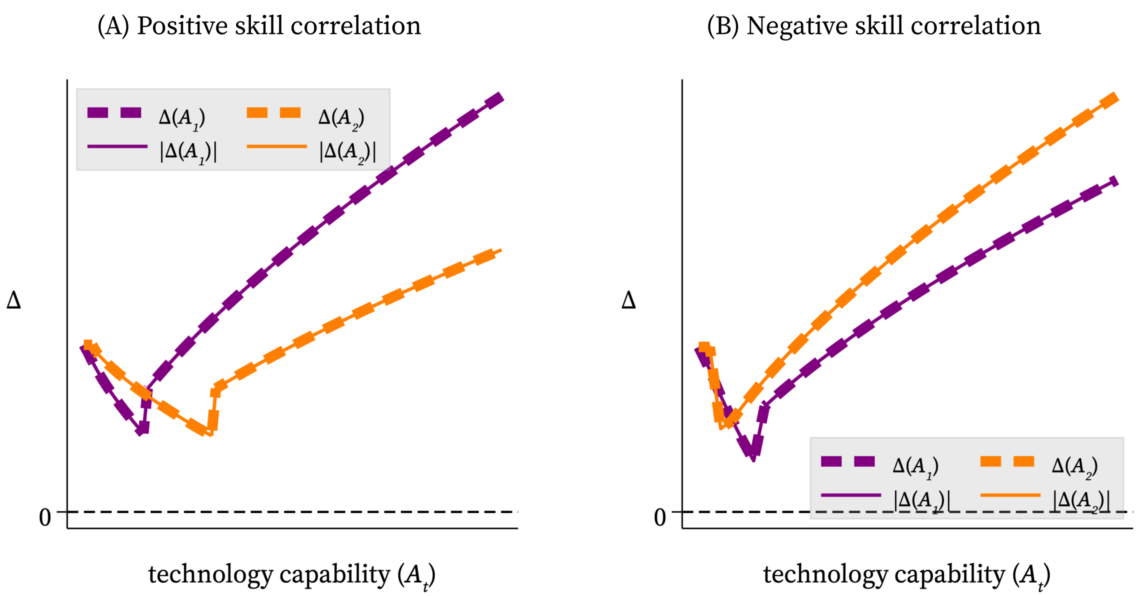
Solving for workers' optimal choices in Equation (A2) using the same scenarios of skill correlations presented in Table 2 and mirroring the analyses presented in Figure 2 of the main text still yields the result of non-monotonicity in inequality effects as automation improves.

In the case of positive skill correlations, Figure A1A illustrates a pattern very similar to the case of inelastic labor shown in Figure 2A. There is an initial decline in output inequality followed by a long-run increase in output inequality that eventually exceeds the original level of inequality. The transition appears to occur more quickly (i.e., at lower levels of technology capability); however, it is difficult to compare the specific values of the objective functions given that our simple, baseline model has workers facing no costs. Overall, the story remains the same.

In the case of negative skill correlations, Figure A1B illustrates some patterns that differ somewhat from our baseline scenario shown in Figure 2B. First, the first-difference and absolute measures of output inequality never diverge in this case of endogenous labor supply. This may partly reflect the specific values of the parameters we use here to illustrate the model — the low type never has enough of an absolute advantage in one task to "overtake" the high type. A second difference is that, in this case, the automation of both tasks leads to non-monotonic inequality paths, whereas in the baseline scenario shown in the main text, automation of task 2 only ever increases inequality.

Overall, these analyses reveal that our focal result of inequality effects being non-monotonic in the technology's capability carries through when workers endogenously allocate their labor across tasks.

FIGURE A1
Inequality and Single-task Automation — Cobb-Douglas Production, Endogenous Labor



Note: Plots the level of inequality across the two groups of workers assuming Cobb-Douglas production per first-difference inequality (Δ; solid line) or absolute inequality ($|\Delta|$; dashed line) depending on skill correlations, the task being automated, and the capability of the technology. Based on the case of endogenous labor supply per Equation (A2).

A.2. Irrational Expectations

In the main model, workers have rational expectations about the automation technology's capability. Thus, worker i uses the technology for task t if and only if the technology's capability exceeds the worker's task-specific skill, which implies an adoption threshold τ:

$$\tau^*_{it} = S_{it} \; . \tag{A3}$$

This assumption keeps the baseline model transparent and makes clear that the key state variable is the technology's capability relative to workers' task skills. However, it also rules out cases in which workers over-estimate or under-estimate whether the technology will improve their performance. This is not merely a theoretical possibility: some automation experiments find workers performing worse with algorithmic assistance than without it (e.g., Dell'Acqua et al. 2025; Otis et al. 2025), and related work on large language models suggests that people may misperceive the boundary of the technology's capabilities (e.g., Steyvers et al. 2025).

One way of relaxing our assumption of certain, rational expectations is to add noise to the adoption threshold (e.g., $\tau_{it} = S_{it} + \epsilon_{it}$, where ϵ is an i.i.d. error). If adoption mistakes are idiosyncratic, drawn from a continuous distribution, and independent of worker type and task skill, they will smooth the sharp adoption probabilities of the baseline model. While such mistakes alter expected output by causing costly premature or delayed delegation, they do not, by themselves, generate a systematic prediction about whether automation should help high- or low-output workers more.

The more interesting case is when adoption decisions are biased in a way that is correlated with worker type or task skill. To capture this in the simplest possible way, suppose workers adopt based on a perceived adoption threshold rather than the true rational-expectations threshold. Let worker *i*'s perceived threshold for using technology on task *t* be:

$$\widetilde{\tau}_{it} = \lambda_{it} \cdot S_{it} \tag{A4}$$

The baseline model is the special case where $\lambda_{it} = 1$ for all workers and tasks. If $\lambda_{it} < 1$, worker i adopts the technology earlier than in the rational-expectations benchmark. This can be interpreted as over-estimating the technology, under-estimating one's own skill, or otherwise being too willing to delegate. If $\lambda_{it} > 1$, worker i adopts later than in the benchmark, which can be interpreted as under-estimating the technology, over-estimating one's own skill, or otherwise being reluctant to delegate.

With irrational expectations, output still depends on the actual technology capability. Thus, after the adoption decision is made, output in the single-task automation Cobb-Douglas case is:

$$Y_i = \left(\left(\mathbf{1}\{\lambda_{it} \cdot S_{it} \geq A_t\} \cdot S_{it} + \mathbf{1}\{\lambda_{it} \cdot S_{it} < A_t\} \cdot A_t \right) \cdot S_{it^-} \right)^{1/2} \tag{A5}$$

where t^- denotes the task that is not automated. This preserves the production side of the model and changes only the adoption rule.

For simplicity, we assume there are early and late adopters corresponding to $\lambda_{it} = \{\lambda_E, \lambda_L\}$, where $\lambda_E < 1 < \lambda_L$; if a worker has a low value of λ, then their threshold for adopting the technology is lower and they adopt (irrationally) earlier.

To study the role of irrational expectations, we consider four structures of correlation between the λ parameter and workers' skill and output levels. Combined with the binary baseline possibility of task-level skills being positively or negatively correlated, this yields eight scenarios that we summarize in Table A1 and plot the absolute inequality trends in

Figures A2 and A3.

TABLE A1
Belief-bias scenarios

Figure	Skill correlation	Bias corr.	Bias timing
A2A	Positive	Type	*Hi* adopts early
A2B	Negative	Type	*Hi* adopts early
A2C	Positive	Type	*Hi* adopts late
A2D	Negative	Type	*Hi* adopts late
A3A	Positive	Skill	high-S_{it} adopts early
A3B	Negative	Skill	high-S_{it} adopts early
A3C	Positive	Skill	high-S_{it} adopts late
A3D	Negative	Skill	high-S_{it} adopts late

Notes: *Hi* and *Lo* denote workers with initially high and low output, "high-S_{it}" denotes worker with higher task-t skill, and "early" means $\lambda_{it} = \lambda_E < 1$ and "late" means $\lambda_{it} = \lambda_L > 1$.

The figures make clear that the non-monotonic inequality effects of automation not only persist, but become even more amplified. On the x axes of Figures A2 and A3, we label the "kinks" when there are discrete, adoption-induced shifts in the inequality trends that depend on the values of the λ (and other skill) parameters. The presence of irrational expectations introduces a sort of "Z"-shaped series of reversals, where inequality goes down-up-down (e.g., when high types adopt early) or up-down-up (e.g., when low types adopt early) before stabilizing to a trend once both types adopt.

Taking a step back, one of the broad messages of this paper is that the external validity of inequality effects estimated in automation experiments — as in, their validity in predicting how inequality will change as the focal technology continues to improve — requires understanding (*i*) the correlation of skills/inputs across workers, and (*ii*) which specific skills/inputs are automated. Without this understanding, it may be the case that a negative inequality effect at current technological capabilities will switch to a positive inequality effect at advanced technological capabilities. This is the "non-monotonicity" we frequently emphasize. This irrational expectations exercise shows that our assumption about rational expectations is not necessary for generating the non-monotonic results. In fact it introduces a parameter that generally amplifies the jaggedness of the inequality trend.

One way for irrational expectations to remove the non-monotonicities is for the two perceived thresholds to coincide exactly:

$$\lambda_{Hi,t} \cdot S_{Hi,t} = \lambda_{Lo,t} \cdot S_{Lo,t} \,. \tag{A6}$$

That is, belief bias *exactly* offsets the task-specific skill gap. This condition is knife-edge and task-specific, so we hypothesize that the presence of irrational expectations only bolsters our message that inequality trends due to increasing technological capabilities will be non-monotonic.

A.3. Use Costs or Skill of Use

We can also use the model to study ex-post differences in the value workers receive from the same automation technology. Rather than changing adoption beliefs, let η_{it} denote worker i's net-use multiplier for task-t automation. This term captures delegation, use, verification, training, or any other "iceberg" style frictions that affect realized effectiveness conditional on use. In the two-task Cobb-Douglas environment used for the figures, the effective input delivered by the technology is $\eta_{it} \cdot A_t$, so the single-task automation production function becomes:

$$Y_i = \left(\max\{S_{it}, \eta_{it} \cdot A_t\} \cdot S_{it^-}\right)^{1/2} \ . \tag{A7}$$

The baseline model is the special case $\eta_{it} = 1$ for all workers and tasks. Values $\eta_{it} < 1$ capture frictions that lower the net value of using the technology, while values $\eta_{it} > 1$ capture training, organizational support, or worker-specific ability to use the technology especially effectively. In this formulation, the relevant capability threshold is $A_t = S_{it}/\eta_{it}$: lower net-use multipliers delay the point at which automation becomes productive for a given worker-task pair, while higher multipliers bring that point forward.

We can then explore the same correlation structures as in the biased-beliefs exercise (Section A.2, above). First, η_{it} may be correlated with initial output type: high-output workers may have higher or lower net-use multipliers than low-output workers. Second, η_{it} may be correlated with skill in the automated task: workers who are better at the task may also be better or worse at using the technology for that task. Figures A4 and A5 plot these cases, crossed with the positive- and negative-skill-correlation environments used throughout the paper. Common use costs mainly rescale effective technology capability, but worker- or task-correlated use costs can change which worker benefits first and how strongly they benefit after adoption. Thus, delegation and verification costs do not generally overturn the paper's core message; they add another reason why the inequality effect of an automation experiment depends on the interaction of technology capability, task-level skill structure, and worker-specific frictions in using the technology.

A.4. Automation (Substitution) or Augmentation (Complementarity)

A useful distinction in the recent literature is between "automation" and "augmentation" technologies. Fügener, Walzner, and Gupta's (2026) analyses of human-AI collaboration are especially helpful on this point because they separate two ideas that are often conflated in discussions of technology-assisted work: using technologies to substitute for human performance on a task, versus using technologies to improve the performance of a human who remains actively involved in the task (Fügener, Walzner, and Gupta 2026).

Fügener, Walzner, and Gupta (2026) introduce two concepts related to augmentation as complementarity. First, they describe "between-task complementarity" as a measure of the correlation between a human's performance and a technology's performance across tasks (see their Equation 6, specifically the b parameter). This is not complementarity in the production-function sense of the word, where the central question is how substitutable two inputs are inside a production function. In our context, this is related to the question of which task gets automated first, which we explore in depth throughout the paper.

The more relevant concept from Fügener, Walzner, and Gupta (2026) is "within-task complementarity" — the notion that, when a human and a technology *both* complete the task, the result may be more output than if either completed the task in isolation (see their Equation 1). Our baseline model does not admit this possibility since we assume within-task perfect substitution.

To consider the possibility of within-task complementarity, it is important to be clear about what it means in the context of our worker-level production function. We can envision three possibilities.

Augmentation as Sub-tasks. If a task is itself composed of lower-level sub-tasks, then augmentation can be modeled as the substitution of technologies for labor within these sub-tasks. For example, if task 1 consists of subtasks k and k^{-}, one could write the worker-level production function as:

$$Y_i = \left((\max\{A_{1k}, s_{i1k}\} \cdot s_{i1k^{-}})^{1/2} \cdot S_{i2} \right)^{1/2} . \tag{A8}$$

This is a nested task model, and it allows for within-task complementarities, but it does not introduce any fundamentally different economic forces from our baseline model. It simply shifts the level at which substitution occurs from tasks to sub-tasks.

Augmentation as New Tasks. Another possibility is that the technology makes a new task feasible. For example, if a third task becomes feasible only after the technology is introduced, one could write the worker-level production function as:

$$Y_i = (S_{i1} \cdot S_{i2} \cdot \max\{A_3, 1\})^{1/3} . \tag{A9}$$

This case is conceptually different because the job itself has changed. In this simple version, there is no adoption threshold tied to worker skill in the new task, and the effect of automation on output differences is smooth and monotone. This is an interesting boundary case, but it is not the main setting we have in mind when interpreting short-run automation experiments. Most technologies in these experiments do not *only* create new tasks; they also change the value of tasks workers were already performing.

Nested Production Functions. The third possibility is perhaps closest to the connotation of "augmentation" as applied to within-task complementarity: the human and the technology jointly perform the same task, and the joint mode may dominate both human- and technology-alone production. A compact way to write this in our two-task Cobb-Douglas environment is:

$$Y_i = \left(\max\left\{S_{it},\, A_t,\, \omega_{it}\cdot(S_{it}\cdot A_t)^{1/2}\right\}\cdot S_{it^-}\right)^{1/2}, \tag{A10}$$

where the third term is the augmented task input. The parameter ω_{it} captures the within-task complementarities. For example, in the forecast setting of Fügener, Walzner, and Gupta (2026), these complementarities may come from the worker and technology having different information sets such that a forecast that aggregates the worker's and technology's forecasts is more accurate than either independently. With Equation A10 in mind, our baseline model implicitly assumes: $\omega_{it} \leq 1\ \forall i, t$. Allowing for positive within-task complementarity in this model can be done by allowing for $\omega_{it} > \max\{\sqrt{S_{it}/A_t}, \sqrt{A_t/S_{it}}\}$.

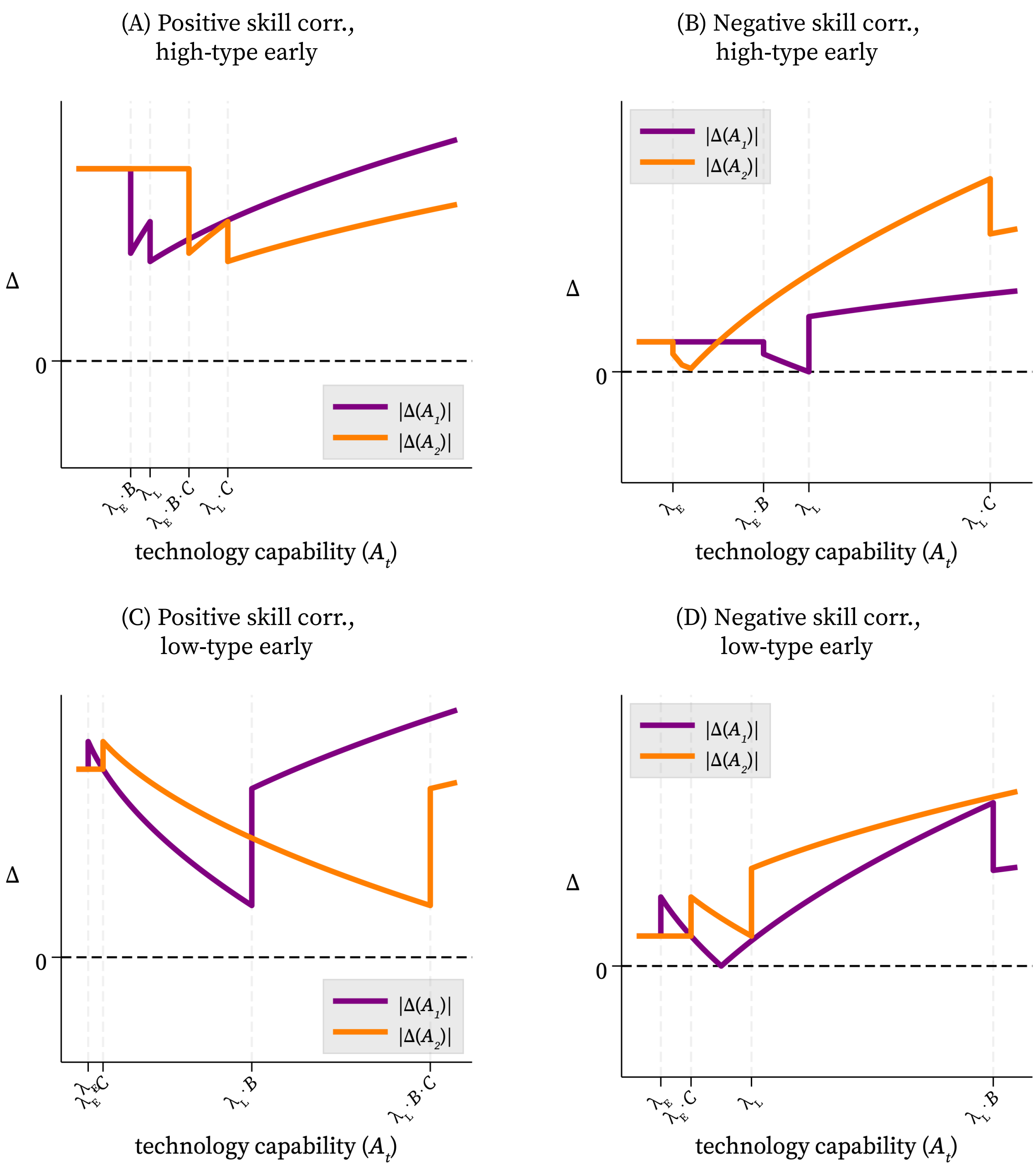
To see how our results change when we model augmentation as these positive, within-task complementarities per Equation A10, we can actually look back to our results above in Section A.3, where we explored models where workers have heterogeneous usage costs/multipliers. To see this, first note that the augmented task-level input can be written as the technology's capabilities (A) multiplied by an effective-use term:

$$\omega_{it}\cdot(S_{it}\cdot A_t)^{1/2} = \omega_{it}\cdot\left(\frac{S_{it}}{A_t}\right)^{1/2}\cdot A_t \equiv \eta^{\text{aug}}_{it}\cdot A_t\,. \tag{A11}$$

Thus, an augmentation model is closely related to the use-cost model in Section A.3, Equation A7. The difference is that the effective multiplier ($\eta_i t$) is now microfounded by the structure of the within-task complementarities.

Rather than solving every possible schedule of ω_{it}, the figures in Section A.3 isolate the first-order effects of augmentation via η^{aug}_{it} in the cases where effective use is higher for the worker who is more skilled in the automated task. These are the cases shown in Panels A5A and A5B. Again, the main conclusion is unchanged. Allowing for human-technology augmentation does not imply that an inequality effect measured at one level of technology capability should be extrapolated monotonically to another. It adds another skill object: the degree of augmentation (per ω). Thus, augmentation is an important extension of the baseline substitution model, and yet it reinforces the need to model task-level skills and technology capability jointly when determining inequality effects of automation experiments.

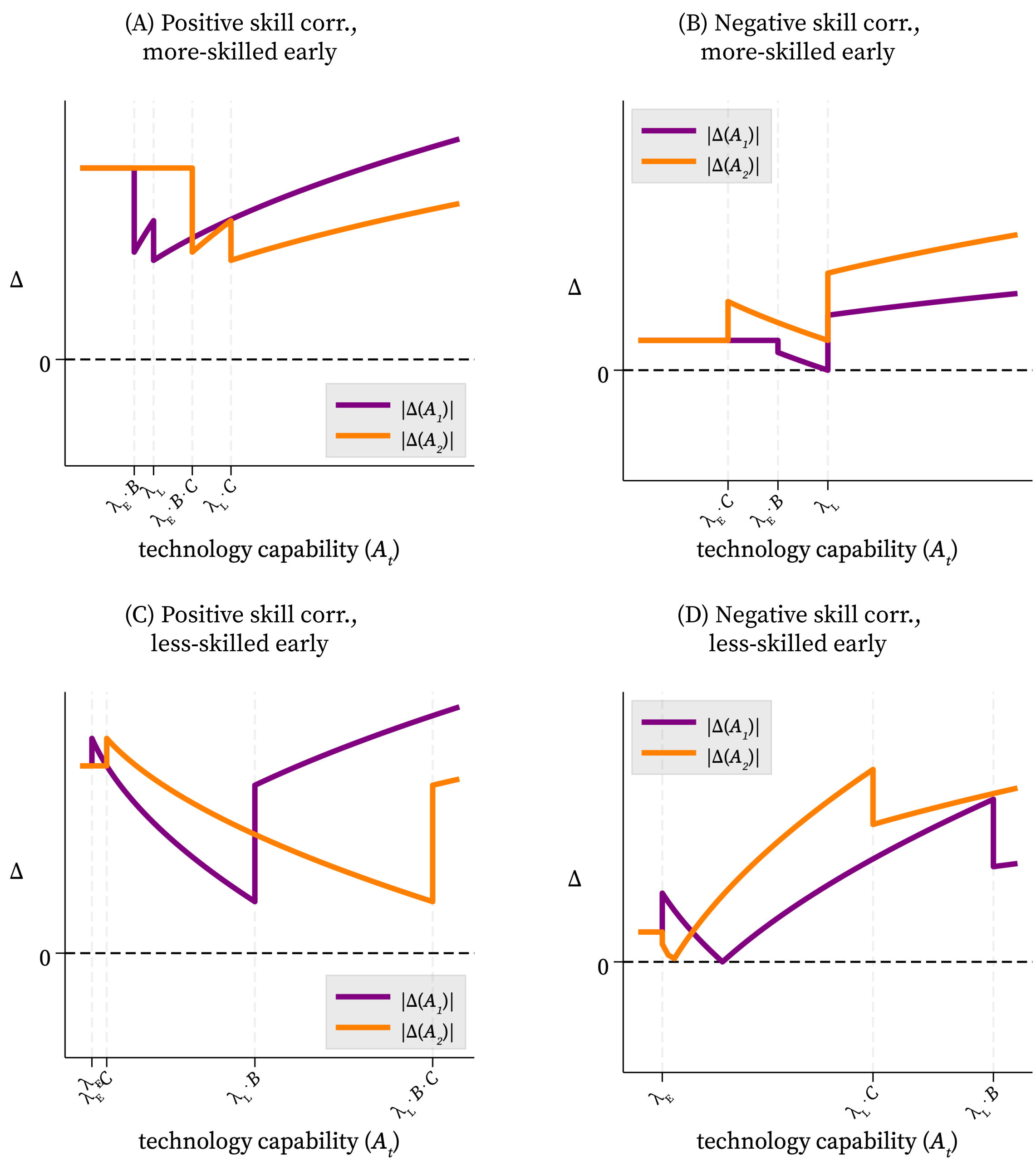
FIGURE A2
Irrational Adoption — Type-correlated Optimism



Note: This figure plots absolute inequality, $|\Delta(A_t)|$, for task 1 and task 2 automation with irrational adoption. Workers adopt task-t automation when $A_t > \lambda_{it} \times S_{it}$, but realized output after adoption uses actual capability A_t. All parameters besides λ match the original, rational figure.

FIGURE A3

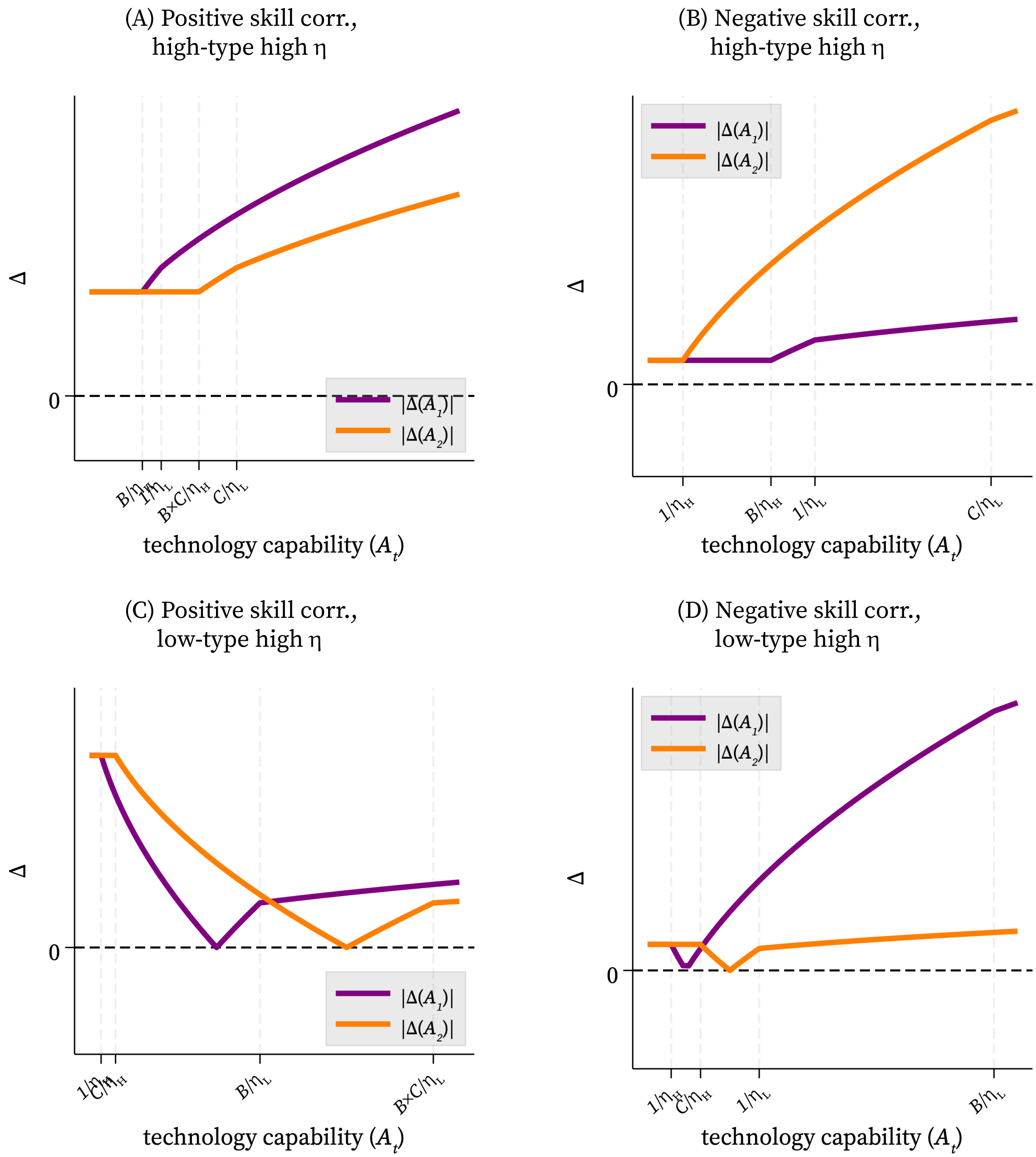
Irrational Adoption — Skill-correlated Optimism



Note: This figure plots absolute inequality, $|\Delta(A_t)|$, for task 1 and task 2 automation with irrational adoption. Workers adopt task-t automation when $A_t > \lambda_{it} \times S_{it}$, but realized output after adoption uses actual capability A_t. All parameters besides λ match the original, rational figure.

FIGURE A4

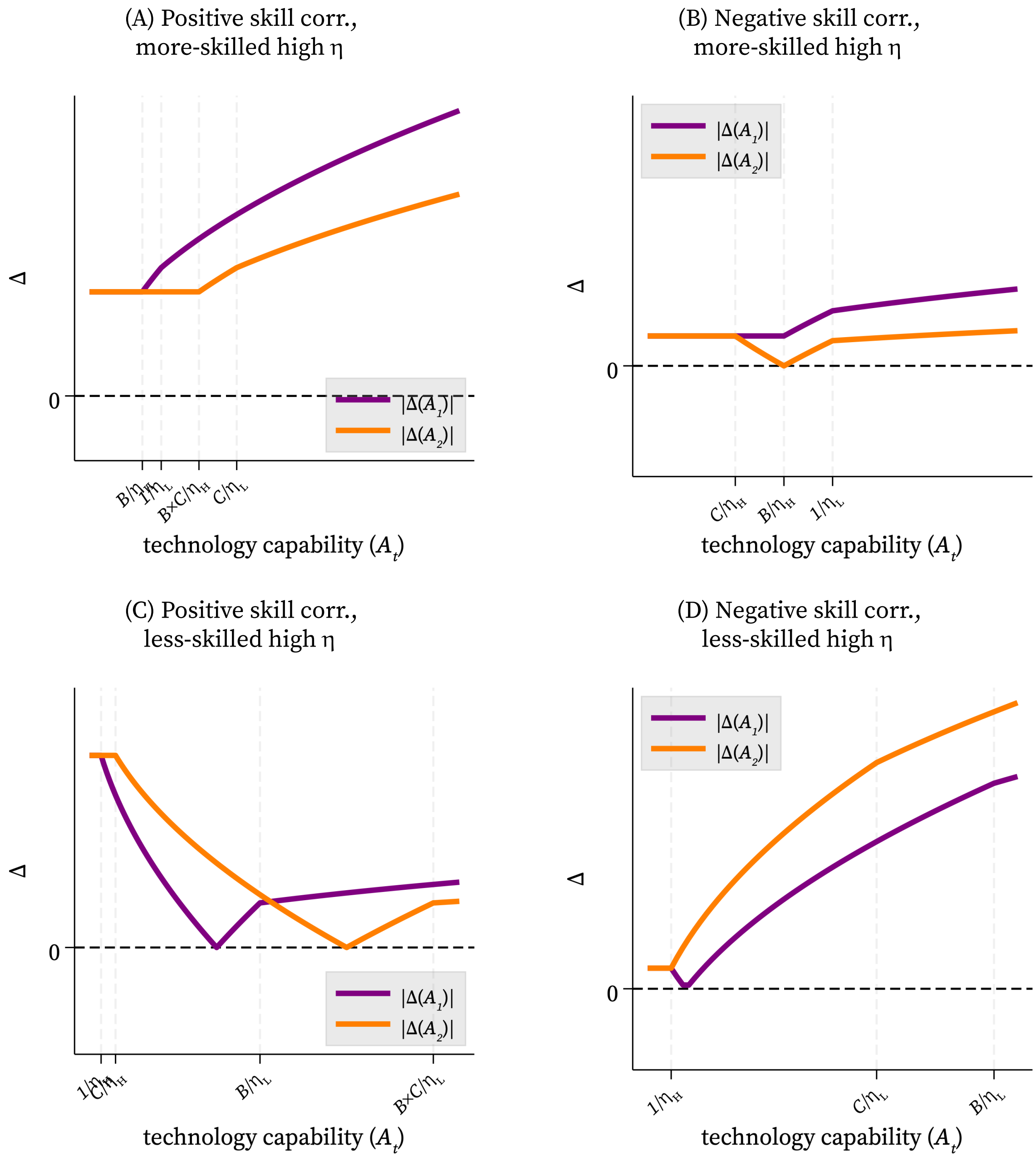
Delegation and Verification Costs — Type-correlated Effective Use



Note: This figure plots absolute inequality, $|\Delta(A_t)|$, for task 1 and task 2 automation when workers differ in the ex-post effectiveness of using the technology. The effective automated-task input is $\eta_{it}A_t$, so output in the automated task is governed by $\max\{S_{it}, \eta_{it}A_t\}$. Higher η_{it} means lower delegation, use, or verification costs. The panels vary according to whether the initially high-output or low-output worker type has the higher net-use multiplier. All parameters besides η match the original, rational figure.

FIGURE A5

Delegation and Verification Costs — Skill-correlated Effective Use



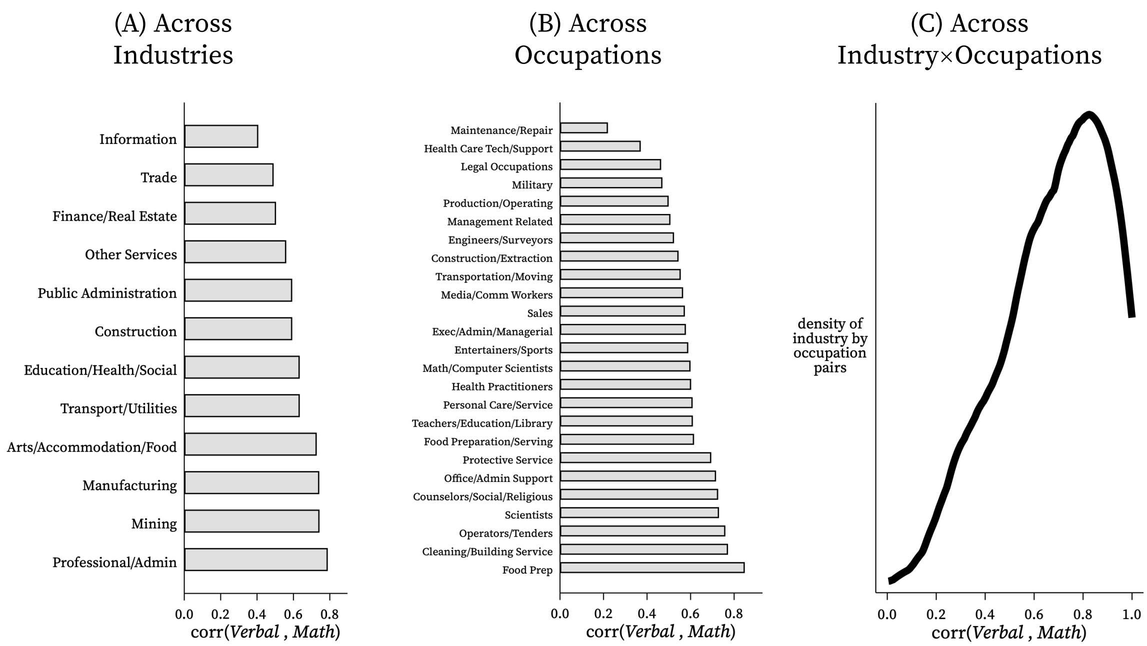
Note: This figure plots absolute inequality, $|\Delta(A_t)|$, for task 1 and task 2 automation when workers differ in the ex-post effectiveness of using the technology. The effective automated-task input is $\eta_{it}A_t$, so output in the automated task is governed by $\max\{S_{it}, \eta_{it}A_t\}$. Higher η_{it} means lower delegation, use, or verification costs. The panels vary according to whether the worker who is more skilled in the automated task has the higher or lower net-use multiplier. All parameters besides η match the original, rational figure.

Appendix B. Skill Correlations per SAT Scores

The National Longitudinal Survey of Youth (NLSY) reports participants' SAT scores for verbal and math components separately (BLS 2024). We use the most commonly listed industry and occupation codes for all employed individuals. The public data yield 1,618 observations with both math and verbal SAT scores as well as non-missing industry and occupation codes.[21]

Overall, math and verbal SAT scores in our sample exhibit a correlation of 0.61, which is squarely in line with the estimates summarized in the previous subsection. But our primary goal is to shed some initial light on portions of the economy where skills (at least per this metric) may be more or less correlated.

FIGURE B1
Math and Verbal SAT Score Correlations across the Economy



Note: Plots the correlation of individuals' SAT scores on the verbal and math components from the NLSY97. Panels (A–B) plot average correlations at the industry- (Panel A) or occupation-level (Panel B). Panel (C) plots the distribution of correlations averaged at the industry-occupation-level.

As shown in Figure B1, workers in some places exhibit a tight link between these skills, with estimated correlations approaching 0.8 (e.g., professional and administrative services, food preparation, cleaning and building services), while in other places the association is

[21] For simplicity and to preserve sample size, we aggregate the industries and occupations into levels slightly broader than listed in the data. SATs are graded out of 800 for each of the two components. For confidentiality, the scores are aggregated into 100-point bins in the public data, which are what we use in this analysis. All correlations are reported per the inclusion of the NLSY97 sample weights, although this makes little practical difference.

much more modest, on the order of 0.3 to 0.5 (e.g., information sector, legal services) going as low as 0.2. At the granular level of industry–occupation pairs, we find substantial dispersion in math–verbal correlations (see Panel C).

Notably, we do not observe any industry and/or occupation where the correlation is estimated to be negative. Still, our model suggests that the heterogeneity in skill correlations across the economy would also imply heterogeneous effects of automation on inequality. That is, a single automation technology deployed in these different sectors could yield significantly different changes in inequality.

More speculatively, the uniformly positive correlations help reconcile the fact that, as of this paper, the majority of automation experiments are documenting inequality declines. Presently, many of the technologies in question are both (*i*) not yet "super-human" in that they are better than the worst workers, but not better than the best workers; and (*ii*) focused on cognitive tasks, in particular, connected to software. Combined with positive (cognitive) skill correlations, our model predicts an inequality reduction in this scenario.

However, the positive math-verbal skill correlations we document here reflect sorting into broad economic sectors and may mask negative correlations at finer levels of analysis within firms or teams where complementary specialization drives task allocation.

Appendix C. Investment in Training and/or Automation

The following is a stylized model of a manager's decision to invest in (*i*) "training" to improve lower-skilled workers (S_{it}), and/or (*ii*) "engineering" to improve the capabilities of the automation technology (A_t). We make a number of simplifications here and develop this model primarily to show that our focal model of jobs-as-production-functions is a key ingredient for the manager's investment problem.

C.1. Setup

Worker-level production and notation are as in the main text. We focus on the two-task, equal-share Cobb–Douglas case. The manager has an investment budget normalized to one and chooses training investment $X_S \geq 0$ and engineering investment $X_A \geq 0$, subject to $X_S + X_A \leq 1$. Unused budget is allowed. Both investments target task 1. Training raises every worker's task-1 skill to $1 + \kappa_S\sqrt{X_S}$, unless the worker's original skill is already higher. Engineering raises the task-1 technology's initial capability from A to $A + \kappa_A\sqrt{X_A}$. The parameters $\kappa_S, \kappa_A > 0$ describe the effectiveness of the two investments; the square-root functions impose diminishing returns.

After investment, worker i produces

$$Y_i(X_S, X_A) = \left(\max\left\{S_{i1}, 1 + \kappa_S \cdot \sqrt{X_S}, A + \kappa_A \cdot \sqrt{X_A}\right\} S_{i2}\right)^{1/2}. \tag{C1}$$

The manager places weight $\delta \in [0, 1]$ on total output and weight $1 - \delta$ on absolute output inequality. Thus, their objective function is:

$$\max_{X_S, X_A} \delta \cdot (Y_{Hi} + Y_{Lo}) - (1 - \delta) \cdot |Y_{Hi} - Y_{Lo}| \quad \text{s.t. } X_S + X_A \leq 1. \tag{C2}$$

When one worker has higher output, the manager places an effective weight of $2 \cdot \delta - 1$ on that worker and a weight of one on the lower-output worker. Thus, $\delta = 1/2$ makes the manager maximize the output of the lower-output worker. When $\delta < 1/2$, the manager is willing to sacrifice total output to avoid increasing the output of the worker who is already ahead.

For the remainder of the analysis, let $q \equiv \max\{1 + \kappa_S\sqrt{X_S}, A + \kappa_A\sqrt{X_A}\}$ denote the common task-1 capability produced by the two investments. The largest attainable capability is $q_{\max} \equiv \max\{1 + \kappa_S, A + \kappa_A\}$. Splitting the budget cannot produce a capability above either all-training or all-engineering investment because the two inputs enter task-1 production through a max operator.

We restrict attention to the empirically relevant, "non-superhuman" range:

$$\max\{1, A\} \leq \frac{B}{C} < q_{\max} \leq B. \tag{C3}$$

The first two inequalities imply that investment can reach, but has not already passed, the capability B/C. The final inequality means that neither investment can raise the common capability above the high type's task-1 skill.

Because unused budget is allowed, a target capability q can be implemented using at least one investment type, provided it fits within the budget. The required expenditures are

$$X_S(q) = \left(\frac{q-1}{\kappa_S}\right)^2, \qquad X_A(q) = \left(\frac{q-A}{\kappa_A}\right)^2. \tag{C4}$$

An expenditure is feasible when it lies in $[0,1]$. At least one is feasible for every target considered below. When several allocations produce the same welfare, we report the one that uses the least total budget. When both investments can reach $q > A$, training uses less budget if and only if $\kappa_S/\kappa_A > (q-1)/(q-A)$.

C.2. Positive Skill Correlation

Using the positive-correlation skill assignments from the main text, the restriction $q \leq B$ means that investment does not change the high type's output: $Y_{Hi} = B \cdot \sqrt{C}$, while $Y_{Lo} = \sqrt{C \cdot q}$. The high type remains ahead, while the low type's output is strictly increasing in q. A higher capability therefore increases total output and reduces inequality. Both components of the manager's objective favor $q = q_{\max}$ for every δ.

The associated investment choice is a corner. If $1 + \kappa_S > A + \kappa_A$, all investment goes to training. If the inequality is reversed, all investment goes to engineering. If the two maximum capabilities are equal, either corner is optimal.

C.3. Negative Skill Correlation

Using the negative-correlation skill assignments from the main text, output over the restricted capability range is $Y_{Hi} = \sqrt{B}$ and $Y_{Lo} = \sqrt{C \cdot q}$. The workers have equal output at $q = B/C$. Below this point, a higher capability raises the low type's output, increasing total output and reducing inequality. Every manager therefore prefers to reach at least B/C.

Above B/C, the initially low-output worker moves ahead while the initially high-output worker's output remains fixed. The manager's objective is then $\sqrt{B} + (2 \cdot \delta - 1) \cdot \sqrt{C \cdot q}$. It decreases in q when $\delta < 1/2$, is constant when $\delta = 1/2$, and increases in q when $\delta > 1/2$. Therefore, using the minimum-spending rule when $\delta = 1/2$, the manager chooses

$$q^* = \begin{cases} B/C, & \delta \leq 1/2, \\ q_{\max}, & \delta > 1/2. \end{cases} \tag{C5}$$

When $\delta \leq 1/2$ and $A < B/C$, training is the cheaper way to reach equality when $\kappa_S/\kappa_A >$

$(B/C-1)/(B/C-A)$. If the inequality is reversed, engineering is cheaper. If the two sides are equal, either investment can implement equality at the same cost. If $A = B/C$, the existing technology already implements equality and no new investment is required. When $\delta > 1/2$, the manager instead selects the same maximum-capability corner as under positive skill correlation.

C.4. Summary

Table C1 summarizes the solution. Each condition determining the investment instrument is reported in a separate row. The reported allocation minimizes spending among allocations that produce the same value of the manager's objective.

TABLE C1
Optimal training and engineering investment

Skill correlation	Output weight	q^*	Condition	Optimal (X_S^*, X_A^*)
Positive	Any δ	$q_{\max}$	$1+\kappa_S \geq A+\kappa_A$	$(1,0)$
		$q_{\max}$	$1+\kappa_S \leq A+\kappa_A$	$(0,1)$
Negative	$\delta \leq 1/2$	B/C	$A = B/C$	$(0,0)$
		B/C	$A < B/C$ and $\frac{\kappa_S}{\kappa_A} \geq \frac{B/C-1}{B/C-A}$	$\left(\left(\frac{B/C-1}{\kappa_S}\right)^2, 0\right)$
		B/C	$A < B/C$ and $\frac{\kappa_S}{\kappa_A} \leq \frac{B/C-1}{B/C-A}$	$\left(0, \left(\frac{B/C-A}{\kappa_A}\right)^2\right)$
Negative	$\delta > 1/2$	$q_{\max}$	$1+\kappa_S \geq A+\kappa_A$	$(1,0)$
		$q_{\max}$	$1+\kappa_S \leq A+\kappa_A$	$(0,1)$

With positive correlation, investment raises the low type's output while leaving the high type's output unchanged. Greater capability therefore raises output and reduces inequality, so all managers choose the largest attainable capability.

With negative correlation, the same investment first closes the output gap and then reverses it. Managers who place relatively more weight on equality stop at B/C, while managers who place relatively more weight on total output continue to $q_{\max}$.

Conditional on the desired capability, the choice between training and engineering is separate and simple. The manager compares the spending needed to reach that capability through each investment. Training is preferred when its relative effectiveness κ_S/κ_A is sufficiently high; engineering is preferred when the technology begins sufficiently close to the target. Thus, workers' task-level skill correlation determines the capability the manager wants, while the investment technologies determine the cheapest way to achieve it.

Appendix D. Derivation of Within-Firm Skill Correlations

To summarize the O-ring model setup, each firm employs n workers, and each worker i performs two tasks using their task-specific endowments of skill: $S_{i1}, S_{i2} \in (0, 1]$. Worker i's production function is Cobb-Douglas and uses equal shares of both tasks. Following Kremer (1993), firm-level output (Z) is the product of worker-level output: $Z = \prod_{i=1}^{n} Y_i$, which generates positive assortative matching of workers into firms. Skills are distributed such that:

$$\begin{pmatrix} \ln S_{i1} \\ \ln S_{i2} \end{pmatrix} \sim \mathcal{N}\left(\begin{matrix} \mu \\ \mu \end{matrix}, \begin{matrix} \sigma^2 & corr_{\text{pop}}\sigma^2 \\ corr_{\text{pop}}\sigma^2 & \sigma^2 \end{matrix} \right),$$

where $corr_{\text{pop}} \in (-1, 1)$ is the population correlation of logged skills.

With Cobb-Douglas production, $Y_i = (S_{i1} \cdot S_{i2})^{1/2}$. Within each firm k, all workers produce output Y_k, so $S_{i1} \cdot S_{i2} = Y_k^2$ for all workers i in firm k. Taking logarithms of the constraint yields: $\ln S_{i1} + \ln S_{i2} = 2 \cdot \ln Y_k$. All within-firm variation in logged skills comes from the difference in logged skill levels. Assuming workers are randomly allocated among firms requiring their specific output level Y_k, we can write:

$$\begin{aligned} \text{Var}(\ln S_{i1} - \ln S_{i2} \mid Y_k) &= \text{Var}(\ln S_{i1} - \ln S_{i2}) \\ &= 2 \cdot \sigma^2 \cdot (1 - corr_{\text{pop}}) . \end{aligned} \tag{D1}$$

Within each firm, we can express each logged skill level as: $\ln S_{i1} = \ln Y_k + (\ln S_{i1} - \ln S_{i2})/2$ and $\ln S_{i2} = \ln Y_k - (\ln S_{i1} - \ln S_{i2})/2$. Therefore:

$$\begin{aligned} \text{Var}(\ln S_{i1} \mid Y_k) &= \text{Var}(\ln Y_k + (\ln S_{i1} - \ln S_{i2})/2 \mid Y_k) \\ &= \text{Var}((\ln S_{i1} - \ln S_{i2})/2) \\ &= \frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}, \end{aligned} \tag{D2}$$

and likewise for $\text{Var}(\ln S_{i2} \mid Y_k)$. Now, transforming back to levels, we have $S_{i1} = Y_k \cdot \exp((\ln S_{i1} - \ln S_{i2})/2)$ and $S_{i2} = Y_k \cdot \exp(-(\ln S_{i1} - \ln S_{i2})/2)$, where $(\ln S_{i1} - \ln S_{i2})/2 \sim \mathcal{N}(0, \sigma^2 \cdot (1 - corr_{\text{pop}})/2)$. Therefore:

$$\text{Var}(S_{i1} \mid Y_k) = Y_k^2 \cdot \exp\left(\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right)\left(\exp\left(\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right) - 1\right), \tag{D3}$$

and likewise for $\text{Var}(S_{i2} \mid Y_k)$. The covariance is then:

$$\text{Cov}(S_{i1}, S_{i2} \mid Y_k) = Y_k^2 \cdot \left(1 - \exp\left(\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right)\right), \tag{D4}$$

and, finally, the within-firm correlation is:

$$\begin{aligned}
corr_{\text{firm}} &= \frac{\text{Cov}(S_{i1}, S_{i2} \,|\, Y_k)}{\sqrt{\text{Var}(S_{i1} \,|\, Y_k) \cdot \text{Var}(S_{i2} \,|\, Y_k)}} \\
&= \frac{1 - \exp\left(\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right)}{\exp\left(\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right)\left(\exp\left(\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right) - 1\right)} \\
&= -\exp\left(-\frac{\sigma^2 \cdot (1 - corr_{\text{pop}})}{2}\right).
\end{aligned} \tag{D5}$$