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FROM HERE FORWARD NOT FOR PUBLICATION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) (1)	Decision Intensity (weighted) (2)	Employment Share (3)	Share with BA (4)	Wage and Salary Income (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 - Higher Marginal Cost of Information Predicts Lower Income

	(1)	(2)	(3)	(4)	(5)
Marginal Cost of Information	-123	-98	-85	-87	-135
	[42]	[40]	[41]	[41]	[51]
Nonverbal IQ (Ravens)			2,066	2,175	1,736
			[1,456]	[1,530]	[1,601]
Cognitive Reflection Test				740	1,148
				[1,725]	[1,897]
Berlin Numeracy Test				-1,358	-1,986
				[1,685]	[1,797]
Demographic Controls		X	X	X	X
ACS Weights					X
R-Squared	0.016	0.183	0.185	0.185	0.198
Sample Size	1,005	1,005	1,005	1,005	1,005

Notes: Table A2 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 U.S. survey sample. The marginal cost of information is derived using the analytic solution to the model developed in the Measurement Appendix. 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 A3 - ED Skill Does Not Predict Employment

Danish Registry Sample

	(1)	(2)	(3)
ED Skill (AG Score)	0.010	0.009	0.009
	[0.006]	[0.006]	[0.007]
Demographic Controls		X	X
Population Weights			X
R-Squared	0.001	0.027	0.027
Sample Size	2,822	2,822	2,822

Notes: Table A3 presents estimates from a regression of the probability of being employed (e.g. having nonzero earned income as of 2022) on economic decision-making skill and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The regression is estimated in the Danish Registry sample. The Assignment Game score is normalized to have mean zero and standard deviation one. Demographic controls include gender, indicators for vocational and tertiary degrees, age and age squared, and whether the respondent is married and their number of children. Column 3 weights the data to be nationally representative of the Danish population, see Table 1 for details.

Table A4 - ED Skill Does Not Predict Contractual Work Hours

Danish Registry Sample

	(1)	(2)	(3)
ED Skill (AG Score)	10.2	23.7	9.7
	[12.0]	[11.5]	[13.1]
Demographic Controls		X	X
Population Weights			X
R-Squared	0.001	0.134	0.143
Sample Size	2,297	2,297	2,297

Notes: Table A4 presents estimates from a regression of annual contractual work hours on economic decision-making skill and the additional covariates indicated in each column. Robust standard errors are shown in brackets. The regression is estimated in the Danish Registry sample. The Assignment Game score is normalized to have mean zero and standard deviation one. Demographic controls include gender, indicators for vocational and tertiary degrees, age and age squared, and whether the respondent is married and their number of children. Column 3 weights the data to be nationally representative of the Danish population, see Table 1 for details.

Table A5 - Correlation between Cognitive Assessments and Income

	(1)	(2)	(3)	(4)
ED 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

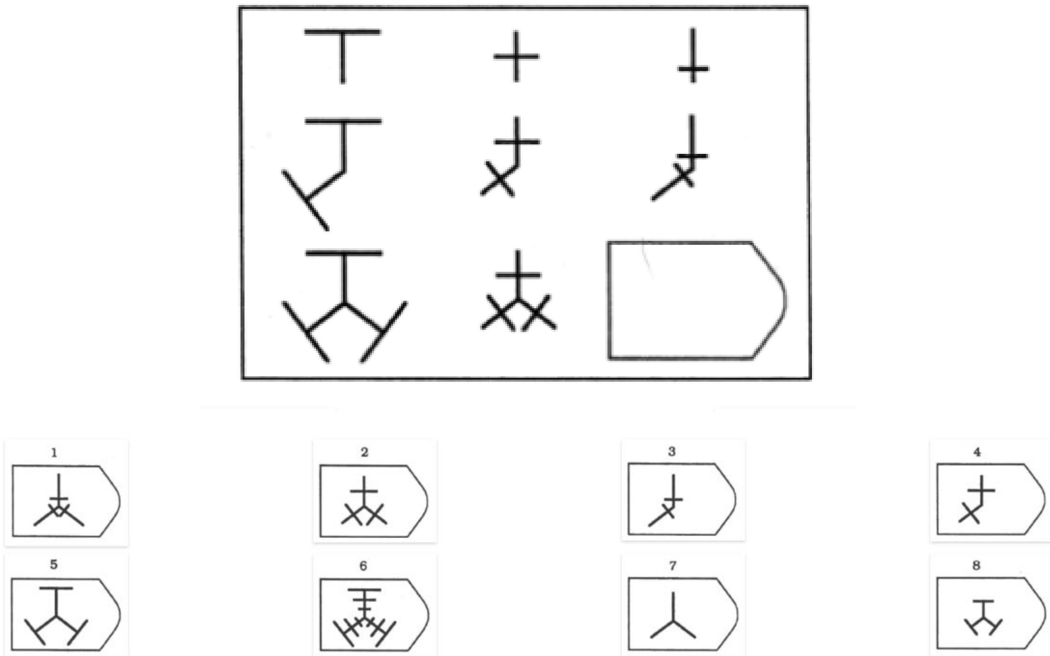
Notes: Table A5 presents estimates from a regression of wage and salary income on each cognitive assessment and demographic controls, which include indicators for gender, race and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Robust standard errors are shown in brackets. The regression is estimated in our U.S. survey sample. The Assignment Game score and all other cognitive assessments are normalized to have mean zero and standard deviation one.

Table A6 - Heterogeneous Relationships between ED Skill and Income

Panel A - U.S. Survey Sample	*Female*	*Male*	*No BA*	*BA*	*Age<=40*	*Age>40*
	(1)	(2)	(3)	(4)	(5)	(6)
ED Skill (AG Score)	1,734	6,808	2,970	5,602	5,152	4,062
	[1,720]	[1,785]	[1,404]	[1,850]	[1,537]	[2,328]
Demographic Controls	X	X	X	X	X	X
R-Squared	0.133	0.216	0.067	0.102	0.140	0.232
Sample Size	372	636	315	693	663	345
Panel B - Danish Registry Sample	*Female*	*Male*	*No BA*	*BA*	*Age<=40*	*Age>40*
	(1)	(2)	(3)	(4)	(5)	(6)
ED Skill (AG Score)	3,932	3,352	3,595	4,355	1,246	6,202
	[867]	[1,048]	[728]	[1,236]	[858]	[926]
Demographic Controls	X	X	X	X	X	X
R-Squared	0.209	0.240	0.234	0.197	0.265	0.162
Sample Size	1,130	1,167	1,098	1,199	1,136	1,161

Notes: Table A6 presents estimates from a regression of wage and salary income on economic decision-making skill and the additional covariates indicated in each column. Panel A presents estimates from the U.S. survey sample, and Panel B presents estimates from the Danish registry sample. Robust standard errors are shown in brackets. The samples vary and are listed in italics. The Assignment Game score is normalized to have mean zero and standard deviation one. Demographic controls in Panel A include indicators for gender, race and ethnicity, and whether the participant has a bachelor's degree, as well as age and age squared. Demographic controls in Panel B include gender, indicators for vocational and tertiary degrees, age and age squared, and whether the respondent is married and their number of children.

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]

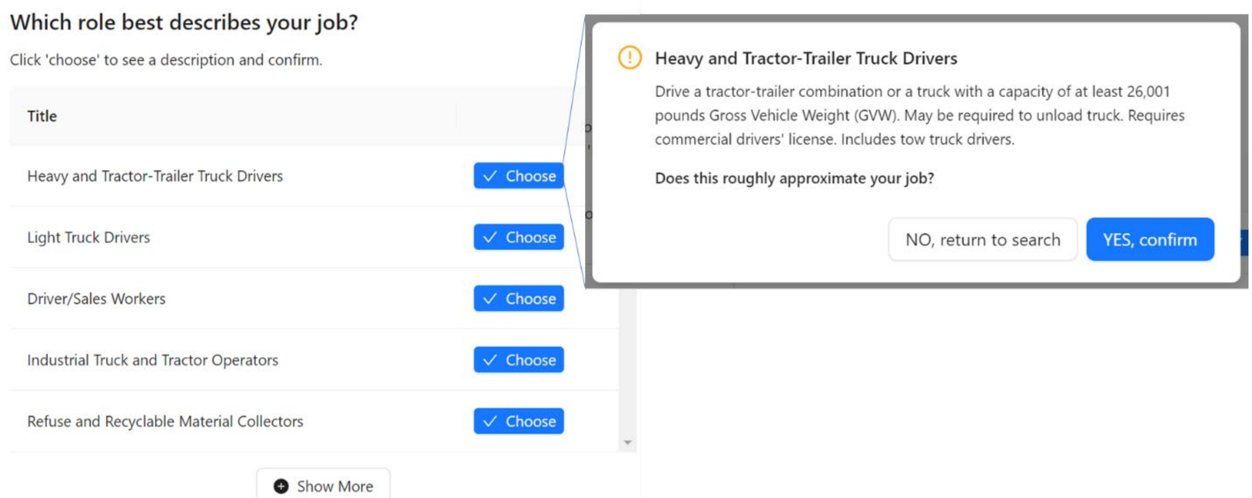
Notes: questions from an expanded version of the Cognitive Reflection Test (CRT). The original CRT is a simple test designed to assess a participant's ability to 'reflect on a question and resist reporting the first response that comes to mind' (Frederick 2005). The original test has 3 questions, and 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.

Theory Appendix to "Economic Decision-Making Skill Predicts Income in Two Countries"

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

April 2024

The first section in this Appendix provides a more extended discussion of the analogy between attention costs in the decision-maker's allocation problem and input costs in standard production theory, followed by a full proof of Theorem 1 in the paper. The second section develops an analytical solution to the model and presents a maximum likelihood estimator that allows us to derive the marginal product of attention for participants using their scores on multiple decision problems.

1 Proof of Theorem 1

Caplin and Dean (2015) establish that the properties of ideal state-dependent stochastic choice data are equivalent to the rational inattention model with a general cost function. They show that any data set with any cost function can be rationalized by one with three special properties:

1. **Weakly increasing in the Blackwell order:** Effectively this means that information can be thrown away for free.

2. **Inattention is free:** Define the **inattentive posterior distribution** as Q_I: it sets $Q_I(\mu) = 1$. Inattention is free is the assumption that $K(Q_I) = 0$. It is just a normalization and implies that costs are always non-negative, since the inattentive posterior is uniquely the least Blackwell informative.

3. **Mixture Feasibility:** This is a much more sophisticated idea but one that is valuable for making the mapping to production theory particularly tight. Caplin and Dean (2015) show that if there exists any cost function that can rationalize data, one exists that allows for mixing of attention strategies at linear cost. They call this *mixture feasibility.* This turns out to be very useful. So we will assume this in what follows.

   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,*

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

   What Caplin and Dean (2015) show is that it is without loss of generality to assume that mixing is possible at the corresponding weighted average cost so that the minimized cost of these posteriors is correspondingly bounded above,

$$K(Q_\lambda) \leq \lambda K(Q_0) + (1-\lambda)K(Q_1). \tag{2}$$

It is this condition that defines the cost function as satisfying mixture feasibility.

Axiom 1 ***Inattention is Free, Weakly Increasing in Blackwell Order, and Mixture Feasibility:*** *The cost function satisfies three conditions:*

1. *It is weakly increasing in the Blackwell order.*

2. *The inattentive distribution costs nothing.*

3. *Mixture Feasibility is satisfied.*

Another important perspective on costs is provided by defining it on a different domain: state dependent stochastic choice (SDSC) functions $P(a, \omega)$. Matějka and McKay (2015) show that one can write the cost function in a given decision problem as a function of the joint distribution of actions and states $P(a, \omega)$. To define this use associate with any $P(a, \omega)$ a **revealed information structure** Q_P by identifying for each chosen action a the corresponding posterior which we denote γ_P^a with,

$$\gamma_P^a(\omega) = \frac{P(a, \omega)}{P(a)} \tag{3}$$

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). 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 on the domain $\mathcal{P}$ comprising all possible $P(a, \omega)$: I use the notation $\hat{K}$ to emphasize the fact that there is an implicit optimization underneath the hood: it describes the cost of learning the least Blackwell informative method of producing the data.

$$\hat{K}(P) \equiv K(Q_P) \tag{4}$$

There are times at which the domain $P(a, \omega)$ is more useful to think about, in particular in establishing existence of optimal strategies etc. The main point here is that the domain $\mathcal{P}$ is a compact subset of $\mathbb{R}^{|A| \times |\Omega}$. The fact that this is closed under the taking of limits is worthwhile. It allows us to define a notion of continuity of costs that is of value.

Definition 2 *Cost function K is continuous if the induced cost function of SDSC $\hat{K}(P)$ is continuous in P (regarding this as a specification of all action and state probabilities in $\mathbb{R}^{|A| \times |\Omega}$*

Axiom 2 *The induced cost function $\hat{K}(P)$ is continuous.*

All reasonable cost functions satisfy this.

1.1 Results

We now establish the key properties of interest that make the analogy with standard production theory essentially complete.

Proposition 1 *With A2, the attention production possibility set is closed.*

Proof. The first point to note is that we can define the attention production possibility set equivalently using the SDSC formulation, noting that given $P \in \mathcal{P}$ expected output is simply $f(P) = \sum_{(a,\omega) \in A \times \Omega} f(a,\omega)P(a,\omega)$ so that,

$$\mathcal{Y} \equiv \{(x,y) \in \mathbb{R}^2 | \exists P \in \mathcal{P} \text{ s.t. } \hat{K}(P) \leq x, f(P) \geq y, \}, \tag{5}$$

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

$$\lim_{n \longrightarrow \infty} P_n(a,\omega) = \bar{P}(a,\omega), \tag{6}$$

all (a, ω).

With A1, we know that costs converge, with the same being true for utilities directly,

$$\begin{aligned} \hat{K}(\bar{P}) &= \lim_{n \longrightarrow \infty} \hat{K}(P_n) \leq \bar{x}; & (7) \\ f(\bar{P}) &= \lim_{n \longrightarrow \infty} f(P_n) \geq \bar{y}, & (8) \end{aligned}$$

establishing that indeed $(\bar{x}, \bar{y}) \in \mathcal{Y}$ and completing the proof. ■

As is standard for production, optima can be found equivalently by identifying points on the boundary of the production possibility set that maximize $g(x) - cx$ and then identifying the learning strategies that produce these points and the corresponding net utilities.

Definition 3 *Given $c > 0$, corresponding optimal strategies $\hat{Q} \in \hat{\mathcal{Q}}(c)$ are those for which*

$$g(K(\hat{Q})) - cK(\hat{Q}) \geq g(x) - cx \tag{9}$$

all $y \geq 0$. Given an optimal strategy $\hat{Q} \in \hat{\mathcal{Q}}(c)$ we define optimal input of attention and gross expected output respectively as $K(\hat{Q}), f(K(\hat{Q}))$. We define the optimal level(s) of attention and gross expected output for c as the union of these across $\hat{Q} \in \hat{\mathcal{Q}}(c)$. We let $(\hat{x}(c), \hat{y}(c))$ denote all pairs of levels of attention and expected output across corresponding optimal strategies. We let $\hat{X}(c)$ denote the set of attentional inputs optimal at c.

Proposition 2 *With A1 and A2, the attention production possibility set is convex and the attention production function is concave, is bounded below by the value of the inattentive strategy Q_I,*

$$g(0) \geq \hat{f}(Q_I) \equiv \max_{a \in A} \sum_{\omega} g(a, \omega)\mu(\omega), \tag{10}$$

and is bounded above by the value of the fully attentive strategy Q_F,

$$\lim_{x \longrightarrow \infty} g(x) \leq \hat{f}(Q_F) = \sum_{\omega} \max_{a \in A} f(a, \omega)\mu(\omega). \tag{11}$$

Proof. Given $(x_0, y_0), (x_1, y_1) \in \mathcal{Y}$ we know that there exists $P_0, P_1 \in \mathcal{P}$ such that:

$$K(P_i) \leq x_i; \tag{12}$$

$$f(P_i) \equiv \sum_a \sum_\omega u(a.\omega)P_i(a, \omega) \geq y_i; \tag{13}$$

$i = 0, 1$. Convexity of the attention production set is immediate since $P_\lambda \in \mathcal{P}$ and by direct computation,

$$f(P_\lambda) = \sum_a \sum_\omega u(a.\omega)P_\lambda(a, \omega) = \lambda f(P_0) + (1 - \lambda)f(P_1). \tag{14}$$

Turning now to the attention production function, with A1 we know an optimal strategy exists that uses all available input, so that given two levels of input use x_0, x_1 there exist SDSC $Q_0, Q_1 \in \mathcal{Q}(\mu)$ such that:

$$K(Q_i) = x_i; \tag{15}$$

$$\hat{f}(Q_i) = g(x_i); \tag{16}$$

Define the mixture strategy as above,

$$Q_\lambda(\gamma) = \lambda \hat{Q}_0(\gamma) + (1-\lambda)\hat{Q}_1(\gamma). \tag{17}$$

It is known (see Caplin and Dean 2015) that maximized expected output is precisely the corresponding weighted average,

$$\hat{f}(Q_\lambda) = \lambda \hat{f}(\hat{Q}_0) + (1-\lambda)\hat{f}(\hat{Q}_1) = \lambda g(x_0) + (1-\lambda) g(x_1) \tag{18}$$

Mixture feasibility (A2) implies that costs are no higher than the corresponding weighted average,

$$K(Q_\lambda) \leq \lambda K(\hat{Q}_0) + (1-\lambda) K(\hat{Q}_1) = \lambda x_0 + (1-\lambda) x_1. \tag{19}$$

Combining these we see that

$$(\lambda x_0 + (1-\lambda)x_1, \lambda g(x_0) + (1-\lambda) g(x_1)) \in \mathcal{Y}, \tag{20}$$

which establishes that,

$$g(\lambda x_0 + (1-\lambda)x_1) \geq \lambda g(x_0) + (1-\lambda) g(x_1), \tag{21}$$

and with it concavity of the attention production function.

With regard to upper and lower bounds, to establish that $g(0) \geq \hat{f}(Q_I)$ note by inattention is free, $K(Q_I) = 0$. Hence $(0, \hat{f}(Q_I)) \in \mathcal{Y}$ implying directly that $g(0) \geq \hat{f}(Q_I)$.

To establish that $\lim_{x \longrightarrow \infty} g(x) \leq \hat{f}(Q_F)$ 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 for any production function. ■

Proposition 3 *With A1 and A2, optimal strategies exist for all $c > 0$.*

Proof. We have shown that $\mathcal{Y}$ is convex and it is closed by assumption. Given that $g(x)$ is non-decreasing in y and bounded above, for any $c > 0$ we can find an upper bound $Y(c)$ on the level of attentional input above which the marginal return to attention is lower than c: given $y_1 > y_2 \geq Y(c)$

$$f(y_1) < f(y_2) + (y_1 - y_2)c. \tag{22}$$

It this were not true we would break the upper bound on expected output.

Hence for purposes of optimization, we can restrict attention to the compact set of attention levels $[0, Y(c)]$. 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. ■

Proposition 4 *Given A1 and A2, the attention production function $g(x)$ is non-decreasing in x:*

$$x_1 > x_2 \geq 0 \implies g(x_1) \geq g(x_2). \tag{23}$$

Proof. Pick $x_1 > x_0 \geq 0$, and find an optimal strategy $\hat{Q}_0 \in \mathcal{Q}$ such that $\hat{f}(Q_0) = g(x_0)$ and $K(Q_0) \leq x_0$. This see that this implies

$$g(x_1) \geq g(x_0) = \hat{f}(Q_0) \tag{24}$$

note that the LHS is the maximum feasible $\hat{f}(Q)$ with $K(Q) \leq x_1$, while the RHS is feasible since $K(Q_0) \leq x_0 < x_1$.

■

Proposition 5 *Given $c_1 > c_2 \geq 0$, let x_1 be an optimal input choice for c_1 and x_2 for c_2. Then*

$$x_1 \leq x_2 \implies g(x_1) \leq g(x_2).$$

Proof. Find respective optimizing strategies $\hat{Q}_1$ and $\hat{Q}_2$ that achieve the maximum in that $\hat{f}(\hat{Q}_1) = g(x_1)$ and $\hat{f}(\hat{Q}_2) = g(x_2)$. By definition:

$$\begin{aligned} g(x_1) - c_1x_1 &\geq g(x) - c_1x \text{ for all } x \geq 0; \\ g(x_2) - c_2x_2 &\geq g(x) - c_2x \text{ for all } x \geq 0; \end{aligned}$$

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 maximized gross expected output unchanged. But total costs change from $c_1x_1 + c_2x_2$ to $c_2x_1 + c_2x_1$. Now look at the difference in costs from the switch.

$$(c_2 - c_1)x_1 + (c_1 - c_2)x_2 = (c_2 - c_1)[x_1 - x_2]. \tag{25}$$

Since $c_1 > c_2 \geq 0$ 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 gross expected output is invariant, we conclude that $x_1 \leq x_2$, establishing the claimed monotonicity. Given monotonicity of f, this further implies that $g(x_1) \leq g(x_2)$, completing the proof.

■

There is some interest in conditions under which the attention production possibility set is convex without invoking AI which allows for mixture strategies. It turns out that the now standard class of **posterior separable** (PS) cost functions (CDL) ensure this. This is a far reaching generalization of the Shannon cost function. 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,

$$K(\mu, Q) = \sum_{\gamma \in \text{supp } Q} Q(\gamma)T(\gamma) - T(\mu). \tag{26}$$

where supp stands for the set of possible posteriors. It is **uniformly posterior separable** if the same T function applies for all priors on Ω with full support. Shannon is the special UPS cost function in which $T(\gamma) = \sum_\omega \gamma(\omega) \ln \gamma(\omega)$.

The following result shows how well this class of models is suited to the production analogy. When the function T is continuous and bounded we establish closedness and convexity of the attention production possibility set and hence concavity of the attention production function.

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

Proof. It is direct that mixture feasibility (A1) is satisfied. By definition of K being PS, there exists a convex function T such that

$$K(Q_0) = \sum_{\gamma \in \text{supp } Q_0} Q_0(\gamma)T(\gamma) - T(\mu); \tag{27}$$

$$K(Q_1) = \sum_{\gamma \in \text{supp } Q_1} Q_1(\gamma)T(\gamma) - T(\mu); \tag{28}$$

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

$$\begin{aligned} K(Q_\lambda) &= \sum_{\gamma \in \text{supp } Q_0} \lambda Q_0(\gamma)T(\gamma) + \sum_{\gamma \in \text{supp } Q_1} (1-\lambda)Q_1(\gamma)T(\gamma) - T(\mu) && (29) \\ &= \lambda K(Q_0) + (1-\lambda)K(Q_1). && (30) \end{aligned}$$

In light of Proposition 2, whereby A2 implies closedness of $\mathcal{Y}$, closedness follows if we establish that continuity of T implies A2, continuity of K in P. Consider $P_n \in \mathcal{P}$ with $\lim_{n \longrightarrow \infty} P_n(a, \omega) = \bar{P}(a, \omega)$ all (a, ω). Define the corresponding sequence $Q_n = Q_{P_n} \in \mathcal{Q}$

and $\bar{Q} = Q_{\bar{P}} \in \mathcal{Q}$. Correspondingly define the unconditional action probabilities

$$P_n(a) \equiv \sum_{\nu \in \Omega} P_n(a, \nu) \tag{31}$$

where recall that: revealed posteriors.

$$\gamma_n^a(\omega) \equiv \frac{P_n(a, \omega)}{P_n(a)}, \tag{32}$$

What we need to show is that the corresponding weighted average T functions converge:

$$\lim_{n \longrightarrow \infty} \sum_{\gamma \in \text{ supp } Q_n} Q_n(\gamma) T(\gamma) = \sum_{\gamma \in \text{ supp } \bar{Q}} \bar{Q}(\gamma) T(\gamma). \tag{33}$$

One key operation is to separate out by chosen actions rather than posteriors. It is a general proposition that one can define posteriors associated with $P(a, \omega)$ by adding up across chosen actions as according to revealed posteriors. Given $P \in \mathcal{P}$, pick possible actions $a \in A$ with $P(a) > 0$ and define the revealed posteriors,

$$\gamma_P^a(\omega) \equiv \frac{P(a, \omega)}{P(a)}. \tag{34}$$

It is clear that the distribution of posteriors and the revealed posteriors are precisely the same distribution as Q_P, so there is in fact no difference in that sense. All that has happened is that, if more than one action has **the same** revealed posterior in the data, then the action based specification separates them out by association with the corresponding action, while the action-based specification merges by posterior. For current purposes the proof is easier in the action-based formulation. Define $A_n \subset A$ as all actions with $P_n(a) > 0$ and $\bar{A} \subset A$ as all actions with $\bar{P}(a) > 0$. What we need to show then is that:

$$\lim_{n \longrightarrow \infty} \sum_{a \in A_n} P_n(a) T(\gamma_n^a) = \sum_{a \in \bar{A}} \bar{P}(a) T(\bar{\gamma}_n^a), \tag{35}$$

where $\gamma_n^a, \bar{\gamma}_n^a$ are the revealed posteriors of actions chosen in P_n (set A_n) and of actions chosen in $\bar{P}$ (set $\bar{A}$). Given that T is continuous, it suffices to establish that the unconditional action probabilities and revealed posterior converge. Note further that in terms of the limit, we can restrict attention to actions chosen in $A_n \cap \bar{A}$

$$\lim_{n \longrightarrow \infty} \sum_{a \in A_n} P_n(a) T(\gamma_n^a) = \lim_{n \longrightarrow \infty} \sum_{a \in \bar{A}} P_n(a) T(\gamma_n^a), \tag{36}$$

since the probability of actions out of $\bar{A}$ falls to zero in the limit and T is bounded.

We can use standard limit arguments to show that the unconditional action probabilities and revealed posteriors also converge to their corresponding limit values. Given $a \in \bar{P}$ it is clear that the unconditional and conditional action probabilities converge:

$$\lim_{n \to \infty} P_n(a, \omega) = \bar{P}(a, \omega), \tag{37}$$

which holds also for their weighted sum, the unconditional choice probabilities,

$$\lim_{n \to \infty} P_n(a) = \bar{P}(a) > 0 \tag{38}$$

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

$$\lim_{n \to \infty} \gamma_n^a(\omega) = \bar{\gamma}_n^a(\omega), \tag{39}$$

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

$$\lim_{n \longrightarrow \infty} \sum_{a \in \bar{A}} P_n(a) T(\gamma_n^a) = \sum_{a \in \bar{A}} \bar{P}(a)(\gamma) T(\gamma). \tag{40}$$

completing the proof. ■

2 Solving the Model Analytically

The allocation problem as specified fits the definition of symmetry outlined in Bucher and Caplin (2021). To show this formally, partition the state space Ω^M into equivalence classes $\{\Omega_h^M\}_{1 \leq h \leq H}$ satisfying two symmetry conditions:

1. **State Equivalence:** Given $\bar{\omega}, \bar{\omega}' \in \Omega_h^M$, $\mu(\bar{\omega}) = \mu(\bar{\omega}')$ and there exists a bijection $\alpha : A \to A$ such that
$$f(\alpha(a), \bar{\omega}') = f(a, \bar{\omega}) \tag{41}$$

2. **Action Equivalence:** Given $a, b \in A$ and $1 \leq h \leq H$, there exists a bijection $\rho : \Omega_h^M \to \Omega_h^M$ such that
$$f(b, \rho(\bar{\omega})) = f(a, \bar{\omega}). \tag{42}$$

For the equivalence classes we define $\bar{\omega}, \bar{\omega}' \in \Omega_h^M$ if and only if there exists a bijection

$\beta : \{1, \cdots, M\} \longrightarrow \{1, \cdots, M\}$ of workers such that

$$\omega'(\beta(m)) = \omega(m). \tag{43}$$

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 $\mu(\bar{\omega}) = \mu(\bar{\omega}')$. Our task now is to construct the bijection $\alpha : A \to A$ such that 41 holds,

$$f(\alpha(a), \bar{\omega}') = f(a, \bar{\omega}). \tag{44}$$

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

$$f(a, \bar{\omega}) \equiv F\left(\omega_1(1), \cdots, \omega_M(M)\right). \tag{45}$$

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

$$[\alpha(a)]^{-1}(n) = \beta(n). \tag{46}$$

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

$$f(\alpha(a), \bar{\omega}') = F\left(\omega_1'([\alpha(a)]^{-1}(1), \cdots, \omega_M'([\alpha(a)]^{-1}(M)\right) = f(a, \bar{\omega}), \tag{47}$$

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 (46) and then the definition of $\beta(n)$ in equation (51)

$$\omega_n'([\alpha(a)]^{-1}(n)) = \omega_n'(\beta(n)) = \omega_n(n), \tag{48}$$

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 $f(a, \bar{\omega})$ as,

$$f(a, \bar{\omega}) \equiv F\left(\omega_1(1), \cdots, \omega_M(M)\right). \tag{49}$$

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

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

so that the productivity of worker $b^{-1}(n)$ in state $\rho^{\bar{\omega}}$ is the same as that of worker n in state ω: note that we superscripted the functional dependence on $\bar{\omega}$. 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 $\bar{\omega}$ specified in equation (51),

$$\rho^{\bar{\omega}}(\beta(m)) = \omega(m). \tag{51}$$

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 $\rho^{\bar{\omega}}$ as with allocation a in state $\bar{\omega}$

$$\rho_n^{\bar{\omega}}(b^{-1}(n)) = \omega_n(n) \tag{52}$$

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

$$\rho^{\bar{\omega}}(b^{-1}(n)) = \omega(n), \tag{53}$$

which is direct from the definition in equation (50). This completes the proof of action equivalence and with it the theorem.

A virtue of symmetry is that it gives rise to a simple expression for the maximum likelihood estimator of the Shannon cost parameter. The key to the interest in these decision problems is that the unweighted logit is an optimal strategy: all actions are equally likely ex ante and the probaility of each action in each state is defined by the unweighted logit formula of Matějka and McKay (2015),

$$P(a|\bar{\omega}) = \frac{\exp\{f(a,\bar{\omega})y\}}{\sum_{b\in A}\exp\{f(b,\bar{\omega})y\}} \tag{54}$$

where $y > 0$ is the reciprocal of the marginal cost of reducing posterior expected entropy in the Shannon cost function, so that higher values correspond to lower such costs. If $\{\exp\{f(b,\bar{\omega})y\} : b \in A\}$ are affine independent, this is the unique solution. In addition to symmetry, this shows that in the Shannon model probabilities depend only on what *is* true, not on what else *might have been* true.

2.1 Maximum Likelihood Estimator

In our experiment we consider a finite sequence of symmetric decision problems $1 \leq k \leq K$, each with choice set A_k and states Ω_k. Each decision problem k corresponds to a test item.

We begin by generalizing equation (54) to express it as a product of choice probabilities over multiple decision problems:

$$P\left(\alpha_j\right) = \prod_{k=1}^{K} \left[\frac{\exp\left\{f\left(a_k, \omega_k\right)\alpha_j\right\}}{\sum_{b_k \in A_k} \exp\left\{f\left(b_k, \omega_k\right)\alpha_j\right\}} \right] \tag{55}$$

To simplify notation, we let $f_k = f\left(a_k, \omega_k\right)$ be the realized reward in the numerator. We then take logs and define the log likelihood as:

$$L\left(\alpha_j\right) = \sum_{k=1}^{K} f_k \alpha_j - \sum_{k=1}^{K} \ln\left\{ \sum_{b_k \in A_k} \exp\left\{f\left(b_k, \omega_k\right)\alpha_j\right\} \right\} \tag{56}$$

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

$$L^{'}\left(\alpha_j\right) \sim \sum_{b \in B} \sum_{k=1}^{K} d_k\left(b\right) \exp\left\{ \sum_{i=1}^{K} -d_k\left(b\right)\alpha_j \right\} \tag{57}$$

where we define $d_k\left(b\right) = f_k - f\left(b_k, \omega_k\right)$ as the output difference between the agent's actual assignment and any other counterfactual assignment b in decision problem k. Equation (57) 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. $\sum_{k=1}^{K} d_k\left(b\right) \geq 0$ 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 $\alpha_j^* \to \infty$, 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 $G\left(\bar{b}\right) \equiv \sum_{k=1}^{K} d_k\left(b\right) > 0$ as the set of assignments that yields higher utility than counterfactual assignments and $H\left(\bar{b}\right) \equiv -\sum_{k=1}^{K} d_k\left(b\right) > 0$ 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

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

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:

$$\sum_{\left\{\bar{b} \mid \sum_k d_k(b_k) \geq 0\right\}} G(b) \exp\left(-G(b)\, \alpha_j^*\right) = \sum_{\left\{\bar{b} \mid \sum_k d_k(b_k) < 0\right\}} H(b) \exp\left(G(b)\, \alpha_j^*\right) \tag{58}$$

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. 2

Caplin, A. and Dean, M.: 2015, Revealed preference, rational inattention, and costly information acquisition, *American Economic Review* **105**(7), 2183–2203. 1, 3, 3, 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, 2