Allocative Skill

Andrew Caplin, New York University and NBER
David J. Deming, Harvard University and NBER
Soren Leth-Petersen, University of Copenhagen
Ben Weidmann, Harvard University*

September 2023

Abstract

Jobs increasingly require good decision-making. Workers are valued not only for how much they can do, but also for their ability to decide what to do. In this paper we develop a theory and measurement paradigm for assessing individual variation in the ability to make good decisions about resource allocation, which we call allocative skill. We begin with a model where agents strategically acquire information about factor productivity under time and effort constraints. Conditional on such constraints, agents' allocative skill can be defined as the marginal product of their attention. We test our model in a field survey where participants act as managers assigning fictional workers with heterogeneous productivity schedules to job tasks and are paid in proportion to output. Allocative skill strongly predicts full-time labor earnings, even conditional on IQ, numeracy, and education, and the return to allocative skill is greater in decision-intensive occupations.

*Corresponding author email: david_deming@harvard.edu. This paper subsumes some prior results from Deming (2021). We thank the Sloan Foundation for supporting this research under the "Cognitive Economics at Work" grant. Deming and Weidmann gratefully acknowledge funding from Schmidt Futures for the Harvard Skills Lab. Caplin and Leth-Petersen thank the NOMIS Foundation for supporting this research under the Firm and Worker Adaptation to Unexpected Events grant and the Center for Economic Behavior and Inequality (CEBI), a center of excellence at the University of Copenhagen financed by grant DNRF134 from the Danish National Research Foundation. We thank Daron Acemoglu, Max Bazerman, Tomas Jagelka, Jesse Shapiro, and Michael Woodford for insightful comments. We also thank Josh Benhamou for outstanding technical work on the experimental platform. All errors are our own.

1 Introduction

Most jobs require decision-making, in part because routine physical and information processing tasks are increasingly automated (e.g. Autor et al. 2003, Deming 2021).[1] Workers are valued not only for *how much* they can do, but also for their ability to *decide what to do*. Human capital theory traditionally emphasizes productive efficiency, in which people with more skill or education produce more output per unit of time (Mincer 1958, Becker 1962, Mincer 1974). Yet firms invest in managerial talent and emphasize problem-solving as the most desirable quality in new hires, suggesting that they also greatly value decision-making skills (Welch 1970, NACE 2020).

Figure 1 shows the rising importance of decision-making across the entire U.S. economy, using job vacancy data to measure employer skill demands.[2] We form a consistent definition of decision-making over time using data from Atalay et al. (2020) and from Burning Glass Technologies (BGT), covering the 1960-2000 and 2007-2019 periods respectively.[3] The share of all jobs requiring decision-making increased from 6 percent in 1960 to 34 percent in 2018, with nearly half of the increase occurring since 2007.[4]

Figure 2 presents a scatterplot of average wage and salary income against decision in-

[1]In settings ranging from retail banks to manufacturing plants, firms that automate routine tasks also delegate more decision-making authority to employees (Autor et al. 2002, Bresnahan et al. 2002, Bartel et al. 2007).

[2]Atalay et al. (2020) collect the text of classified ads placed in the *New York Times,* the *Wall Street Journal,* and the *Boston Globe* and map them to work activities from the Occupational Information Network (ONET) data, among other measures. We use their keyword mapping to the three ONET work activities Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work - see the appendix to Atalay et al. (2020) for details. BGT classify vacancy text into thousands of unique job skills, and we use job skills (and related strings) with the key words and phrases above to create a consistent definition over time.

[3]To ensure representativeness, we weight the job ad data by the actual distribution of occupations in each year. To reduce classification error from narrowly defined occupations (some of which only exist in certain years of the data), we aggregate occupations to the 3 digit SOC level using occupation crosswalks and compute weights using Census and American Community Survey (ACS) data to make the job vacancy data representative of the actual occupation distribution in each year. We then apply a 5-year moving average of the share of ads requiring decision-making, to account for gaps between Census years and to reduce noise.

[4]The grey lines show the same trend but controlling for occupation fixed effects, which diminishes the impact only slightly, implying that most of the shift toward decision-making is occurring within rather than between occupations. Excluding management occupations diminishes the growth only slightly, suggesting that growing demand for decision-making is an economy-wide phenomenon.

tensity, with labels attached to selected occupations. We measure decision intensity using data from the Occupational Information Network (ONET), a survey administered by the U.S. Department of Labor to a random sample of U.S. workers in each occupation.[5] Decision intensity is strongly correlated with average earnings. Not surprisingly, managers are among the most decision-intensive occupations, as well as scientists, engineers, doctors and lawyers.[6] Personal services and clerical occupations are the least decision-intensive. Occupations that pay above median wages but have low decision intensity include media and communications workers, sales representatives, and production supervisors. Occupations with above average decision intensity that pay below median wages include counselors, social workers and K-12 teachers.[7]

This paper develops a theory and measurement paradigm for assessing individual variation in the ability to make good decisions about resource allocation, which we call *allocative skill*. We begin with a simple model where agents assign factors of production to different roles to maximize total output. This could be a manager assigning workers to jobs, or workers allocating their own effort to job tasks. Factors have heterogeneous productivity schedules, so the agent must compare hypothetical assignments and choose the one with the highest expected output. Agents acquire costly information about payoffs to different actions, deploying their attention endogenously (e.g. Sims 2003, Caplin and Dean 2015, Matějka and

[5]Our measure of decision-making intensity is a simple average of the three work activities also used in Figure 1 - making decisions and solving problems, developing objectives and strategies, and planning and prioritizing work. Since the raw ONET values have no cardinal meaning, we transform the decision intensity variable into a 0-10 scale that reflects each occupation's percentile rank in the labor supply-weighted distribution of employment from the 2018-2019 ACS. We use the "level" variable in ONET, and apply separate scalings for each SOC code disaggregation (6 digit, 5 digit, etc).

[6]However, food services managers and retail store managers are large occupation categories that have below average decision intensity. Appendix Table A1 presents decision intensity, average educational attainment, and earnings for all three digit occupations using the 2018 Standard Occupational Classification (SOC) codes and data from the 2018 and 2019 American Community Survey.

[7]We also measure decision intensity with the job vacancy data from Burning Glass Technologies (BGT) used in Figure 1. Following Deming and Noray (2020), we exclude vacancies with missing employers and employ a pruning algorithm to create unique employer IDs. We then construct a vacancy-level measure of decision-making intensity by creating an indicator variable that is equal to one if the vacancy includes one of the several key words or phrases that relate to decision-making. This generally yields very similar results, so we use the ONET in the analyses below to increase transparency and replicability. The labor supply-weighted occupation-level correlation between the ONET and BGT measures of decision-making intensity is 0.83.

McKay 2015, Maćkowiak et al. 2023).

We define and estimate individual-specific attention costs using an approach that is analogous to the role of input costs in standard production theory. In a competitive labor market, workers with higher earnings per unit of time (e.g. wages) have a higher marginal product of labor. In our model, agents with higher levels of allocative skill choose more efficient allocations, holding constant information complexity and time constraints. We thus define allocative skill as the *marginal product of attention*. Allocative skill captures total processing bandwidth, but also the ability to strategically pay attention to important information and to understand comparative advantage.

We measure allocative skill by creating a novel task we call the Assignment Game. Participants are managers who assign fictional workers to jobs to maximize output. They observe multiple draws from workers' productivity schedules over tasks and then make an assignment. Participants are scored based on each worker's mean output in the task to which they were assigned. The Assignment Game requires participants to process information quickly and to assign workers to their highest value task given the skills of the others. Since we know the efficient assignment, we can measure allocative efficiency precisely for each participant.

We test the implications of our model in a sample of more than one thousand full-time U.S. workers ages 25-55. Participants were recruited on the research platform Prolific and paid for performance on several tests of cognitive skill, including the Assignment Game. Partcipants also completed a demographic and labor market survey, which elicited information about current income and occupation. The survey sample is more educated than average but otherwise fairly representative, and we reweight results from the survey to match the distribution of U.S. full-time employed workers.

We find that allocative skill is strongly associated with income, even after controlling for IQ, numeracy, education, occupation, and other covariates. Participants with one standard deviation higher allocative skill have 7 percent higher earnings even after controlling for a rich set of covariates. The magnitude of the association between allocative skill and income is nearly twice as large as the association with IQ, and is the largest of the four cognitive assessments we use when all are included in the same model.[8]

We also find that the association between allocative skill and income is significantly greater in decision-intensive occupations, which we define using task data from the 2019 Occupational Information Network (ONET) as in Figures 1 and 2 above. For managerial and professional jobs at the 75th percentile of decision intensity and above, a one standard deviation increase in allocative skill increases earnings by more than 10 percent. In contrast, allocative skill is unrelated to income in jobs at the 25th percentile of decision intensity and below. We also find that participants with higher allocative skill are slightly more likely to work in decision-intensive occupations.

Our paper contributes to human capital theory by formalizing and testing the value of allocative efficiency in the labor market. Modeling allocative skill requires us to treat attention as a scarce resource that some people deploy more effectively than others (e.g. DellaVigna 2009, Bordalo et al. 2012, Gabaix 2019). As a consequence, we must use the tools of information theory to measure individual differences in labor productivity. Our measure of allocative skill requires both raw information processing capacity and an ability to use information strategically by exploiting comparative advantage. Since the value of allocative skill is grounded in economic theory and it strongly predicts economic success even conditional on IQ and other measures, the Assignment Game can be viewed as an "economic IQ" test.

A large literature explores how management practices and managerial decision-making affect firm productivity (e.g. Bloom and Van Reenen 2010, Bandiera et al. 2020, Bertrand and Schoar 2003, Hoffmann et al. 2020, Minni 2022, Adhvaryu et al. 2022, 2023, Metcalfe et al. 2023). This paper contributes to the management literature by focusing on and empirically measuring one aspect of management skill, the ability to assign factors of production their

[8] Allocative skill is positively correlated with nonverbal IQ ($\rho = 0.38$), numeracy ($\rho = 0.31$), and cognitive reflection ($\rho = 0.29$). It is positively correlated with having a bachelor's degree ($\rho = 0.11$), negatively correlated with age ($\rho = -0.13$).

best use. There are many other aspects of being a good manager, such as social skills, leadership, and other factors that are not captured by our approach (e.g. Deming 2017, Hansen et al. 2021).

An older literature in economics studies "allocative ability", with a particular focus on technology adoption and decision-making in agriculture (e.g. Nelson and Phelps 1966, Welch 1970, Huffman 1977). Standard competitive theory rules out allocative ability as an important driver of outcomes through the assumption of perfect information, yet evidence of allocative inefficiency - or "X-inefficiency" - is everywhere (Leibenstein 1966, Stigler 1976). A few other papers study decision-making skill in healthcare, where doctors vary in both procedural skill and diagnostic skill (Currie and MacLeod 2017, Chan Jr et al. 2019, Chandra and Staiger 2020). Goldfarb and Xiao (2011) and Hortaçsu et al. (2019) find that education and other proxies for skill improve managerial decision-making. Our results are also related to the literature on misallocation as a constraint on economic growth (e.g. Restuccia and Rogerson 2017).

There is a large literature in psychology on the determinants of effective decision-making, including some evidence that general intelligence and numeracy in particular predict effective decision-making among those who are not domain experts (e.g. Baron 2000, Stanovich and West 2000, Cokely et al. 2018). Finally, there is growing evidence that cognitive ability improves learning in new environments and reduces behavioral biases and decision errors, which provides supporting evidence for the idea that decision-making is a skill (e.g. Dohmen et al. 2010, Benjamin et al. 2013, Rustichini 2015, Gill and Prowse 2016).

The paper proceeds as follows. Section 2 develops the model. Section 3 describes the survey data and measurement constructs. Section 4 presents the results, and Section 5 concludes.

2 Model

We consider the problem of a risk-neutral agent assigning factors of production to different tasks to maximize total output. The agent could be a manager assigning workers to jobs, or a worker allocating their own effort over job tasks. The agent's problem is to choose the optimal assignment when factor productivity varies over roles and information about productivity is costly to observe.

A finite set of M factors, indexed by $1 \leq m \leq M$, must be assigned to M tasks, indexed by $1 \leq n \leq M$. We assume a 1:1 mapping of factors to tasks only for simplicity; this can be relaxed. An *assignment* is a one-to-one onto function $a : \{1, ..., M\} \to \{1, ..., M\}$, with $a(m)$ denoting the task to which factor m is assigned and $a^{-1}(n) \in \{1, ..., M\}$ denoting the factor assigned to task n.

Factors have heterogeneous productivity over tasks. There is a finite set of possible productivity types, where a type specifies the potential output of factor m in all M tasks. The state $\omega \in \Omega$ specifies all factor productivities in all tasks. We denote productivity type of factor m in state ω as $\omega(m) = (\omega_1(m), ..., \omega_M(m)) \in \mathit{\Omega}^M$, and $\omega_n(m)$ is factor m's productivity type in task n.

Since there is uncertainty about productivity types and assignments, an expected production function $\mathcal{Y}$ maps task levels measured as real numbers $y_n \geq 0$ into output, e.g. $\mathcal{Y} : \mathbb{R}^M_+ \to \mathbb{R}$ with $\mathcal{Y}$ increasing in all its arguments. In principle this could the production function for a firm rather than for an individual agent.

Putting this all together, we can compute the agent's expected output for any assignment of factors $a \in A$ in any state $\omega \in \Omega$ as:

$$f(a, \omega) \equiv \mathcal{Y}\left(\omega_1\left(a^{-1}(1)\right), ..., \omega_M\left(a^{-1}(M)\right)\right) \tag{1}$$

2.1 The Marginal Cost of Attention

If worker productivity types are perfectly observed, the optimal assignment of workers to jobs is a linear programming problem that maximizes equation (1) as in Koopmans and Beckmann (1957). Agents simply evaluate all possible assignments and choose the one that maximizes output. In practice, agents face an almost infinite number of possible decisions about how to deploy workers and other factors of production, and acquiring information about productivity is costly in terms of time and attention.

We model the agent's optimal information acquisition strategy when attention is costly, following the rational inattention literature (e.g. Maćkowiak et al. 2023). Agents begin with prior beliefs about productivity types, which we denote as $\mu(\omega)$. They next choose a set of signals (e.g. which workers to monitor and for how long, asking the right questions) that help them refine their beliefs about the state of the world ω. We call this choice of signals the agent's *attention strategy*. After receiving signals, they form a posterior belief $\gamma(\omega)$ and choose an assignment a that maximizes expected output in equation (1) given their beliefs. Since we assume risk neutrality, output refers to expected output.

We now show how to characterize agents' attention costs. First we define the Bayes' consistent distribution of posterior beliefs as $Q \in Q(\mu)$, a function which assigns probabilities to posterior beliefs that average back to the agent's prior beliefs, e.g.

$$\sum_{\gamma} \gamma Q(\gamma) = \mu \tag{2}$$

where $Q(\gamma)$ is the unconditional probability of posterior belief γ (Kamenica and Gentzkow 2011). Define the optimal value of a posterior belief as:

$$\hat{f}(\gamma) = \max_{a \in A} \sum_{\omega} f(a, \omega) \gamma(\omega) \tag{3}$$

and the optimal value of the distribution of posterior beliefs as $\hat{f}(Q) = \sum_{\gamma} Q(\gamma) \hat{f}(\gamma)$.

With those definitions in place, we can define the set of production outputs and attention inputs that are jointly feasible given some prior belief:

$$\mathcal{Y} \equiv \left\{ (x, y) \in \mathbb{R}^2 | \exists Q \in \mathcal{Q}(\mu) \ s.t. \ \hat{f}(Q) \geq y, K(Q) \leq x \right\} \tag{4}$$

where y is the output level, x is the attention input, and $K(Q)$ is an attention cost function.

We call $\mathcal{Y}$ the *attention production set,* because it maps the space of possible outputs the agent with cost function $K(Q)$ can achieve for any fixed amount of attention x. We also define an attention production function $g(x)$, which equals the supremum of output levels in $\mathcal{Y}$ for attention inputs of x or below.

We now consider the impact of a change in the marginal cost of attention. Consider a family of potential attention cost functions $K(Q)$ that can be scaled up or down linearly by some multiple $c > 0$. Conceptually, c indexes the marginal cost of attention and its reciprocal can be interpreted as a technology term that augments attention inputs (e.g. the marginal product of attention).

Define the net value function of Q for $c > 0$ as:

$$V(c, Q) = \hat{f}(Q) - cK(Q) \tag{5}$$

Agents adopt attention strategies that maximize net value in equation (5), achieving $\hat{V}(c) \equiv \sup_Q V(c, Q)$, with $\hat{Q}(c)$ denoting an optimal attention strategy.[9] In the Theory Appendix we provide general conditions under which optimal strategies exist for all $c > 0$ and in which optimal output $\hat{y}(c)$ is weakly decreasing in c and attention inputs $\hat{x}(c)$ are increasing in c.[10]

We can see this most clearly for the broad class of *posterior separable* cost functions, meaning attention costs are additively separable in the agent's production function (Caplin et al. 2022).

[9]Formally, we have $Q \in \hat{Q}(c) \iff V(c, Q) \geq V(c, Q') \, \forall Q' \in \mathcal{Q}(\mu)$.

[10]We consider a broad class of continuous cost functions in which mixed strategies are possible.

Theorem 1: *For any posterior separable cost function of the general form* $K(\mu, Q) = \sum_{\gamma} Q(\gamma) T(\gamma) - T(\mu)$ *for convex bounded and continuous function* $T : \Delta(\Omega) \to \mathbb{R}$, *the attention production set* $\mathcal{Y}$ *is convex and the production function* $g(x)$ *is concave.*

See the Theory Appendix for a proof. Theorem 1 establishes that $\mathcal{Y}$ is a convex set and thus the production function $g(x)$ is concave for a broad class of attention cost functions $K(Q)$. This implies that the output supply curve is upward sloping in the marginal cost of attention c.

Figure 3 presents a visual illustration of how the attention production set $\mathcal{Y}$ and the production function $g(x)$ relate attention to output. The lefthand panel shows the concavity of the production function, with the lowest output arising from an inattentive strategy (e.g. random guessing) and output asymptoting as attention increases. The righthand panel shows the impact of a decrease in the marginal cost of attention c. When c declines from 1 to 0.5, the slope of the tangent line becomes flatter, and the agent optimally pays more attention and produces higher expected output. Of course the precise impact of a change in c depends on the shape of the attention production function $g(x)$.

It is convenient for our purposes to re-express equations (3) and (5) in terms of a set of state-contingent assignment probabilities $P(a \mid \omega) \geq 0$ which must sum to one in each state.[11] We can then write the agent's problem as:

$$V(a, \omega) = \max_{P \in P(A)} \sum_{a} \sum_{\omega} y(a, \omega) P(a \mid \omega) \mu(\omega) - cK(P) \tag{6}$$

subject to the constraints that $P(a \mid \omega) \geq 0$ and $\sum_{a} P(a \mid \omega) = 1$.

In words, the agent develops a joint attention-action strategy - reflected in the term $P(a \mid \omega) \mu(\omega)$ - that maximizes expected output in any possible state, taking into account their prior beliefs and the cost of acquiring information $cK(P)$. The choice of state-contingent

[11] For any cost function in which more informative attention strategies are more costly, acquiring different signals and posterior beliefs that lead to the same action is an inefficient use of costly information (Blackwell 1953, Kamenica and Gentzkow 2011, Maćkowiak et al. 2023). Thus optimal manager behavior requires that there is a unique mapping between the choice of signals and the choice of actions, allowing us to represent these two steps with a single joint distribution $y(a, \omega)$.

actions satisfies Bayesian rationality (e.g. consistency of prior and posterior beliefs) and directly incorporates the optimal attention strategy $\hat{Q}(c)$.

2.2 Allocative Skill

Having defined the marginal cost of attention c, we now study variation in attention costs across agents. We start by adding individual-specific subscripts j to the agent's problem:

$$V_j(a,\omega) = \max_{P_j \in P(A)} \sum_a \sum_\omega y_j(a,\omega) P_j(a \mid \omega) \mu_j(\omega) - c_j K(P_j) \tag{7}$$

where $c_j > 0$ is the agent's marginal cost of attention.[12] We refer to the inverse of c_j as *allocative skill,* $\alpha_j = \frac{1}{c_j}$, which is equivalent to the marginal product of attention. Our main hypothesis is that allocative skill is an individual trait. For example, the righthand panel of Figure 3 could represent optimal output for two individuals with different amounts of allocative skill. All else equal, agents with higher allocative skill will process information more efficiently, flattening the slope of the tangency with $\mathcal{Y}$ and achieving higher optimal output.

Measuring agents' allocative skill requires us to hold fixed several key aspects of the decision problem. First, we must define the set of possible assignments ($a \in A$). In the real world, agents face a near-infinite set of choices about what actions to undertake. Thus it is important to choose a context where the set of possible assignments can be clearly enumerated. Second, agents' output should map cleanly to their utility, so that we can compare performance differences across individuals. Third, we must account for agents' prior beliefs, so that differences in output can be attributed to differences in c rather than differences in preexisting knowledge about the setting.

After accounting for the set of possible actions, the mapping between output and utility, and agents' prior beliefs, the only remaining variation across individuals in (7) is the state-

[12] And where the cost of P_j can be computed from $K(Q)$ as the least Blackwell informative distribution of posteriors that can produce P_j.

dependent assignment probabilities $P_j(a \mid \omega)$ and the marginal cost of attention c_j (with allocative skill α_j as its inverse). Thus we can identify differences in agents' allocative skill by observing $P_j(a \mid \omega)$ and the associated output $V_j(a, \omega)$.

2.3 Solving the Model

The results above establish that we can identify ordinal rank differences in c_j and α_j across individuals, using data from an appropriately specified assignment problem. To solve the model analytically, we assume that the cost function takes on the Shannon mutual information form:

$$K_j(c_j, P_j) = c_j I(a; \omega) = c_j \left(\sum_a \left[H(\mu) - H(\gamma^a) \right] \sum_\omega P(a, \omega) \right) \tag{8}$$

with $I(a; \omega)$ defined as the mutual information between actions and states. $H(\mu)$ and $H(\gamma^a)$ are the entropy of the agent's prior and posterior beliefs associated with the various chosen actions respectively, with $H(p) = -\sum_\omega p(\omega) \ln p(\omega)$ so that learning reduces entropy. Mutual information $I(a; \omega)$ represents the amount of information about the state in the action choices and is bounded below at zero when actions are completely uninformative.

Using the Shannon cost function in (8), Matějka and McKay (2015) identify necessary conditions for optimality of the weighted logit form,

$$P_j(a \mid \omega) = \frac{\exp\left(\alpha_j y_j(a, \omega) + \ln P_j(a)\right)}{\sum_{b \in A} \exp\left(\alpha_j y_j(b, \omega) + \ln P_j(b)\right)} \tag{9}$$

where $P_j(a)$ is the agent's unconditional probability of choosing assignment a (Matějka and McKay 2015). These conditions show that rich data on prior beliefs is required to separately identify allocative skill from preexisting differences in available information.

We can overcome this identification challenge if agents have no reason to believe that any particular worker or job task is different from the others - or more formally, if the agent's problem is *symmetric*, following Bucher and Caplin (2021). Symmetry requires that the prior

μ is exchangeable, which means that prior beliefs satisfy:

$$\mu(\omega(1), ..., \omega(M)) = \mu(\omega(\beta(1)), ..., \omega(\beta(M))) \tag{10}$$

for all bijections $\beta : \{1, ..., M\} \rightarrow \{1, ..., M\}$ of workers. Intuitively, symmetry holds when workers are initially seen as equivalent to each other, even if they are revealed to be heterogeneous ex post. The Theory Appendix shows that with this assumption of exchangeability, the symmetry conditions of Bucher and Caplin (2021) are satisfied. That result converts equation (9) into the simple unweighted logit formula:

$$P_j(a \mid \omega) = \frac{\exp(\alpha_j y_j(a, \omega))}{\sum_{b \in a} \exp(\alpha_j y_j(b, \omega))} \tag{11}$$

If $\{\exp(\alpha_j y_j(b, \omega)) : b \in A\}$ are affine independent, then equation (11) is the unique solution to the agent's problem.

Equation (11) relates the agent's observed state-dependent choice probabilities directly to their allocative skill α_j. If we can observe the true state $\omega \in \Omega$ and the agents' chosen assignments $a \in A$, we can measure *ex post* output $y_j(a, \omega)$ as well as the counterfactual outputs from every other choice $y_j(b, \omega)$. In other words, if we impose symmetry then we can derive α_j for every participant using data on observed assignments and outputs.

2.4 Mapping Theory to Data

Our empirical setting closely matches the theory above. We administer an assessment to survey participants which asks them to play the role of managers assigning fictional workers to tasks. Participants observe information about workers' heterogeneous productivity schedules, and they assign exactly exactly M workers to exactly M tasks. Thus the action set is known and finite, because there are $M!$ possible assignments.

We also fix the information received by participants, the amount of time available to make assignments, and the overall difficulty of the problems. Participants are paid for performance and recruited from an online platform where payment amounts are known in advance, which ensures that utility maps cleanly to output. Payments are small enough that risk aversion is unlikely to be a concern. Finally, workers and tasks are given general labels (e.g. workers 1 and 2, tasks A and B) to ensure that workers and job tasks are seen as equivalent ex ante, which satisfies the symmetry condition.

Thus we map theory to data by treating our survey participants as managers who are solving equation (7) by receiving a set of information signals (worker productivity schedules) and choosing an assignment of workers to tasks that maximizes output given their posterior beliefs. Because of the restrictions we impose on the setting, variation in performance across participants arises from variation in their marginal cost of information c_j or its inverse, allocative skill α_j.

We show above that we can identify ordinal differences between agents in α_j given relatively weak assumptions about the structure of the cost function. Thus in principle, the ranking of output across survey participants should give us a valid measure of allocative skill.

However, we can also solve the model analytically using data from an agent's performance on a series of problems like equation (7). We develop a maximum likelihood estimator that derives the agent's marginal cost of information by comparing their actual assignment to all possible assignments for each decision problem. The estimator also accounts for the nonlinearity of the logit model and for the baseline score expected from inattentive strategies such as random guessing. See the Theory Appendix for a detailed derivation.

2.5 Allocative Skill and Decision-Making

While all jobs in principle require some combination of production and allocation, allocative skill should be relatively more valuable in jobs that are more decision-intensive.

How should we think about the returns to allocative skill across occupations? For simplicity, consider a competitive economy where workers are paid according to a combination of their productive skill (e.g. the marginal product of labor) and their allocative skill (e.g. the marginal product of attention):

$$Y_{ij} = (\theta_i \alpha_j + (1 - \theta_i) z_j) l_j \tag{12}$$

where j indexes individuals (as above) and i indexes occupations. α_j is allocative skill, z_j is productive skill, l_j is worker j's labor supply, and θ_i is the decision intensity of an occupation. In principle we could disaggregate l_j at the task level, with workers supplying effort to different tasks to maximize output Y_{ij} through a combination of production and allocation. For pure production jobs (e.g. when $\theta_i = 0$), allocative skill is unimportant and output equals productive efficiency times labor supply as in Becker (1962). In all other cases, allocative skill increases earnings.

An important caveat is that we do not measure productive skill z_j. If productive and allocative skill both have impacts on earnings, workers will sort into the occupation where they are most productive overall (e.g. Borjas 1987, Heckman and Honore 1990, Hsieh et al. 2019). Thus people with higher allocative skill may not sort into the most decision-intensive occupations. However, conditional on Roy-type occupational sorting, the economic return to allocative skill should be higher in decision-intensive occupations, which we can verify by taking the derivative of (16) with respect to α_j. Thus in an earnings regression we should expect to see a positive coefficient on the interaction between allocative skill and decision intensity. Depending on the correlation between productive skill and allocative skill among respondents in our data, we may also see that workers with higher allocative skill sort into decision-intensive occupations.

3 Measuring Allocative Skill

We recruited a study sample from the online research website Prolific, a platform that is specifically designed for academic research. Douglas et al. (2023) compares sample participation, representativeness, and other measures of data quality across research website and find that Prolific outperforms MTurk, Qualtrics and other competitors. We restricted our study sample to prime-age (age 25 to 55) U.S. residents who spoke fluent English and were employed full-time (at least 35 hours per week). We imposed two additional restrictions to maximize the representativeness of our sample relative to the U.S. working age population. First, we conducted the study on weekends so that people with full-time jobs could participate. Second, we asked respondents what share of their total income is derived from Prolific and other research sites, and we excluded from our analysis sample the 8.6 percent of respondents who reported a share greater or equal to 10 percent. However, our results are not sensitive to this sample restriction.

We recruited a total of 1,250 participants. 9 participants failed to complete the experiment, 79 failed one of the attention or effort checks during the survey, 108 reported earning 10 percent or more of their income through Prolific, and 40 did not report income data. This left us with a core analysis sample of 1,014 respondents, all of whom were full-time employed U.S. residents between the ages of 25 and 55 with valid income data.

The survey had three parts. First, we administered our main assessment (the Assignment Game), described in more detail in Section 4.1. Second, we administered several widely-used and psychometrically validated skills assessments, including an IQ test and a numeracy test. These are described in more detail in Section 4.2. Third, we administered a short survey, with questions about demographics and other important characteristics such as income, education, and occupation. These are described in more detail in Section 4.3.

All participants who passed the attention checks were paid $12 for their time. To ensure a fair comparison across assessments, participants were paid for their performance on each cognitive assesment, including the Assignment Game. Participants spent an average of 44 minutes on the study. They were informed that their performance directly influenced their pay, and that outstanding performance would more than double their pay (bonuses were between $0 and $14). Ultimately, the maximum bonus was $13.85, and the minimum was

$0.01. The mean bonus payment was $4.45.

3.1 The Assignment Game

Before starting the Assignment Game, participants completed a tutorial that explains how the game works and includes a practice problem.[13] In the Assignment Game, participants play the role of a manager who assigns a set of tasks to a set of fictional workers. In each assignment problem there are 3 or 4 tasks and a matching number of workers. Each worker must be assigned to only 1 task, and all tasks must be assigned. The manager's goal is to assign workers to the right tasks to maximize the total output of the team. Conceptually, the Assignment Game requires participants to find patterns in complex, non-verbal information stored in working memory, which makes it similar in some respects to an IQ test. However, it also requires the ability to allocate attention strategically and to understand comparative advantage. In that sense, we can think of the Assignment Game as an economic IQ test.

In the first phase of the Assignment Game, participants observe worker output. Each worker has a productivity schedule over the tasks, and participants are shown multiple draws of workers' output for each task. Participants are told that "workers have good days and bad days" and that as manager their job is to figure out "how good workers are at different tasks ON AVERAGE". Figure 4 presents screenshots of the information provided to participants. Participants are initially shown multiple draws of each worker's output by worker (the outputs worker 1, then worker 2 etc; see top panel of Figure 4). A worker's output for all tasks on a given day is displayed for 1-2 seconds. Next, there is a review period in which information about all workers is presented simultaneously (see bottom panel of Figure 4; note that the review repeats information that participants have already seen). Each review table is presented for 2-3 seconds.

Participants must assign exactly one worker to exactly one task. They can assign workers

[13] A shortened version of the Assignment Game is available for public use at `https://www.skillslab.dev/assignment-game`. The site asks visitors to enter an ID, which can be any combination of letters and numbers.

at any point during the game, including during the observation period - see the top panel of Figure 5, which shows a screenshot of a participant making an initial assignment for all three workers on day 4 of worker 3's observation period. Participants can change their assignments at any time. After the observation period ends, participants have 10 seconds to finalize their assignments. They lose access to worker productivity information during this period - see the bottom panel of Figure 5 for a screenshot.

Scores are based on the average productivity of workers. For example, suppose the left panel of Figure 6 gives the average worker productivity across the observation period. If a participant chooses the assignment on the right panel of the figure (Task A to worker 3, Task B to worker 1, and Task C to worker 2) the raw score would be 16 = 4+10+2. Participant scores can be compared to two thresholds: a ceiling score (i.e. the optimal assignment) and a floor (calculated as the expected score from random guessing).

Assignment Game items can vary in complexity, from trivially easy to nearly impossible. In our version, easier items have 3 tasks and 3 workers while harder items have 4 tasks and 4 workers. To illustrate, there are $n!$ possible assignments for an nxn item, so a 4x4 item is at least four times more difficult than a 3x3 item (6 vs. 24 possible assignments). Item difficulty is also increasing in the variance of productivity draws over the observation period. To make the 3x3 and 4x4 items as comparable as possible, each 4x4 item has a 3x3 embedded within it.[14] Our Assignment Game assessment consisted of 16 items - 8 were 3x3, and 8 were 4x4. The maximum score of 84 was achieved by 7 (0.67 percent) of the 1,014 participants. The mean score on the Assignment Game was 68, and the standard deviation was 9.4.

To assess the reliability of the Assignment Game score, we randomly split the items into two samples 5,000 times, calculated a Spearman-Brown adjusted correlation between each half, and took the mean of the 5,000 estimates. The split sample reliability of the Assignment Game score is 0.75, which is similar to the reliabilities we calculate for the more common

[14]To disguise this fact, we jumbled task labels and changed the levels of various worker outputs by adding or subtracting a constant. However, we left relative productivities untouched, thus retaining the structure of the 3x3 item when increasing it to 4x4.

assessments we use in our survey.[15]

3.2 Other Assessments

We administered three other widely-used assessments of cognitive skills - the Raven's Advanced Progressive Matrices (Ravens), the Cognitive Reflection Test (CRT), and the Berlin Numeracy Test (BNT).

The Ravens test measures participants' pattern recognition and spatial reasoning, and is widely interpreted as a measure of IQ (e.g. Ravens 2003). Participants observe a pattern and determine "what comes next" - see Appendix Figure A.1 for an example item. Our Ravens test included 14 items. The maximum score of 14 was achieved by only 1 participant. 11 participants scored 0 out of 14. The mean score on the Ravens IQ test was 5.7, and the standard deviation was 2.7.

The CRT is a simple test designed to assess a participant's ability to 'reflect on a question and resist reporting the first response that comes to mind' (Frederick 2005). The original test has 3 questions, and some researchers have suggested that the items might have become too well known and are now subject to floor effects (Toplak et al. 2014). We thus add the revised test reported in Toplak et al. (2014) to the original version, which gives us 6 total items (listed in Appendix Figure A.2). 194 participants answered all 6 CRT items correctly, and 107 scored 0 out of 6. The mean score on the CRT was 3.47, and the standard deviation was 1.96.

Finally, we use the original version of the BNT from Cokely et al. (2012), which contains 4 questions (listed in Appendix Figure A.3). The BNT is a validated test of statistical numeracy that has been taken by over 100,000 participants across a large number of countries and professions (Cokely et al. 2018). The BNT helps us account directly for numerical fluency, which is an important sub-component of allocative skill. Moreover, existing research finds

[15] The split sample reliabilities of the Ravens test (a measure of nonverbal IQ), the Cognitive Reflection Test, and the Berlin Numeracy test are 0.72, 0.76, and 0.65 respectively. As a result, adjusting for differential measurement error across assessments has no substantive impact on our main results.

that performance on the Berlin Numeracy Test predicts decision-making quality independent of fluid intelligence, working memory and cognitive reflection (Cokely et al. 2018). The mean score on the BNT was 1.75, and the standard deviation was 1.32.

3.3 Demographics and Other Characteristics

We collected basic demographic information from participants, including gender, race and ethnicity, age, and educational attainment. Participants reported their income in ranges of $20k USD up to $200k (0-$20k, $20k-$40k,...,$180k-$200k). We code income as the midpoint in the range. There were two categories for high earners: $200k-$250k (coded as $225k) and "Over $250k" (coded as $300k; only 9 participants reported income over $250k).

We also elicited information about participant's current occupation, which we mapped to Standard Occupation Classification (SOC) codes. This was a three step process. First, participants provided their current job title and a 1-sentence description of their role. Second, they were asked to select the job category that most closely matched their current job from a dropdown list that was based on the ONET-SOC taxonomy of major and minor occupation groups. Participants were then presented with the top 5 options generated from their selections by the ONET online tool Autocoder and asked to make a selection (or they could see more options if they requested). See Appendix Figure A.4 for a screenshot. Finally, participants were shown a brief description of the job category they chose and were asked to confirm whether or not this adequately described their current job. If not, they were asked to repeat steps 2 and 3 using a modified job description. This process yielded a valid SOC code for all but three survey participants. We link participants' reported occupations to ONET and ACS data using the most detailed level available, up to six digits whenever possible. There are 95 three digit SOC codes and 759 six digit SOC codes.

Table 1 lists summary statistics for our analysis sample (n=1,014) alongside the average characteristics of the US full-time employed population age 25 to 55, calculated from the 2018-2019 ACS. Our sample is 76 percent white compared to 72 percent nationally, with slightly fewer Black respondents (8 percent vs. 13 percent) but otherwise fairly representative. Our sample is also slightly more male (64 percent vs. 56 percent). However, our sample is much more educated than the U.S. average, with 67 percent having obtained a bachelor's degree versus 41 percent nationally. The occupations held by our sample participants are at the 65th percentile nationally in terms of decision intensity, compared to the 55th percentile in the full-time employed ACS sample. Wage and salary income are nearly identical across the two samples ($71,784 versus $71,528, both in 2022 dollars).

Table 2 presents correlations between the Assignment Game score and other assessments, as well as selected demographics. Assignment Game score is positively correlated with non-verbal IQ ($\rho = 0.38$) and with the CRT and the BNT ($\rho = 0.31$ and $\rho = 0.29$ respectively). In general, all the cognitive assessments are modestly positively associated with each other. Having a bachelor's degree is also modestly positively correlated with Assignment Game score ($\rho = 0.11$) and with other cognitive assessments. Age is negatively correlated with performance on the Assignment Game ($\rho = -0.13$), and men score slightly higher ($\rho = 0.06$).

4 Results

Our first main hypothesis is that allocative skill - as measured by the Assignment Game - is positively associated with income. To facilitate comparison we normalize the Assignment Game and all the cognitive assessments to have a mean of zero and a standard deviation of one. We also present results that use the scaling implied by the analytic solution developed in Section 3.4, where the marginal cost of attention is derived from the logit formula.

Table 3 presents regressions of income on Assignment Game score, controlling for demographics, other cognitive assessments, and other variables. Column 1 shows the bivariate association. A one standard deviation increase in allocative skill is associated with a $6,006 increase in annual income, which is statistically significant at the less than one percent level and equivalent to about 8 percent of the sample mean. Column 2 adds controls for gender, race/ethnicity, age and age squared, and educational attainment. The coefficient drops to \$4,480 but still remains significant at the less than one percent level.

Column 3 adds controls for IQ. The coefficient on allocative skill falls further to \$3,816 ($p = 0.005$), but is twice the magnitude of the relationship between IQ and income.[16] Column 4 adds controls for the CRT and the BNT, the other two cognitive assessments. This increases the coefficient on allocative skill slightly to \$3,955 ($p = 0.004$). Notably, the coefficients on the other assessments are all smaller in magnitude, and none are statistically distinguishable from zero.[17] The relationship between allocative skill and income is positive, highly statistically significant, and robust to controlling for multiple other cognitive assessments.

Column 5 reweights the data from Prolific to match the full-time employed prime age labor force in the 2018-2019 ACS. This increases the magnitude of the coefficient on allocative skill by about 25 percent, to \$5,087 ($p < 0.001$) or 7.1 percent of average annual income in the ACS.

Column 6 adds fixed effects for 3-digit SOC codes, which asks whether the association between allocative skill and income holds within occupations. A one standard deviation in allocative skill is associated with a \$4,420 increase in annual income, even after controlling for other cognitive assessments, demographics, and current occupation at the 3-digit level. Thus the relationship between allocative skill and income holds within occupations. As before, allocative skill is a stronger predictor of income than nonverbal IQ, the Cognitive Reflection test, or the Berlin Numeracy test. However, it is important to note that AG score and IQ are strongly related, and in many specifications we cannot reject the hypothesis that they have the same magnitude. We do not argue that IQ is irrelevant, only that AG score is on

[16] In a bivariate regression of income on IQ, the coefficient is \$4,852 ($p = 0.001$), which is about 80 percent as large as the coefficient on AG score and 6.8 percent of the sample mean. A recent study in Finland, where nearly all men are conscripted into the army and given an IQ test, finds that a one standard deviation increase in nonverbal IQ increases earnings at ages 30-34 by $1,390$, which is about 6 percent of the sample mean (Jokela et al. 2017).

[17] The negative coefficient on the Berlin Numeracy Test in Table 3 is an artifact of the high degree of collinearity between tests. To show this directly, Appendix Table A2 presents separate regressions of income on each cognitive assessment plus demographic covariates. Allocative skill is most strongly related to income, but all the coefficients are positive.

average a stronger predictor of economic outcomes and also that it is more firmly grounded in economic theory.

Appendix Table A3 presents an analogous set of results using the marginal cost of attention measure derived from the analytic solution to the model in Section 2.4. The results are qualitatively very similar - attention costs are negatively associated with income even after conditioning on demographics, other cognitive assessments, and occupation. Thus we adopt the standardized allocative skill score as our main result so that we can compare the magnitude directly to other cognitive assessments.

Table 4 studies occupational sorting. We regress the decision intensity of a participant's current occupation on allocative skill, controlling for other cognitive assessments and demographics. Column 1 shows the bivariate association, which suggests that a one standard deviation increase in allocative skill is associated with an increase in decision intensity of about 3.1 percentile ranks ($p < 0.001$). This association falls to 2.2 percentile ranks when adding controls for demographics in Column 2. Column 3 adds controls for measured IQ, which lowers the association further to 1.9 percentile ranks ($p = 0.020$). Column 4 adds controls for the other cognitive assessments, which lowers the association to 1.5 percentile ranks ($p = 0.074$). Finally, Column 5 adds ACS weights, which increases the association slightly to 1.6 percentile ranks ($p = 0.136$).

Overall, we find a small, borderline statistically significant positive relationship between allocative skill and the decision intensity of a participant's current occupation. As discussed in Section 3, the impact of allocative skill on occupational sorting is ambiguous and depends on the correlation in our sample between allocative skill and productive skill in various tasks (which we do not observe). However, it is notable that the AG score is more predictive of occupation decision intensity than any of the other cognitive measures.

Our model unambiguously predicts that the association between allocative skill and income should be increasing in decision intensity. Table 5 tests this prediction by regressing income on allocative skill, decision intensity of current occupation, and the interaction between the two.[18] Column 1 shows this relationship without any additional controls. We find strong evidence that allocative skill is more important in decision intensive occupations. The coefficient on the interaction term is large and statistically significant, and suggests that the impact of a one standard deviation increase in allocative skill increases by $1,115 for every 10 percentile rank increase in the decision intensity of a worker's occupation ($p = 0.025$).[19] For occupations at the 75th percentile of decision intensity, this translates to an increase in annual earnings of about 10 percent.

Column 2 adds demographic controls, which slightly increases the coefficient on the interaction term to $1,177 ($p = 0.012$). Adding controls for nonverbal IQ and other cognitive assessments increases it a bit more, to $1,254 ($p = 0.008$). Column 4 adds interactions between decision intensity and the other cognitive assessments. This lowers the coefficient slightly to $1,122 ($p = 0.019$). Overall, adding controls for other cognitive assessments and their interaction with decision intensity has no substantive impact on the estimated relationship between income, Assignment Game score, and the decision intensity of a respondent's occupation.

Notably, we find no evidence that the economic returns to IQ, cognitive reflection, or numeracy are increasing in decision intensity. All of the coefficients are smaller in magnitude than the coefficient on allocative skill, and none are statistically significant. This strongly suggests that allocative skill is particularly important in occupations that require more decision-making. Finally, Column 5 adds ACS weights, which has no substantive impact on the estimates.

Appendix Table A4 presents heterogeneous impacts of allocative skill by gender, age, and educational attainment. The association between allocative skill and income is positive and statistically significant at the less than 5 percent level for all subgroups. However, it is larger

[18]We de-mean the decision intensity variable when interacting it with each cognitive assessment, so that the main effect captures the impact of a one standard deviation increase in each skill measure for workers in jobs of average decision intensity. This makes the mean effect easier to interpret, but has no effect on the magnitude or the precision of the interaction terms themselves.

[19]The results are also robust to splitting respondents by terciles or quartiles of occupation decision intensity and interacting the Assignment Game score separately with each quantile.

in magnitude for men and for college-educated workers.

5 Conclusion

This paper develops a theory and measurement paradigm for assessing individual differences in decision-making, which we call *allocative skill*. We first show that modern work increasingly requires decision-making. We then develop a simple model where agents assign factors of production to different tasks to maximize total output. This could be a manager assigning workers to jobs, or a worker assigning her own effort over job tasks. Factors have heterogeneous productivity over tasks, and productivity information is costly to observe. Holding complexity and time constraints fixed, skilled agents possess more total attention and allocate it more efficiently, achieving higher total output. Thus we can think of allocative skill as the marginal product of attention.

We measure allocative skill with a novel task we call the Assignment Game, where participants are managers who assign fictional workers to jobs to maximize output. We administer the Assignment Game to more than a thousand full-time, prime-age U.S. workers along with a survey that collects demographic information as well as data on employment, occupational choice, and earnings. Allocative skill is strongly associated with income, even after controlling for IQ, numeracy, education, occupation, and other covariates. The association between earnings and allocative skill is twice as large as the association with IQ, conditional on demographics. We also find that the association between allocative skill and income is significantly greater in decision-intensive occupations, which is consistent with our theoretical framework.

Our paper contributes to human capital theory by formalizing and testing the idea that good decision-making is rewarded in the labor market. Modeling decision-making and measuring individual differences in allocative skill requires us to take seriously the idea that attention is a scarce resource. Agents with greater allocative skill deploy their attention more efficiently, and thus are better able to use available information to make complex decisions. Put another way, to understand labor productivity in the information age, we must use the tools of information theory. Good decision-making is likely to be increasingly important in the labor market as routine information processing tasks are increasingly automated.

References

Adhvaryu, A., Kala, N. and Nyshadham, A.: 2022, Management and shocks to worker productivity, *Journal of Political Economy* **130**(1), 1–47.

Adhvaryu, A., Nyshadham, A. and Tamayo, J.: 2023, Managerial quality and productivity dynamics, *The Review of Economic Studies* **90**(4), 1569–1607.

Atalay, E., Phongthiengtham, P., Sotelo, S. and Tannenbaum, D.: 2020, The evolution of work in the united states, *American Economic Journal: Applied Economics* **12**(2), 1–34.

Autor, D. H., Levy, F. and Murnane, R. J.: 2002, Upstairs, downstairs: Computers and skills on two floors of a large bank, *ILR Review* **55**(3), 432–447.

Autor, D. H., Levy, F. and Murnane, R. J.: 2003, The skill content of recent technological change: An empirical exploration, *The Quarterly journal of economics* **118**(4), 1279–1333.

Bandiera, O., Prat, A., Hansen, S. and Sadun, R.: 2020, Ceo behavior and firm performance, *Journal of Political Economy* **128**(4), 1325–1369.

Baron, J.: 2000, *Thinking and deciding*, Cambridge University Press.

Bartel, A., Ichniowski, C. and Shaw, K.: 2007, How does information technology affect productivity? plant-level comparisons of product innovation, process improvement, and worker skills, *The quarterly journal of Economics* **122**(4), 1721–1758.

Becker, G. S.: 1962, Investment in human capital: A theoretical analysis, *Journal of political economy* **70**(5, Part 2), 9–49.

Benjamin, D. J., Brown, S. A. and Shapiro, J. M.: 2013, Who is 'behavioral'? cognitive ability and anomalous preferences, *Journal of the European Economic Association* **11**(6), 1231–1255.

Bertrand, M. and Schoar, A.: 2003, Managing with style: The effect of managers on firm policies, *The Quarterly journal of economics* **118**(4), 1169–1208.

Blackwell, D.: 1953, Equivalent comparisons of experiments, *The annals of mathematical statistics* pp. 265–272.

Bloom, N. and Van Reenen, J.: 2010, Why do management practices differ across firms and countries?, *Journal of economic perspectives* **24**(1), 203–24.

Bordalo, P., Gennaioli, N. and Shleifer, A.: 2012, Salience theory of choice under risk, *The Quarterly journal of economics* **127**(3), 1243–1285.

Borjas, G. J.: 1987, Self-selection and the earnings of immigrants, *The American Economic Review* pp. 531–553.

Bresnahan, T. F., Brynjolfsson, E. and Hitt, L. M.: 2002, Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence, *The Quarterly Journal of Economics* **117**(1), 339–376.

Bucher, S. F. and Caplin, A.: 2021, Inattention and inequity in school matching, *Technical report*, National Bureau of Economic Research.

Caplin, A. and Dean, M.: 2015, Revealed preference, rational inattention, and costly information acquisition, *American Economic Review* **105**(7), 2183–2203.

Caplin, A., Dean, M. and Leahy, J.: 2022, Rationally inattentive behavior: Characterizing and generalizing shannon entropy, *Journal of Political Economy* **130**(6), 1676–1715.

Chan Jr, D. C., Gentzkow, M. and Yu, C.: 2019, Selection with variation in diagnostic skill: Evidence from radiologists, *Technical report*, National Bureau of Economic Research.

Chandra, A. and Staiger, D. O.: 2020, Identifying sources of inefficiency in healthcare, *The quarterly journal of economics* **135**(2), 785–843.

Cokely, E. T., Feltz, A., Ghazal, S., Allan, J., Petrova, D. and Garcia-Retamero, R.: 2018, Decision making skill: From intelligence to numeracy and expertise, *Cambridge handbook of expertise and expert performance* pp. 476–505.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S. and Garcia-Retamero, R.: 2012, Measuring risk literacy: The berlin numeracy test, *Judgment and Decision making* **7**(1), 25–47.

Currie, J. and MacLeod, W. B.: 2017, Diagnosing expertise: Human capital, decision making, and performance among physicians, *Journal of labor economics* **35**(1), 1–43.

DellaVigna, S.: 2009, Psychology and economics: Evidence from the field, *Journal of Economic literature* **47**(2), 315–372.

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

Deming, D. J.: 2021, The growing importance of decision-making on the job, *Technical report*, National Bureau of Economic Research.

Deming, D. J. and Noray, K.: 2020, Earnings dynamics, changing job skills, and stem careers, *The Quarterly Journal of Economics* **135**(4), 1965–2005.

Dohmen, T., Falk, A., Huffman, D. and Sunde, U.: 2010, Are risk aversion and impatience related to cognitive ability?, *American Economic Review* **100**(3), 1238–60.

Douglas, B. D., Ewell, P. J. and Brauer, M.: 2023, Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona, *Plos one* **18**(3), e0279720.

Frederick, S.: 2005, Cognitive reflection and decision making, *Journal of Economic perspectives* **19**(4), 25–42.

Gabaix, X.: 2019, Behavioral inattention, *Handbook of behavioral economics: Applications and foundations 1*, Vol. 2, Elsevier, pp. 261–343.

Gill, D. and Prowse, V.: 2016, Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis, *Journal of Political Economy* **124**(6), 1619–1676.

Goldfarb, A. and Xiao, M.: 2011, Who thinks about the competition? managerial ability and strategic entry in us local telephone markets, *American Economic Review* **101**(7), 3130–61.

Hansen, S., Ramdas, T., Sadun, R. and Fuller, J.: 2021, The demand for executive skills, *Technical report*, National Bureau of Economic Research.

Heckman, J. J. and Honore, B. E.: 1990, The empirical content of the roy model, *Econometrica: Journal of the Econometric Society* pp. 1121–1149.

Hoffmann, F., Lee, D. S. and Lemieux, T.: 2020, Growing income inequality in the united states and other advanced economies, *Journal of Economic Perspectives* **34**(4), 52–78.

Hortaçsu, A., Luco, F., Puller, S. L. and Zhu, D.: 2019, Does strategic ability affect efficiency? evidence from electricity markets, *American Economic Review* **109**(12), 4302–42.

Hsieh, C.-T., Hurst, E., Jones, C. I. and Klenow, P. J.: 2019, The allocation of talent and us economic growth, *Econometrica* **87**(5), 1439–1474.

Huffman, W. E.: 1977, Allocative efficiency: The role of human capital, *The Quarterly Journal of Economics* pp. 59–79.

Jokela, M., Pekkarinen, T., Sarvimäki, M., Terviö, M. and Uusitalo, R.: 2017, Secular rise in economically valuable personality traits, *Proceedings of the National Academy of Sciences* **114**(25), 6527–6532.

Kamenica, E. and Gentzkow, M.: 2011, Bayesian persuasion, *American Economic Review* **101**(6), 2590–2615.

Koopmans, T. C. and Beckmann, M.: 1957, Assignment problems and the location of economic activities, *Econometrica: journal of the Econometric Society* pp. 53–76.

Leibenstein, H.: 1966, Allocative efficiency vs." x-efficiency", *The American economic review* **56**(3), 392–415.

Maćkowiak, B., Matějka, F. and Wiederholt, M.: 2023, Rational inattention: A review, *Journal of Economic Literature* **61**(1), 226–273.

Matějka, F. and McKay, A.: 2015, Rational inattention to discrete choices: A new foundation for the multinomial logit model, *American Economic Review* **105**(1), 272–298.

Metcalfe, R. D., Sollaci, A. B. and Syverson, C.: 2023, Managers and productivity in retail, *Technical report*, National Bureau of Economic Research.

Mincer, J.: 1958, Investment in human capital and personal income distribution, *Journal of political economy* **66**(4), 281–302.

Mincer, J.: 1974, Schooling, experience, and earnings. human behavior & social institutions no. 2.

Minni, V.: 2022, Making the invisible hand visible: Managers and the allocation of workers to jobs, *Job Market Paper* .

NACE: 2020, *Job Outlook 2020*, National Association of Colleges and Employers.

Nelson, R. R. and Phelps, E. S.: 1966, Investment in humans, technological diffusion, and economic growth, *The American economic review* **56**(1/2), 69–75.

Restuccia, D. and Rogerson, R.: 2017, The causes and costs of misallocation, *Journal of Economic Perspectives* **31**(3), 151–174.

Rustichini, A.: 2015, The role of intelligence in economic decision making, *Current Opinion in Behavioral Sciences* **5**, 32–36.

Sims, C. A.: 2003, Implications of rational inattention, *Journal of monetary Economics* **50**(3), 665–690.

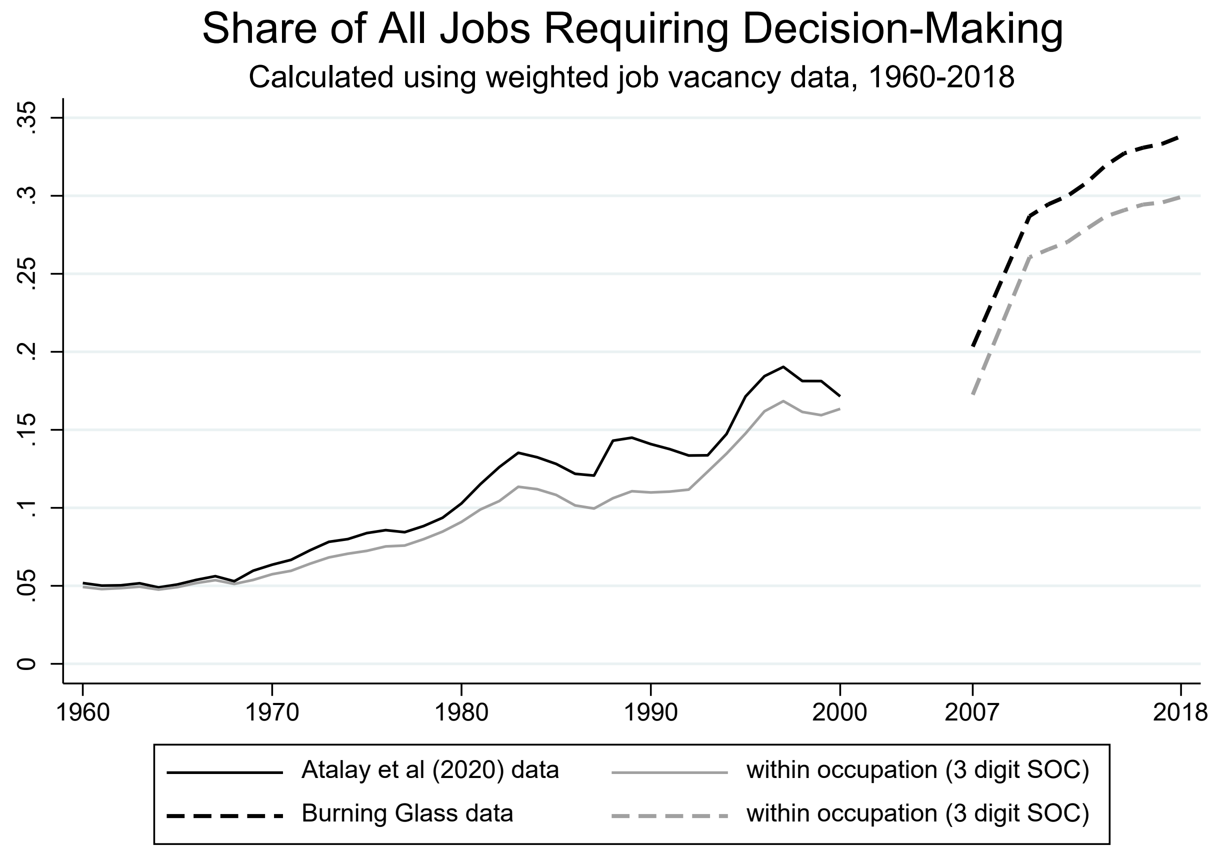
Stanovich, K. E. and West, R. F.: 2000, Individual differences in reasoning: Implications for the rationality debate?, *Behavioral and brain sciences* **23**(5), 645–665.

Stigler, G. J.: 1976, The xistence of x-efficiency, *The American Economic Review* **66**(1), 213–216.

Toplak, M. E., West, R. F. and Stanovich, K. E.: 2014, Assessing miserly information processing: An expansion of the cognitive reflection test, *Thinking & reasoning* **20**(2), 147–168.

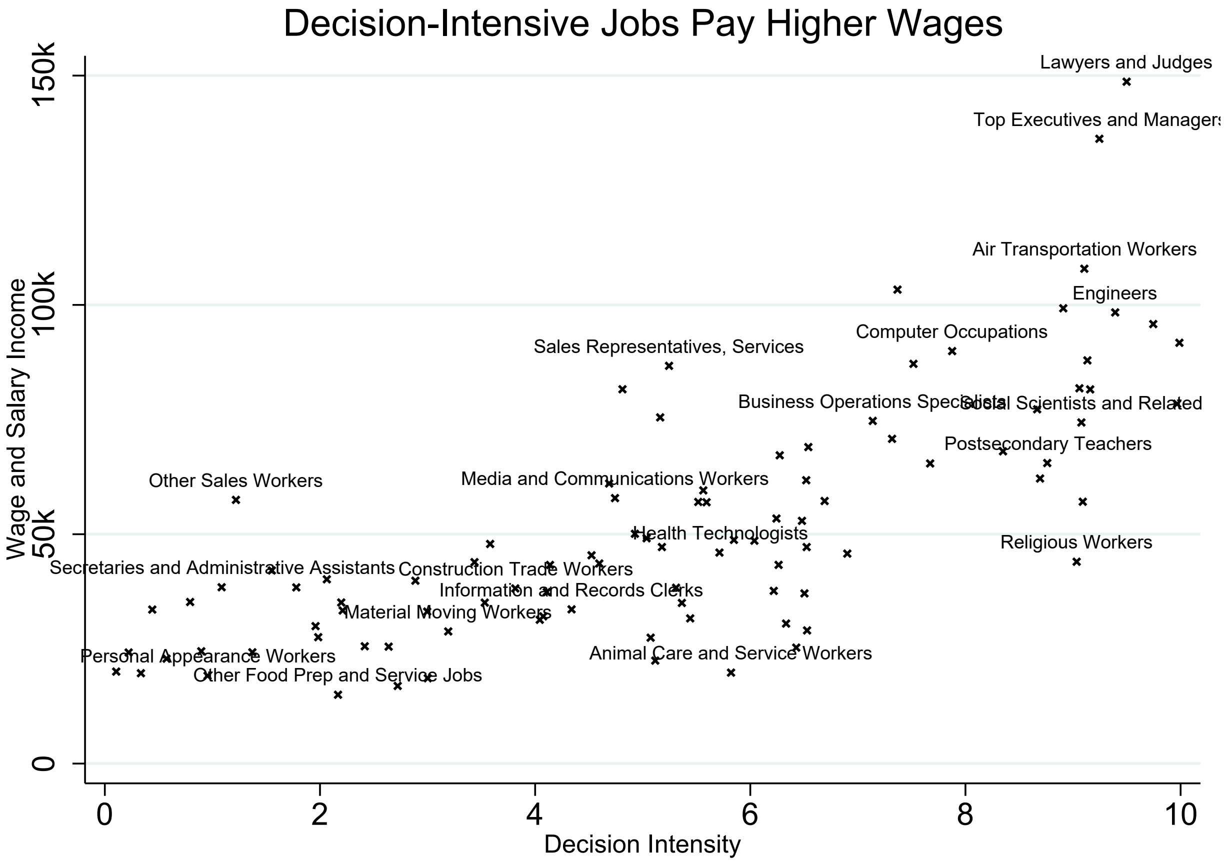
Welch, F.: 1970, Education in production, *Journal of political economy* **78**(1), 35–59.

Figure 1



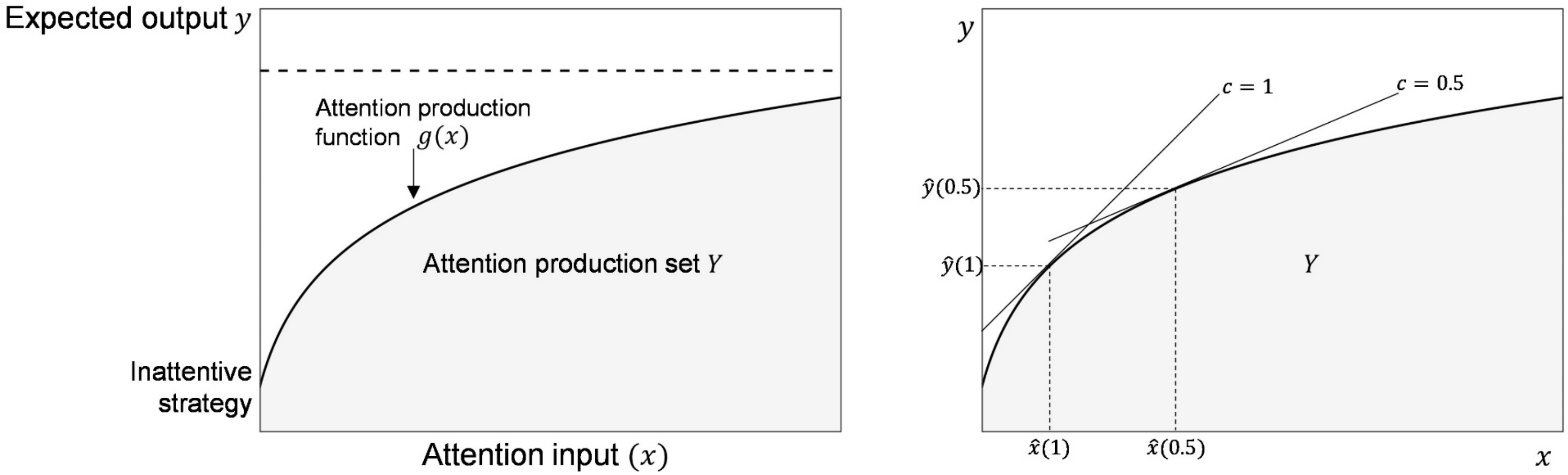
Notes: This figure computes the labor supply-weighted share of all job vacancies that include key words and phrases signaling a demand for worker decision-making – see the text for detailed definitions. The solid line uses classified ad data collected by Atalay et al (2020) over the 1960-1999 period, while the dashed line uses Burning Glass Technologies data from 2007 and 2010-2018. The data are weighted by the actual occupation distribution in the nearest Census and ACS years and are smoothed using a five-year moving average. The grey lines below present the same series except controlling for occupation fixed effects at the three-digit Standard Occupation Classification (SOC) level. We convert Census occupation codes to SOC codes using a crosswalk developed by Atalay et al (2020).

Figure 2



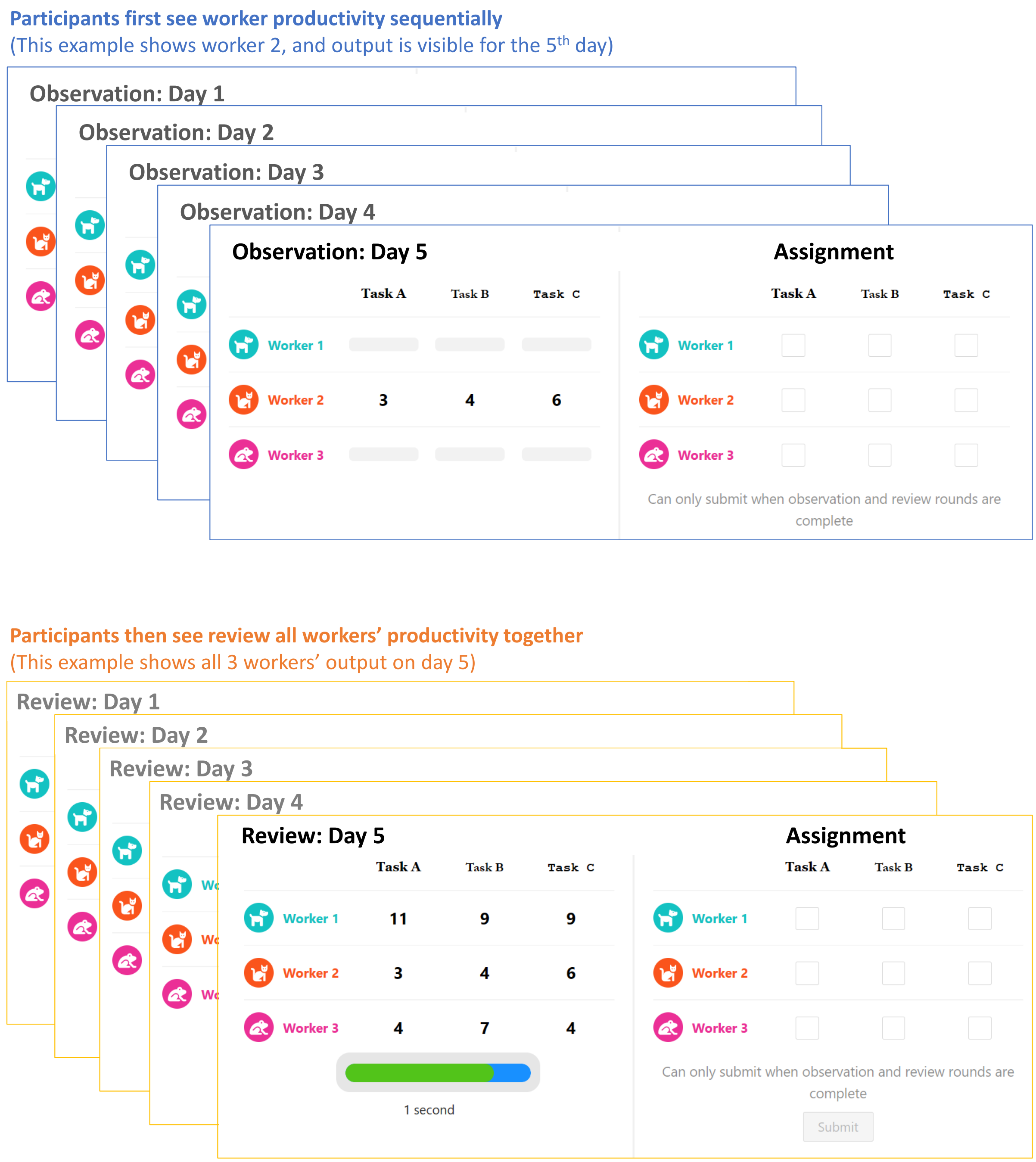
Notes: This figure plots average wage and salary income in the 2018 and 2019 American Community Survey against the average decision intensity of occupations at the 3-digit Standard Occupation Classification (SOC) code level, with selected occupations labeled. Occupation decision intensity is represented on a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the full 2018-2019 ACS sample. Income is reported in 2022 dollars. We construct the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details.

Figure 3



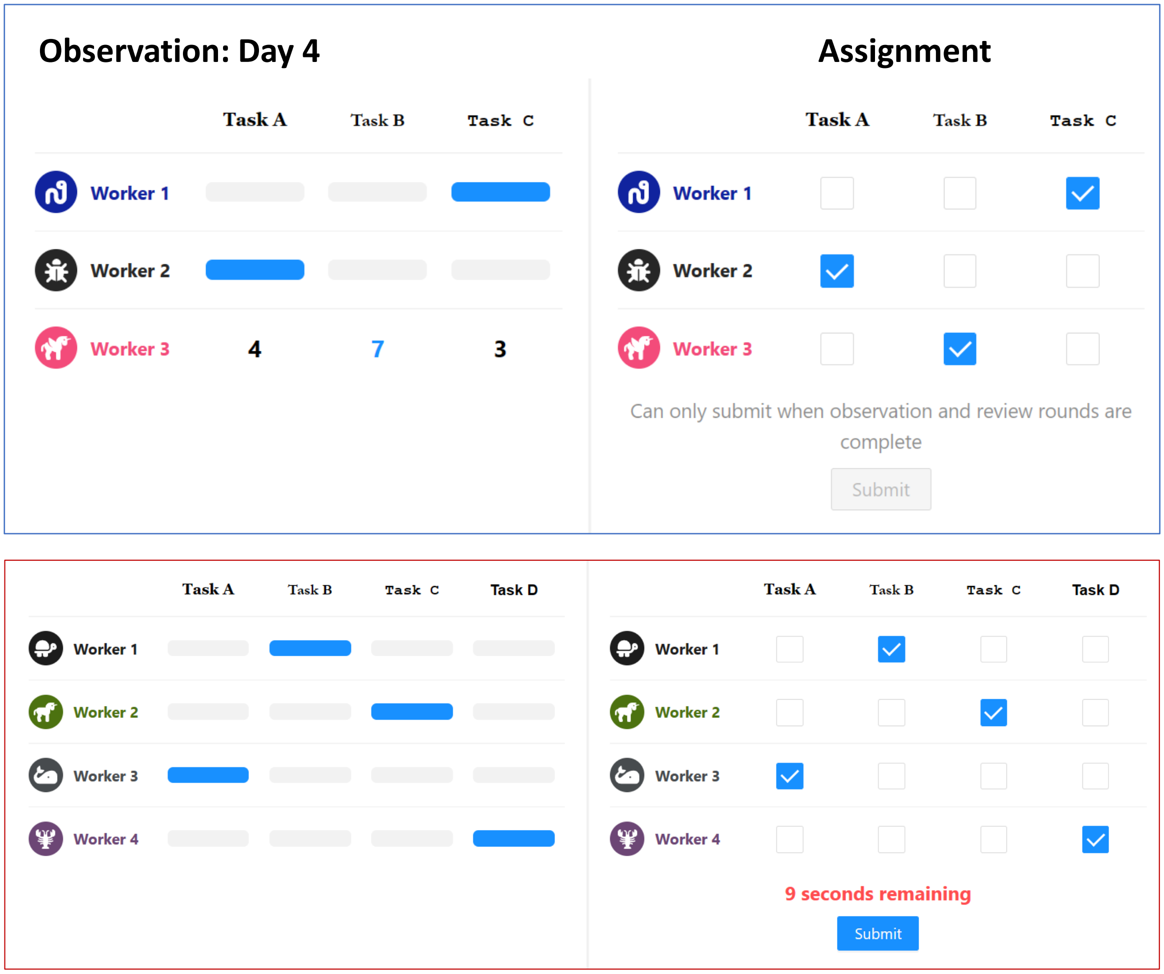
Notes: This figure presents a graphical illustration of the attention production set Y from equation (4), which maps the space of possible outputs the agent can achieve for any fixed amount of attention x. The vertical axis intercept corresponds to output under a fully inattentive strategy (e.g. random guessing.) The production function $g(x)$ maps the frontier of expected output for any given input. The righthand panel depicts the impact of a decrease in the marginal cost of attention from $c = 1$ to $c = 0.5$, which flattens the slope of the tangency line and causes the agent to optimally pay more attention and produce higher expected output. See Section 2.1 of the paper for details.

Figure 4



Notes: This figure shows screenshots from the Assignment Game. The top panel illustrates how participants initially see each worker's productivity individually and sequentially. The bottom panel illustrates how participants are then shown a review where all workers productivity schedules are shown simultaneously.

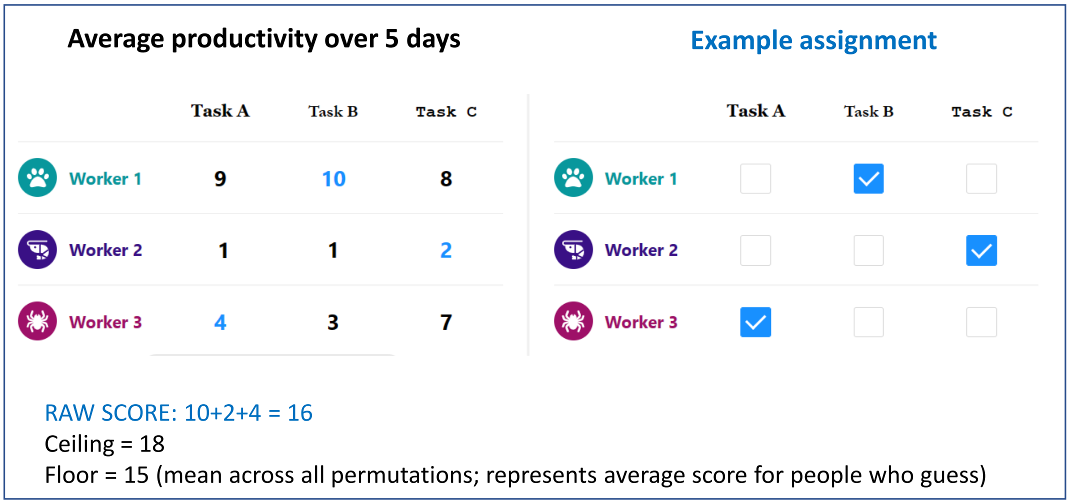
Figure 5



Notes: This figure shows another screenshot from the Assignment Game. The top panel shows a 3x3 puzzle and demonstrates how participants are able to make assignments at any point in the game (i.e. they can start assigning from the observation period onwards). The bottom panel shows a 4x4 puzzle and illustrates the final 10 second 'submission period'. During this time participants lose access to productivity information and need to make their final assignments before hitting 'Submit'. If participants fail to hit submit, we still record any assignments that have been made and give people partial credit.[1]

[1] In this scenario, unassigned workers receive a score of 0

Figure 6



Notes: This figure demonstrates how the raw scores for the assignment game are calculated. The table on the left represents the average productivity schedule across all 5 days of a puzzle. Note that this **would not** be shown to participants. Combining the productivity schedules on the left with the assignment on the right, we see that the participants score is 16 (10+2+4). The optimal solution on this problem would score 18 (worker 1 -> task B; worker 2 -> task A; worker 3 -> task C).

Table 1 - Summary Statistics

	Prolific	2018-2019 ACS
	(1)	(2)
Male	0.637	0.564
White	0.760	0.717
Black	0.076	0.125
Asian	0.080	0.070
Other Race / Not Reported	0.080	0.088
Age	37.8	39.3
Bachelor's Degree	0.667	0.412
Occupation Decision Intensity	6.543	5.537
Wage and Salary Income	71,784	71,528
Sample Size	1,014	1,446,680

Notes: Table 1 presents summary statistics for our Prolific survey sample and compares them to the combined 2018 and 2019 American Community Survey. Column 2 is weighted to be nationally representative of the full-time working population age 25 to 55. Occupation decision intensity is represented on a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the full 2018-2019 ACS sample. Income is reported in 2022 dollars. We construct the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details.

Table 2 - Correlations between Allocative Skill and Other Variables

	Allocative Skill (AG Score)	Nonverbal IQ (Ravens)	Cognitive Reflection Test	Berlin Numeracy Test	Male	Age
Allocative Skill (AG Score)	1					
Nonverbal IQ (Ravens)	0.381	1				
Cognitive Reflection Test	0.313	0.430	1			
Berlin Numeracy Test	0.293	0.355	0.598	1		
Male	0.064	0.108	0.126	0.108	1	
Age	-0.132	-0.162	-0.026	-0.098	-0.047	1
Bachelor's Degree	0.109	0.119	0.149	0.099	-0.011	-0.003

Notes: Table 2 presents correlations between our measure of allocative skill (the Assignment Game) and other cognitive assessments and demographics. All tests are normalized to have mean zero and standard deviation one. The data come from our Prolific survey sample, N=1,014. See the text for a more detailed description of the cognitive assessments.

Table 3 - Allocative Skill Predicts Higher Income

	(1)	(2)	(3)	(4)	(5)	(6)
Allocative Skill (AG Score)	6,006	4,480	3,816	3,955	5,087	4,420
	[1,423]	[1,312]	[1,358]	[1,380]	[1,548]	[1,580]
Nonverbal IQ (Ravens)			1,954	2,094	1,963	2,209
			[1,499]	[1,558]	[1,657]	[1,676]
Cognitive Reflection Test				672	1,056	711
				[1,739]	[1,929]	[1,890]
Berlin Numeracy Test				-1,415	-2,315	-4,386
				[1,643]	[1,769]	[1,861]
Demographic Controls		X	X	X	X	X
ACS Weights					X	X
Occupation FE						X
R-Squared	0.0175	0.1824	0.1840	0.1845	0.1991	0.3182
Sample Size	1,014	1,014	1,014	1,014	1,014	1,014

Notes: Table 3 presents estimates from a regression of wage and salary income on allocative skill and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The regression is estimated in our Prolific survey sample. The Assignment Game score (our measure of allocative skill) and all other cognitive assessments are normalized to have mean zero and standard deviation one. Average income in the sample is $71,728. Demographic controls include indicators for gender, race and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Column 5 weights the data to be nationally representative according to the 2018-2019 ACS sample, see Table 1 for details. Column 6 adds fixed effects for 3-digit occupation codes from the Standard Occupation Classification (SOC) - see Table A.1 for a complete list.

Table 4 - Occupational Sorting on Allocative Skill

	(1)	(2)	(3)	(4)	(5)
Allocative Skill (AG Score)	0.311	0.219	0.188	0.147	0.157
	[0.077]	[0.076]	[0.081]	[0.082]	[0.105]
Nonverbal IQ (Ravens)			0.090	0.030	0.081
			[0.081]	[0.086]	[0.098]
Cognitive Reflection Test				-0.027	0.014
				[0.104]	[0.122]
Berlin Numeracy Test				0.261	0.291
				[0.096]	[0.114]
Demographic Controls		X	X	X	X
ACS Weights					X
R-Squared	0.0149	0.1353	0.1363	0.1442	0.1643
Sample Size	1,034	1,034	1,034	1,034	1,034

Notes: Table 4 presents estimates from a regression of occupation decision intensity on allocative skill and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The regression is estimated in our Prolific survey sample. Occupation decision intensity is represented on a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the full 2018-2019 ACS sample. Income is reported in 2022 dollars, We construct the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. The Assignment Game score (our measure of allocative skill) and all other cognitive assessments are normalized to have mean zero and standard deviation one. Demographic controls include indicators for gender, race and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Column 5 weights the data to be nationally representative according to the 2018-2019 ACS sample, see Table 1 for details.

Table 5 - Allocative Skill Predicts Income More in Decision-Intensive Occupations

	(1)	(2)	(3)	(4)	(5)
Allocative Skill (AG Score)	4,200	3,758	3,583	3,622	5,402
	[1,381]	[1,318]	[1,391]	[1,399]	[1,648]
* Decision Intensity (demeaned)	1,115	1,177	1,254	1,122	1,154
	[497]	[467]	[470]	[477]	[509]
Decision Intensity (O*NET)	5,793	4,031	4,112	4,116	4,016
	[468]	[456]	[462]	[462]	[483]
Nonverbal IQ (Ravens)			2,119	2,097	1,501
			[1,547]	[1,525]	[1,609]
* Decision Intensity (demeaned)				227	430
				[530]	[595]
Cognitive Reflection Test			646	766	1,098
			[1694]	[1,748]	[2,005]
* Decision Intensity (demeaned)				706	635
				[536]	[625]
Berlin Numeracy Test			-2,760	-2,832	-3,686
			[1,627]	[1,635]	[1,832]
* Decision Intensity (demeaned)				-611	-1,045
				[510]	[561]
Demographic Controls		X	X	X	X
ACS Weights					X
R-Squared	0.1210	0.2287	0.2322	0.2337	0.2477
Sample Size	1,004	1,004	1,004	1,004	1,004

Notes: Table 5 presents estimates from a regression of wage and salary income on allocative skill and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The regression is estimated in our Prolific survey sample. Average income in the sample is $71,728. Occupation decision intensity is represented on a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the full 2018-2019 ACS sample. Income is reported in 2022 dollars, We construct the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. The Assignment Game score (our measure of allocative skill) and all other cognitive assessments are normalized to have mean zero and standard deviation one. The interaction terms multiply each cognitive assessment times a demeaned version of the decision intensity variable for ease of comparison. Demographic controls include indicators for gender, race and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Column 5 weights the data to be nationally representative according to the 2018-2019 ACS sample, see Table 1 for details.

APPENDIX TABLES AND FIGURES

FROM HERE FORWARD NOT FOR PUBLICATION

Table A1 - Complete list of Occupation Codes by Decision Intensity

SOC Code	Occupation Category	Decision Intensity (O*NET)	Decision Intensity (weighted)	Employment Share	Share with BA	Wage and Salary Income
		(1)	(2)	(3)	(4)	(5)
111	Top Executives and Managers	76	9.24	0.015	0.592	136,234
112	Advertising, PR, Sales Managers	67	7.37	0.008	0.717	103,350
113	Operations Specialties Managers	71	8.91	0.021	0.601	99,273
119	Other Managers	70	8.35	0.063	0.502	68,091
131	Business Operations Specialists	67	7.14	0.034	0.639	74,723
132	Financial Specialists	68	7.52	0.022	0.774	87,163
151	Computer Occupations	70	7.88	0.032	0.683	89,941
152	Mathematical Science Occupations	80	9.99	0.002	0.812	91,759
171	Architects and Surveyors	73	9.06	0.002	0.871	81,838
172	Engineers	76	9.39	0.014	0.820	98,375
173	Drafters and Engineering Technicians	61	5.59	0.005	0.212	56,988
191	Life Scientists	76	9.16	0.002	0.989	81,591
192	Physical Scientists	76	9.13	0.003	0.984	87,921
193	Social Scientists and Related	73	9.08	0.002	0.977	74,382
194	Life/Phys/Soc Science Technicians	59	5.18	0.002	0.402	47,222
195	Occupational Health & Safety Specialists	70	8.67	0.000	0.520	77,260
211	Counselors and Social Workers	67	6.90	0.014	0.754	45,785
212	Religious Workers	72	9.03	0.004	0.716	44,014
231	Lawyers and Judges	78	9.50	0.007	0.977	148,680
232	Legal Support Workers	63	6.24	0.004	0.463	53,446

Notes: Table A1 presents summary statistics by 3-digit occupation codes from the Standard Occupation Classification (SOC) system. The data come from the combined 2018 and 2019 American Community Survey and are weighted to be nationally representative. Column 1 presents the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. Column 2 rescales occupation decision intensity to a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the 2018-2019 ACS sample. Column 3 presents the labor supply weighted employment share for each occupation category, and Column 4 reports the share of workers in each occupation who have a bachelor's degree or more. Column 5 reports average income by occupation, in 2022 dollars.

Table A1 - Complete list of Occupation Codes by Decision Intensity

SOC Code	Occupation Category	Decision Intensity (O*NET)	Decision Intensity (weighted)	Employment Share	Share with BA	Wage and Salary Income
		(1)	(2)	(3)	(4)	(5)
251	Postsecondary Teachers	70	8.76	0.008	0.913	65,536
252	K-12 Teachers	62	6.04	0.035	0.877	48,601
253	Other Teachers and Instructors	55	4.07	0.006	0.529	32,117
254	Librarians and Archivists	57	4.52	0.002	0.753	45,428
259	Other Education Occupations	63	6.43	0.009	0.340	25,292
271	Art and Design Workers	58	5.04	0.007	0.597	49,138
272	Entertainers and Performers	56	4.14	0.005	0.552	43,315
273	Media and Communications Workers	57	4.74	0.005	0.743	57,893
274	Media/Comms Equipment Workers	60	5.31	0.002	0.501	38,356
291	Healthcare Practitioners	78	9.74	0.042	0.773	95,841
292	Health Technologists	61	5.71	0.019	0.220	46,014
299	Other Healthcare Occupations	74	9.09	0.001	0.719	57,100
311	Home Health and Personal Care Aides	48	1.37	0.022	0.105	24,275
312	Occ and Physical Therapy Aides	55	4.11	0.001	0.291	37,416
319	Other Healthcare Aides	63	6.33	0.010	0.167	30,540
331	Supervisors, Protective Services	79	9.97	0.002	0.374	78,441
332	Firefighting and Prevention Workers	67	7.32	0.002	0.236	70,814
333	Law Enforcement Workers	69	7.67	0.009	0.343	65,446
339	Other Protective Service Workers	60	5.36	0.008	0.185	35,042
351	Supervisors, Food Prep Workers	60	5.44	0.007	0.136	31,667

Notes: Table A1 presents summary statistics by 3-digit occupation codes from the Standard Occupation Classification (SOC) system. The data come from the combined 2018 and 2019 American Community Survey and are weighted to be nationally representative. Column 1 presents the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. Column 2 rescales occupation decision intensity to a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the 2018-2019 ACS sample. Column 3 presents the labor supply weighted employment share for each occupation category, and Column 4 reports the share of workers in each occupation who have a bachelor's degree or more. Column 5 reports average income by occupation, in 2022 dollars.

Table A1 - Complete list of Occupation Codes by Decision Intensity

SOC Code	Occupation Category	Decision Intensity (O*NET)	Decision Intensity (weighted)	Employment Share	Share with BA	Wage and Salary Income
		(1)	(2)	(3)	(4)	(5)
352	Cooks and Food Prep Workers	41	0.33	0.021	0.058	19,716
353	Food and Beverage Serving Workers	38	0.10	0.021	0.129	20,067
359	Other Food Prep and Service Jobs	49	2.17	0.005	0.059	15,018
371	Supervisors, Grounds Cleaning/Maintenance	64	6.50	0.003	0.149	37,115
372	Building Cleaning and Pest Control	44	0.57	0.026	0.058	22,882
373	Grounds Maintenance Workers	59	5.12	0.008	0.069	22,464
391	Supervisors, Personal Care and Services	50	2.20	0.001	0.231	35,128
392	Animal Care and Service Workers	61	5.82	0.002	0.211	19,842
393	Entertainment Attendants	39	0.22	0.002	0.185	24,233
394	Funeral Service Workers	65	6.52	0.000	0.302	47,219
395	Personal Appearance Workers	47	0.95	0.009	0.081	19,158
396	Baggage Porters and Bellhops	53	3.00	0.001	0.182	33,154
397	Tour and Travel Guides	54	3.00	0.000	0.345	18,614
399	Other Personal Care and Service Workers	52	2.72	0.011	0.246	16,941
411	Supervisors, Sales Workers	66	6.69	0.028	0.298	57,258
412	Retail Sales Workers	50	2.41	0.039	0.148	25,589
413	Sales Representatives, Services	60	5.24	0.011	0.537	86,730
414	Sales Representatives, Wholesale and Mfg	57	4.81	0.009	0.485	81,621
419	Other Sales Workers	48	1.22	0.009	0.461	57,505
431	Supervisors, Office and Admin Support	60	5.52	0.008	0.360	57,060

Notes: Table A1 presents summary statistics by 3-digit occupation codes from the Standard Occupation Classification (SOC) system. The data come from the combined 2018 and 2019 American Community Survey and are weighted to be nationally representative. Column 1 presents the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. Column 2 rescales occupation decision intensity to a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the 2018-2019 ACS sample. Column 3 presents the labor supply weighted employment share for each occupation category, and Column 4 reports the share of workers in each occupation who have a bachelor's degree or more. Column 5 reports average income by occupation, in 2022 dollars.

Table A1 - Complete list of Occupation Codes by Decision Intensity

SOC Code	Occupation Category	Decision Intensity (O*NET)	Decision Intensity (weighted)	Employment Share	Share with BA	Wage and Salary Income
		(1)	(2)	(3)	(4)	(5)
432	Communications Equipment Operators	43	0.44	0.000	0.199	33,589
433	Financial Clerks	49	2.06	0.016	0.229	40,188
434	Information and Records Clerks	56	4.34	0.035	0.236	33,638
435	Scheduling and Dispatching Workers	48	1.55	0.014	0.164	42,119
436	Secretaries and Administrative Assistants	48	1.08	0.018	0.264	38,466
439	Other Office and Admin Support Workers	46	0.79	0.018	0.267	35,245
451	Farming, Fishing, and Forestry Workers	63	6.26	0.000	0.149	43,335
452	Agricultural Workers	51	2.64	0.005	0.070	25,502
453	Fishing and Hunting Workers	49	1.98	0.000	0.105	27,571
454	Forestry and Logging Workers	65	6.53	0.000	0.086	29,045
471	Supervisors, Construction and Extraction	70	8.69	0.005	0.103	62,167
472	Construction Trade Workers	55	3.81	0.045	0.055	38,197
473	Helpers, Construction Trades	59	5.07	0.000	0.060	27,462
474	Other Construction Workers	54	3.58	0.002	0.122	47,895
475	Extraction Workers	60	5.56	0.001	0.064	59,582
491	Supervisors, Installation and Repair	63	6.27	0.002	0.143	67,231
492	Electrical and Electronic Equipment Repair	62	5.85	0.003	0.157	48,792
493	Vehicle and Mobile Equipment Repair	57	4.60	0.013	0.045	43,640
499	Other Install, Maintenance and Repair Workers	58	4.93	0.015	0.078	50,121
511	Supervisors, Production	57	4.69	0.006	0.165	61,078

Notes: Table A1 presents summary statistics by 3-digit occupation codes from the Standard Occupation Classification (SOC) system. The data come from the combined 2018 and 2019 American Community Survey and are weighted to be nationally representative. Column 1 presents the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. Column 2 rescales occupation decision intensity to a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the 2018-2019 ACS sample. Column 3 presents the labor supply weighted employment share for each occupation category, and Column 4 reports the share of workers in each occupation who have a bachelor's degree or more. Column 5 reports average income by occupation, in 2022 dollars.

Table A1 - Complete list of Occupation Codes by Decision Intensity

SOC Code	Occupation Category	Decision Intensity (O*NET)	Decision Intensity (weighted)	Employment Share	Share with BA	Wage and Salary Income
		(1)	(2)	(3)	(4)	(5)
512	Assemblers and Fabricators	54	3.53	0.008	0.062	35,109
513	Food Processing Workers	49	1.96	0.005	0.072	29,981
514	Metal and Plastics Workers	54	3.44	0.011	0.039	43,934
515	Printing Workers	62	6.22	0.001	0.102	37,667
516	Textile Workers	46	0.90	0.003	0.075	24,503
517	Woodworkers	55	4.04	0.001	0.080	31,360
518	Plant and System Operators	65	6.54	0.002	0.178	69,020
519	Other Production Occupations	53	2.89	0.021	0.105	39,888
531	Supervisors, Transport and Material Moving	64	6.48	0.002	0.163	52,939
532	Air Transportation Workers	74	9.10	0.002	0.603	107,912
533	Motor Vehicle Operators	48	1.78	0.031	0.086	38,425
534	Rail Transportation Workers	59	5.16	0.001	0.132	75,493
535	Water Transportation Workers	65	6.52	0.001	0.183	61,792
536	Other Transportation Workers	50	2.21	0.002	0.101	33,387
537	Material Moving Workers	54	3.19	0.038	0.064	28,830

Notes: Table A1 presents summary statistics by 3-digit occupation codes from the Standard Occupation Classification (SOC) system. The data come from the combined 2018 and 2019 American Community Survey and are weighted to be nationally representative. Column 1 presents the decision intensity variable as the unweighted average of three task measures in the 2019 O*NET - Making Decisions and Solving Problems, Developing Objectives and Strategies, and Planning and Prioritizing Work. See the text for further details. Column 2 rescales occupation decision intensity to a 0 to 10 percentile scale, where 5 represents occupations at the 50th percentile of decision intensity according to the 2018-2019 ACS sample. Column 3 presents the labor supply weighted employment share for each occupation category, and Column 4 reports the share of workers in each occupation who have a bachelor's degree or more. Column 5 reports average income by occupation, in 2022 dollars.

Table A2 - Correlation between Cognitive Assessments and Income

	(1)	(2)	(3)	(4)
Allocative Skill (AG Score)	4,480			
	[1,312]			
Nonverbal IQ (Ravens)		3,267		
		[1,448]		
Cognitive Reflection Test			1,721	
			[1,296]	
Berlin Numeracy Test				633
				[1,273]
Demographic Controls	X	X	X	X
R-Squared	0.1824	0.1781	0.1747	0.1735
Sample Size	1,008	1,008	1,008	1,008

Notes: Table A2 presents estimates from a regression of wage and salary income on each cognitive assessment and demographic controls, which include indicators for gender, race and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Robust standard errors are shown in brackets. The regression is estimated in our Prolific survey sample. The Assignment Game score (our measure of allocative skill) and all other cognitive assessments are normalized to have mean zero and standard deviation one. Average income in the sample is $71,728.

Table A3 - Higher Marginal Cost of Information Predicts Lower Income

	(1)	(2)	(3)	(4)	(5)	(6)
Marginal Cost of Information	-123	-98	-85	-87	-135	-130
	[42]	[40]	[41]	[41]	[51]	[53]
Nonverbal IQ (Ravens)			2,066	2,175	1,736	1,949
			[1,456]	[1,530]	[1,601]	[1,619]
Cognitive Reflection Test				740	1,148	855
				[1,725]	[1,897]	[1,870]
Berlin Numeracy Test				-1,358	-1,986	-4,259
				[1,685]	[1,797]	[1,922]
Demographic Controls		X	X	X	X	X
ACS Weights					X	X
Occupation FE						X
R-Squared	0.0155	0.1828	0.1845	0.1851	0.1978	0.3191
Sample Size	1,005	1,005	1,005	1,005	1,005	1,005

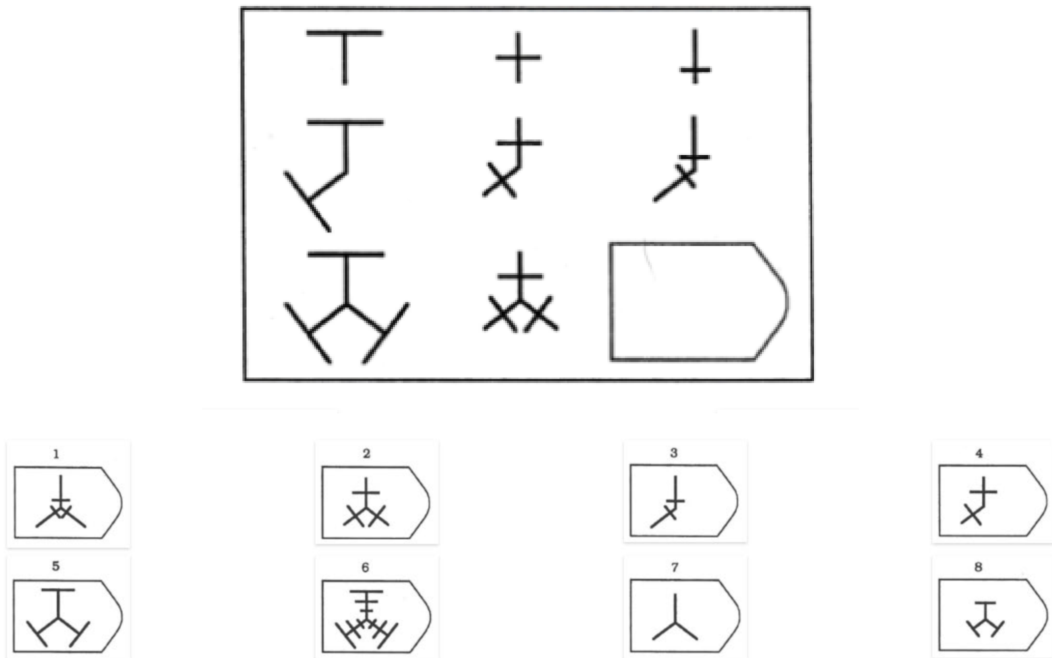
Notes: Table A3 presents estimates from a regression of wage and salary income on the marginal cost of information and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The regression is estimated in our Prolific survey sample. The marginal cost of information is derived using the analytic solution to the model developed in Section 2.4 - see text for details. All other cognitive assessments are normalized to have mean zero and standard deviation one. Average income in the sample is $71,728. Demographic controls include indicators for gender, race, and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Column 5 weights the data to be nationally representative according to the 2018-2019 ACS sample, see Table 1 for details. Column 6 adds fixed effects for 3-digit occupation codes from the Standard Occupation Classification (SOC) - see Table A.1 for a complete list.

Table A4 - Heterogeneous Relationships between Allocative Skill and Income

	Female	*Male*	*No BA*	*BA*	*Age<=37*	*Age>37*
	(1)	(2)	(3)	(4)	(5)	(6)
Allocative Skill (AG Score)	2,372	5,474	2,916	4,819	4,248	4,079
	[1,928]	[1,836]	[1,471]	[1,951]	[1,798]	[2,075]
Nonverbal IQ (Ravens)	-2,105	4,786	1,995	2,382	1,429	3,191
	[2,274]	[2,086]	[1,812]	[2,108]	[2,197]	[2,189]
Cognitive Reflection Test	1,098	-375	-2,576	2,154	4,060	-3,349
	[2,436]	[2,288]	[2,041]	[2,385]	[2,236]	[2,704]
Berlin Numeracy Test	-1,040	-829	-531	-2,224	-1,860	-596
	[2,494]	[2,087]	[1,718]	[2,318]	[2,107]	[2,568]
Demographic Controls	X	X	X	X	X	X
R-Squared	0.1359	0.2221	0.0772	0.1054	0.1458	0.2257
Sample Size	372	636	315	693	528	480

Notes: Table A4 presents estimates from a regression of wage and salary income on allocative skill and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The samples vary and are listed in italics. The regression is estimated in our Prolific survey sample. The Assignment Game score (our measure of allocative skill) and all other cognitive assessments are normalized to have mean zero and standard deviation one. Average income in the sample is $71,728. Demographic controls include indicators for gender, race, and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared.

Figure A1



1/14

Notes: example of an item from the Ravens Advanced Progressive Matrices test. Participants are asked to look for patterns in rows and columns and find the missing piece of the puzzle.

Figure A2

1. Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are in the class? ______ students
   [correct answer = 29 students; intuitive answer = 30]
2. A bat and a ball cost $1.10 in total. The bat costs a dollar more than the ball. How much does the ball cost? ____ cents
   [Correct answer 5 cents; intuitive answer 10 cents]
3. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? ____ minutes
   [Correct answer 5 minutes; intuitive answer 100 minutes]
4. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? ____ days
   [Correct answer 47 days; intuitive answer 24 days]
5. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? _____ days
   [correct answer = 4 days; intuitive answer = 9]
6. A man buys a pig for $60, sells it for $70, buys it back for $80, and sells it finally for $90. How much has he made? _____ dollars
   [correct answer = $20; intuitive answer = $10]

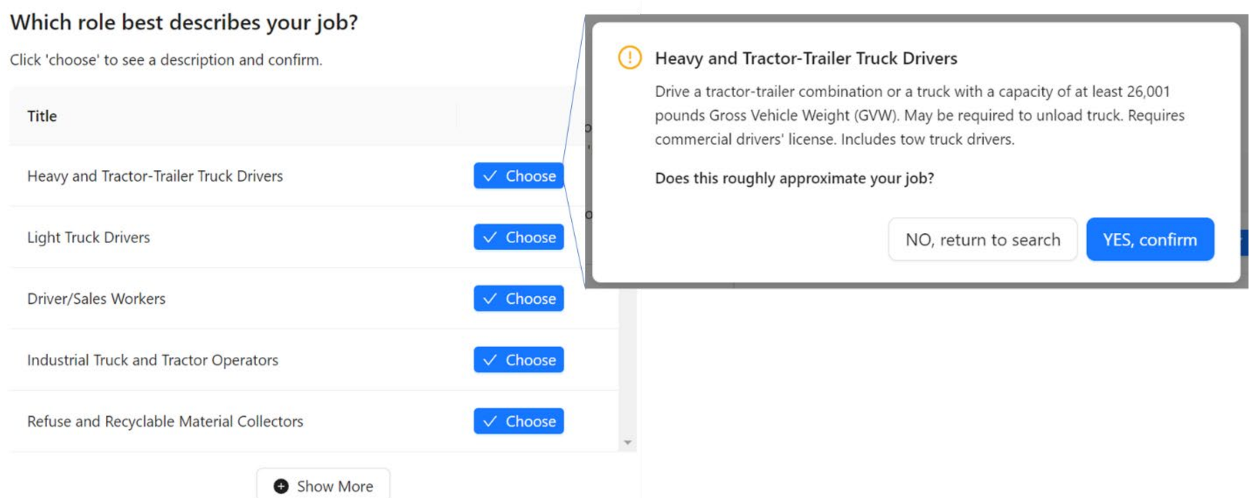
Notes: questions from an expanded version of the Cognitive Reflection Test (CRT). The original CRT is a simple test designed to assess a participant's ability to 'reflect on a question and resist reporting the first response that comes to mind' (Frederick 2005). The original test has 3 questions, and some researchers have suggested that the items might have become too well known and are now subject to floor effects (Toplak, West, and Stanovich 2014). We complement the original 3-question test with the revised test reported in Toplak et al. 2014.

Figure A3

1. Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)?
   ________ out of 50 throws.
2. Out of 1,000 people in a small town 500 are mem- bers of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? (please indicate the probability in percent).
   ___________ %
3. Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, how many times would the die show the number 6?
   ___________ out of 70 throws.
4. In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red?
   _____________%

Notes: questions for the Berlin Numeracy Test. We use the traditional Berlin Numeracy Test (Cokely et al. 2012) containing 4 questions (listed above). This is a validated test of statistical numeracy that has been taken by over 100,000 participants across a large number of countries and professions (Cokely et al. 2018).

Figure A4



Notes: screenshot illustrating the elicitation of occupational codes, which included three steps. First, participants provided their current job title and a 1-sentence description of their role. Second, participants were asked to select the job category (ONET 2019-8) that most closely matched their current job (as per the screenshot). Participants were presented with the top 5 options from O*NET's autocoder, based on their job description (see screenshot; participants were also able to view options 6-10 from the O*NET autocoder by clicking the 'show more' button). Third, participants were shown a brief description of the job category they had chosen and were asked to confirm whether or not this was an approximate description of their job. If not, they were asked to repeat steps 2 and 3 using a refined set of keywords to search for their job.

1 Theory Appendix

We open by restating key definitions to allow this Appendix to be self-contained. We define the agent's expected output for any assignment of factors $a \in A$ in any state $\omega \in \Omega$ as $f(a, \omega)$. Agents' prior beliefs about productivity types are $\mu(\omega)$. We define the Bayes' consistent distributions of posterior beliefs $Q \in Q(\mu)$ with

$$\sum_{\gamma} \gamma Q(\gamma) = \mu \tag{1}$$

Define the optimal value of a posterior belief as:

$$\hat{f}(\gamma) = \max_{a \in A} \sum_{\omega} f(a, \omega) \gamma(\omega) \tag{2}$$

and the optimal value of the distribution of posterior beliefs as $\hat{f}(Q) = \sum_{\gamma} Q(\gamma) \hat{f}(\gamma)$.

The attention production set is,

$$\mathcal{Y} \equiv \left\{ (x, y) \in \mathbb{R}^2 | \exists Q \in Q(\mu) \ s.t. \ \hat{f}(Q) \geq y, K(Q) \leq x \right\} \tag{3}$$

where y is the output level, x is the attention input, and $K(Q)$ is an attention cost function. The attention production function $g(x)$ is defined as the supremum of output in $\mathcal{Y}$ for posterior distributions with attention inputs of x or below,

$$g(x) = \sup_{\{Q \in Q(\mu) | K(Q) \leq x\}} \hat{f}(Q) = \sup_{\{(z,y) \in \mathcal{Y} | z \leq x\}} y \tag{4}$$

1.1 Optimal Production with Costly Attention

Our first general result establishes the basic properties of the attention production function. The attention production function has three characteristics. First, it is non-decreasing: having higher attention input available cannot lower maximum output. The second and third provide general lower and upper bounds on $g(x)$. The lower bound is defined by the optimal strategy at the inattentive posterior distribution that assigns probability 1 to the prior, $Q_I(\mu) = 1$. We normalize to $K(Q_I) = 0$. Note that optimization is straightforward,

$$\hat{f}(Q_I) = \max_{a \in A} \sum_{\omega} f(a, \omega) \mu(\omega) \tag{5}$$

The upper bound is defined by the fully attentive posterior distribution Q_F. It sets $Q_F(e_\omega) = \mu(\omega)$, where $e_\omega \in \Delta(\omega)$ is the belief that state ω is certain. Note that:

$$\hat{f}(Q_F) = \sum_{\omega} \max_{a \in A} f(a, \omega)\mu(\omega). \tag{6}$$

Proposition 1 *The attention production function $g(y)$ satisfies three conditions:*

1. *It is non-decreasing in the attention input x:*

$$x_1 > x_0 \geq 0 \implies g(x_1) \geq g(x_0). \tag{7}$$

2. *It is bounded below and satisfies $g(0) \geq \hat{f}(Q_I)$.*

3. *It is bounded above and satisfies $\lim_{y \longrightarrow \infty} g(y) \leq \hat{f}(Q_F)$.*

Proof. Given $x_1 > x_0 \geq 0$, all Bayes' consistent distributions with $K(Q) \leq x_0$ also satisfy $K(Q) \leq x_1$, so that the corresponding supremum is at least as high, which establishes that $g(x_1) \geq g(x_0)$. To establish that $g(0) \geq \hat{f}(Q_I)$ note that $K(Q_I) = 0$. Hence $(\hat{f}(Q_I), 0) \in \mathcal{Y}$ implying directly that $q(0) \geq \hat{f}(Q_I)$. Finally, to establish that $\lim_{y \longrightarrow \infty} g(x) \leq \hat{f}(Q_F)$ note that any Bayes consistent distribution of posteriors is a garbling (in the sense of Blackwell) of the fully informed distribution, hence by Blackwell's theorem has no higher expected output. ■

We can also define attention costs in a different domain: state dependent stochastic choice (SDSC) functions $P(a, \omega)$, as in Matějka and McKay (2015). Following the logic of Caplin and Dean (2015) there is no reason to use more than one posterior for a given action, given that costs are weakly increasing in the Blackwell order. One can associate with any $P(a, \omega)$ consistent with the prior a **revealed information structure** $Q_P \in Q(\mu)$ by identifying for each chosen action a the corresponding posterior, which we denote γ_P^a:

$$\gamma_P^a(\omega) = \frac{P(a, \omega)}{P(a)} \tag{8}$$

where $P(a) > 0$ is the unconditional choice probability. Adding these up across chosen a defines the revealed information structure Q_P which assigns to each of the revealed posteriors the sum of the unconditional probabilities with which it is chosen (this allows more than one action to have the same revealed posterior). Caplin and Dean (2015) show that this is the least Blackwell informative form of learning that allows $P(a, \omega)$ to be chosen. Hence we can use it to define the induced cost function

$$\hat{K}(P) \equiv K(Q_P) \tag{9}$$

on all SDSC $P \in \mathcal{P}(\mu)$ consistent with the prior

This formulation is useful because the domain $\mathcal{P}(\mu)$ is a compact subset of $\mathbb{R}^{|A|\times|\Omega|}$, which makes for a simple definition of the cost function as continuous. Continuity is useful since it implies that the attention production set is closed and that optimal strategies exist for all levels of marginal cost. It is not easy to think of interesting cost functions for which this condition would fail. One example would be if it was very expensive to rule out any state but costless to arrive at any interior distribution of posteriors. In this case costs would jump up at full knowledge of any state and there would be no optimal strategy since learning more about each state is always beneficial, yet learning everything impossible.

Axiom 1 *The induced cost function $\hat{K}(P)$ is continuous on $P \in \mathcal{P}(\mu)$.*

With this we establish closedness of the attention production set and that supremum in the attention production function is always achieved.

Proposition 2 *If Axiom 1 holds, the attention production set Y is closed and, given $x \geq 0$ there exist strategies $\hat{Q}(x) \in Q(\mu)$ that achieve the supremum.*

$$g(x) = \hat{f}(\hat{Q}(x)) \tag{10}$$

Proof. We first define the attention production set using the SDSC formulation,

$$\mathcal{Y} \equiv \{(x, y) \in \mathbb{R}^2 | \exists P \in \mathcal{P}(\mu) \text{ s.t. } f(P) \geq y, \hat{K}(P) \leq x\}, \tag{11}$$

where $f(P) = \sum_{(a,\omega)\in A\times\Omega} f(a,\omega)P(a,\omega)$. Now take a convergent sequence $(x_n, y_n) \in \mathcal{Y}$. We must show that their limit, which we denote (x, y), is in $\mathcal{Y}$. The first step is to find corresponding $P_n \in \mathcal{P}(\mu)$ such that $f(P_n) \geq y_n$ and $K(P_n) \leq x_n$. Since $\mathcal{P}(\mu)$ is a compact subset of $\mathbb{R}^{|A|\times|\Omega}$, we know that there is a convergent subsequence with limit $P \in \mathcal{P}(\mu)$: for notational simplicity we remove others from the sequence and relabel so that,

$$\lim_{n\longrightarrow\infty} P_n(a, \omega) = P(a, \omega), \tag{12}$$

for all (a, ω). Given Axiom 1, we know that costs converge, with the same being true for output by standard limit arguments,

$$\hat{K}(P) = \lim_{n\longrightarrow\infty} \hat{K}(P_n) \leq x; \tag{13}$$
$$f(P) = \lim_{n\longrightarrow\infty} f(P_n) \geq y, \tag{14}$$

establishing that indeed $(x, y) \in \mathcal{Y}$ and completing the proof that the attention production set is closed.

To establish that there exist strategies $\hat{Q}(x) \in Q(\mu)$ that achieve the supremum for any $x \geq 0$ we again work with the SDSC definition to rewrite the definition of $g(x)$ as:

$$g(x) = \sup_{\{P\in\mathcal{P}(\mu)|\hat{K}(P)\leq x\}} f(P). \tag{15}$$

Take a sequence $P_n \in \mathcal{P}(\mu)$ heading to the supremum:

$$K(P_n) \leq x;$$
$$\hat{f}(P_n) \geq g(x) - \frac{1}{n}$$

and select a convergent subsequence heading to limit $P^L \in \mathcal{P}(\mu)$. Note by continuity of $\hat{K}(P)$ that $K(P^L) \leq x$ and by continuity of $f(P)$ that $\hat{f}(P^L) = g(x)$. Existence of $Q^L \in Q(\mu)$ with $Q^L_P = P^L$ follows directly from the definitions. ■

It is simplest to consider cost functions that give rise to convex attention production sets to make the analogy with standard production theory tight. This is not universally true: think of a standard search model. But it turns out that there is a simple condition that is in a key sense without loss of generality that guarantees convexity. Mixture feasibility allows for mixing of attention strategies at linear cost.

Definition 1 *Given any two attention strategies Q_0, Q_1 and $\lambda \in (0, 1)$ define the corresponding mixture strategy Q_λ by taking the appropriate weighted average of the probabilities of the posteriors. For any posterior γ possible only in Q_0, assign $Q_\lambda(\gamma) = \lambda Q_0(\gamma)$, for any posterior γ possible only in Q_1, assign $Q_\lambda(\gamma) = (1-\lambda)Q_1(\gamma)$, while for posteriors γ possible in both assign the corresponding weighted average probability,*

$$Q_\lambda(\gamma) = \lambda Q_0(\gamma) + (1-\lambda)Q_1(\gamma). \tag{16}$$

*The cost function satisfies **mixture feasibility** if*

$$K(Q_\lambda) \leq \lambda K(Q_0) + (1-\lambda)K(Q_1). \tag{17}$$

Axiom 2 *The cost function satisfies mixture feasibility.*

Caplin and Dean (2015) show that mixture feasibility is without loss of generality in data that is consistent with rational inattention theory. It is also not strictly necessary for much that follows. On the surface, for example, mixture feasibility rules out standard search theory if one uses the assumption that the decision to search is deterministic. But if one allows a coin flip to decide, fixed search costs can be accommodated. Even if one rules out mixing, assuming it is possible still identifies all optimal strategies: whenever there is an optimal strategy that searches probabilistically, there is also one that does so deterministically. What mixture feasibility does is to ensure convexity of key sets, which cuts out a lot of throat clearing but allows optimal strategies to be identified regardless of whether or not mixing is allowed in the model of costly learning. We now show how this links with the shape of the attention production set and the attention production function.

Proposition 3 *Given Axioms 1 and 2, the attention production function is concave and the attention production set is convex.*

Proof. Note that the attention production set is the lower epigraph of the attention production function, hence its convexity follows if we establish concavity of the attention production function. Take $x_1 > x_0 \geq 0$ and note that Proposition 2 shows that with Axiom 1 there exist $\hat{Q}_0, \hat{Q}_1 \in \mathcal{Q}(\mu)$ such that

$$\hat{f}(\hat{Q}_i) = g(x_i); \tag{18}$$

$$K(\hat{Q}_i) \leq x_i; \tag{19}$$

Define the mixture strategy as above,

$$Q_\lambda(\gamma) = \lambda \hat{Q}_0(\gamma) + (1-\lambda)\hat{Q}_1(\gamma). \tag{20}$$

The maximized output is precisely the corresponding weighted average,

$$\hat{f}(Q_\lambda) = \lambda \hat{f}(\hat{Q}_0) + (1-\lambda)\hat{f}(\hat{Q}_1) = \lambda g(x_0) + (1-\lambda) g(x_1) \tag{21}$$

Mixture feasibility (Axiom 2) implies that costs are no higher than the corresponding weighted average,

$$K(Q_\lambda) \leq \lambda K(\hat{Q}_0) + (1-\lambda) K(\hat{Q}_1) \leq \lambda x_0 + (1-\lambda) x_1. \tag{22}$$

Combining these we see that

$$(\lambda x_0 + (1-\lambda)x_1), \lambda g(x_0) + (1-\lambda)g(x_1)) \in \mathcal{Y}, \tag{23}$$

which establishes that,

$$g(\lambda x_0 + (1-\lambda)x_1) \geq \lambda g(x_0) + (1-\lambda)g(x_1), \tag{24}$$

and with it concavity of the attention production function. This completes the proof. ■

Recall from the body of the paper that $c > 0$ indexes the marginal cost of attention. We define the net value function of Q for $c > 0$ as:

$$V(c, Q) = \hat{f}(Q) - cK(Q) \tag{25}$$

Agents adopt attention strategies that maximize net value in equation (5), achieving thereby $\hat{V}(c) \equiv \sup_Q V(c, Q)$, with $\hat{Q}(c)$ denoting an optimal attention strategy. We let $\hat{x}(c)$ and $\hat{f}(c)$ respectively denote optimal expected output and input of attention at the corresponding cost.

Proposition 4 *With Axioms 1 and 2, optimal strategies exist for all $c > 0$.*

Proof. Propositions 1 through 3 show that these axioms imply that $\mathcal{Y}$ is closed and convex, and that $g(x)$ is non-decreasing in x and bounded above. Hence for any $c > 0$ we can find an upper bound $B(c)$ on the level of attentional input above which the marginal return to attention is lower than c: given $x_1 > x_2 \geq B(c)$

$$f(x_1) < f(x_2) + (x_1 - x_2)c. \tag{26}$$

If this were not true we would break the upper bound on expected output identified in proposition 1.

Hence for purposes of optimization, we can restrict attention to the compact set of attention levels $[0, B(c)]$. We have also shown that the attention production function is concave and hence continuous on the interior of its domain. Hence an optimum exists by the standard continuous function on compact set argument. ■

We seek to establish that higher levels of the marginal cost of attention lower both attentional input and expected output. This is precisely analogous to a general proof that the supply curve of a competitive firm is upward sloping.

Proposition 5 *Given $c_1 > c_2 \geq 0$, optimal attention and output are no higher with c_1 than* *with c_2:*

$$\hat{x}(c_1) \le \hat{x}(c_2);$$
$$\hat{y}(c_1) \le \hat{y}(c_2),$$

Proof. Given $c_1 > c_2 \ge 0$ identify optimal inputs $\hat{x}(c_1), \hat{x}(c_2)$ whose existence is established in proposition 4. Find Bayes' consistent distributions $\hat{Q}_1$ and $\hat{Q}_2$ such that $\hat{f}(\hat{Q}_1) = g(x_1)$ and $\hat{f}(\hat{Q}_2) = g(x_2)$. By definition:

$$\begin{aligned} g(\hat{x}_1) - c_1\hat{x}_1 &\ge g(x) - c_1 x \text{ for all } x \ge 0; \\ g(\hat{x}_2) - c_2\hat{x}_2 &\ge g(x) - c_2 x \text{ for all } x \ge 0; \end{aligned}$$

Now consider the following strategy switch: switch $\hat{Q}_2$ to be chosen when the cost is c_1, $\hat{Q}_1$ to be chosen when the cost is c_2. By definition, this switch leaves the sum of maximized outputs unchanged. But the sum of costs change from $c_1\hat{x}_1 + c_2\hat{x}_2$ to $c_2\hat{x}_1 + c_2\hat{x}_1$. Now look at the difference in costs from the switch,

$$(c_2 - c_1)\hat{x}_1 + (c_1 - c_2)\hat{x}_2 = (c_2 - c_1)[\hat{x}_1 - \hat{x}_2]. \tag{27}$$

Since $c_1 > c_2 \ge 0$ we know that the first term in the product on the RHS is strictly negative. Hence to avoid a contradiction in which total costs fall while total output is invariant, we conclude that $\hat{x}_1 \le \hat{x}_2$, establishing the claimed monotonicity. Given monotonicity of g, this further implies that $g(\hat{x}_1) \le g(\hat{x}_2)$, completing the proof.

■

1.2 Posterior Separability and Theorem 1

Given state space Ω and strictly positive prior $\mu \in \Delta(\Omega)$, a cost function is **posterior separable** if there exists a convex function $T(\gamma)$ of posteriors $\gamma \in \Delta(\Omega)$ such that,

$$K(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma)T(\gamma) - T(\mu), \tag{28}$$

where supp Q stands for the set of possible posteriors. With this we are in position to prove theorem 1 from the main text.

Theorem 1 *For any posterior separable cost function $K(\mu, Q) = \sum_\gamma Q(\gamma)T(\gamma) - T(\mu)$ with T bounded and continuous, the attention production set $\mathcal{Y}$ is closed and convex, and the attention production function $g(x)$ is concave.*

Proof. Given propositions 1 through 3 it suffices to show that continuity (Axiom 1) and mixture feasibility (Axiom 2) are satisfied. The latter is direct. By definition, given $Q_0, Q_1 \in Q(\mu)$,

$$K(Q_0) = \sum_{\gamma \in \text{supp } Q_0} Q_0(\gamma)T(\gamma) - T(\mu); \tag{29}$$

$$K(Q_1) = \sum_{\gamma \in \text{supp } Q_1} Q_1(\gamma)T(\gamma) - T(\mu); \tag{30}$$

It is direct from the definition that any mixture strategy Q_λ has precisely the corresponding weighted average cost,

$$K(Q_\lambda) = \sum_{\gamma \in \text{supp } Q_0} \lambda Q_0(\gamma)T(\gamma) + \sum_{\gamma \in \text{supp } Q_1} (1-\lambda)Q_1(\gamma)T(\gamma) - T(\mu) \tag{31}$$

$$= \lambda K(Q_0) + (1-\lambda)K(Q_1). \tag{32}$$

To complete the proof we show continuity of $\hat{K}(P)$ in P. Consider $P_n \in \mathcal{P}(\mu)$ with $\lim_{n \longrightarrow \infty} P_n(a,\omega) = P(a,\omega)$ all (a,ω). Define the corresponding sequence $Q_n = Q_{P_n} \in Q(\mu)$ and $Q = Q_P \in Q(\mu)$. Correspondingly define the unconditional action probabilities,

$$P_n(a) \equiv \sum_{(\overline{\nu}) \in \Omega} P_n(a, \overline{\nu});$$

$$P(a) \equiv \sum_{\overline{\nu} \in \Omega} P(a, \overline{\nu})$$

Define $A_n \subset A$ as all actions with $P_n(a) > 0$ and $A_P \subset A$ as all actions with $P(a) > 0$ and define corresponding revealed posteriors:

$$\gamma_n^a(\omega) \equiv \frac{P_n(a,\omega)}{P_n(a)};$$

$$\gamma_P^a(\omega) \equiv \frac{P(a,\omega)}{P(a)}.$$

To establish continuity of $\hat{K}(P)$ in P we show that,

$$\lim_{n \longrightarrow \infty} \sum_{a \in A_n} P_n(a)T(\gamma_n^a) = \sum_{a \in \bar{A}} P(a)T(\gamma_P^a), \tag{33}$$

In making this limit argument we can restrict attention to actions chosen in A_P since the probability of actions out of A_P falls to zero in the limit and T is bounded

$$\lim_{n\longrightarrow\infty}\sum_{a\in A_n}P_n(a)T(\gamma_n^a)=\lim_{n\longrightarrow\infty}\sum_{a\in A_P}P_n(a)T(\gamma_n^a), \tag{34}$$

Given $a \in A_P$ it is clear that the unconditional action probabilities converge to their limit values

$$\lim_{n\to\infty}P_n(a)=P(a)>0$$

Likewise convergence of revealed posteriors to their limit values is implied by convergence of the limit of the ratio with a non-zero denominator to the ratio of the limits,

$$\lim_{n\to\infty}\gamma_n^a(\omega)=\gamma_P^a(\omega), \tag{35}$$

all $a \in \bar{A}$. Given continuity of T and the limit behavior of action probabilities and revealed posteriors,

$$\lim_{n\longrightarrow\infty}\sum_{a\in A_P}P_n(a)T(\gamma_n^a)=\sum_{a\in A_P}P(a)T(\gamma_P^a). \tag{36}$$

completing the proof. ■

1.3 Theorem 2: Symmetry of the Decision Problem

We now prove that the allocation problem with a symmetric prior satisfies the definition of symmetry due to Bucher and Caplin (2021), which implies that there is an optimal solution with all actions equiprobable and that it is generically unique.

By definition, symmetry of the prior requires that prior beliefs satisfy:

$$\mu\left(\omega\left(1\right),....,\omega\left(M\right)\right)=\mu\left(\omega\left(\beta\left(1\right)\right),....,\omega\left(\beta\left(M\right)\right)\right) \tag{37}$$

for all bijections $\beta : \{1, ..., M\} \to \{1, ..., M\}$ of workers.

To establish the conditions of Bucher and Caplin (2021), is to identify a partition of the state space Ω into equivalence classes $\{\Omega_h\}_{1\leq h\leq H}$ satisfying two symmetry conditions:

1. **State Equivalence:** Given $\omega, \omega' \in \Omega_h$, $\mu(\omega) = \mu(\omega')$ and there exists a bijection $\alpha : A \to A$ such that
$$f(\alpha(a),\omega')=f(a,\omega) \tag{38}$$
2. **Action Equivalence:** Given $a, b \in A$ and $1 \leq h \leq H$, there exists a bijection ρ :

$\Omega_h \to \Omega_h$ such that

$$f(b, \rho(\omega)) = f(a, \omega). \tag{39}$$

For the equivalence classes we define $\omega, \omega' \in \Omega_h$ if and only if there exists a bijection $\beta : \{1, \cdots, M\} \longrightarrow \{1, \cdots, M\}$ of workers such that

$$\omega'(\beta(m)) = \omega(m). \tag{40}$$

We now confirm that this partition has the defining properties that make a decision problem symmetric.

1. **State Equivalence:** Since $\beta : \{1, \cdots, M\} \longrightarrow \{1, \cdots, M\}$ is a bijection exchangeability directly implies that $\mu(\omega) = \mu(\omega')$. Our task now is to construct the bijection $\alpha : A \to A$ such that 38 holds,

$$f(\alpha(a), \omega') = f(a, \omega). \tag{41}$$

To simplify notation, we first relabel tasks and the attention production function so that a is the identify map, assigning worker n to task n so that outputs satisfy,

$$f(a, \omega) \equiv \mathcal{Y}\left(\omega_1(1), \cdots, \omega_M(M)\right). \tag{42}$$

We now define bijection α to allocate worker $\beta(n)$ rather than worker n to task n,

$$[\alpha(a)]^{-1}(n) = \beta(n). \tag{43}$$

Note that $\alpha(a)$ is defined on all workers $1 \leq m \leq M$ since β is a bijection. For precisely this same reason, it is itself a bijection. To establish that

$$f(\alpha(a), \omega') = \mathcal{Y}\left(\omega'_1([\alpha(a)]^{-1}(1), \cdots, \omega'_M([\alpha(a)]^{-1}(M)\right) = f(a, \omega), \tag{44}$$

it suffices to establish that all tasks $1 \leq n \leq M$ are performed at the same level with assignment $\alpha(a)$ in state ω' as in assignment a in state ω. This follows from first applying the definition of $[\alpha(a)]^{-1}$ in equation (43) and then the definition of $\beta(n)$ in equation (48)

$$\omega'_n([\alpha(a)]^{-1}(n)) = \omega'_n(\beta(n)) = \omega_n(n), \tag{45}$$

as required.

2. **Action Equivalence:** To prove action equivalence start with two allocations $a, b \in A$.

Again to simplify we label tasks so that in a worker m is assigned to task m thereby defining $f(a,\omega)$ as,

$$f(a,\omega) \equiv \mathcal{Y}\left(\omega_1(1),\cdots,\omega_M(M)\right). \tag{46}$$

We now specify mapping $\rho : \Omega \longrightarrow \Omega$ by,

$$\rho^{\omega}(b^{-1}(n)) = \omega(n), \tag{47}$$

so that the productivity of worker $b^{-1}(n)$ in state ρ^{ω} is the same as that of worker n in state ω: note that we superscripted the functional dependence on ω. Note that this is defined on all m since b is a bijection, and that is a bijection for the same reason. It is also clear that for any h it maps Ω_h into itself, since $b^{-1}(n)$ is itself a bijection that satisfies the defining property of belonging in the same equivalence class as ω specified in equation (48),

$$\rho^{\omega}(\beta(m)) = \omega(m). \tag{48}$$

What is left is to establish that all tasks $1 \leq n \leq M$ are performed at the same level with allocation b in state ρ^{ω} as with allocation a in state ω

$$\rho^{\omega}_n(b^{-1}(n)) = \omega_n(n) \tag{49}$$

For this it further suffices that the types of the corresponding workers are the same,

$$\rho^{\omega}(b^{-1}(n)) = \omega(n), \tag{50}$$

which is direct from the definition in equation (47).

1.4 Maximum Likelihood Estimator for the Marginal Cost of Information

Consider a finite sequence of symmetric decision problems $1 \leq k \leq K$, each with choice set A_k and states $\mathit{\Omega}_k$. This could be decisions made in the real world by managers or other agents. In our experiment, each decision problem k corresponds to a test item.

We begin by generalizing equation (11) from the main text to express it as a product of choice probabilities over multiple decision problems:

$$P\left(\alpha_j\right) = \prod_{k=1}^{K}\left[\frac{\exp\left\{y\left(a_k,\omega_k\right)\alpha_j\right\}}{\sum_{b_k\in A_k}\exp\left\{y\left(b_k,\omega_k\right)\alpha_j\right\}}\right] \tag{51}$$

We then take logs and define the log likelihood as:

$$L\left(\alpha_j\right)=\sum_{k=1}^{K} y_k \alpha_j-\sum_{k=1}^{K} \ln \left\{\sum_{b_k \in A_k} \exp \left\{y\left(b_k, \omega_k\right) \alpha_j\right\}\right\} \tag{52}$$

Taking the derivative and simplifying yields the following expression for the first order condition:[1]

$$L^{'}\left(\alpha_j\right) \sim \sum_{b \in B} \sum_{k=1}^{K} d_k(b) \exp \left\{\sum_{i=1}^{K}-d_k(b) \alpha_j\right\} \tag{53}$$

where we define $d_k(b)=y_k-y\left(b_k, \omega_k\right)$ as the output difference between the agent's actual assignment and any other counterfactual assignment b in decision problem k. Equation (53) sums over all $b_k \in A_k$ and over $1 \leq k \leq K$ decision problems to yield $L^{'}\left(\alpha_j\right)$, the marginal product of attention and our estimate of allocative skill.

We develop a maximum likelihood estimator for α_j that accounts for three possible cases. First, if an agent always chooses the optimal allocation, e.g. $\sum_{k=1}^{K} d_k(b) \geq 0$ for all $b \in \prod_{k=1}^{K} A_k$, then the likelihood function is strictly increasing in α_j. Thus the best estimate is $\alpha_j^* \rightarrow \infty$, e.g. the marginal cost of attention for agent j is zero.

Second, we consider the case where agents do no better than randomly guessing the correct assignment. Define $G\left(\bar{b}\right) \equiv \sum_{k=1}^{K} d_k(b)>0$ as the set of assignments that yields higher utility than counterfactual assignments and $H\left(\bar{b}\right) \equiv-\sum_{k=1}^{K} d_k(b)>0$ as the set of assignments that yields lower output than counterfactual choices. If $\sum_b G\left(\bar{b}\right) \leq \sum_b H\left(\bar{b}\right)$, the agent has on average done no better than what they could have achieved without acquiring any information. In that case, the best estimate is $\alpha_j^*=0$, e.g. the marginal cost of attention is infinite (or alternatively, the marginal product of attention is zero).

In all other cases, there is a unique solution $\alpha_j^*>0$ that satisfies:

$$\sum_{\left\{\bar{b} \mid \sum_k d_k\left(b_k\right) \geq 0\right\}} G(b) \exp \left(-G(b) \alpha_j^*\right)=\sum_{\left\{\bar{b} \mid \sum_k d_k\left(b_k\right)<0\right\}} H(b) \exp \left(G(b) \alpha_j^*\right) \tag{54}$$

[1] The derivative of (52) with respect to α_j is $L^{'}\left(\alpha_j\right)=\sum_k y_k-\sum_k\left[\frac{\sum_{b_k \in A_k} y\left(b_k, \omega_k\right) \exp \left\{y\left(b_k, \omega_k\right) \alpha_j\right\}}{\sum_{b_k \in A_k} \exp \left\{y\left(b_k, \omega_k\right) \alpha_j\right\}}\right]$. Multiplying through by the grand product of denominators $\prod\left(\alpha_j\right)=\prod_k\left[\sum_{b_k \in A_k} \exp \left\{y\left(b_k, \omega_k\right) \alpha_j\right\}\right]$ and dividing all terms by $\exp \left\{\sum_{i=1}^{K} y_i \alpha_j\right\}$ gives us the expression $L^{'}\left(\alpha_j\right) \sim \sum_{b \in B} \sum_{k=1}^{K}\left(y_k-y\left(b_k, \omega_k\right)\right) \exp \left\{\sum_{i=1}^{K}\left(y\left(b_k, \omega_k\right)-y_k\right) \alpha_j\right\}$. We then define $d_k(b)=y_k-y\left(b_k, \omega_k\right)$.

Empirically, we first compute $d_k(b)$ for all counterfactual assignments and over all decision problems. We then divide them into gains and losses and sum them up to obtain $G(b)$ and $H(b)$ respectively. Finally. we solve equation (54) for the only remaining unknown variable, our estimate of allocative skill α_j^*.

References

Bucher, S. F. and Caplin, A.: 2021, Inattention and inequity in school matching, *Technical report*, National Bureau of Economic Research. 1.3, 1.3

Caplin, A. and Dean, M.: 2015, Revealed preference, rational inattention, and costly information acquisition, *American Economic Review* **105**(7), 2183–2203. 1.1, 1.1, 1.1

Matějka, F. and McKay, A.: 2015, Rational inattention to discrete choices: A new foundation for the multinomial logit model, *American Economic Review* **105**(1), 272–298. 1.1