Inattention in School Choice: Theory and Evidence

Bennett Smith-Worthington
Miguel Urquiola [1]

December 2, 2024

Preliminary draft, not for circulation

Abstract

This paper explores the role of inattention in the context of school choice, addressing why informational interventions in education markets often yield limited improvements in household beliefs or actions. We first present a series of empirical results which use both descriptive and experimental data from households in Romania during the high school choice process. These results demonstrate a) households' beliefs about educational tracks' value-added and selectivity exhibit systematic, non-random errors and b) their beliefs reflect partial attention in the form of anchoring and adjustment. We then analyze the results of an experimental intervention which provides information to households about educational tracks' value-added, and identify that households only respond to information when they express uncertainty about their own beliefs. We rationalize these empirical findings using a model of rational inattention, which formalizes households' process of weighing the costs and benefits of learning information. This model's key predictions explain the empirical results, which serves to emphasize the importance of understanding cognitive attention mechanisms to design effective interventions for improving educational outcomes.

[1]Smith-Worthington: Harvard University; bennettsmithworthington@fas.harvard.edu. Urquiola: Columbia University; msu2101@columbia.edu. Kate Kennedy-Moulton and Juan Martin Facal provided excellent research assistance.

1 Introduction

In seminal work, Friedman (1955) suggested that a fundamental market intuition—that competition among firms improves product quality—extends to the market for education. Following this simple argument, many governments promoted competition using tools like school vouchers and centralized school choice. Surprisingly, the results have often been disappointing.

Recent research suggests that this is because a central dimension of school quality—value-added—is difficult for households to observe (Ainsworth et al. 2023). Since value-added can be opaque, it is not surprising that households do not always prefer higher value-added schools (e.g., Abdulkadiroglu et al. 2014, 2017). However, if households do not prioritize value-added, then the demand for a school's services will generally be inelastic to its value-added, blunting schools' incentive to invest in this attribute (Rothstein 2006).

An obvious solution to this problem is to inform consumers by distributing data on school value-added. However, that has also produced disappointing results. The impacts of information provision are not homogeneous nor always intuitive. Abdulkadiroglu et al. (2020) and Beuermann et al. (2023) attribute this to the fact that even if households can observe school value-added, they discount it in favor of other dimensions of school quality.

This paper empirically and theoretically studies an alternative possibility—that when households choose schools, they are *rationally inattentive*. That is, in deciding to learn, retain, and use information, households trade off the costs and perceived benefits of information. This helps explain why households do not absorb all data provided, even in a high-stakes setting. Further, it helps predict *who* will respond to *what* information.

Specifically, we revisit data from Ainsworth et al. (2023), and present three sets of new results:

1. Descriptive results — While households hold mistaken beliefs about schools' quality, rather than making random errors, they seem to weigh the cost and value of information acquisition. Further, their beliefs exhibit reliance on coarse heuristics and display "anchoring and adjustment".
2. Experimental results — While information on value-added can lead households to change their beliefs and preferences, households' *uncertainty*—likely reflecting the cost of information—determines whether this happens. Information only has an impact among: a) households that express uncertainty about their choices, and b) schools about which households did not report beliefs (presumably at least in part due to uncertainty).
3. Theoretical results — We follow Gabaix and Graeber (2024) to build a model in which rationally inattentive households choose over a discrete number of schools. The descriptive results inform how we build the model, and the model yields predictions that account for the role of uncertainty

in the experimental results.

The use of a model of rational inattention to study household responses to information in a school choice setting helps to resolve a persistent puzzle in the education literature. The fact that parents typically don't have accurate beliefs about value-added, and as a result do not choose schools based on value-added, is one of the central barriers that prevents school competition from improving educational outcomes. Education economists have hoped that information provision could resolve this problem, but the underperformance of these interventions demonstrates that more must be done to understand the mechanisms behind attention to information, rather than simply modifying the availability of said information. Just as properties of supply and demand curves need to be known in order to calculate the welfare impacts of taxes or subsidies, so too must the supply curve for *attention* be understood in order to properly understand the impacts of information provision.

The first empirical section makes use of data from a survey with Romanian households before the start of the high school choice period. In Romania, students attend educational tracks, which are 'schools within schools' in the sense that they are self-contained units with a specific curricular focus (like mathematics or natural science) that are located within a larger high-school. Having collected households' baseline beliefs about various dimensions of track quality, such as value-added and selectivity, we find that parents' beliefs display sizable and systematic errors. Additionally, these errors are large even for dimensions of quality for which information is publicly available. However, the magnitude of these errors is decreasing in true quality: their beliefs are less erroneous for the more prestigious educational tracks. This is consistent with a scenario in which better tracks are more well-known, and as a result information is more available in comparison to lower-performing tracks. We also present evidence of 'anchoring and adjustment', whereby households' beliefs about track quality are correlated with coarse proxies for track quality, like the school's average track quality.

Finally, we also make use of the fact that households do not submit beliefs about all tracks in the baseline survey—they choose which tracks about which to submit beliefs, implicitly delivering a measure of how confident they are in their beliefs. Although they could surely produce some guess of the selectivity of a track they only vaguely know, their decision to not report a belief demonstrates a high level of uncertainty about its quality. Matching the results just discussed, we also find that more households report beliefs about the truly better tracks; this again matches a story whose key explanatory variable is informational costs. This first empirical section demonstrates that households make systematic and large errors about the quality of educational tracks in their town.

The experimental section then goes on to analyze the impacts of an informational intervention in which parents in the treatment group are given information about the value-added of the tracks in their town. The empirical results here take those in Ainsworth et al. (2023) as a starting point: when one looks at all tracks and all households together, the treatment seems to have had very little impact on beliefs, and very little impact on endline preferences. We demonstrate that the treatment has an impact on both beliefs and preferences when the sample is restricted to those who exhibit high levels of uncertainty about their beliefs. We find that individuals' perceptions become more accurate with treatment if they were uncertain about the quality of the track, and we find that households change their preference rankings to prioritize value-added for tracks about which they were uncertain and that they previously thought were low-quality. Thus, the experimental section demonstrates that a key predictor of response to information is households' uncertainty about their own beliefs.

We explain the mechanisms that would produce the empirical results in the descriptive and experimental sections with a model of rational inattention. This model of rational inattention (taken from Gabaix and Graeber 2024) demonstrates how individuals will weigh the costs and benefits of information, where the benefits of information are demonstrated to be related to the degree to which the information can reduce uncertainty. We model the experimental intervention as a shock to the informational costs, and derive predictions about which households and about which tracks the treatment will improve attention and would lead to higher preference rankings. The key forces are uncertainty, costs of information, and prior preferences. The model predicts that individuals will respond most to information about the tracks which were very costly to observe in baseline, but become relatively cheaper with the treatment. Further, individuals will increase their preference ranking for tracks for which a) they are paying attention to information but also b) they had underestimated its true quality in baseline.

Our paper is not the first to study "behavioral" subjects who respond imperfectly to information or incentives in a choice environment. However, many of the prominent papers in the literature (e.g., Dean and Neligh 2023) make use of specialized choice settings that generate tight predictions but link less directly to real-world decision making. We study an experimental intervention in a real, high-stakes choice environment. Further, the main discussion and analysis of the role of uncertainty in predicting attention to information fits strongly with a quickly growing literature on 'cognitive uncertainty', which is defined as the degree to which individuals are unsure about their utility-maximizing decision (see Enke and Graeber 2023). Finally, (to our knowledge) ours is the first application of rational inattention to school choice and the first to use the model of Gabaix and Graeber (2024) to derive and empirically test predictions.

Our work also relates to papers on how informed households are about school quality (Arteaga et al. 2022), and to those that study the effects of information on school markets. Several papers show positive effects from information related to schools' *absolute* achievement (Hastings and Weinstein 2008, Andrabi et al. 2017, Ajayi et al. 2017, Corcoran et al. 2018; Allende et al. 2019) but limited impacts from information related to value-added (Mizala and Urquiola 2013, Imberman and Lovenheim 2016, Agte et al. 2024).

The rest of the paper is structured as follows. Section 2 provides background on the empirical setting. Section 3 provides descriptive results on households' beliefs about school quality. Section 4 presents the experimental results, and section 5 proposes a model to explain the empirical findings. Section 6 concludes.

2 Setting, surveys, and experiment

Romanian students completing middle school apply to high school *tracks*. While a single high school building may contain multiple tracks, each track is a self-contained unit (a "school-within-a-school"). Students in a given track take all their classes together; they never receive instruction with students from another track.

One reason for this is that each track has a specific curricular orientation. For example, one may be focused on mathematics, one on literature, one on social studies, etc. Related to this, tracks tend to place students on paths toward specific types of higher education (e.g., students completing math or science tracks are more likely to enter a STEM college major later).[1]

The track application process consists of the following steps:

1. Students find out their (middle to high-school) "transition score." This is the unweighted average of their middle school GPA and their score on a national high school admissions exam.
2. Students submit a ranked list of the tracks they wish to attend; this list can be long.[2]
3. The Ministry of Education uses a serial dictatorship mechanism to assign students to tracks. The mechanism takes the highest-scoring student and assigns her to her most preferred track. It then takes the second-highest-scoring student and does the same, and so on. Eventually, it reaches a student whose first choice is full. It attempts to assign her to her second-most preferred; if that is also full, it looks for space at her third-most preferred, and so on.

This process is incentive-compatible: it is rational for students to state their preferences truthfully. It is also our sense that the process is widely understood. For example, it is explained to parents during annual "application night" events that will be relevant below.

This process also induces sorting by ability, as higher-scoring students tend to congregate in

[1] For further institutional details, see Ainsworth et al. (2024) and Malamud et al. (2024).

[2] The list may contain up to 287 entries.

given tracks.[3] As a result, each track can be characterized by a cutoff or minimum transition score (MTS)—the minimum score required for admission. We rely on those cutoffs to measure track selectivity. They are also a proxy for the demand tracks experience; a track that is in high demand will fill early and have a high MTS.

Finally, at the end of high school, students take the national *Baccalaureate* exam. This is a high-stakes test; passing is required to enroll in academic higher education and to access many desirable careers. Passing the Baccalaureate is not trivial for many students—the national pass rate is about 53%.[4] Thus, passing is a salient outcome for high schoolers.

2.1 Baseline survey

Our empirical analyses rely on data collected by Ainsworth et al. (2023) using a baseline and an endline survey. They administered these surveys in 48 towns that each had between 7 and 28 tracks—towns large enough to provide households with a significant choice among tracks but also small enough for households to know a reasonable fraction of the tracks available. Parents from 3,898 households received the baseline survey at 194 middle school-sponsored "application nights." These annual sessions provide parents with information a few weeks before their children submit their ranked track choices.

As sessions ended, parents received survey questionnaires asking them to list (in ranked order by preference) the tracks they expected to include; for the rest of this paper, we call the rankings they gave each track (if any) their *preferences*. They also had to indicate their certainty that their listing would not change before submission.

We will interpret this self-reported certainty as a proxy for households' certainty about their beliefs regarding track quality. In other words, we will assume that their answer indicates their certainty about tracks' traits as opposed to certainty about what traits they care about. For example, parents may prefer schools with excellent teachers, but they may be unsure about which tracks have such teachers; as a result, they report uncertainty about their track ranking.

In addition, parents were asked to assign 1-5 scores to each track along dimensions including:

- *Selectivity*—households were asked to score each track in response to the prompt: "This track attracts academically gifted students,"
- *Value-added*—households were asked to score each track in response to the prompt: "This track will help my child pass the Baccalaureate exam."

[3] Theoretical models predict stratification by ability in this type of setting, e.g., Epple and Romano 1998, MacLeod and Urquiola 2015.

[4] This is the rate for the cohorts we study. The rate has varied over time (Borcan et al. 2017).

Their responses to these questions will be called their *beliefs* about value-added or selectivity throughout the rest of this paper. These questions presented households with tasks of different difficulty. First, selectivity is relatively well understood in Romania—most households are aware that some tracks have higher cutoffs/minimum transition scores (MTS), and the application sessions cover this. Further, many tracks' cutoffs are fairly stable year-to-year. In addition, the Ministry of Education publicizes tracks' prior-year MTS values. In short, a household willing to do some basic research could score track selectivity in its town relatively accurately.

By contrast, value-added is a significantly harder concept to understand. Value-added is typically defined as the causal effect of a track on a child's performance. As such, it requires establishing counterfactuals about a child's outcomes at different tracks. This is hard to do, even for researchers using extensive data. Further, the Ministry does not produce data on track value-added. As a result, a household wishing to establish a track's value-added would face a relatively significant cost. Beyond that, this cost likely varies across tracks. For instance, some tracks may have well-known reputations or be known to a household's friends; others may be harder to research.

2.2 Experiment

An experiment followed the baseline survey. At randomly selected control middle schools, parents received a flyer containing links to publicly available information about track selectivity. At treatment middle schools, these flyers had two additional ingredients:[5]

- a brief explanation/definition of value-added, and
- a ranking of the town's tracks by value-added (with an explanation that this was calculated by economists in the U.S.).

The treatment induces a reduction in informational costs. In particular, while a given household would typically find it costly to ascertain many tracks' value-added, the treatment suddenly reduces this cost to a level common across tracks. This commonality (within a household) is a feature of the experiment we will leverage in the model below.[6]

2.3 Endline survey

Once high school assignments were complete, households were contacted for a follow-up phone survey. They were asked for the final track preference rankings they had submitted. In addition,

[5] The middle schools were assigned to treatment and control groups based on a clustered randomization process. One criterion used to select towns for the survey and experiment was how predictable value-added was among its tracks. See Ainsworth et al. (2023) for details.

[6] We assert that costs are equalized across tracks but not necessarily across households. Different households may experience the shock differently, e.g., they may have different capacities to absorb concepts or information.

interviewers asked them to again score tracks on a scale of 1 to 5, in this case about value-added only. The follow-up scores allow us to see how the treatment affected the number of tracks that households ranked and the accuracy of their beliefs about the tracks' value-added.

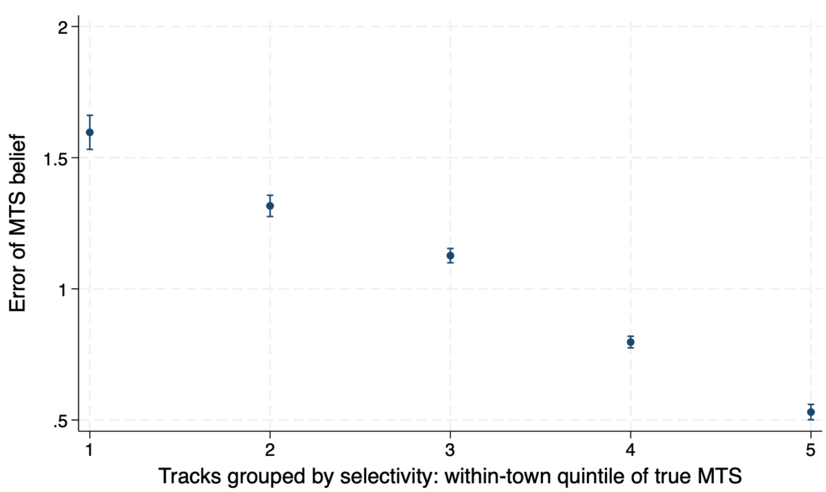
3 Descriptive results

Two sets of empirical results motivate our model: descriptive and experimental. This section covers the first, showing that households hold erroneous beliefs about track quality. These errors are not random—they display systematic patterns consistent with households weighing the costs and value of information acquisition. In addition, we show that behavior consistent with "anchoring and adjustment" arises in this high-stakes, non-laboratory setting. These empirical regularities suggest the existence of imperfect attention, which motivates the model in Section 5.

3.1 Costs of information

Figure 1 (page 7) begins to characterize the errors in households' beliefs. The x-axis groups the tracks in each town into quintiles of selectivity. Quintile 1 (extreme left) contains the least selective tracks—the 20 percent with the lowest cutoffs or minimum transition scores (MTS). Quintile 5 (extreme right) contains the most selective.

Figure 1: Errors in beliefs about selectivity/MTS



To build the y-axis, we begin with the 1-to-5 scores households gave tracks during the baseline survey. In Romania, respondents often interpret such scales as quintiles. While the survey did not ask households to group tracks into equal-sized bins, respondents' answers approximately exhibit that interpretation—the frequencies of each numerical score are close to 0.2. Thus, we interpret the scores to correspond to a household's approximate expectation about a track's quintile.[7] Given this interpretation, we define the error of household i's belief about track k's selectivity to be $\mathcal{E}_{ik,\text{MTS}} \equiv \left|x_{k,\text{MTS}} - x^s_{ik,\text{MTS}}\right|$, where $x_{k,\text{MTS}}$ denotes the true quintile of selectivity of track k and $x^s_{ik,\text{MTS}}$ denotes the $s$ubjective perception (i.e., belief) that household i holds about track k's selectivity.

The y-axis presents the average error observed for each quintile. For example, the maximum error—a value of 4—would arise if a household gave all tracks in quintile 1 the highest selectivity score, a 5.

Two facts stand out. First, overall, the errors are significant: on average, households are off by more than a quintile. This is not necessarily to be expected, since track selectivity (as proxied by MTS) is salient in Romania, and information on tracks' previous-year MTS is publicly available. Fully attentive households could easily form correct beliefs about track selectivity. Thus, households are not (on average) fully attentive. Second, the errors in households' beliefs decline markedly (and statistically significantly) as track selectivity increases. For the least-selective quintile, households are off by about 1.6 points; for the most-selective, by less than 0.6.

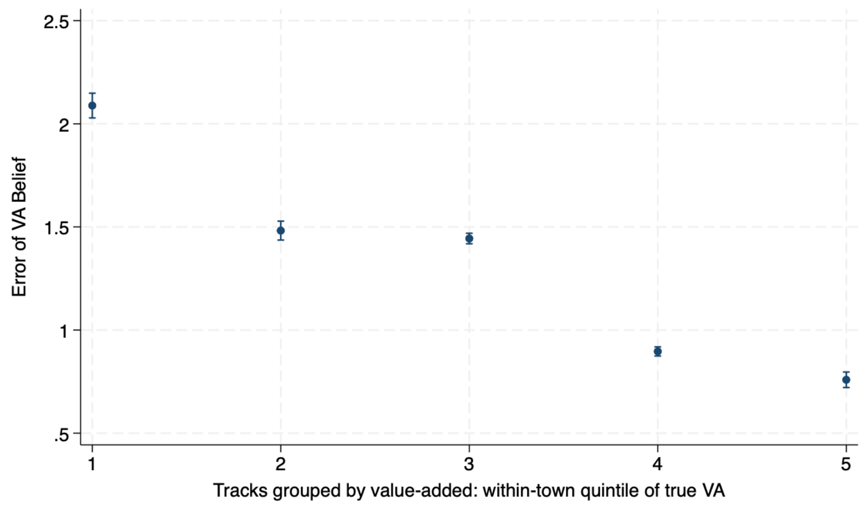
The latter is consistent with commonly observed patterns about the costs of information acquisition in educational markets. Namely, it is often less costly for households to get information about selective, "elite" schools. To illustrate, in New York City, it is difficult for students and parents—even those not in high-achievement middle schools—to avoid hearing that Stuyvesant is the most selective high school in town. Similarly, at the college level, it is difficult for families not to hear about Harvard's extreme selectivity. By contrast, learning about the selectivity of lower-ranking schools may require more deliberate action.

Figure 2 (page 9) presents information in the same format as Figure 1. The only difference is that rather than selectivity/MTS, the quality dimension is value-added (as stated, with respect to passing the high-stakes Baccalaureate exam). Two points emerge.

First, on average, households are more mistaken about tracks' value-added than their selectivity (an average error of 1.3 quintiles for value-added vs. 1.0 for selectivity). This is again consistent with casual observation regarding the costs of obtaining information. Value added is a more complex concept than selectivity, and information on it is not publicly available in Romania. Thus, it is not shocking that households achieve less precision on this dimension. Second, as before, households make the largest errors for the lowest-quality schools. Again, this is consistent

[7] This is approximate because households assign somewhat higher scores; see Appendix Table X, page Y.

Figure 2: Errors in value-added beliefs



with higher costs of obtaining information about these schools.

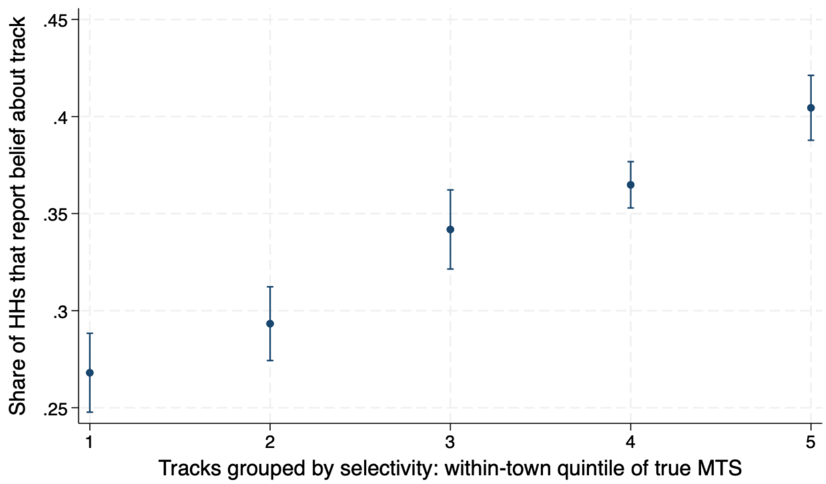
Further, knowing more about a harder-to-access track is not necessarily more valuable than knowing more about a mid-ranked track. By definition, the most selective tracks are "out of reach" for many households. Appendix Figure 5 (page 32) shows that the relationship between erroneous beliefs and true quality is similar whether one considers households with children above or below the median in transition scores. Additionally, it is similar controlling for households fixed effects (Appendix Figure 6, page 32).

While having correct beliefs requires effort, the same holds for having beliefs at all. For example, Arteaga et al. (2022) show that Chilean households are sometimes unaware of schools that are physically close to their home. In such cases, they are not likely to express beliefs when asked. Consistent with this possibility, in Romania, a substantial proportion of households do not report beliefs about some tracks in their town.

Figure 3 (page 10) shows that these patterns are not random but rather are consistent with variation in the costs of obtaining information. Like the previous figure, it groups tracks into quintiles of selectivity (on the x-axis). The y-axis shows the share of households that reported beliefs about tracks' selectivity in that quintile. The pattern is clear: households are more likely to have beliefs about more selective tracks. This is consistent with households being sensitive to information costs. Appendix Figure 7 (page 33) presents an analogous figure for value-added.

Importantly, Figure 3 is also consistent with households having varying confidence in their beliefs about tracks' quality. Although, if pressured, households could surely provide some belief about every track, their decision not to report beliefs about some track is endogenous and may

Figure 3: Share of households with beliefs about selectivity by true selectivity



reflect how much they feel they know about that track. Thus, whether a household provides beliefs about a given track plausibly proxies for its certainty at the track level. In addition, households were asked how certain they were about their overall preference ranking, thus providing an overall household-level indication of certainty. Both levels of certainty (track-specific vs. overall) will be relevant later on.

3.2 Anchoring and adjustment

Households' responses are also consistent with "anchoring and adjustment," in which agents start from a default belief—possibly based on easier-to-observe characteristics—and insufficiently adjust towards the truth.

To illustrate, recall that households in Romania choose *tracks*, self-contained units within high schools. This section presents suggestive evidence that households tend to infer the attributes of tracks from the schools that contain them—i.e., households attach excessive weight to school traits to infer track quality. An analogy would be university applicants inferring the quality of specific programs (e.g., master's programs in anthropology or engineering) based on the overall quality of the university that houses them.

We can test for anchoring and adjustment by comparing the sensitivity of beliefs to true quality. Call household i's belief about track k's quality type n (either value-added or selectivity) $x^s_{ik,n}$ where as before s denotes that it is a subjective belief, and denote the true quality $x_{k,n}$. In regressions below, we test for the presence of anchoring by adding another covariate, $\overline{x}_{k,n}$, which is the average of all tracks' quality type n in the *school* to which track k belongs:

$$x^s_{ik,n} = \beta_1 x_{k,n} + \beta_2 \overline{x}_{k,n} + \epsilon_{ik,n} \tag{1}$$

As a benchmark, Table 1 (page 11), column (1) first presents a univariate regression at the household-track level, excluding $\overline{x}_{k,n}$. The dependent variable is the 1-5 selectivity score that a household assigned to a track k in its town. The independent variable is that track's actual quintile of selectivity, labeled "True track quintile$_k$". If households perceived selectivity perfectly, one might expect the coefficient in column (1) to approximate one: a one-quintile increase in actual selectivity would translate into a one-point increase in scores.

Table 1: Beliefs about selectivity and value-added, all households

	Belief about selectivity$_{ik}$			Belief about value-added$_{ik}$		
	(1)	(2)	(3)	(4)	(5)	(6)
True track quintile$_k$	0.486***	0.171***	0.291***	0.327***	0.0793***	0.207***
	(0.0106)	(0.0101)	(0.00821)	(0.0102)	(0.00828)	(0.00775)
Avg. track quintile in school$_s$		0.490***			0.512***	
		(0.0166)			(0.0191)	
Avg. track quintile in school$_{s\setminus k}$			0.371***			0.383***
			(0.0122)			(0.0142)
Observations	13951	13951	13951	13512	13512	13512
Adjusted R^2	0.248	0.306	0.307	0.116	0.188	0.188
Clusters	2380	2380	2380	2353	2353	2353

The variable True track quintile$_k$ represents track k's quintile of value-added in its town. Avg. track quintile in school$_s$ is the average (over all tracks) quintile of value-added in the school s, where track $k \in s$. Avg. track quintile in school$_{s\setminus k}$ is the same as the previous variable but excludes track k's own value-added in the average. Standard errors are clustered at the household level. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

The estimated coefficient is 0.49 and highly significant. *Prima facie*, this is consistent with households having, on average, a reasonable sense of track selectivity. The fact that the magnitude is less than 1 might reflect noise, or that households were not asked precisely about MTS or to place tracks into quintiles, etc.

Column (2) runs a "horse race," adding $\overline{x}_{k,n}$: the average quintile of selectivity of the *school*, each track belongs to (labeled "Avg. track quintile in school$_s$"). One might expect this additional variable to have no explanatory power—for its coefficient to be zero. In Romania, selectivity is a track-specific characteristic; the selectivity of the school a track belongs to is irrelevant to the track's cutoff score. In addition, previous-year track-specific cutoff scores are publicly available. In short, the school-level variable should not explain households' beliefs.

In stark contrast to this expectation, the coefficient on the school-level selectivity is highly significant and much larger than that on track-level selectivity (0.49 vs. 0.17). This is consistent with households using *school*-level selectivity to infer *track*-level selectivity, i.e., using school-level traits as an "anchor."

Column (3) underlines this point. Rather than include the overall selectivity of the school to which the track belongs (Avg. track quintile in school$_s$), it includes the school's selectivity, *excluding* the track being scored (the third row, labeled "Avg. track quintile is school$_{s\backslash k}$"). This is strictly extraneous information and should not matter for the households' assessment of track selectivity. However, as before, this school-level quality measure better explains households' beliefs about track selectivity than track selectivity itself (a coefficient of 0.37 vs. 0.29).

Columns (4)-(6) replicate the exercise for value-added rather than selectivity. First, the R^2 is significantly lower than for selectivity—less than half the value in column (1). This reflects that, as stated, value-added is likely harder to understand/observe, and information on it is not publicly available. Related, the magnitude of coefficients is almost always smaller for value-added than for selectivity. Second, despite this difference in accuracy, there is still evidence consistent with anchoring and adjustment. Namely, households' beliefs about track value-added seem to be driven by beliefs about the schools the tracks belong to.[8]

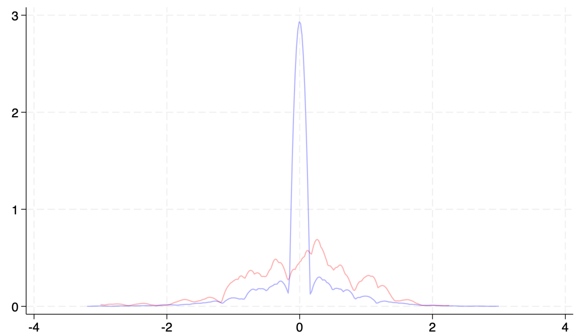
These results imply that households will underestimate the variance of track quality within a school. As stated, this is as if, in evaluating a university, households assumed that all majors within a university ranked similarly relative to those at other schools. This would lead them to underestimate the true across-major variance in quality within the university.

Figure 4 (page 13) shows the data are consistent with this prediction. Panel (a) refers to selectivity, and Panel (b) to value added. In each case, the blue line plots the distribution of the standard deviation of households' beliefs about track-level traits within a given school. For example, at the extreme, many households believe there is no variation in track selectivity within schools (the density of observations at exactly zero). The red line plots the actual within-school, across-track standard deviation.

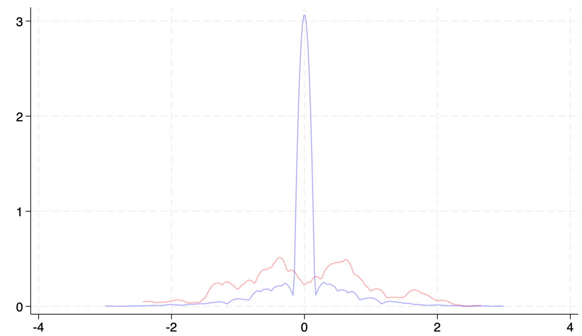
Together, the density plots illustrate that households underestimate the dispersion of selectivity and value-added within schools throughout the distribution. This is consistent with anchoring and adjustment.

[8] Other results (not displayed) show that a similar pattern arises when we consider how households evaluate the differences between two given tracks. The data are consistent with them putting substantial weight on the differences between the schools these tracks belong to.

Figure 4: Dispersion of beliefs vs. true quality



(a) Selectivity



(b) Value-added

Notes: This figure shows the dispersion of beliefs about value-added and selectivity compared to true quality, illustrating the variance in perceptions. For each household i and school S, we construct the average belief about tracks in the school: $\overline{x}^s_{iSn} = \frac{1}{K_S}\sum_{k\in S} x^s_{ikn}$ and then we plot the distribution of $x_{ikn} - \overline{x}^s_{iSn}$ in blue. In red, we construct the average true quality of tracks in the school $\overline{x}_{Sn} = \frac{1}{K_S}\sum_{k\in S} x_{ikn}$ and then plot $x_{kn} - \overline{x}_{Sn}$ in red.

4 Experimental results

The experiment reveals the causal impact of providing households with information on value-added. To begin, this section restates the key results in Ainsworth et al. (2023). Then we note two new findings. All will serve to provide a test for our model's predictions.

Ainsworth et al. first note a descriptive fact: at baseline, about 95 percent of households expected their child to be admitted to at least one of its top two choices. They then highlight that the intervention causally increased:

- the accuracy of households' beliefs about value-added for tracks outside their top two choices;
- the likelihood that households prefer high value-added tracks—outside their top two choices.

Their bottom line is that households were willing to absorb and act on information—but only for tracks they expected to be irrelevant. This result is somewhat puzzling; despite the high-stakes setting, households ignore seemingly useful information.

Our new empirical results begin to address this puzzle by noting that baseline *uncertainty*—likely arising from informational costs—influenced when the treatment had a statistically significant impact. We revisit the treatment's impacts relative to two outcomes: the error of households' beliefs and the extent to which households' track preferences prioritize value-added.

First, we find that information improved the accuracy of beliefs but did so primarily for:

- households that expressed overall uncertainty, and
- tracks for which households were uncertain, as evidenced by the fact that they did not submit

baseline beliefs about them (likely because finding out about them was costly).

To illustrate these results, we begin with the following regression:

$$\mathcal{E}_{ik,\mathrm{VA}} = \beta_1 \mathbb{1}\left\{\mathrm{Treated}_i\right\} + \epsilon_{ik} \tag{2}$$

where $\mathcal{E}_{ik,\mathrm{VA}} \equiv \left|x_{k,\mathrm{VA}} - x^s_{ik,\mathrm{VA}}\right|$ is the error of household i's belief about track k's value-added.

Table 2 (page 14) again presents regressions at the household-track level, splitting the sample in a way that highlights the role of uncertainty. All households reported a level of uncertainty so that we can split the entire sample based on this trait. In addition, we can split the sample based on whether or not the household submitted a belief about the value-added of the track in the baseline survey. Column (1) shows that the point estimate (β_1 in the equation above) is close to zero for households that stated they were very certain about their baseline choices. By contrast, column (2) shows a much larger (if statistically insignificant) reduction in endline errors for households that stated they were not very certain. These may be households that were more open to information.

Table 2: Splitting sample by certainty of household's ranking

	Error of endline belief$_{ik}$			
	All tracks		Unreported beliefs at baseline	
	Very certain	Not very certain	Very certain	Not very certain
	(1)	(2)	(3)	(4)
$\mathbb{1}\left(\mathrm{Treated}_i\right)$	-0.00636	-0.0774	-0.0404	-0.184**
	(0.0592)	(0.0514)	(0.0839)	(0.0778)
Observations	2189	2705	978	1253
Clusters	75	76	70	74

Columns (1) and (3) restrict to households who stated they were very certain of their preference ranking, and columns (2) and (4) restrict to all other (i.e. did not state they were very certain) households. Columns (1)-(2) do not restrict the sample based on whether or not the household submitted a belief about the value-added of track k in baseline; columns (3)-(4) restrict to the subsample of household-tracks where the household did not submit a belief about the track's value-added in baseline. Standard errors are clustered at the middle school level. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

Columns (3) and (4) restrict the sample to tracks for which households did not report beliefs at baseline. Not stating beliefs may reflect a "deep" form of uncertainty, e.g., not knowing enough about the track to score it, possibly because exploring it was costly. These columns additionally distinguish by whether households indicated certainty about their choices overall. Columns (3) and (4) show that essentially all the increase in accuracy is concentrated among households that indicated they were not certain overall.

To summarize, Table 2 shows that uncertainty—possibly influenced by the costs of acquiring information—plays a significant role in households' responses to information provision. Appendix Table 4 splits the sample based on different certainty thresholds, suggesting a qualitatively similar conclusion.

The next result relates to how strongly households prioritized value-added in their track rankings. We show that households increase the preference ranking of tracks with more value-added when they receive the informational treatment, but again this effect is strongest in the presence of uncertainty. Specifically, we find that the effect of information provision (on the extent to which preference rankings prioritize value added) is concentrated in the same subsamples that demonstrated the greatest reduction in erroneous beliefs.

To get at this, we construct a household's preference ranking as a percentile within their choice set: the most preferred track takes the maximal value, and the least preferred track(s) takes the minimal value. Denoting the percentile of household i's preference ranking for track k in the baseline survey as $\hat{a}^d_{ik}$, and its preference ranking in endline as $\hat{a}_{ik}$, the change in its ranking is given by $\Delta\hat{a}_{ik} \equiv \hat{a}_{ik} - \hat{a}^d_{ik}$. We run a regression of the form:

$$\Delta\hat{a}_{ik} = \beta_1 \mathbb{1}\left\{\text{Treated}_i\right\} + \beta_2 x_k + \beta_3 \mathbb{1}\left\{\text{Treated}_i\right\} \times x_k + \epsilon_{ik} \tag{3}$$

Where x_k is the true value-added (also measured in percentiles within-town to match the outcome variable) of track k. The coefficient of interest is β_3, which captures if households that receive treatment increase their weight on high value-added tracks by more. In other words, β_3 being positive suggests that value-added is more relevant for increasing preference rankings in the treatment than in the control group. Since the information only shifts perceived utility from a track k if households pay attention to said information, we partition the subsample as in Table 2 to suggest that the same mechanisms that cause changes in beliefs cause changes in preferences.

Specifically, in Table 3 (page 16) columns (1)-(4) partition the sample in the same manner as Table 2, whereas columns (5) and (6) are new. These last two columns refer to households that were uncertain, and tracks for which households did not submit a belief about value-added in baseline. Within that subsample, column (5) restricts to the set of tracks included in a household's preference ranking in baseline, and column (6) restricts to the set of tracks not ranked in baseline. We see that the treatment effect is significant for the unranked tracks, but not the ranked ones: this suggests that although high uncertainty is necessary for a change in preferences, it is not sufficient. Column (6) demonstrates that the treatment only impacts preferences if households perceived the track to be low-quality in baseline, as evidenced by their decision to not rank the track at all.

Table 3: Splitting sample by certainty and type of household's ranking

	Increase in preference$_{ik}$					
	All tracks		Unreported beliefs at baseline			
	Very certain	Not very certain	Very certain	Not very certain		
					Ranked in baseline	Not ranked in baseline
	(1)	(2)	(3)	(4)	(5)	(6)
$\mathbb{1}$ (Treated$_i$)	-0.00970	-0.0438*	-0.0153	-0.0630**	0.00963	-0.0537***
	(0.0257)	(0.0239)	(0.0236)	(0.0248)	(0.0604)	(0.0163)
True track percentile$_k$	-0.103***	-0.134***	-0.0633*	-0.109***	-0.0258	0.0622**
	(0.0338)	(0.0278)	(0.0345)	(0.0335)	(0.0733)	(0.0273)
$\mathbb{1}$ (Treated$_i$) × True track percentile$_k$	0.0316	0.0756**	0.0351	0.102**	-0.0632	0.134***
	(0.0382)	(0.0349)	(0.0431)	(0.0457)	(0.0876)	(0.0434)
Observations	8780	10919	5281	6373	1600	4318
Clusters	75	76	73	75	72	75

The variable True track percentile$_k$ is the track's value-added measured in its percentile within the town. The outcome variable is the change in their (reverse) percentile of ranking, meaning that a higher value of the outcome means that the track was more preferred in endline than in baseline. The first two columns do not restrict based on whether or not the household submitted a belief about the track's value-added in baseline; columns (3)-(6) restrict to the household-tracks where the household did not submit a belief about the track's value-added in baseline. The first four columns feature no restriction on whether or not the track was ranked in baseline; column (5) restricts to the household-tracks where the household ranked the track in baseline, and column (6) restricts to those where the household did not rank the track in baseline. Standard errors are clustered at the middle school level. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

To summarize, these experimental results demonstrate the importance of uncertainty in determining the impact of the treatment on beliefs and preferences. Beliefs about a track k only become more accurate with treatment when a) the household is generally uncertain and b) they are particularly uncertain about track k. The same two requirements extend to the impact of treatment on preference rankings: treatment changes preference rankings to prioritize value-added more only when the household is uncertain about a track and if they thought the track was low-value-added in baseline. The next section attempts to explain these empirical findings with a model of rational inattention.

5 Model

Motivated by these empirical results, we take as a starting point the model of rational inattention from Gabaix and Graeber (2024).[9] The idea in this type of model is that inattention—individuals

[9] The model used here differs from their model in that we set $\beta = 1$ and thus ignore possible (between person) heterogeneity in action-based frictions; in addition, we stress that x^d and σ^2 we interpret parameters of an individual's subjective prior, rather than being the true mean and variance of a random variable. The difference in

not paying full attention to some dimension of the world—is rational in that it results from a trade-off between the value and the cost of information. We derive predictions consistent with the role that uncertainty plays in the experimental results, and demonstrate how this model can deliver beliefs resembling those found in the baseline survey.

Most rational inattention models consider continuous choices, e.g., how much to invest or how many units to produce. In our context, however, households submit a preference ranking of tracks in a strategy-proof environment, which calls for a discrete action space. We, therefore, build on Gabaix and Graeber (2024) where the choice is not discrete in the usual sense but instead requires putting weights on various discrete options.[10] These weights add to one and represent the intensity of the preference for each option; they can be interpreted as a mixed strategy, i.e., the probability of choosing each option.[11]

5.1 Perceived utility

Suppose household i's child must attend a high school track. The household's town contains $k \in K$ tracks, and the household's task is to submit a ranked listing of the tracks it desires. Each track k has multiple attributes (e.g., location, selectivity) that households care about.

We assume that some of these attributes are easy to observe. For example, a household might easily ascertain a track's distance from its home, whether siblings use that track, etc. We denote the utility from track k's bundle of easy-to-observe attributes as μ_{ik}. As the i index indicates, this term is household-specific (e.g., a track may be located close to one household but far from another).

In addition, tracks have $n = 1, ..., N$ attributes that are hard to observe. Let $\theta_{i,n}$ denote household i's preference for dimension n; for example, some households may mostly value selectivity, whereas others prioritize value-added. By definition, households are uncertain about these hard-to-observe attributes; to denote this, we let $x^s_{ik,n}$ be household i's subjective perception of track k's dimension of quality n.

Assuming linearity, household i's perceived utility from track k is:

$$v_{ik} = \mu_{ik} + \sum_{n=1}^{N} \theta_{in} x^s_{ik,n}$$

interpretation makes no difference to the mathematical derivations.

[10] Gabaix and Graeber (2024) themselves build on Matejka and McKay (2015).

[11] The action variable being weights rather than ranked choices means that there is some difference in the exact microfoundation of the model and our empirical setting, but it is trivial to translate ranked preferences to weights that sum to one; this allows us to make qualitative predictions using a tractable model which can be tested with an empirical analysis whose discrete features are a close analogue to the model.

For expositional clarity, and given that the experiment distributed information only on value-added, we consider the case where value-added is the only non-trivial track attribute. This simplifies the notation and does not affect the key results. As a result, we suppress the subscripts n and the summation from the above equation, such that perceived utility is:

$$v_{ik} = \mu_{ik} + \theta_i x_{ik}^s \tag{4}$$

5.2 Strategy profiles

Based on its perceived utility $v_{ik}(x_{ik}^s)$, a household assigns a weight a_i^k to each track k. That is, it forms a strategy profile $a_i = (a_i^1, a_i^2, ..., a_i^K)$ such that $\sum_k a_i^k = 1$. A track with $a_i^k \approx 1$ is one that the household favors strongly for its child. For example, this might be a track that other household members use or have used. By contrast, a household attaches little value to a track with $a_i^k \approx 0$.

In practice, households form a strategy profile a_i in a continuous process, making slight adjustments as different tracks' perceived utility changes. For analytical purposes, we suppose this happens over two periods.

1. An initial, potentially long, *pre-application period* over which households gradually form priors—some distribution with a mean and variance—about each track's value added. Households express uncertainty about the quality of the mean of their prior through the variance. As will be discussed later, the variance of their prior is increasing in the costs of observing information about a track, i.e., households will be uncertain about more costly tracks. At the end of this period, a household's default beliefs $x_i^d = (x_{i1}^d, \ldots, x_{iK}^d)$ translate into a default strategy profile $a_i^d = (a_i^{1,d}, a_i^{2,d}, \ldots, a_i^{K,d})$.
2. An *application period* during which households search more intensely and finalize their strategy profile, $a_i = (a_i^1, a_i^2, ..., a_i^K)$. We assume that the baseline and endline surveys take place during this period. This is reasonable since the baseline happened at events where middle schools "kick off" the formal application process and the endline survey after households had submitted their official track rankings. These surveys give us two snapshots of households' beliefs, x_{ik}^s, and their strategy profiles, a_i.

To motivate this in a different context, consider a boy who will someday apply to college in the U.S. For him, the pre-application period happens as a child/teenager. During these years, his household receives bits of information and forms default preferences. For example, as a child, he and his parents might form an expectation that he will attend the University of Michigan—perhaps his mother and a neighbor went there, he lives in the Midwest, and he follows Michigan's teams. His household gradually gets information consistent with Michigan being a good option.

By contrast, this student hardly thinks of Duke University. It is far away, and his household hears little about it; e.g., they do not know people who went there. Michigan will receive a high $a_i^{k,d}$ in this student's choice set; Duke a low one. Importantly, the household would likely express high uncertainty in its beliefs about Duke, in part because of the cost of acquiring information on it.

For this hypothetical student, the "application period" would typically happen between the junior and senior years of high school. During these months, families more actively search for colleges. They might look at rankings, visit campuses, ask friends, etc. During this period, our example child might visit Duke and realize it is a good option for him. If this happened, for Duke (denoted as track k) we would observe that $a_i^k > a_i^{k,d}$.

5.3 Utility

With perceived utility and strategy profiles defined, we can specify that the household will maximize the following utility function:

$$U_i(a_i) = \sum_k a_i^k \cdot v_{ik} - D(a_i, a_i^d) \tag{5}$$

where $\sum\limits_k a_i^k \cdot v_{ik}$ captures the expected perceived utility from strategy profile a_i, and $D(a_i, a_i^d)$ is the Kullback-Leibler divergence between the chosen profile and the default. Intuitively, the divergence introduces action-based frictions; households face a penalty for deviating from their default plan. In our example above, $D(a_i, a_i^d)$ enters with a negative sign because the student will experience an adjustment cost in switching weight from Michigan to Duke. For example, he will realize that he will not share the Michigan experience with his mother; he may have to switch athletic team allegiances.

One virtue of this formulation is that the optimal weights a_i^k have a logit form:

$$a_i^k \left(a_i^{k,d}, v_{ik} \right) = \frac{a_i^{k,d} \exp v_{ik}(x_{ik}^s)}{\sum\limits_{k'} a_i^{k',d} \exp v_{ik'}(x_{ik'}^s)} \tag{6}$$

where k' indexes all tracks. Thus, during the application period, the weights are a function of two factors. First, the default weights, a_i^{kd} which capture the weight the household places on a given track during the pre-application period. Second, the utility the household believes it will receive from a given track, and therefore, the beliefs it has about that track's value-added, $v_{ik}(x_{ik}^s)$.

5.4 Attention and uncertainty

Expression (6) illustrates that a key input into households' decisions is x^s_{ik}, their subjective perception of track value-added. To capture that this is difficult to observe, we suppose that household i perceives it as a convex combination of the true track value, x_k, and a "default" or "anchor" belief, x^d_{ik}:

$$x^s_{ik} = m_{ik} x_k + (1 - m_{ik})\, x^d_{ik} \tag{7}$$

The parameter $m_{ik} \in [0, 1]$ is the attention household i devotes to investigating track k's value-added. When $m_{ik} = 0$, $x^s_{ik} = x^d_{ik}$, i.e., a household that pays no attention simply believes that the track has the same quality as expected. When $0 < m_{ik} < 1$, households start from some default belief x^d_{ik} and insufficiently adjust to the true value x_k. This is in the spirit of Romanian households "anchoring" track quality on school quality (see Section 3). Supposing that the anchor is constant within-school, an inattentive ($m_{ik'} \approx 0\ \forall k'$) household might conclude all tracks in a school are the same—that is a low-cost cognitive procedure. An attentive household will incur a track-specific cost (developed below) to adjust that belief.[12]

As explained in the appendix, the above subjective perception is microfounded as being the result of a signal extraction problem, in which an individual receives a noisy signal $\tilde{x}_{ik}$ about x_k with noise ϵ_{ik} that is mean-zero and has variance σ^2_{ik}. The subjective perception written in equation (7) is $\mathbb{E}_{\epsilon_{ik}}[\tilde{x}_{ik}]$; it is the average (over mean-zero noise) signal received. We do not suppose that individuals realize that their default beliefs x^d are erroneous, but instead assume that individuals are uncertain about the quality of the x^d, which is captured by σ^2_{ik}. We can write uncertainty to have a household-level and track-level component, to emphasize that uncertainty can vary on the person and track level:

$$\sigma^2_{ik} = \overline{\sigma}^2_i + \hat{\sigma}^2_{ik}$$

Where $\overline{\sigma}_i$ is the person's average uncertainty over all tracks (e.g. how uncertain they may state they are about their entire preference ranking, as in the empirical context) and a track-level deviation from the person-level average $\hat{\sigma}^2_{ik}$ (whether or not household i submits a belief about track k in the baseline survey), which can be positive or negative. The weakness of the prior σ^2_{ik} is key in determining attention to a dimension of the world, as it represents the degree to which individuals are unsure about their own beliefs.

[12] Another way to view this formulation is as follows. Households may understand that a school S has a distribution of track quality over its multiple tracks k, where $x_k \sim F_S(x)$. Then, the mean of that distribution is simply the school's average level of track quality, and the true value of x_k (i.e. the true quality at track k in the school) can be normalized as its deviation from the school mean, i.e. $x'_k = x_k - \overline{x}$. In this case, x'_k is now mean zero (by definition) and so if we define x^d to be the average value of x'_k, then $x^s = m x'_k = m\,(x_k - \overline{x})$ since the default is (by construction) zero.

5.5 The value of information

As households enter the application period, they decide how much attention to devote to each track. This depends on the value of information about track k, i.e., how much more utility the household would get if it dedicated attention to that track. As Lemma 1 in Gabaix and Graeber (2024) shows, this is given by:

$$V_{ik} = \frac{1}{2} a_i^{k,d} \left(1 - a_i^{k,d}\right) \sigma_{ik}^2 \tag{8}$$

where $a_i^{k,d}$ is the default weight on track k, and σ_{ik}^2 is the variance of the household's prior belief about the value of x_k.

This expression implies that the value of information is increasing in σ_{ik}^2, which is a proxy for uncertainty and the track-specific cost of information in the pre-application period—households will have high σ_{ik}^2 for tracks it was difficult for them to find out about. Households uncertain about a track's true value-added, x_k, would see information about that track as more valuable. These households have less certainty about x_k's true value being close to the prior mean x_{ik}^d.

Equation (8) also implies that the value of information is lower when $a_i^{k,d}$ is extreme, i.e., close to zero or to one—in those cases, the product $a_i^{k,d}\left(1 - a_i^{k,d}\right)$ approaches zero. Intuitively, tracks with extreme values are essentially "no-brainers." The household is unlikely to change its mind about them, so devoting effort to investigating them is wasteful. To return to our example, our student who strongly expects to apply to Michigan might not invest much in researching Michigan. Conversely, if the student wishes to remain in the Midwest, he might not invest effort into researching schools that are distant.

5.6 Cognitive labor and attention

With the value of information, V_{ik} defined, the household's problem is clear: it must decide how much cognitive labor, L_{ik}, to devote to each track k. It solves the following constrained maximization problem:

$$\max_{L_{i1},\cdots,L_{iK}} \sum_k V_{ik} m_{ik}(L_{ik}) \quad \text{s.t.} \quad \sum_k L_{ik} \leq \overline{L}_i \tag{9}$$

where $\overline{L}_i$ is the constraint on cognitive labor. This illustrates that households choose attention to minimize the losses from inattention—they direct effort to gain the most from what they think about.

Equation (9) determines cognitive labor, L_{ik}, which in turn translates to attention via a Cobb-

Douglas production function:

$$m_{ik} = \min\left(\left(\frac{L_{ik}}{c_{ik}}\right)^{1-\alpha}, 1\right)$$

The denominator of this expression includes c_{ik}, a track-specific cost of cognitive effort. Our descriptive results suggest this is natural; some tracks are likely easier to investigate than others.

We can now describe the optimal cognitive labor and attention that result from solving (9):[13]

$$L^*_{ik} = \frac{\overline{L}_i}{\Gamma_i} \cdot \left(\frac{V_{ik}}{c^{1-\alpha}_{ik}}\right)^{\frac{1}{\alpha}} \tag{10}$$

$$m^*_{ik} = \left(\frac{\overline{L}_i}{\Gamma_i}\right)^{1-\alpha} \left(\frac{V_{ik}}{c_{ik}}\right)^{\frac{1-\alpha}{\alpha}} \tag{11}$$

where $\Gamma_i \equiv \sum_{k'} \left(\frac{V_{ik'}}{c^{1-\alpha}_{ik'}}\right)^{\frac{1}{\alpha}}$.

In words, a household's optimal cognitive labor and attention depend upon how a track's cost-adjusted value of information, $V_{ik}/c^{1-\alpha}_{ik}$, compares *relative* to the same quantity for other tracks. Tracks with relatively high cost-adjusted value of information receive more attention.

Note that this is broadly consistent with the descriptive patterns in Section 3. Namely, households' baseline beliefs about tracks' selectivity are more accurate than their beliefs about value-added (figures 1 and 2). This is expected since in Romania selectivity is much cheaper to observe. Similarly, in either dimension, beliefs are more accurate for higher-ranked tracks. This is expected if, as discussed, more easily accessible information exists for elite than non-elite schools.

5.7 Current uncertainty and past costs

With the main elements of the model set out, we can explain why uncertainty and the costs of information are related. Denote the pre-application period as $t-1$ and the post-baseline survey, deliberative period as t. To make matters stark, suppose that households begin in total ignorance—that is, they do not discern any differences in track quality and hence are indifferent across tracks. Formally, $x^d_{ik} = \overline{x}^d_i$ and $\sigma^2_{ik,t-1} = \overline{\sigma}^2_{i,t-1}$ for all k, $a^d_{ik,t-1} = 1/K_i$ (where K_i is the number of tracks available in household i's town).

While all tracks begin from the same place, assume—crucially—that costs are *not* homogeneous across tracks: households can hear about some tracks (e.g., elite ones) easily but have a harder

[13] These solutions' derivations (which are restatements of those in Gabaix and Graeber 2024) can be found in Appendix B.1.

time learning about more obscure tracks. According to equation (11), household attention to a track k is:

$$m^*_{ik,t-1} = \left(\frac{\overline{L_i}}{\Gamma_{i,t-1}}\right)^{1-\alpha} \left(\frac{b_i\left(K_i - 1\right)}{2K_i^2 c_{ik}}\right)^{\frac{1-\alpha}{\alpha}} = \overline{L_i}^{1-\alpha} \left[\sum_{k'} \left(\frac{c_{ik}}{c_{ik'}}\right)^{\frac{1}{\alpha}}\right]^{\alpha-1} \tag{12}$$

So, as the cost c_{ik} is larger, $m^*_{ik,t-1}$ is lower, and thus beliefs will be less accurate. In other words, at baseline, individuals will exhibit less accurate beliefs about more costly tracks. This matches the baseline empirical results discussed in Section 3.

The explanations of the results in the empirical sections suggested a relationship between uncertainty and costs of information: we now formalize this. Very simply, we can say that there is some relationship between the costs pre-baseline survey, $c_{ik,t-1}$, and the sense of uncertainty in the deliberation period $\sigma^2_{ik,t}$, related by some increasing function $\sigma^2_{ik,t} = f(c_{ik,t-1})$. In short, individuals will be more uncertain about tracks that were costlier to observe in baseline.

5.8 Experiment and predictions

The model frames how to think about the experiment and interpret its results. The timeline and the intervention are as follows:

1. Over the pre-application season, a household forms priors with means x_i^d and variances $\sigma_i^{k,d}$, and default weights, $a_i^{k,d}$.
2. During the application season, a household receives a noisy signal about track k's value-added and can devote attention, m_{ik}, to increasing its precision. It chooses attention trading off the value and cost of information. This leads to a subjective perception of value-added, x_k^s. Based on this, the household develops a preference weight a_i^k.
3. The baseline survey collects beliefs and intended preference rankings; these proxy x_{ik}^s and a_i^k, respectively.
4. The experiment provides a random subsample of households with value-added estimates for each track. This lowers the cost of information, c_{ik}, to a common $\overline{c}_i$.
5. The endline survey collects final beliefs and preference rankings (we also refer to these as x_{ik}^s and a_i^k, but will make clear whether they are measured at baseline or endline).

For clarity, we state the predictions of the model upfront. We focus here on the predictions most linked to the experiment. The next subsection presents their development, along with some other implications.

Prediction 1. ***How treatment will impact beliefs***

The impact of treatment on the accuracy of beliefs will be largest for tracks:
a) which were relatively costly in baseline (high $c_{ik}/c_{ik'}$ for $k' \neq k$)
b) had non-extreme prior weights (i.e., $a_i^{kd} \neq 0$ and $a_i^{kd} \neq 1$)
c) about which individuals were uncertain (high σ_{ik}^2)

Prediction 2. ***How treatment will impact preference weights***
The tracks that will receive more weight due to treatment are those for which the perceived quality increases the most. The only tracks whose perceived quality will change are those to which individuals pay attention, and the tracks for which quality increases are those which individuals thought were low-quality at baseline. Following Prediction 1, this implies that the tracks whose perceived quality changes the most will be those:
a) about which households exhibit uncertainty, and
b) which households considered to be low-quality at baseline.

These predictions help interpret our experimental results. First, in our framework, high-cost tracks will be those for which households express a relatively high degree of uncertainty. Thus, predictions 1a) and 1c) are consistent with the results in Table 2 (page 14). In that table, the larger effects were for households that expressed overall uncertainty and tracks that they had not scored at baseline. Assuming the latter is related to high costs, these results are qualitatively in line with predictions 1a) and 1c).

Prediction 1b) is also verified by the results in Table 2, but its interpretation requires considering the joint distribution of prior weights $a_i^{k,d}$ and uncertainty σ_{ik}^2. The tracks ranked very high, by definition, have extreme (close to one) values of $a_i^{k,d}$; empirically, however, we also find that these tracks tend to be those for which σ_{ik}^2 is low. This means that attention will be low for those tracks. Even if households receive new information, they may be inattentive to it. This is broadly consistent, for example, with households leaving top choices (e.g., the very top two) unchanged.

By contrast, tracks that were unranked at baseline also have extreme (close to zero) values of $a_i^{k,d}$; but empirically we find that they also tend to have high σ_{ik}^2. The fact that households tend to be more uncertain about tracks they did not rank in baseline means that it is possible that attention could be high for them. High uncertainty amplifies attention, and that force could overwhelm the dampening effect that comes from very low $a_i^{k,d}$.

Prediction 2 relates to how individuals' preference rankings change as a result of the treatment. Clearly, perceived quality can only increase if households are attentive. From Prediction 1, we know that this means that the only tracks whose preference rankings can change are the tracks for which individuals are uncertain. Further—and distinct from the first prediction—the tracks which will *improve* in ranking are those whose perceived quality increases greatly; a large increase in perceived quality requires that the households did not believe the track was high quality in baseline. Table 3 (page 16) demonstrates this in columns (5) and (6): conditional on household uncertainty at the track level and overall, the tracks which experience the greatest increase are those which were not ranked (i.e., not preferred) at baseline. Given that those tracks were not ranked at baseline, we infer that households did not perceive them to be high quality.[14] Thus, Prediction 2 builds on Prediction 1 via 2a) but additionally demonstrates its validity through a distinct prediction about underestimated tracks.

The interaction term in Table 3 is the key piece of information to understand the relation between the second prediction and the empirical evidence. If true value-added leads to larger increases in preference ranking for the treated group, then it must be because the treated group's perceptions of track quality change more as a result of the decrease in costs. Since perceived quality of unranked (and uncertain) tracks will increase when true value-added is higher, then the interaction term represents the relation between the treatment and the increase in perceived quality. Since Prediction 2 establishes that large increases in perceived quality lead to higher preference weights, then the interaction term's sign and significance are evidence in favor of Prediction 2.

The next subsection, 5.9, works through deriving these predictions. We then return to discussing how they align with the empirical results in Section 3.

5.9 Derivation of predictions

The experiment reduces household i's cost of information about track k's value-added from some heterogeneous set, $c_i = (c_{i1}, \dots, c_{iK})$, (which varies across tracks) to a common, lower cost $\overline{c}_i$. The following derives predictions for how attention, beliefs, and preference rankings change due to this intervention.

Following from the solutions in (10) and (11), the optimal labor and attention when costs are equal across tracks (within person) are:

$$L_{ik}(\overline{c}) = \overline{L_i} \cdot \frac{V_{ik}^{\frac{1}{\alpha}}}{\sum_{k'} V_{ik'}^{\frac{1}{\alpha}}} \tag{13}$$

[14]Note that households may have high uncertainty σ^2_{ik}, but also high default beliefs x^d_{ik}, which leads them to rank tracks in baseline but not submit beliefs about them. The presence of this behavior is demonstrated by the number of observations in column 5 of Table 3, in which individuals did not submit a belief about the track's value-added in baseline, but ranked the track anyway.

$$m_{ik}(\bar{c}_i) = \left(\frac{\overline{L_i}}{\bar{c}}\right)^{1-\alpha} \left(\frac{V_{ik}^{\frac{1}{\alpha}}}{\sum_{k'} V_{ik'}^{\frac{1}{\alpha}}}\right)^{1-\alpha} \tag{14}$$

These solutions follow immediately from Equations 10 and 11; they are derived in Appendix B.1. Given that costs are equalized across tracks, the key quantity is now the relative value of information rather than the relative cost-adjusted value. Given that we have the optimal attention and labor in the pre-period (when costs were c_{ik}) and in the post-period (when all costs are $\bar{c}_i$), we can identify the forces that determine how the treatment affects labor (and thus attention).

Denoting the change in labor from household i for track k as $\Delta L_{ik} = L_{ik}(\bar{c}) - L_{ik}(c_{ik})$, we have:

$$\Delta L_{ik} = \overline{L_i} \cdot \frac{V_{ik}^{\frac{1}{\alpha}}}{\sum_{k'} V_{ik'}^{\frac{1}{\alpha}}} - \overline{L_i} \cdot \frac{\left(\frac{V_{ik}}{c_{ik}^{1-\alpha}}\right)^{\frac{1}{\alpha}}}{\sum_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}}} \tag{15}$$

In words, the treatment increases the labor directed to track k when its (cost-adjusted) value of information makes up a larger share of the total when the costs are all $\bar{c}_i$ rather than when each track has its own c_{ik}. Denoting the household-level constant $\zeta_i \equiv \frac{\overline{L_i}}{\sum_k \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}} \cdot \sum_{k'} V_{ik'}^{\frac{1}{\alpha}}}$, then the change in labor can be written as (see Appendix Section B.2.1):

$$\zeta_i \cdot \left(\frac{1}{2} a_i^{k,d} \left(1 - a_i^{k,d}\right) \sigma_{ik}^2\right)^{\frac{1}{\alpha}} c_{ik}^{\frac{1-\alpha}{\alpha}} \left[\sum_{k' \neq k} V_{ik'}^{\frac{1}{\alpha}} \left(\left(\frac{c_{ik}}{c_{ik'}}\right)^{\frac{1-\alpha}{\alpha}} - 1\right)\right] \tag{16}$$

This equation makes the forces of the change in labor more clear. Looking at the term in parenthesis, the value of information, V_{ik}, influences the magnitude of the change in labor. Since V_{ik} is weakly positive, it changes only the size (not the sign) of the change in cognitive labor. Additionally determining the magnitude of the change is the pre-treatment cost of information $c_{ik}^{\frac{1-\alpha}{\alpha}}$. These costs also determine the sign of the change in cognitive labor. If $c_{ik} > c_{ik'}\ \forall k' \neq k$, then the sum in the righmost term (in square brackets) is positive and therefore $\Delta L_{ik} > 0$. This is because all tracks are moving to a single cost $\bar{c}_i$. So the tracks that will enjoy the largest *relative* decrease in cost are those that were relatively costly to begin with—the larger the relative cost decrease, the larger the increase in cognitive labor. Of course, an increase in cognitive labor implies an increase in attention, which implies an increase in the accuracy of beliefs.

In Equation (16), the leftmost term in parenthesis illustrates that the change in attention will be amplified/muted depending on the joint distribution of default weights, $a_i^{k,d}$ and uncertainty, σ^2_{ik}.

An example can help illustrate how these multiple forces play out. Focusing on one household (so dropping i subscript), suppose that there are only two tracks, and one is high cost c_h and the other is low cost $c_\ell < c_h$. Households know that the high-cost tracks are difficult to observe and so do not trust their priors (as discussed in Section 5.7) so individuals feel uncertain about the track h's quality. The other track is low cost and people feel low uncertainty about it. So, one track is characterized by (c_h, σ^2_h) and the other by (c_ℓ, σ^2_ℓ), where $\sigma^2_h > \sigma^2_\ell$. Finally, when individuals are asked about their track beliefs and rankings in baseline, they put a large fraction of their total preference weight on the low-cost option a^d_ℓ and put the rest on the high-cost option $a^d_h = 1 - a^d_\ell$.

In this case, the treatment effect on the cognitive labor towards the high-cost track c_h is given by:

$$\Delta L_h = \zeta \cdot \left(\frac{1}{2}a^{h,d} \cdot \left(1 - a^{h,d}\right)\sigma^2_h\right)^{\frac{1}{\alpha}} c_h^{\frac{1-\alpha}{\alpha}} \left[V_\ell^{\frac{1}{\alpha}} \cdot \left(\left(\frac{c_h}{c_\ell}\right)^{\frac{1-\alpha}{\alpha}} - 1\right)\right] > 0 \tag{17}$$

Despite the fact that the h tracks had low initial weights $a^d_h \approx 0$, cognitive labor still increases for these tracks because they were higher cost than the other track $(c_h > c_\ell)$, and thus the relative change in costs as a result of the treatment is larger for the h track. Finally, the magnitude of the change in labor is amplified by large σ^2_h, and so even if the h tracks were not preferred in baseline, they are the only ones that will exhibit large increases in cognitive labor. In this example, then, the tracks that would enjoy large increases in cognitive labor are those:

- which received low weight $a^{k,d}$ in baseline
- had high cost in baseline c_h, and as a result
- about which individuals' priors were weak (high σ^2_h).

This example illustrates the mechanisms that would determine the results seen in Table 2, in which the only tracks for which treatment improves belief accuracy are those about which individuals are uncertain.

The change in cognitive labor results in a change in attention, which in turn results in a change in beliefs. The change in attention $\Delta m_{ik} \equiv m_{ik}(\bar{c}) - m_{ik}(c_{ik})$ is given by:

$$\Delta m_{ik} = \left(\frac{1}{2}a_i^{k,d}\left(1 - a_i^{k,d}\right)\sigma^2_{ik}\right)^{\frac{1-\alpha}{\alpha}} \Lambda_{ik} \tag{18}$$

where $\Lambda_{ik} \equiv \overline{L}_i^{1-\alpha} \cdot \left[\left(\bar{c}_i \cdot \sum_{k'} V_{ik'}^{\frac{1}{\alpha}}\right)^{\alpha-1} - \left(c_{ik}^{\frac{1}{\alpha}} \cdot \sum_{k'}\left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}}\right)^{\alpha-1}\right]$ measures the decrease in costs for track k. We note that $\Lambda_{ik} > 0$ if $c_{ik} > c_{ik'}$ $\forall k' \neq k$; both Equation 18 and the conditions for positivity are given in Appendix Section B.2.2. Therefore, Δm^*_{ik} is positive if the relative cost of track k decreases and negative otherwise.

The available data measures beliefs in baseline and endline. Thus, it is useful to have a prediction for how beliefs (not just attention) change in response to treatment. Denoting the change in beliefs as a result of treatment as $\Delta x^s_{ik} \equiv x^s_{ik}(m_{ik}(\bar{c})) - x^s_{ik}(m_{ik}(c_{ik}))$, we have that beliefs will change in the following way:

$$\Delta x^s_{ik} = \Delta m_{ik} \cdot \left(x_k - x^d_{ik}\right) \tag{19}$$

This means that beliefs will change proportional to the change in attention, but that change in beliefs will be larger when the "surprise"— the gap between the true value and the baseline belief— is larger. So, if individuals become more attentive to a very good track that they thought was not good, they will increase their perceived belief of the track's quality by a large amount.

Since the change in beliefs inherits the predictions about the change in labor (via the change in attention), we can summarize how treatment impacts the accuracy of beliefs as shown in Prediction 1. A change in costs will travel through the cognitive supply chain to result in a change in beliefs; this change in beliefs will then result in a change in the preference weight put on each track. Some tracks will receive higher weight, and some lower weight, as the beliefs impact the perception of relative quality across tracks.

Denoting the change in weight on track k as a result of the treatment as $\Delta a^k_i \equiv a^k_i(\bar{c}) - a^k_i(c_{ik})$, we can write this change as:

$$\Delta a^k_i = \sum_{k' \neq k} a^{k',d}_i \exp\left\{x^s_{ik'}(\bar{c})\right\}\left[\exp\left\{\omega_{ik'}\right\} - 1\right] \tag{20}$$

For $\omega_{ik'} \equiv x^s_{ik}(\bar{c}) + x^s_{ik'}(c_{ik'}) - (x^s_{ik}(c_{ik}) + x^s_{ik'}(\bar{c}_i))$. See Appendix Section B.4 for the derivation. This formulation makes it such that necessary and sufficient conditions are easy to derive.

We begin with the sufficient condition. Noting that every term in the above sum is positive if $\omega_{ik'} > 0$, then it is clear that track k will receive a higher ranking if the following condition holds for all $k' \neq k$:

$$\begin{gathered} x^s_{ik}(\bar{c}_i) - x^s_{ik}(c_{ik}) > x^s_{ik'}(\bar{c}_i) - x^s_{ik'}(c_{ik'}) \iff \\ \Delta m_{ik} \cdot \left(x_k - x^d_{ik}\right) > \Delta m_{ik'} \cdot \left(x_{k'} - x^d_{ik'}\right) \end{gathered} \tag{21}$$

That is, if household i's perception of track k *improves* more than its perception of all others

$k' \neq k$, then track k will receive more weight. Even if track k is perceived to be worse than track k', the weight a_i^k will still increase with treatment because k's relative perceived quality has improved. We emphasize via the second inequality that the change in perceptions Δx_{ik} depends on the change in attention and whether or not the household overestimated the true track quality.

Returning again to the example of one household for which one track is high-cost and high-uncertainty and the other is the low-cost, low-uncertainty counterpart, (i.e. tracks with (c_h, σ_h^2) and (c_ℓ, σ_ℓ^2) for $\sigma_\ell^2 < \sigma_h^2$ and $c_h > c_\ell$), we ask under what conditions $\Delta a_h > 0$. Although we (given assumptions on costs and uncertainty) know that $\Delta m_h > \Delta m_\ell$, this is not enough to guarantee more weight on a_h: we also need that the household underestimated the true quality of track k, i.e. $x_k^d < x_k$.[15] Therefore, we would only expect increases in preference weight for tracks for which households a) are attentive, so that $\Delta m_{ik} > 0$ and b) underestimated the track in baseline, i.e. $x_k > x_{ik}^d$. This is why Table 3 includes columns (5) and (6), as it emphasizes the importance of underestimation of true quality.

Next, we discuss the necessary conditions for the inequality; this discussion is detailed, but explains why tracks that were unranked (i.e. not preferred) in baseline would exhibit the largest increase in preference weight as a result of the treatment effect. The necessary condition for a track k's weight to increase with the intervention is (by definition):

$$\sum_{k' \neq k} a_i^{k',d} \exp\left\{x_{ik'}^s(\bar{c}_i)\right\} \left[\exp\left\{x_{ik}^s(\bar{c}_i) + x_{ik'}^s(c_{ik'}) - (x_{ik}^s(c_{ik}) + x_{ik'}^s(\bar{c}_i))\right\} - 1\right] > 0$$

Recall that $\omega_{ik'} = \Delta x_{ik}^s - \Delta x_{ik'}^s$, i.e. the difference in the changes in perception of track k and track k'. If $\omega_{ik'} < 0$, the k'-th term of the sum is negative, and if $\omega_{ik'} > 0$, the k'-th term is positive. Suppose that we want Δa_i^k to be positive—under what conditions would this be likely? First, suppose that $\omega_{ik'} < 0$, i.e. that

$$x_{ik}^s(\bar{c}_i) - x_{ik}^s(c_{ik}) < x_{ik'}^s(\bar{c}_i) - x_{ik'}^s(c_{ik'})$$

This means that the k-th track's perceived quality improved less than the k'-th track (including, of course, if the perception of k got worse). So, even if k is perceived as a better track (in levels) than k', what matters is the *change* in its perceived quality. In the case that $\omega_{ik'} < 0$ and as a result the k'-th term of the sum is negative, given that we want Δa_i^k positive, we want the conditions under which the k'-th term is as small as possible, to minimize the impact of this negative term.

[15] Further, if we have that the household overestimated track k' (meaning the default is higher than the true value for track k') then changes in attention to track k' are not relevant to the inequality, as the inequality will be satisfied since the right side will be negative.

Two terms can change the magnitude: $a_i^{k',d}$ and $\exp\{x^s_{ik'}(\overline{c}_i)\}$. So, if k improves less than some k'-th track, we want the default probability of choosing the k'-th track to be low (small $a_i^{k',d}$) and/or the endline quality in *levels* to be perceived to be low (small $x^s_{k'}$). Thus, if k' improves in quality more than k, we want k' to be a 'bad' track, i.e. one that was not preferred in baseline and one that is not seen as high-quality in endline.

Suppose now that $\omega_{ik'} > 0$, i.e. that

$$x^s_{ik}(\overline{c}_i) - x^s_{ik}(c_{ik}) > x^s_{ik'}(\overline{c}_i) - x^s_{ik'}(c_{ik'})$$

which means that the perception of k improved more than the perception of k'; given that when $\omega_{ik'} > 0$ the k'-th term of the sum is positive, we'd like to maximize the size of this term. The conditions under which that happens are simply the opposite of before— we'd like a large $a_i^{k',d}$ or a high perceived quality of k' in endline in *levels*. Thus, if k improves more than k', we want k' to be a 'good' track in the sense that k' was preferred in baseline and is seen as high-quality in endline.

This line of thinking is why we would expect tracks that were given low weight in baseline (low $a_i^{k,d}$) but are actually of high quality to be the ones for which the difference Δa_i^k is largest. Assuming that individuals had some (unreported) prior that the tracks they left unranked were not very good, then when costs go from $c_{ik} \to \overline{c}$, these tracks *improve* in perceived quality more than most other tracks, but critically, they improve in perceived quality more than the tracks that were ranked high in baseline, i.e. tracks that are (in reality) good and also had high $a_i^{k',d}$.

This point can be summarized. Recall that Δa_i^k is positive when the sum over all k' is positive, and that sum is positive when the following two conditions are met:

- when k's quality improves less than some other track k', that track k' should have a low prior probability $a^d_{ik'}$ and should be low quality $x^s_{k'}(\overline{c})$ and
- When the perceived quality of track k improves more than some other track k'', that track k'' should have a high prior probability $a^{k'',d}$ and should be high quality $x^s_{k''}(\overline{c}_i)$.

Thus, the tracks that are most likely to satisfy the requirements are those that were at first perceived to be low-quality but are, in reality, high-quality.

We can coarsely partition the sets of tracks into four groups, based on whether they had high/low perceived quality in baseline, and high/low perceived quality in endline. Those that high perceived quality in baseline and endline as well the tracks that had low perceived quality in baseline and endline will not undergo large changes in perception, and so may receive more or less weight depending on the movement of perception for other tracks. The two types of tracks which could drive changes in weights are tracks for which the perceived quality in baseline is very different than the perceived quality in endline. However, the endline perception is only different than the baseline perception if individuals are attentive to the information, and individuals are not very attentive towards tracks which they were confident were good in baseline. This leaves only the tracks which were believed to be low quality but actually are high quality, since a) individuals will be attentive towards this information (since they were less certain about the track's true quality in baseline) and as a result b) the perceived quality will improve more than the other tracks.

This line of reasoning delivers Prediction 2.

6 Conclusion

The market for education suffers from an important flaw: the most important dimension of quality, value-added, is difficult to observe. This makes it such that parents do not choose schools based on value-added, and thus schools do not invest in it. The main remedy for this problem is to make information about value-added easier to access, thus increasing parents' ability to use the information and therefore pressure schools to invest in value-added.

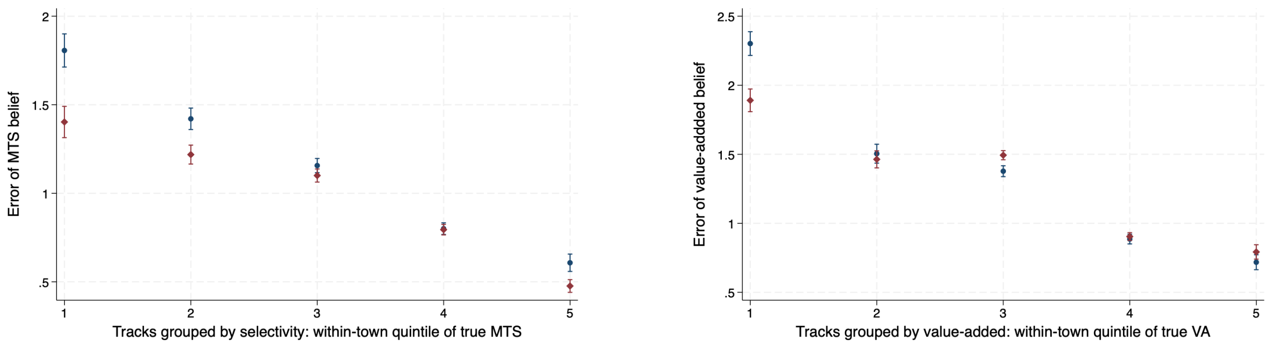
However, households do not costlessly incorporate new information into their decision-making process. This paper has demonstrated that households make a trade-off between the costs and benefits of information, which results in a situation where only households who are uncertain about their own beliefs will respond to information about tracks which were initially high-cost to observe. If other school choice environments exhibit similarly high costs of observing value-added, then interventions which provide information to all households may also, at first glance, exhibit disappointing returns.

If informational interventions can instead target households that express uncertainty about their beliefs—low-income households may be overrepresented in this population, for example—then policymakers who wish to maximize the return on informational interventions may wish to target these subpopulations. Alternatively, policymakers may wish to nudge individuals' sense of certainty in order to stimulate attention to information, which, if incorporated, will increase welfare.

A Additional tables and figures

A.1 Figures

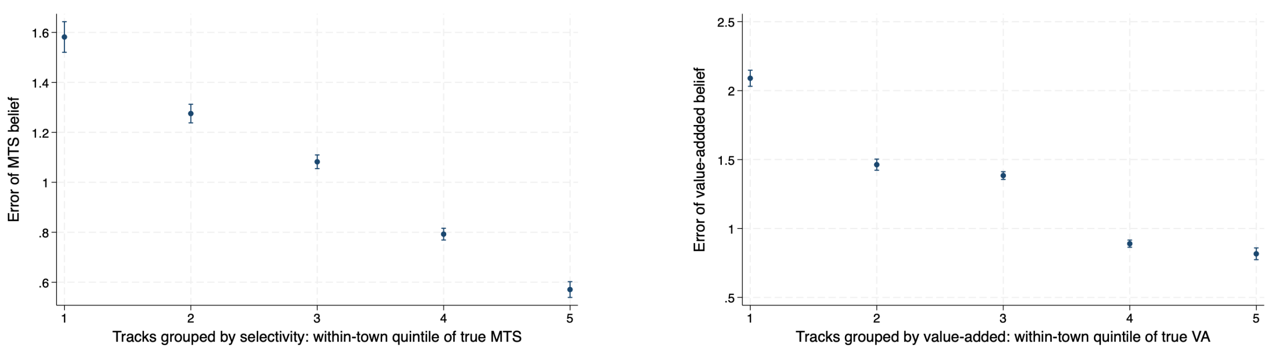
Figure 5: Errors of beliefs vs. true quality, by student ability



(a) Selectivity (b) Value-added

Notes: This figure shows the errors of beliefs about value-added and selectivity compared to true quality, where the blue indicates the below median transition score households, and the red indicates the households with above-median transition scores.

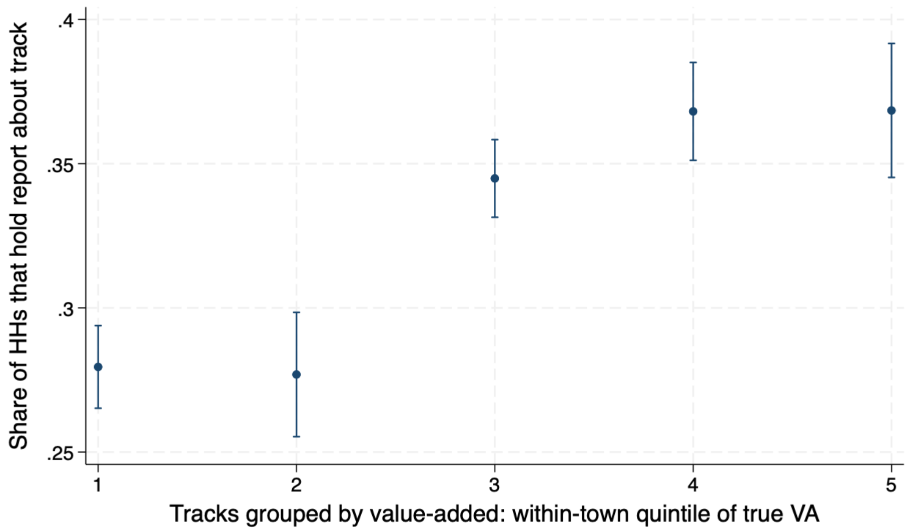
Figure 6: Errors of beliefs vs. true quality with household fixed effects



(a) Selectivity (b) Value-added

Notes: This figure shows the errors of beliefs about value-added and selectivity compared to true quality with household fixed effects.

Figure 7: Share of households with beliefs about value-added by true value-added



A.2 Tables

Table 4: Splitting sample by certainty of household's ranking

	Error of endline belief$_{ik}$			
	All tracks		Unreported beliefs at baseline	
	Any certainty	No certainty	Any certainty	No certainty
	(1)	(2)	(3)	(4)
Treated	-0.0402	-0.211	-0.113*	-0.310
	(0.0455)	(0.138)	(0.0609)	(0.237)
Observations	4630	340	2115	172
Clusters	76	50	74	36

Columns (1) and (3) restrict to households who stated they were either somewhat certain or very certain of their preference ranking, and columns (2) and (4) restrict to all other (i.e. did not state they were very/somewhat certain) households. Columns (1)-(2) do not restrict the sample based on whether or not the household submitted a belief about the value-added of track k in baseline; columns (3)-(4) restrict to the subsample of household-tracks where the household did not submit a belief about the track's value-added in baseline. Standard errors are clustered at the middle school level. * $p < 0.1$, ** $p < .05$, *** $p < 0.01$

B Derivations

B.1 Optimal cognitive labor and attention

Note that the content and proof of these solutions is entirely taken from the proof of Proposition 1 in Gabaix and Graeber (2024). Writing the dual of the problem with shadow cost of labor $(1-\alpha)\,w$, we see that the choice of optimal cognitive labor is characterized by:

$$\max_{L_{i1},\ldots,L_{iK}} \sum_k V_{ik} m_{ik}(L_{ik}) - (1-\alpha)\,w \sum_k L_{ik} \tag{22}$$

The FOCs of the problem yield:

$$L^*_{ik} = \left(\frac{V_{ik}}{w \cdot c_{ik}}\right)^{\frac{1}{\alpha}} c_{ik} \implies \tag{23}$$

$$m^*_{ik} = \left(\frac{V_{ik}}{c_{ik} w}\right)^{\frac{1-\alpha}{\alpha}} \tag{24}$$

Noting that w is a Lagrange multiplier, we see that its value is

$$w = \left[\left(\sum_{k'} V_{ik'}\right) \overline{L}_i^{-\alpha} C^{\alpha-1}\right] \tag{25}$$

Where $C = \left(\sum_{k'} \left(\frac{V_{ik'}}{\sum_{k''} V_{ik''}}\right)^{\frac{1}{\alpha}} c_{ik'}^{\frac{\alpha-1}{\alpha}}\right)^{-\frac{\alpha}{1-\alpha}}$ is the complexity aggregator from Gabaix and Graeber (2024).

So plugging in this value to get optimal cognitive labor, we see:

$$L^*_{ik} = c_{ik} \left(\frac{V_{ik}}{c_{ik}}\right)^{\frac{1}{\alpha}} w^{-\frac{1}{\alpha}} \tag{26}$$

$$= c_{ik} \left(\frac{V_{ik}}{c_{ik}}\right)^{\frac{1}{\alpha}} \left(\left[\left(\sum_{k'} V_{ik'}\right) \overline{L}_i^{-\alpha} C^{\alpha-1}\right]\right)^{-\frac{1}{\alpha}} \tag{27}$$

$$= c_{ik} \left(\frac{V_{ik}}{c_{ik}}\right)^{\frac{1}{\alpha}} \left(\sum_{k'} V_{ik'}\right)^{-\frac{1}{\alpha}} \cdot \overline{L}_i \cdot C^{\frac{1-\alpha}{\alpha}} \tag{28}$$

$$= c_{ik} \left(\frac{V_{ik}}{\sum\limits_{k'} V_{ik'}} \cdot \frac{1}{c_{ik}} \right)^{\frac{1}{\alpha}} \cdot \overline{L}_i \cdot \left(\left(\sum_{k'} \left(\frac{V_{ik'}}{\sum\limits_{k''} V_{ik''}} \right)^{\frac{1}{\alpha}} c_{ik'}^{\frac{\alpha-1}{\alpha}} \right)^{\frac{\alpha}{\alpha-1}} \right)^{\frac{1-\alpha}{\alpha}} \tag{29}$$

$$= c_{ik} \left(\frac{V_{ik}}{\sum\limits_{k'} V_{ik'}} \cdot \frac{1}{c_{ik}} \right)^{\frac{1}{\alpha}} \cdot \overline{L}_i \cdot \left(\sum_{k''} V_{ik''} \right)^{\frac{1}{\alpha}} \left(\sum_{k'} (V_{ik'})^{\frac{1}{\alpha}} c_{ik'}^{\frac{\alpha-1}{\alpha}} \right)^{-1} \tag{30}$$

$$= \left(\frac{V_{ik}}{c_{ik}^{1-\alpha}} \right)^{\frac{1}{\alpha}} \cdot \overline{L}_i \cdot \left(\sum_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}} \right)^{\frac{1}{\alpha}} \right)^{-1} \tag{31}$$

$$\tag{32}$$

Define $\Gamma_i \equiv \sum\limits_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}} \right)^{\frac{1}{\alpha}}$. By definition of $m_{ik} = \min\left((L_{ik}/c_{ik})^{1-\alpha}, 1 \right)$ and the fact that we assume an interior solution (i.e. $m_{ik}^* \in (0,1)$), then we have that attention is given by:

$$m_{ik}^* = \left(\frac{L_{ik}^*}{c_{ik}} \right)^{1-\alpha} \tag{33}$$

$$= \left(\frac{\overline{L}_i}{c_{ik}} \cdot V_{ik}^{\frac{1}{\alpha}} \frac{1}{c_{ik}^{\frac{1}{\alpha}-1}} \cdot \Gamma_i^{-1} \right)^{1-\alpha} \tag{34}$$

$$= \left(\frac{\overline{L}_i}{\Gamma_i} \right)^{1-\alpha} \left(\frac{V_{ik}}{c_{ik}} \right)^{\frac{1-\alpha}{\alpha}} \tag{35}$$

B.2 Change in labor and attention with treatment

B.2.1 Labor

Denote $\Gamma_i(c_i) \equiv \sum\limits_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}} \right)^{\frac{1}{\alpha}}$ and $\Gamma_i(1) \equiv \sum\limits_{k'} (V_{ik'})^{\frac{1}{\alpha}}$ The change in labor $\Delta L_{ik} = L_{ik}^*(\overline{c}_i) - L_{ik}^*(c_{ik})$ can be written as:

$$L_{ik}^*(\overline{c}_i) - L_{ik}^*(c_{ik}) = \overline{L_i} \cdot \frac{V_{ik}^{\frac{1}{\alpha}}}{\sum\limits_{k'} V_{ik'}^{\frac{1}{\alpha}}} - \overline{L_i} \cdot \frac{\left(\frac{V_{ik}}{c_{ik}^{1-\alpha}} \right)^{\frac{1}{\alpha}}}{\sum\limits_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}} \right)^{\frac{1}{\alpha}}} \tag{36}$$

$$\overline{L_i} \cdot V_{ik}^{\frac{1}{\alpha}} \left[\frac{\sum\limits_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}} \right)^{\frac{1}{\alpha}} - \sum\limits_{k'} \left(\frac{V_{ik'}}{c_{ik}^{1-\alpha}} \right)^{\frac{1}{\alpha}}}{\Gamma_i(c_i) \cdot \Gamma_i(1)} \right] \tag{37}$$

$$\frac{\overline{L_i}\cdot V_{ik}^{\frac{1}{\alpha}}}{\Gamma_i(c_i)\cdot\Gamma_i(1)}\left[\sum_{k'}\left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}}-\sum_{k'}\left(\frac{V_{ik'}}{c_{ik}^{1-\alpha}}\right)^{\frac{1}{\alpha}}\right] \tag{38}$$

$$\frac{\overline{L_i}\cdot V_{ik}^{\frac{1}{\alpha}}c_{ik}^{\frac{1-\alpha}{\alpha}}}{\Gamma_i(c_i)\cdot\Gamma_i(1)}\left[\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}\left(\left(\frac{c_{ik}}{c_{ik'}}\right)^{\frac{1-\alpha}{\alpha}}-1\right)\right] \tag{39}$$

$$\frac{\overline{L_i}\cdot V_{ik}^{\frac{1}{\alpha}}c_{ik}^{\frac{1-\alpha}{\alpha}}}{\Gamma_i(c_i)\cdot\Gamma_i(1)}\left[\sum_{k'\neq k}V_{ik'}^{\frac{1}{\alpha}}\left(\left(\frac{c_{ik}}{c_{ik'}}\right)^{\frac{1-\alpha}{\alpha}}-1\right)\right] \tag{40}$$

B.2.2 Attention

Assuming an interior solution (i.e. $m^* \in (0,1)$), then we have that

$$m_{ik}^*(\overline{c}_i)-m_{ik}^*(c_{ik})=\left(\frac{L_{ik}^*(\overline{c}_i)}{\overline{c}_i}\right)^{1-\alpha}-\left(\frac{L_{ik}^*(c_{ik})}{c_{ik}}\right)^{1-\alpha} \tag{41}$$

$$=\left(\frac{\overline{L}_{ik}}{\overline{c}_i}\right)^{1-\alpha}\cdot\left(\frac{V_{ik}^{\frac{1}{\alpha}}}{\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}}\right)^{1-\alpha}-\left(\overline{L}_i\right)^{1-\alpha}\cdot\left(\frac{V_{ik}}{c_{ik}}\right)^{\frac{1-\alpha}{\alpha}}\cdot\left(\sum_{k'}\left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}}\right)^{\alpha-1} \tag{42}$$

$$=\overline{L}_i^{1-\alpha}\cdot V_{ik}^{\frac{1-\alpha}{\alpha}}\left[\left(\overline{c}_i\cdot\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}\right)^{\alpha-1}-\left(c_{ik}^{\frac{1}{\alpha}}\cdot\sum_{k'}\left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}}\right)^{\alpha-1}\right] \tag{43}$$

The above quantity is positive if the following inequality is satisfied; we will show the conditions under which it is true:

$$\overline{c}_i\cdot\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}<c_{ik}^{\frac{1}{\alpha}}\sum_{k'}\left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}} \tag{44}$$

$$\overline{c}_i\cdot\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}<c_{ik}^{\frac{1}{\alpha}}\sum_{k'}\left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}}\right)^{\frac{1}{\alpha}}=\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}\left(\frac{c_{ik}}{c_{ik'}}\right)^{\frac{1}{\alpha}}\cdot c_{ik'} \tag{45}$$

$$\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}<\sum_{k'}V_{ik'}^{\frac{1}{\alpha}}\left(\frac{c_{ik}}{c_{ik'}}\right)^{\frac{1}{\alpha}}\cdot\underbrace{\frac{c_{ik'}}{\overline{c}_i}}_{\geq 1} \tag{46}$$

So, if $c_{ik} > c_{ik'}\ \forall k' \neq k$, then both of the factors to $V_{ik'}^{\frac{1}{\alpha}}$ are greater than one, and thus the inequality is satisfied. Going back to the expression for Δm_{ik}^* in Equation 43, we can denote $\Lambda_{ik} \equiv$

$\overline{L}_i^{1-\alpha} \cdot \left[\left(\overline{c}_i \cdot \sum_{k'} V_{ik'}^{\frac{1}{\alpha}} \right)^{\alpha-1} - \left(c_{ik}^{\frac{1}{\alpha}} \cdot \sum_{k'} \left(\frac{V_{ik'}}{c_{ik'}^{1-\alpha}} \right)^{\frac{1}{\alpha}} \right)^{\alpha-1} \right]$, and we note that $\Lambda_{ik} > 0$ if $c_{ik} > c_{ik'}$ $\forall k' \neq k$. Then, the change in attention can be written as:

$$\Delta m_{ik}^* = V_{ik}^{\frac{1-\alpha}{\alpha}} \Lambda_{ik} \tag{47}$$

B.3 Change in beliefs with treatment

$$\begin{aligned}
x_{ik,t+1}^{s,\tau} - x_{ik,t+1}^{s} &= m_{ik,t+1}^{\tau} \cdot \left(x_{ik} - x_{ik}^{d} \right) + x_{ik}^{d} - \left[m_{ik,t+1} \cdot \left(x_{ik} - x_{ik}^{d} \right) + x_{ik}^{d} \right] && (48)\\
&= \left(m_{ik,t+1}^{\tau} - m_{ik,t+1} \right) \cdot \left(x_{ik} - x_{ik}^{d} \right) && (49)\\
&= \left(m_{ik}(\overline{c}) - m_{ik}(c_k) \right) \cdot \left(x_{ik} - x_{ik}^{d} \right) && (50)\\
&= V_{ik}^{\frac{1-\alpha}{\alpha}} \Lambda_{ik} \cdot \left(x_{ik} - x_{ik}^{d} \right) && (51)\\
&= \left(a_i^{k,d} \cdot \left(1 - a_i^{k,d} \right) \sigma_{ik}^2 \right)^{\frac{1-\alpha}{\alpha}} \Lambda_{ik} \cdot \left(x_{ik} - x_{ik}^{d} \right) && (52)
\end{aligned}$$

B.4 Change in preferences with treatment

The difference in rankings for a given track k when receiving treatment vs. control is given by $\Delta a_i^k = a_i^k(\overline{c}_i) - a_i^k(c_{ik})$. We derive here the way to rewrite this difference expressed in Equation 20:

$$a_i^k(\overline{c}_i) - a_i^k(c_{ik}) = \frac{a_i^{k,d} \exp\left\{ v_{ik}(x_{ik}^s(\overline{c}_i)) \right\}}{\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\}} - \frac{a_i^{k,d} \exp\left\{ v_{ik}(x_{ik}^s(c_{ik}) \right\}}{\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'}) \right\}} > 0 \iff \tag{53}$$

$$\frac{\left(a_i^{k,d} \exp\left\{ v_{ik}(x_{ik}^s(\overline{c}_i)) \right\} \right) \cdot \left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'}) \right\} \right) - \left(a_i^{k,d} \exp\left\{ v_{ik}(x_{ik}^s(c_{ik}) \right\} \right) \cdot \left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\} \right)}{\left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\} \right) \left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'}) \right\} \right)} > 0 \iff \tag{54}$$

$$\left(a_i^{k,d} \exp\left\{ v_{ik}(x_{ik}^s(\overline{c}_i)) \right\} \right) \cdot \left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'}) \right\} \right) - \left(a_i^{k,d} \exp\left\{ v_{ik}(x_{ik}^s(c_{ik}) \right\} \right) \cdot \left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\} \right) > 0 \tag{55}$$

$$\left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'})) + v_{ik}(x_{ik}^s(\overline{c}_i)) \right\} \right) - \left(\sum_{k'} a_i^{k',d} \exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) + v_{ik}(x_{ik}^s(c_{ik}) \right\} \right) > 0 \tag{56}$$

$$\sum_{k'} a_i^{k',d} \left[\exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'})) + v_{ik}(x_{ik}^s(\overline{c}_i)) \right\} - \exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) + v_{ik}(x_{ik}^s(c_{ik}) \right\} \right] > 0 \tag{57}$$

$$\sum_{k'} a_i^{k',d} \left(\exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) + v_{ik}(x_{ik}^s(c_{ik}) \right\} \right) \left[\exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'})) + v_{ik}(x_{ik}^s(\overline{c}_i)) - \left(v_{ik'}(x_{ik'}^s(\overline{c}_i)) + v_{ik}(x_{ik}^s(c_{ik})) \right) \right\} - 1 \right] > 0 \tag{58}$$

$$\sum_{k'} a_i^{k',d} \left(\exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\} \right) \left[\exp\left\{ v_{ik'}(x_{ik'}^s(c_{ik'})) + v_{ik}(x_{ik}^s(\overline{c}_i)) - \left(v_{ik'}(x_{ik'}^s(\overline{c}_i)) + v_{ik}(x_{ik}^s(c_{ik})) \right) \right\} - 1 \right] > 0 \tag{59}$$

$$\sum_{k'} a_i^{k',d} \left(\exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\} \right) \left[\exp\left\{ \omega_{k'} \right\} - 1 \right] > 0 \tag{60}$$

$$\sum_{k' \neq k} a_i^{k',d} \left(\exp\left\{ v_{ik'}(x_{ik'}^s(\overline{c}_i)) \right\} \right) \left[\exp\left\{ \omega_{k'} \right\} - 1 \right] > 0 \tag{61}$$

Where $\omega_{k'} \equiv v_{ik'}(x^s_{ik'}(c_{ik'})) + v_{ik}(x^s_{ik}(\bar{c}_i)) - (v_{ik'}(x^s_{ik'}(\bar{c}_i)) + v_{ik}(x^s_{ik}(c_{ik}))$. The sum switches to $k' \neq k$ because at $k' = k$, $\omega = 0$. Next, note that since $v(x^s) = \mu + \theta x^s$, then we'll have:

$$\begin{aligned}
\omega_{k'} &= v_{ik'}(x^s_{ik'}(c_{ik'})) + v_{ik}(x^s_{ik}(\bar{c}_i)) - (v_{ik'}(x^s_{ik'}(\bar{c}_i)) + v_{ik}(x^s_{ik}(c_{ik})) && (62)\\
&= \mu_{k'} + \theta_i x^s_{ik'}(c_{ik'}) + \mu_k + \theta_i x^s_{ik}(\bar{c}_i) - (\mu_{k'} + \theta_i x^s_{ik'}(\bar{c}_i) + \mu_k + \theta_i x^s_{ik}(c_{ik})) && (63)\\
&= \theta_i \left[x^s_{ik'}(c_{ik'}) + x^s_{ik}(\bar{c}_i) - (x^s_{ik'}(\bar{c}_i) + x^s_{ik}(c_{ik})) \right] && (64)
\end{aligned}$$

Then the $\exp \theta_i$ term can be removed and $w_{k'}$ is as in the main text.

The sufficient condition is immediate: if $\omega_{k'} > 0$ for all $k' \neq k$, then each term in the sum is positive and so the whole sum is positive, which means $a^k_i(\bar{c}_i) > a^k_i(c_{ik})$. The necessary condition is equally immediate.

References

Abdulkadiroglu, A., J. Angrist, and P. Pathak (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica 82*(1), 137–196.

Abdulkadiroglu, A., P. Pathak, J. Schellenberg, and C. Walters (2017). Do parents value school effectiveness? Mimeo, Duke University.

Abdulkadiroglu, A., P. Pathak, and C. R. Walters (2018). Free to choose: Can school choice reduce student achievement? *American Economic Journal: Applied Economics 10*(1), 175–206.

Abdulkadiroglu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2020, May). Do parents value school effectiveness? *American Economic Review 110*(5), 150239.

Agte, P., C. Allende, A. Kapor, C. Neilson, and F. Ochoa (2024). Search and biased beliefs in education markets. Technical report, National Bureau of Economic Research Working Paper No. 32670.

Ainsworth, R., R. Dehejia, C. Pop-Eleches, and M. Urquiola (2023). Why do households leave school value added on the table? the roles of information and preferences. *American Economic Review 113*(4), 1049–1082.

Ajayi, K. F., W. H. Friedman, and A. M. Lucas (2017). The importance of information targeting for school choice. *American Economic Review: Papers and Proceedings 107*(5), 638–643.

Ajayi, K. F. and M. Sidibé (2022). School choice under imperfect information. Mimeo, Boston University, Mimeo.

Allende, C., F. Gallego, and C. Neilson (2019). Approximating the equilibrium effects of information interventions. Mimeo, Mimeo, Princeton University.

Andrabi, T., J. Das, and A. Khwaja (2017). Report cards: The impact of providing school and child test scores on educational markets. *American Economic Review 107*(6), 1535–15635.

Arteaga, F., A. J. Kapor, C. A. Neilson, and S. D. Zimmerman (2022, July). Smart matching platforms and heterogeneous beliefs in centralized school choice. *The Quarterly Journal of Economics 137*(3), 1791–1848.

Beuermann, D. W. and C. K. Jackson (2023). What is a good school, and can parents tell? evidence on the multidimensionality of school output. *Review of Economic Studies 90*, 65–101.

Borcan, O., M. Lindahl, and A. Mitrut (2017). Fighting corruption in education: What works and who benefits? *American Economic Journal: Economic Policy 9*(1), 180–209.

Corcoran, S. P., J. L. Jennings, S. R. Cohodes, and C. Sattin-Bajaj (2018). Leveling the playing field for high school choice: Results from a field experiment of informational interventions. Technical report, National Bureau of Economic Research Working Paper No. 24471.

Dean, M. and N. Neligh (2023). Experimental tests of rational inattention. *Journal of Political Economy 131*(12), 3415–3461.

Epple, D. and R. E. Romano (1998). Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review 88*(1), 33–62.

Friedman, M. (1955). The role of government in education. In R. Solow (Ed.), *Economics and the Public Interest*. NJ: Trustees of Rutgers College.

Gabaix, X. and T. Graeber (2024). The complexity of economic decisions. Technical report, National Bureau of Economic Research Working Paper No. 33109.

Hastings, J. and J. Weinstein (2008, November). Information, school choice, and academic achievement: Evidence from two experiments. *Quarterly Journal of Economics 123*(4), 1373–1414.

Imberman, S. and M. Lovenheim (2016). Does the market value value-added? evidence from housing prices after as public release of school and teacher value added. *Journal of Urban Economics 91*, 104–121.

Immorlica, N., J. Leshno, I. Lo, and B. Lucier (2020). Information acquisition in matching markets: The role of price discovery. Mimeo, SSRN Working Paper 3705049, https://ssrn.com/abstract=3705049.

MacLeod, W. B. and M. Urquiola (2015). Reputation and school competition. *American Economic Review 105*(11), 3471–3488.

MacLeod, W. B. and M. Urquiola (2019). Is education consumption or investment? Implications for the effects of school competition. *Annual Review of Economics 11*, 563–589.

Matejka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *The American Economic Review 105*(1), 272–98.

Mizala, A. and M. Urquiola (2013). Parental choice and school markets: The impact of information on school effectiveness. *Journal of Development Economics 103*, 313–335.

Muralidharan, K. and V. Sundararaman (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in india. *The Quarterly Journal of Economics 130*(3), 1011–1066.

Rothstein, J. (2006). Good principals or good peers: Parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American*

Economic Review 96(4), 1333–1350.