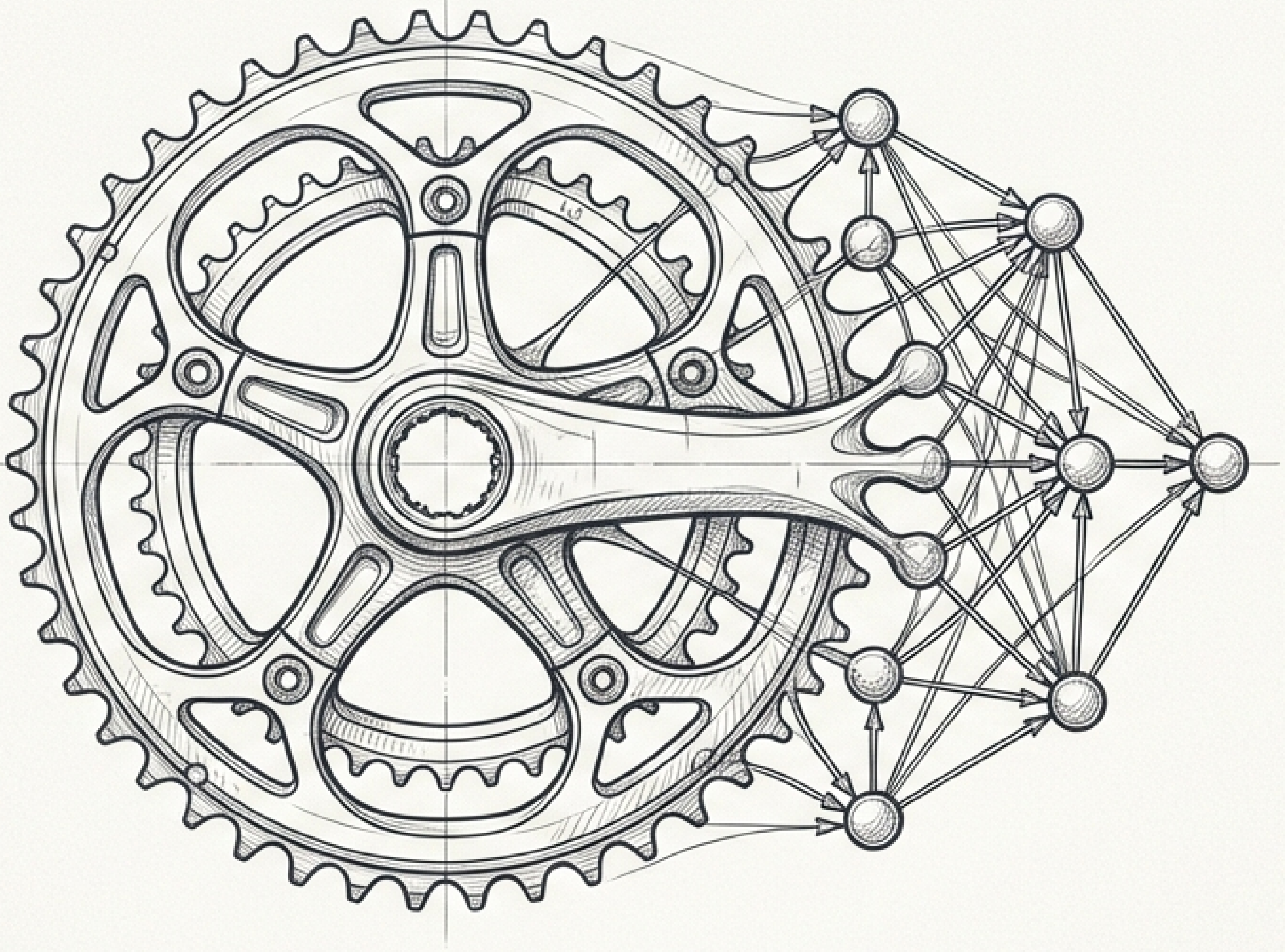


The Economics of Bicycles for the Mind

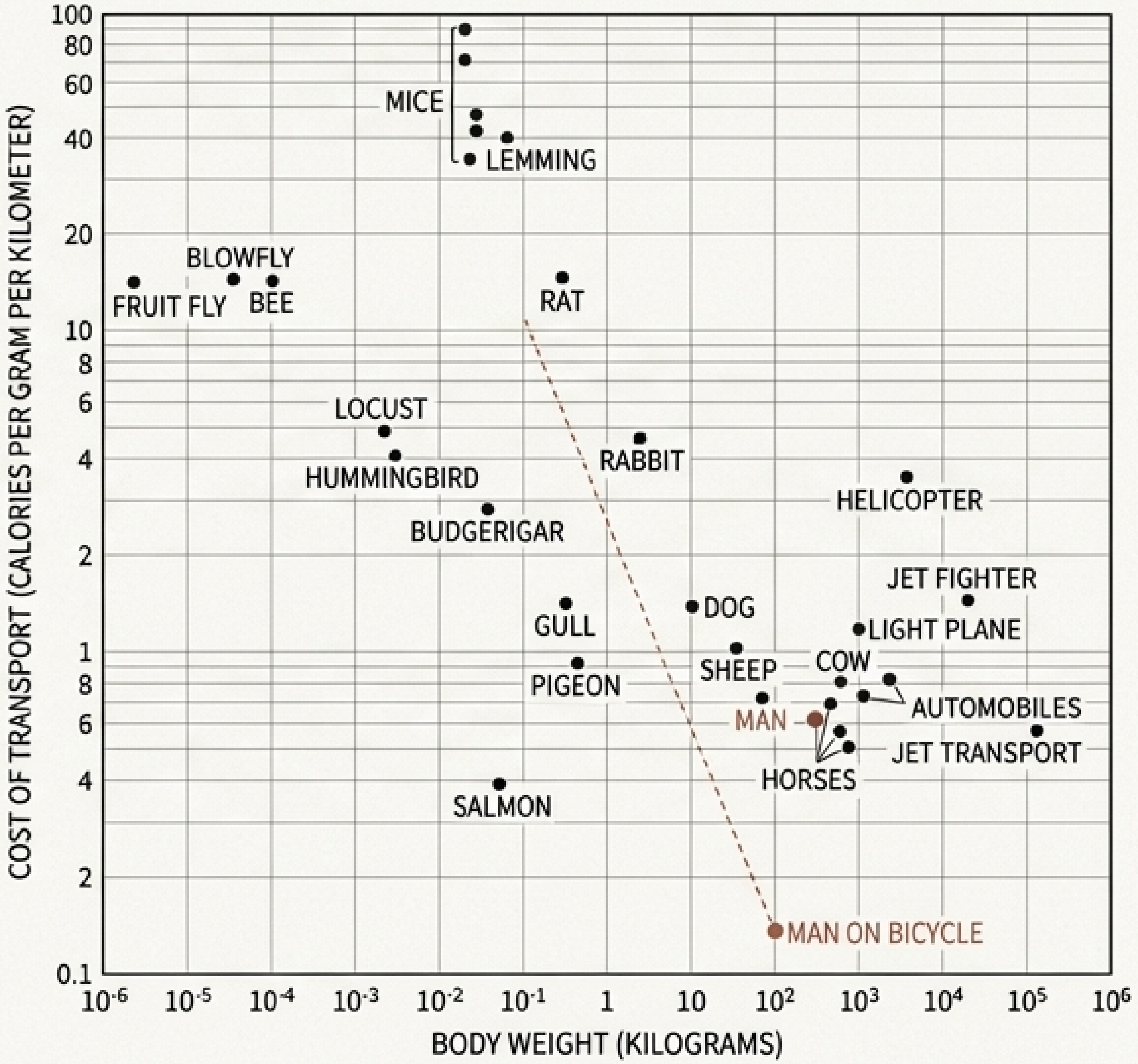
A Framework for Understanding Cognitive Tools, AI, and Inequality

Based on the paper by Ajay Agrawal, Joshua S. Gans, and Avi Goldfarb

The most remarkable tool we have ever come up with.

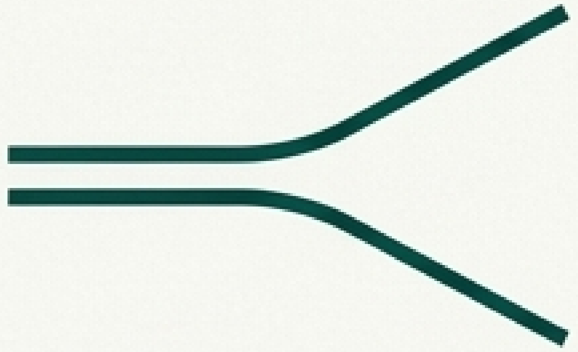
"What a computer is to me is it is the most remarkable tool that we have ever come up with, and it is the equivalent of a bicycle for our minds."

– Steve Jobs, Library of Congress, 1990



MAN ON BICYCLE ranks first in efficiency among traveling animals and machines in terms of energy consumed in moving a certain distance as a function of body weight.

A Puzzle: Cognitive tools have had opposing effects on inequality.

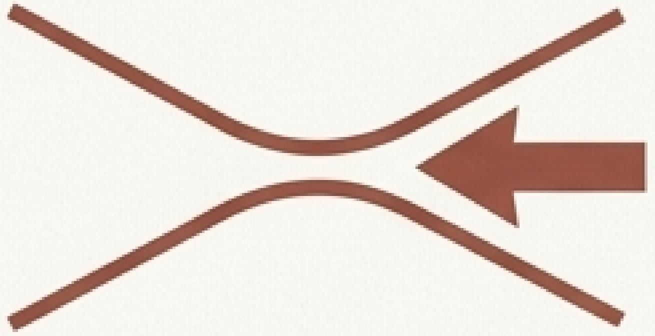


Computers Widened Wage Gaps

Computer adoption since the 1970s has disproportionately benefited high-skilled workers, widening wage inequality.

Mechanism
Primarily complemented non-routine cognitive skills while automating routine implementation tasks.

Autor et al., 2006; Goldin and Katz, 2008.



Early AI Seems to Compress Them

Early evidence on AI tools suggests they compress productivity differences, with the largest gains for lower-performing workers.

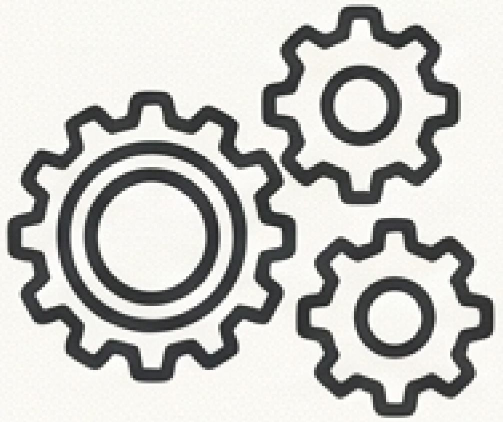
Mechanism:
Tools appear to substitute for experience and implementation skill in settings like call centers and software development.

Brynjolfsson et al., 2025; Noy and Zhang, 2023.

If both computers and AI are ‘bicycles for the mind,” why do they have opposing distributional consequences?

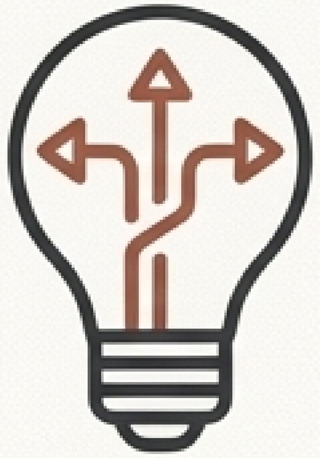
A Framework for Resolution: Deconstructing Cognitive Work

We can resolve the puzzle by distinguishing three inputs to cognitive work:



Implementation Effort & Skill (s)

The mental or physical resources devoted to *executing* improvements.



Opportunity Judgment (Γ)

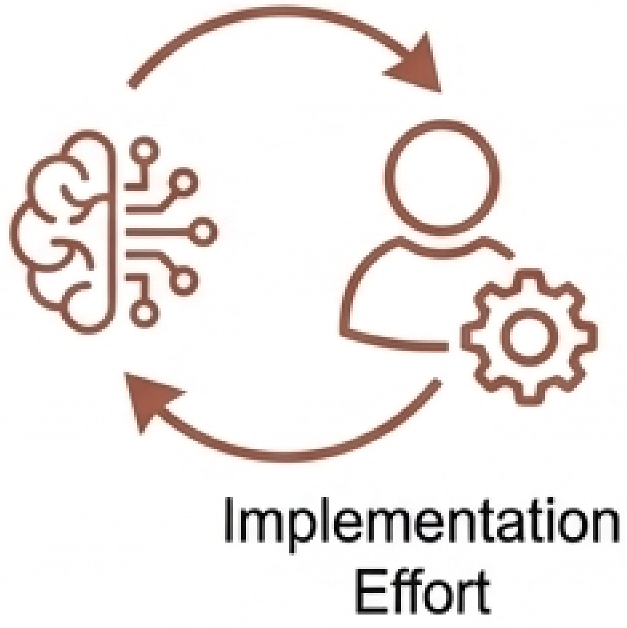
The ability to *recognize* potential improvements or valuable problems to solve.

Payoff Judgment (α)

The ability to *realize value* from improvements, such as translating a technical success into a business outcome.

The Core Insight: Tools Substitute for Implementation but Complement Judgment

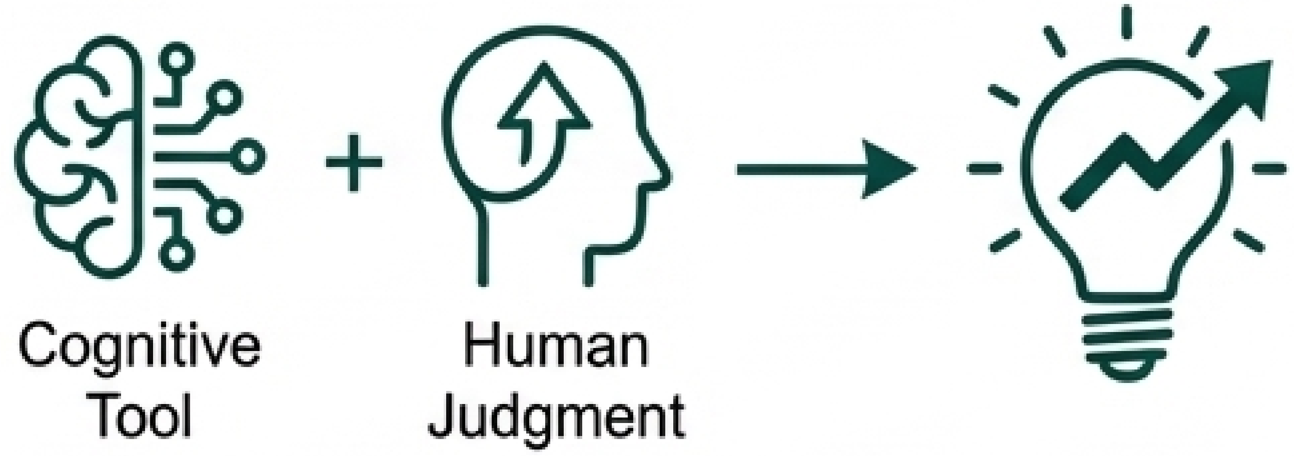
The distributional impact of any cognitive tool depends on a fundamental asymmetry:



Substitution Effect

Tools substitute for human implementation effort. A better tool raises the probability of success for any given effort level, but it *reduces the marginal return to that effort*. The tool does some of the work, so additional human exertion matters less.

This creates an **"inverse skill bias"**: high-skill implementers gain the least.



Complementarity Effect

Tools complement human judgment. Better implementation tools amplify the returns to recognizing the right opportunities (**Opportunity Judgment**) and acting on them correctly (**Payoff Judgment**).

The effect of any tool on inequality depends on the balance of these forces and the relative variance of implementation skill vs. judgment in the population.

The Baseline Model: Iterative Task Improvement

An agent repeatedly perceives opportunities to improve a task over discrete time periods $t = 0, 1, 2, \ldots$

1. **Opportunity Judgment ($\gamma(t)$)**
   The probability the agent perceives an improvement opportunity in period t. Assumed to be weakly declining in t.
2. **Implementation Effort (e_t)**
   Chosen by the agent conditional on perceiving an opportunity. Effort has a cost $c(e)$.
3. **Implementation Skill (s)**
   An agent-specific parameter that enhances the productivity of effort.
4. **Cognitive Tool (θ)**
   A technology that improves the probability of successful implementation, $p(se; \theta)$.
5. **Payoff Judgment (α)**
   The probability that a successful implementation is realized, yielding a payoff increase of Δ.

The Agent's Optimization Problem

Net Benefit Per Opportunity

Conditional on perceiving an opportunity, the agent chooses effort e to maximize the expected net benefit:

$$M(e;\theta) = \underbrace{p(se;\theta)\alpha\Delta}_{\text{Expected Payoff}} - \underbrace{c(e)}_{\text{Cost of Effort}}$$

First-Order Condition

The optimal effort, e^*, equates marginal benefit and marginal cost:

$$\underbrace{p'(se^*;\theta)s^{\alpha\Delta}}_{\text{Marginal Benefit of Effort}} = \underbrace{c'(e^*)}_{\text{Marginal Cost of Effort}}$$

Unconditional Expected Value

The total value to the agent is the net benefit per opportunity, multiplied by the expected number of opportunities.

$$V_0(\theta) = \Gamma M^*(\theta)$$

Captures the expected discounted number of opportunities perceived by the agent.

Where Γ is the discounted opportunity judgment multiplier: $\Gamma = \sum \left(\prod \gamma(i)\right) \delta^t$

What is a Cognitive Tool? A Formal Definition

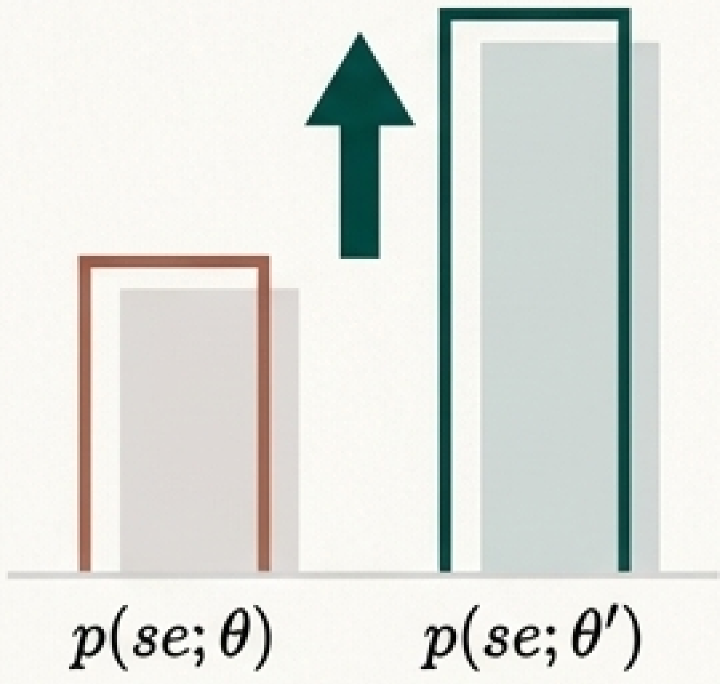
A cognitive tool is characterized by a parameter $\theta \geq 0$ such that:

1. Productivity Enhancement

For any effort e and tool improvement $\theta' > \theta$:

$$p(se;\theta') \geq p(se;\theta)$$

Intuition: Better tools increase the probability of success for any given level of human effort.

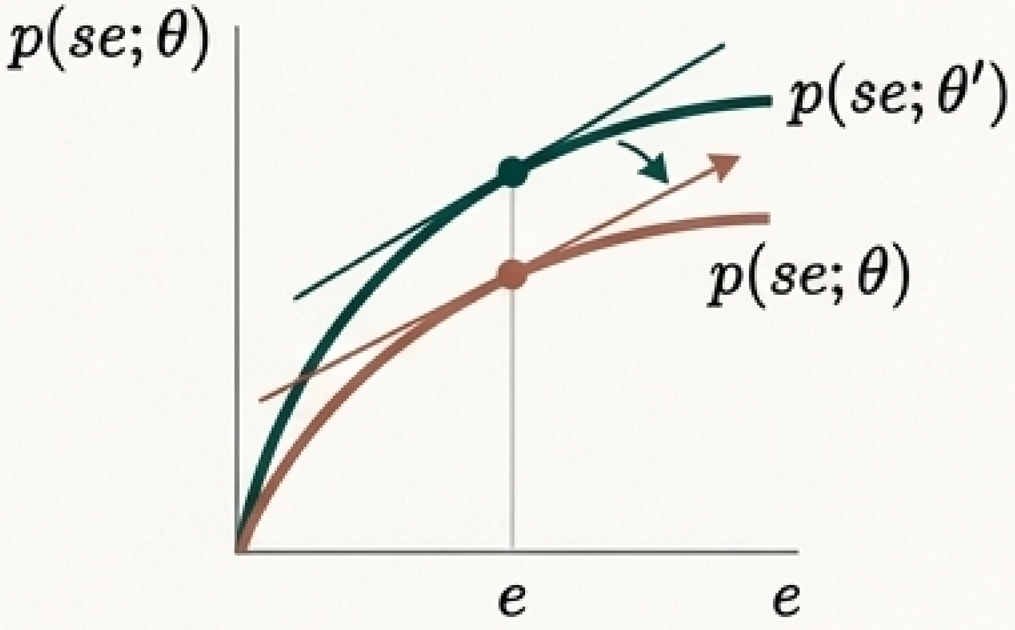


2. Effort Substitution

The marginal product of effort is decreasing in tool quality:

$$p'(se;\theta) \text{ is decreasing in } \theta \text{ for all } s, e > 0.$$

Intuition: As the tool gets better, it reduces the marginal benefit of additional human effort. The tool is doing more of the work.



Three Microfoundations for Cognitive Tools

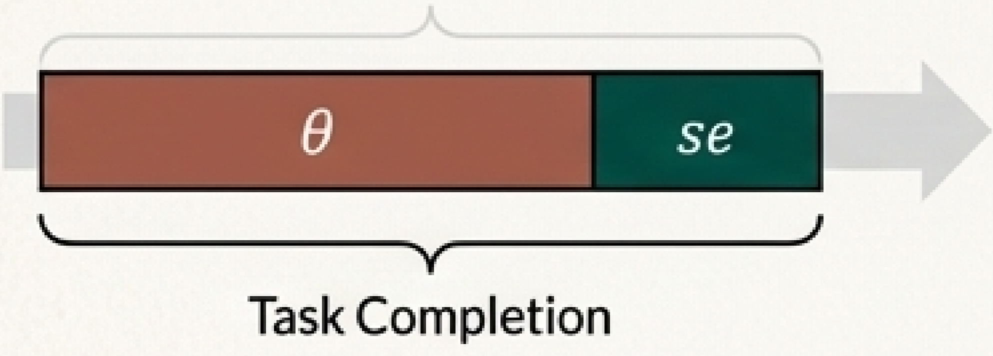
The formal definition is motivated by several distinct ways tools can function:

1. The "Head Start" Model (Additive Input)

Concept: The tool provides an initial input, θ, to which human effort is added. Success is $p(se + \theta)$.

Mechanism: Due to diminishing returns (concave p), the tool's input reduces the marginal effectiveness of subsequent human effort.

Example: A generative AI providing a first draft.

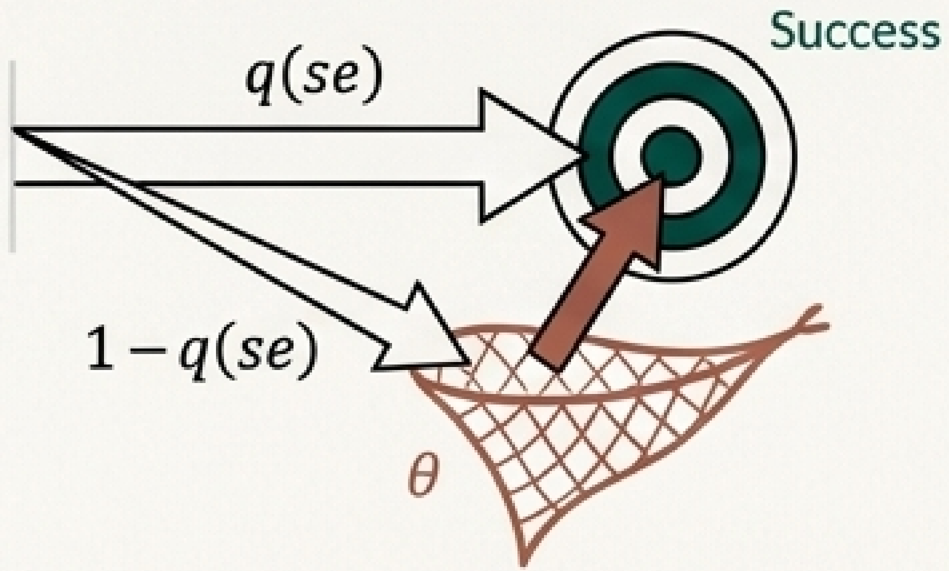


2. The "Safety Net" Model (Error Correction)

Concept: The agent succeeds with probability $q(se)$. If they fail, the tool corrects the error with probability θ.

Mechanism: As the safety net θ improves, the expected cost of human error falls, reducing the marginal value of exerting effort to avoid failure.

Example: An automated spell checker or code testing suite.

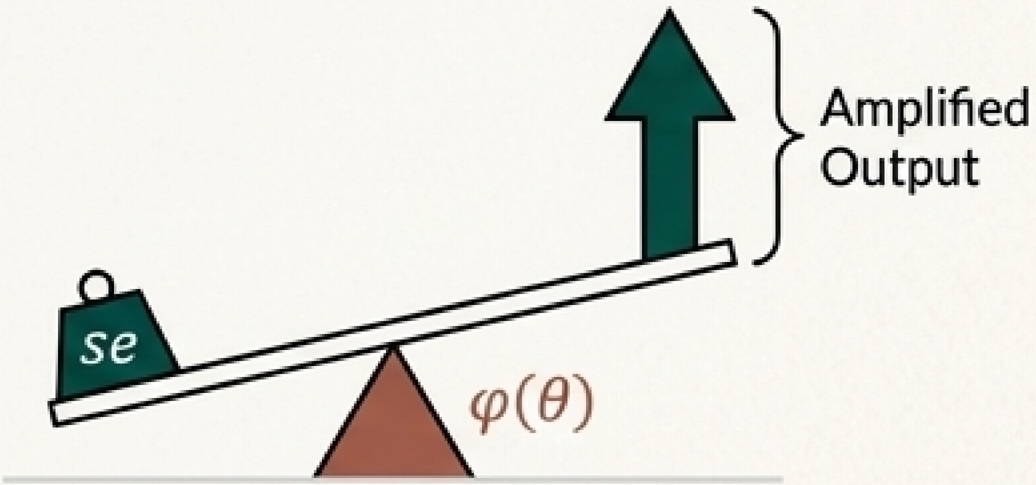


3. The "Leverage" Model (Effort Amplifier)

Concept: The tool amplifies the effectiveness of effort. Effective input is $\varphi(\theta)se$.

Mechanism: Even with amplification, diminishing returns in the production function $p(x)$ can cause the marginal return to raw effort e to fall.

Example: Advanced search tools that help target effort.



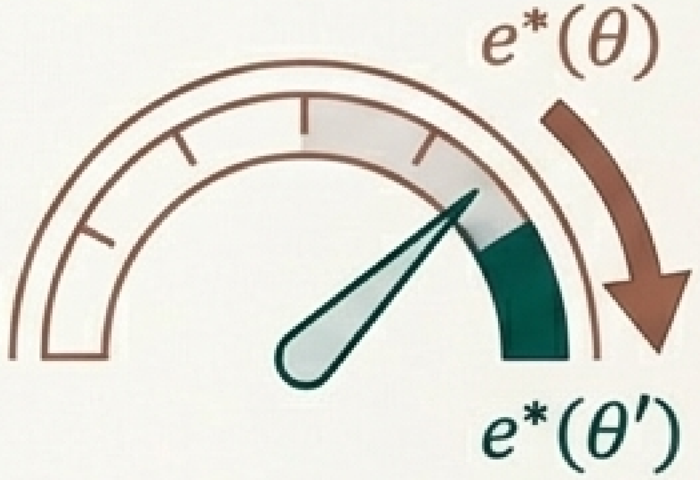
Proposition 1: Better Tools Lead to Less Effort but More Value

For an improvement in tool quality from θ to $\theta' > \theta$:

Effort Reduction

$$e^*(\theta') \leq e^*(\theta)$$

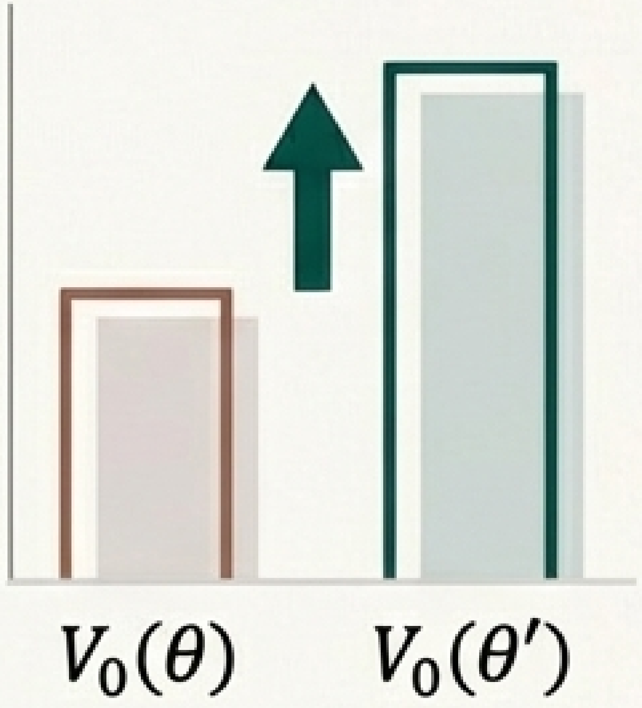
Reason: The tool substitutes for human effort at the margin.



Value Enhancement

$$V_0(\theta') \geq V_0(\theta)$$

Reason: The tool's direct productivity gain outweighs the effect of reduced human effort.



Key Takeaway: "Workers do less but produce more." This corresponds to documented productivity gains from computers and AI.

Proposition 2: Who Gains Most From a Better Tool?

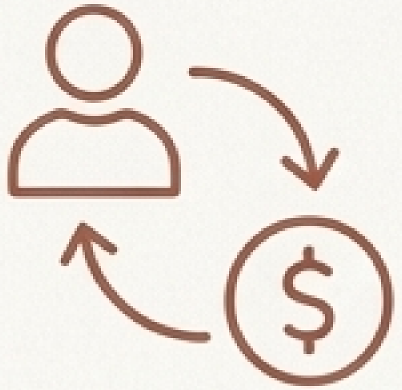
The value gain from adopting a better tool, $V_0(1) - V_0(0)$, is:

- **Non-decreasing in Opportunity Judgment (Γ):** Agents who see more opportunities can apply the tool's benefits more often.
- **Non-decreasing in Payoff Judgment (α):** (If the tool's direct effect dominates effort reduction). Agents who are better at realizing value gain more from a higher probability of success.

- **Weakly DECREASING in Implementation Skill (s):** This is the "**Inverse Skill Bias.**" The tool is a substitute for implementation skill, so it provides the least benefit to those who are already highly skilled at implementation.

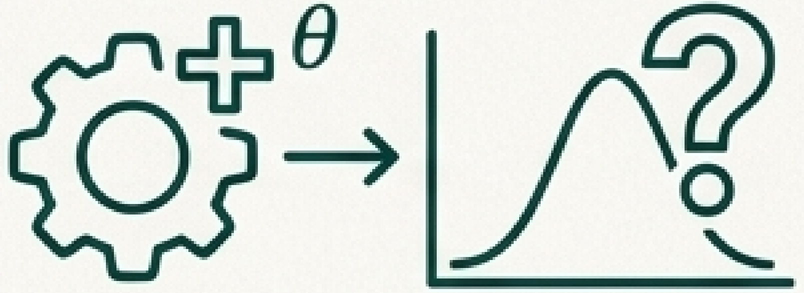
From Individual Productivity to Wage Inequality

The Link



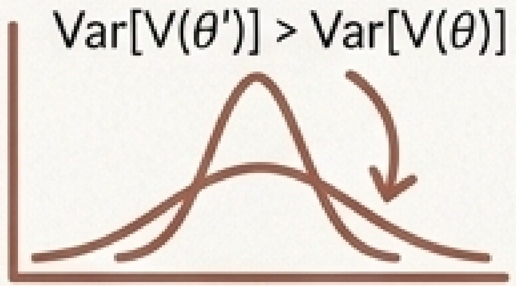
In labor markets, wages reflect marginal productivity. In our framework, this corresponds to the expected value, $V(\theta)$, generated by an agent.

The Question



How does improving a cognitive tool (increasing θ) affect the *distribution* of $V(\theta)$ across a population of agents with heterogeneous skills (α, s, Γ)?

The Analysis



We examine the variance of wages, $\mathrm{Var}[V(\theta)]$, and how it changes as tool quality θ improves.

We will analyze $\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{Var}[V(\theta)]$.

Proposition 3: The Forces Driving Inequality

$$\frac{d}{d\theta}\mathrm{Var}[V] = 2\mathrm{E}[\Gamma^2]\,\mathrm{Cov}(M_*, g) + 2\mathrm{Var}(\Gamma)\mathrm{E}[M_*]\mathrm{E}[g]$$

(where $g = \partial M_* / \partial \theta$ is the marginal gain from the tool)

Cerm 1: The Covariance Effect

$\mathrm{Cov}(M_*, g)$ captures the tension between substitution and complementarity.

Inverse Skill Bias

High implementation skill (s) increases baseline productivity (M*) but *decreases* the marginal gain from the tool (g). This makes the covariance **negative**, compressing inequality.

Judgment Complementarity

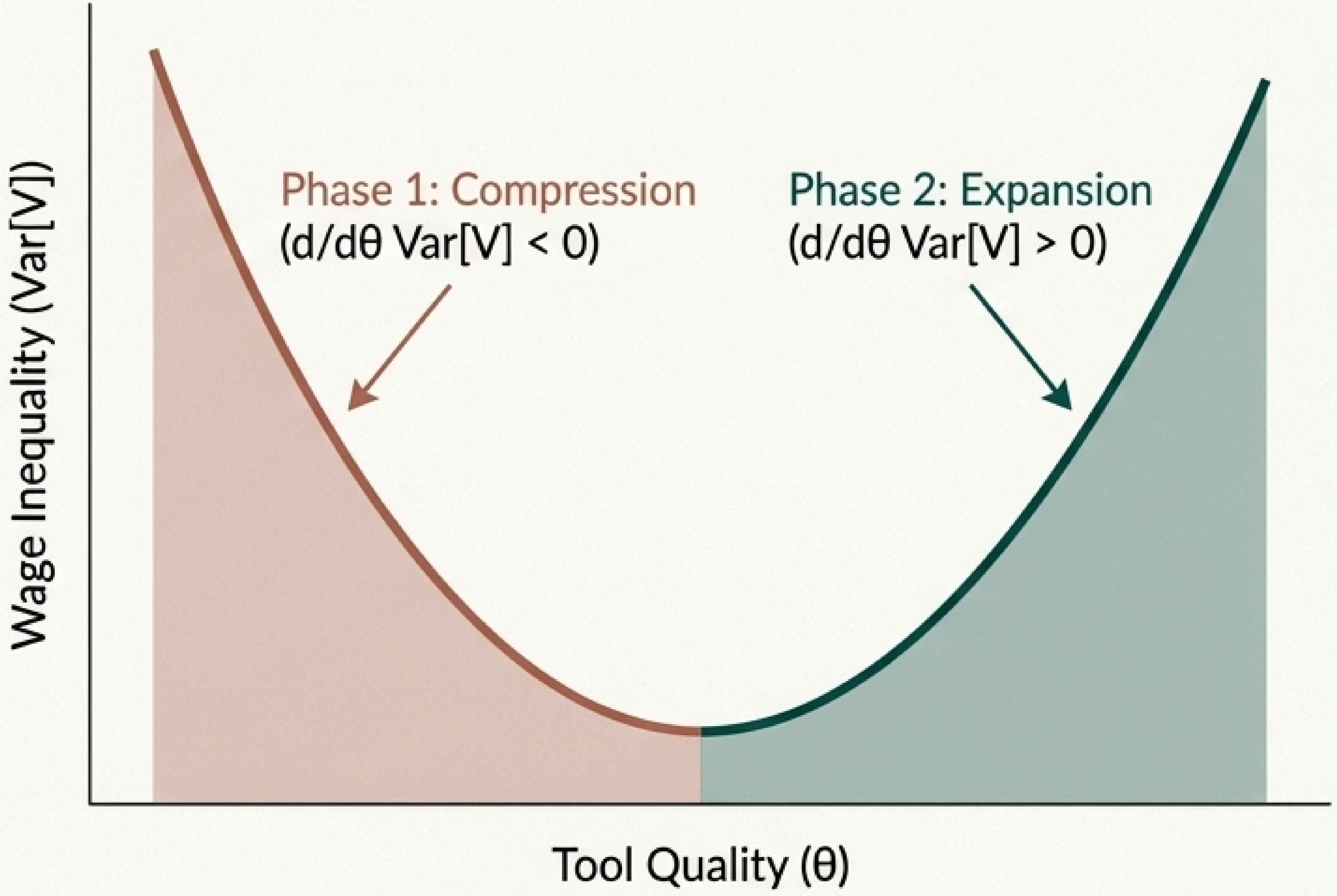
High judgment (α, Γ) increases both M* and g. This makes the covariance **positive**, widening inequality.

Term 2: Opportunity Judgment Amplification

High-Γ agents use the tool more often, multiplicatively scaling up any gains and amplifying baseline differences. This term is always **positive**, widening inequality.

The Predicted U-Shaped Trajectory for AI and Inequality

The Dynamic: The balance of forces changes as tool quality (θ) improves.



For low θ, implementation is the main bottleneck. The tool's primary effect is substituting for implementation skill. The **Inverse Skill Bias** dominates, $\mathrm{Cov}(M^*, g)$ is negative, and inequality falls.

For high θ, implementation becomes largely automated and is no longer a binding constraint. Productivity becomes dominated by heterogeneity in judgment. The **Judgment Complementarity** effect takes over, $\mathrm{Cov}(M^*, g)$ becomes positive, and inequality rises again.

An Illustrative Example: Square-Root Production

Assumptions

- Production Function: $p(se;\theta) = \sqrt{(se+\theta)}$
- Cost Function: $c(e) = e$

Resulting Value Function

The agent's problem yields two regimes:

Augmentation Regime (High s)

- Agent exerts positive effort.

$$M^* = \frac{\alpha^2\Delta^2}{4}\cancel{s} + \frac{\theta}{s}$$

Explicitly shows the inverse skill bias.

Automation Regime (Low s)

- Agent exerts zero effort.

$$M^* = \alpha\Delta\sqrt{\theta}$$

Implementation skill s drops out entirely. The tool creates a productivity floor, rendering skill differences irrelevant.

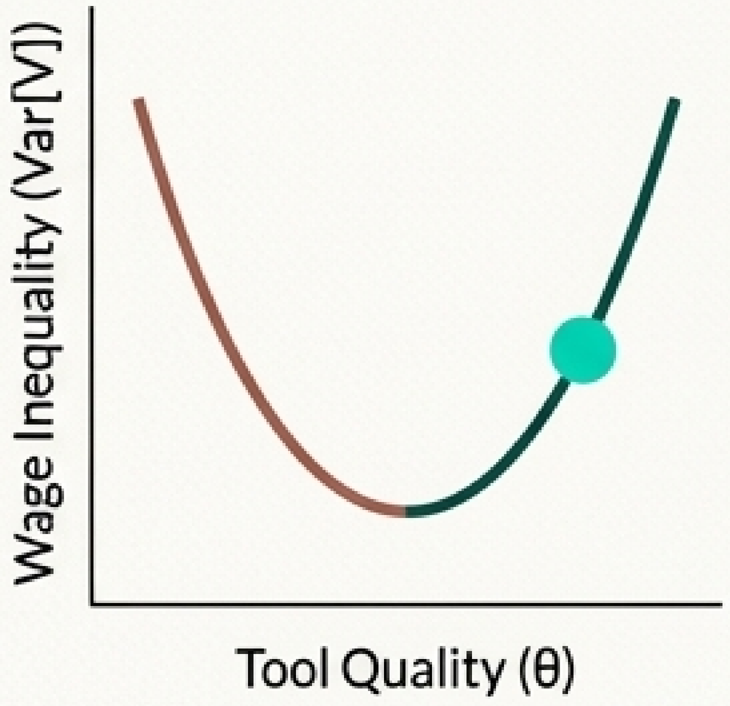
Conclusion

The math confirms the intuition. The $1/s$ term drives compression, but as θ grows large, the Automation Regime becomes more common and differences in judgment (α, Γ) re-emerge as the main drivers of inequality.

Resolving the Puzzle

The model provides a unified explanation for the divergent empirical findings.

Why did computers widen inequality?

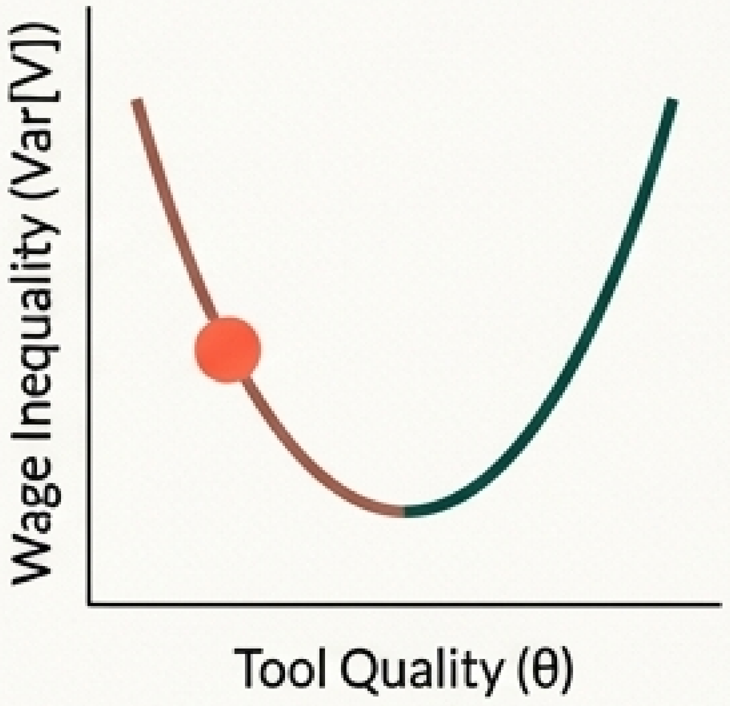


Computer adoption occurred in an environment where the variance of non-routine **judgment skills** was large relative to the variance of routine **implementation skills**.

The **judgment complementarity effect** dominated.

The economy was on the ***upward-sloping*** part of the U-curve.

Why is early AI compressing inequality?



Current AI tools are being deployed in domains (call centers, coding) where there is substantial variance in **implementation skill** and experience.

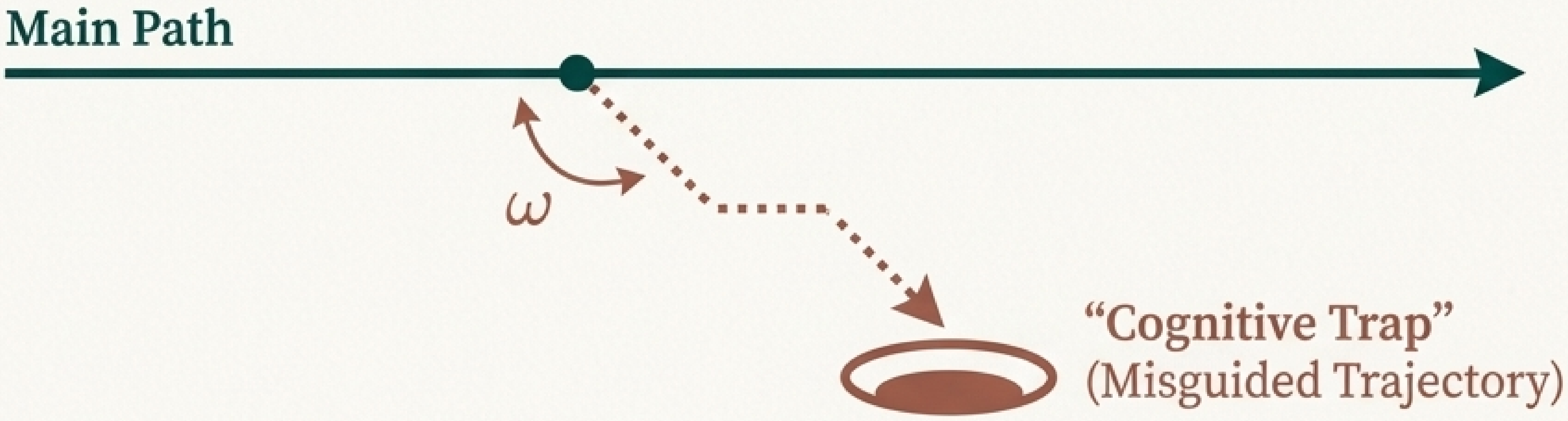
The **inverse skill bias** (substitution) is the dominant force.

The economy is on the ***downward-sloping*** part of the U-curve for these specific tasks.

Key Point: This compression may be temporary. As AI advances, judgment may again become the scarce factor.

Extension 1: Oversight Judgment and “Cognitive Traps”

The Concept: The baseline model assumes agents improve along a given path. But what if the entire path is wrong?



Oversight Judgment (ω): A third type of judgment; the ability to recognize that an entire trajectory is misguided and switch to a better path.

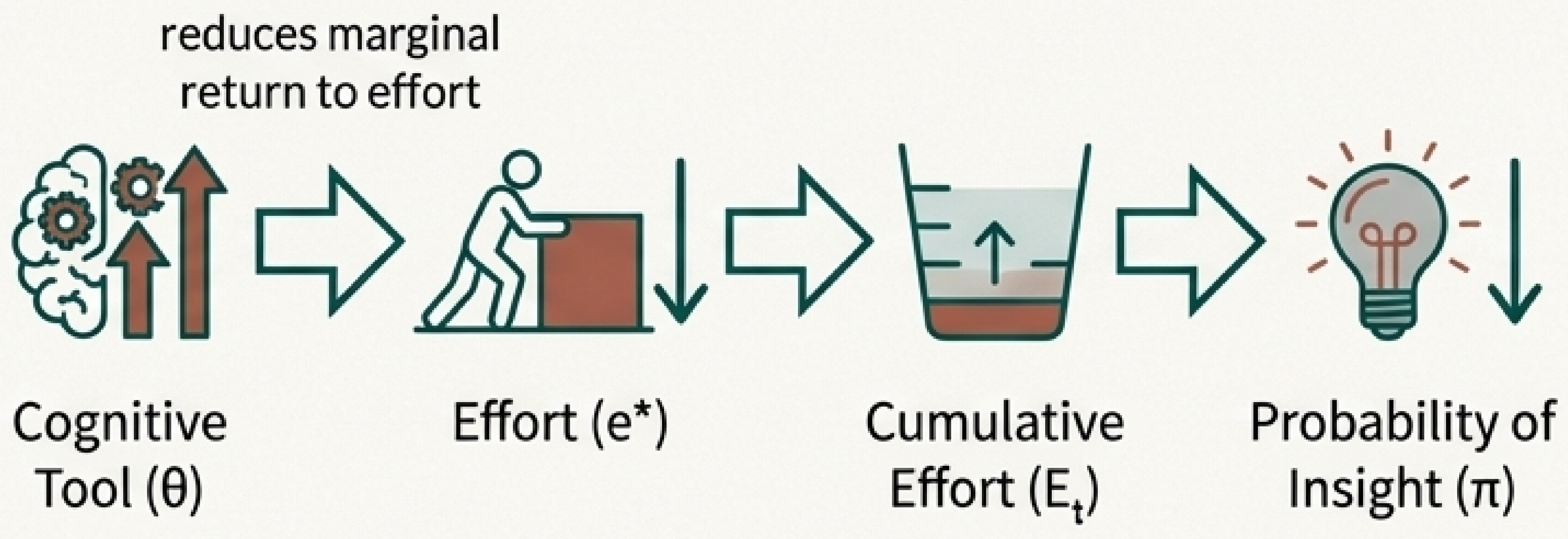
The Key Assumption: Insight is not free. It arises from deep engagement and ‘struggle’ with the task, which we model as being dependent on cumulative implementation effort (E_t).

$$\text{Insight Hazard Rate} = \pi(E_t, \omega)$$

The Insight Inhibition Effect

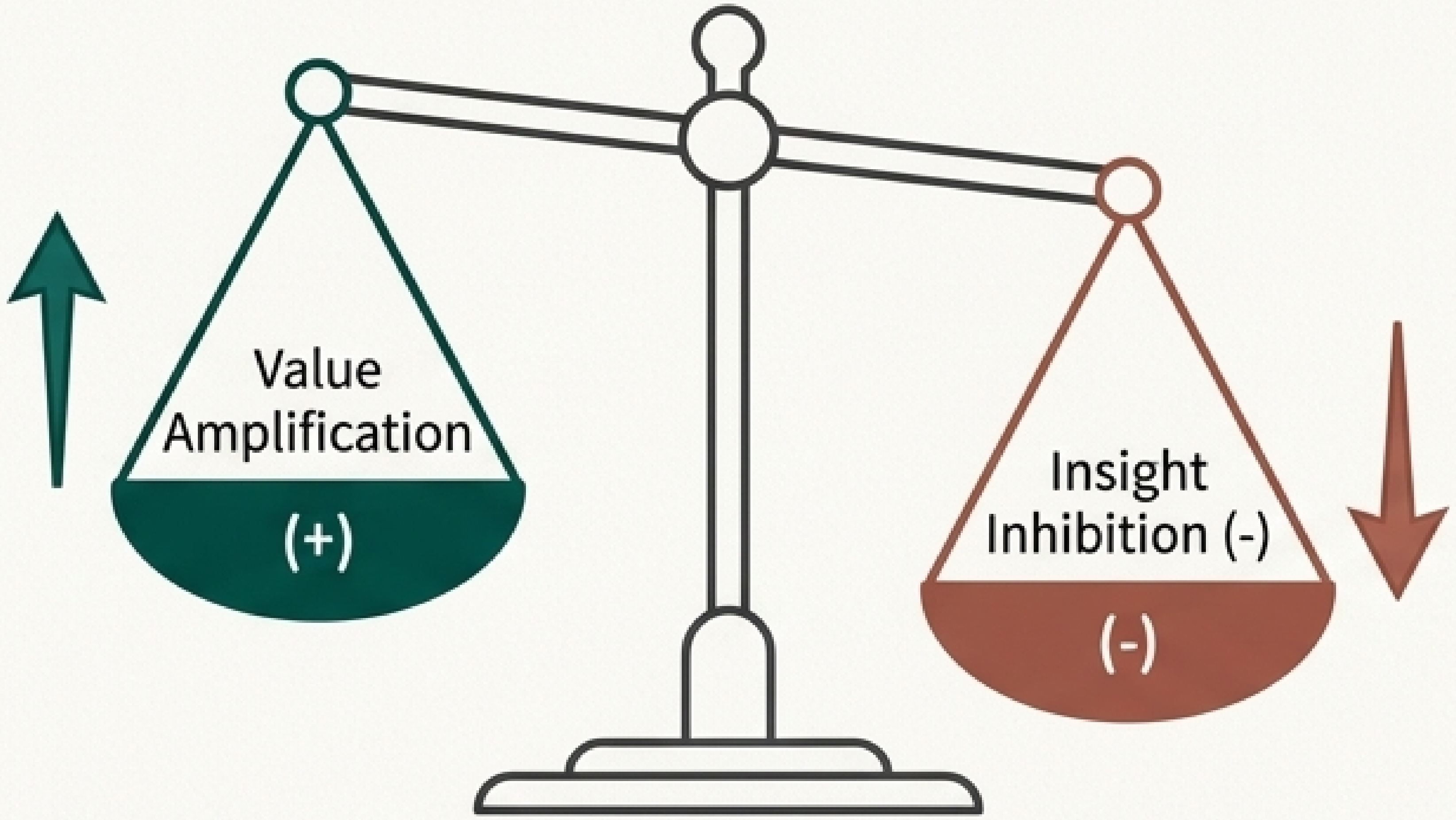
The Mechanism (Proposition 4)

1. Better cognitive tools (higher θ) **reduce optimal effort** ($e^*\downarrow$) because they substitute for it (from Prop 1).
2. Lower per-period effort leads to **slower accumulation of cumulative effort** (E_t).
3. Since insight depends on cumulative effort (π is increasing in E_t), better tools **reduce the probability of gaining strategic insight.**



The 'Cognitive Trap': By making execution easier, cognitive tools can reduce the deep engagement required to identify strategic errors. Agents become highly efficient at executing a flawed plan.

Inverse Oversight Bias: A New Equalizing Force



The Result (Proposition 5)

The relationship between tool quality (θ) and oversight judgment (ω) is ambiguous, balancing two forces:

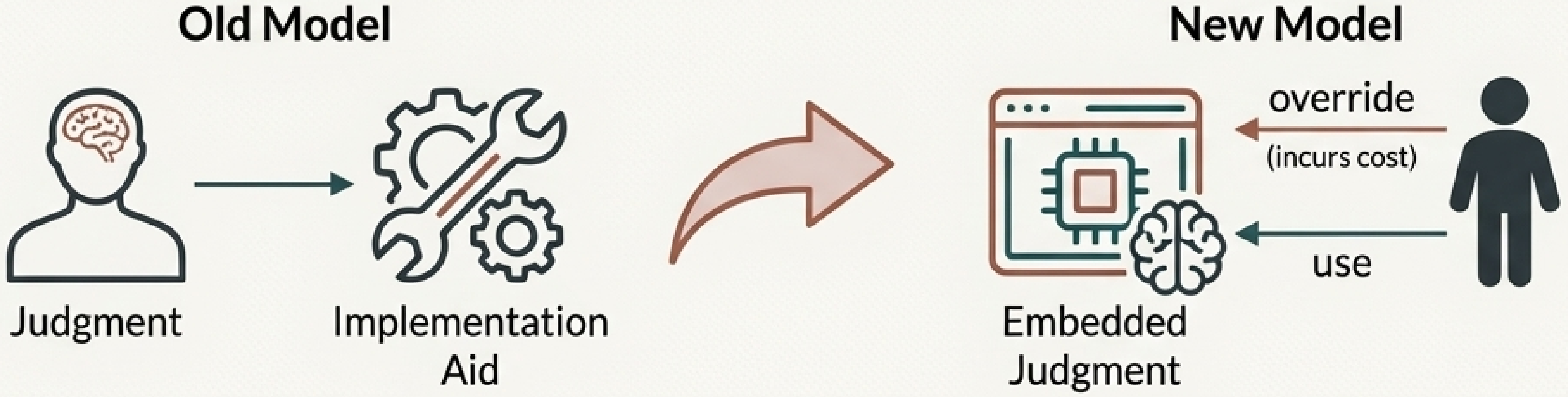
- **Value Amplification (+)**: Better tools increase the gains from being on the right path, amplifying the value of insight.
- **Insight Inhibition (-)**: The reduction in effort disproportionately harms high-ω agents, who benefit most from the learning that effort provides.

Inverse Oversight Bias: If Insight Inhibition dominates, the tool and oversight judgment become **substitutes**. The tool's marginal value is *lower* for agents with higher oversight judgment.

Implication for Inequality (Proposition 6)

This acts as a **new equalizing force**, potentially mitigating the long-run rise in judgment-based inequality predicted by the baseline model.

Extension 2: When the Tool Itself Makes Judgments



The Concept So far, the tool only aids implementation. But sophisticated AI can **embed judgment** directly.

Mechanism The tool suggests actions or filters opportunities based on judgment calls made by its designers and codified into the system.

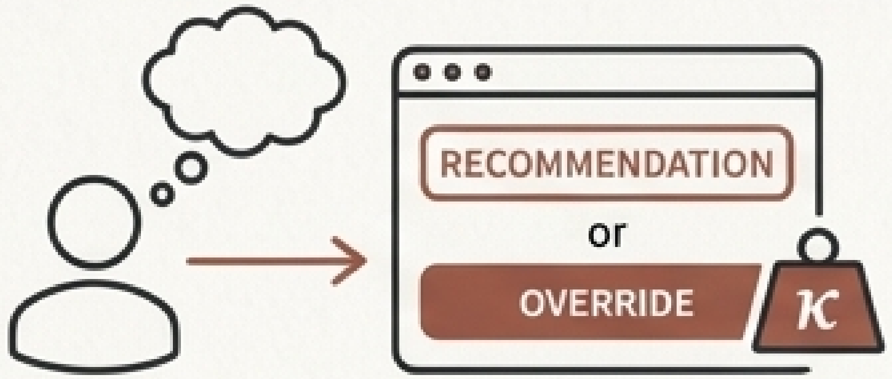
The Trade-off

- ☞ **Low-Judgment Users:** Can delegate to the tool, benefiting from judgment superior to their own and raising their productivity floor.
- ☞ **High-Judgment Users:** Face a choice: accept the standardized recommendation or incur a cost (cognitive or otherwise) to override it and exercise their own superior judgment.

Three Channels for Embedded Judgment and Their Effects

The paper models **three** ways judgment can be embedded:

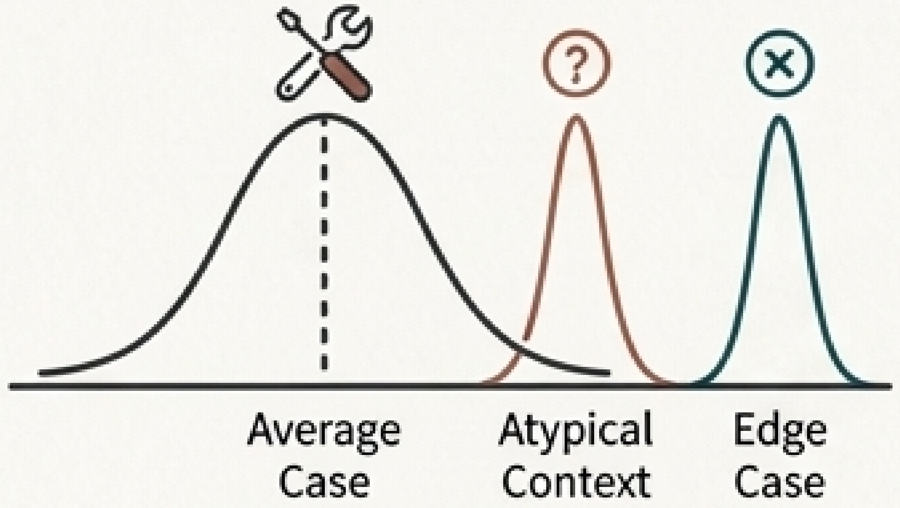
Cognitive Cost of Overriding



Users must pay a mental cost κ to ignore the tool's payoff judgment recommendation.

Creates a threshold; **low-α** users **delegate**.

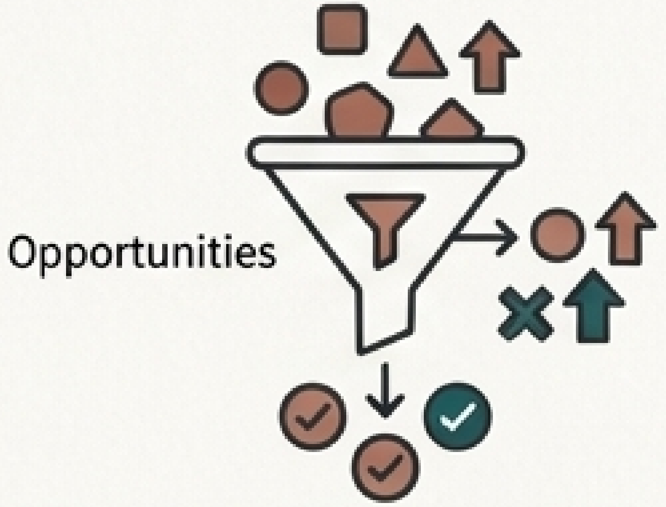
Contextual Mismatch



The tool's judgment is optimized for the average case and performs poorly in atypical contexts.

High-α users override in a **wider** range of '**edge cases**."

False-Positive Filtering



The tool screens opportunities, eliminating bad ones but also some high-risk, high-reward ones.

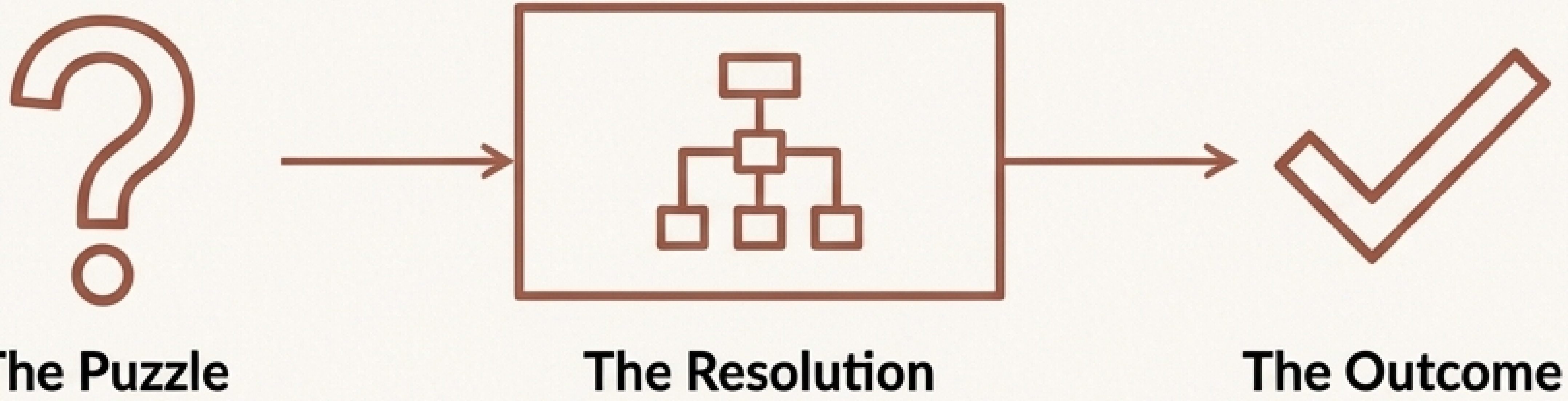
High-γ users prefer to screen manually to **preserve upside**.

Consistent Outcome:

Across all mechanisms, embedded judgment tends to **compress inequality** in **routine contexts** by **raising the floor for low-judgment** users. However, it can **preserve** or **widen gaps** in **atypical contexts** where **high-judgment** users retain their advantage.

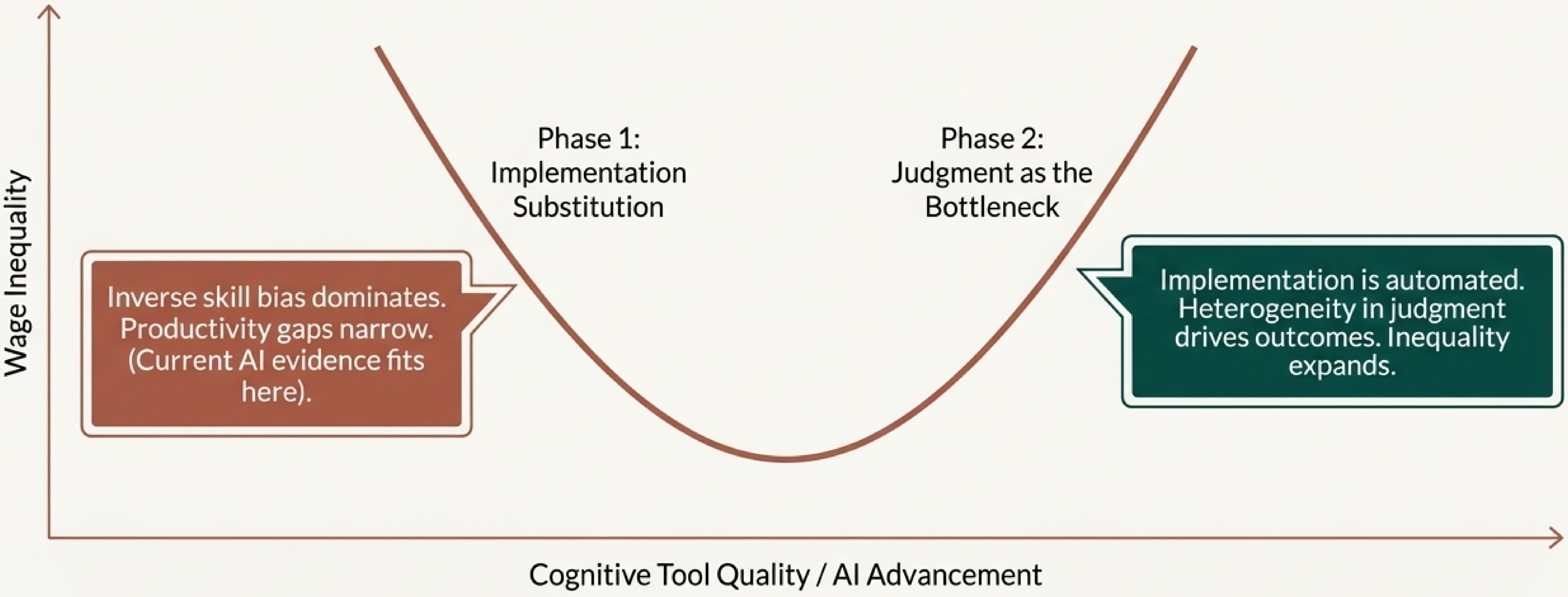
A Unified Framework for a Complex Reality

- **The Puzzle**: Seemingly contradictory effects of cognitive tools (computers vs. AI) on wage inequality.
- **The Mechanism**: The effect of any tool depends on the balance between **substitution** for implementation skill (compressing) and **complementarity** with judgment (expanding).



- **The Resolution**: A single model that distinguishes between implementation skill and human judgment can explain these divergent outcomes.
- **The Implication**: The empirical patterns we observe reflect different points on the same underlying technological trajectory, driven by the relative variance of skills in the affected populations.

AI and the Future of Inequality: A U-Shaped Path



The current compression of productivity differences by AI may be the first phase of a dynamic that will eventually reverse as judgment becomes the truly scarce economic factor.

As Cognitive Tools Improve, the Economic Importance of Human Judgment Rises, Not Falls

The automation of implementation does not devalue human contribution; it shifts the locus of value creation. The most critical skills in an age of advanced AI will be the uniquely human abilities to recognize new opportunities, understand complex trade-offs, and provide strategic oversight.