Information Goods

Cameron Taylor[*]

March 30, 2020

Abstract

The main goal of this paper is to understand how people will change their information acquisition strategies as information sources become more or less costly. To do this, I develop a model of information acquisition in the spirit of traditional consumer theory that treats information sources, which are distinct dimensions of the state space, as different consumption goods. A general form of the model shows that as information becomes more costly, people will demand less of it, and also characterizes when information sources are substitutes or complements. The models insights are extensively analyzed in two settings: determining the optimal firm recruiting strategy when considering technical and social skills, and determining the optimal way to evaluate students using testing and assessing creativity. Other insights into dating and media consumption are also discussed.

1 Introduction

Data and information are important inputs into decision-making. Consumers, firms and policymakers rely on information to make the best choices. Importantly, though, analyzing data and gathering information are costly - they take up time and resources. Moreover, information from different sources or mediums may be correlated, affecting how one optimally chooses to acquire information. This paper presents a model where an agent gathers costly information. The model treats information sources as distinct "goods" that are priced as in traditional models.

The model is able to answer fundamental questions in information acquisition, such as a "law of demand" for information and characterizing when information goods are substitutes or complements. These questions are important for understanding how agents may change their information acquisition behaviors over time as some types of data and information become more or less costly.

[*]Stanford Graduate School of Business. Email contact: cntaylor@stanford.edu. Thank you to Matthew Gentzkow, Eddie Lazear and Andy Skrzpacyz for helpful discussions and suggestions.

The model treats the state space as multidimensional with each dimension representing a different "information good" and allows signals to only fully reveal the dimensions with independent probabilities. Within the standard model, this implies a significant restriction on the support of posterior beliefs. However, it allows the costs of information to be modeled in a manner closer to prices of the goods.

This set-up, combined with a quasi-linear assumption on utility and costs allows for a first-order condition to fully characterize the agent's information acquisition problem. This first-order condition contains important economic intuition for a consumer's demand for information and serves as the basis for comparative statics and closed-form solutions of the model. I analyze the 2-dimensional case of the model to keep the math tractable and transparent.

In the first part of this paper, I derive two natural economic predictions analogous to standard consumer theory results. First, information goods cannot be Giffen goods. Second, there is a necessary and sufficient condition for two information goods to be substitutes and complements. This condition turns out to be extremely similar to the definition provided in Börgers et al. (2013). Importantly though, I derive this condition whereas they treat it as a definition and characterize it in more general settings.

The second part of the paper focuses on specific applications of the model to firm hiring and education testing. Both applications are of independent interest to labor and education economists, but are also useful in exhibiting the simplicity and transparency of the model. The applications allow further clarify the insights of the two results in the general model, allow for closed form demand function expressions, and highlight the importance of modeling costs and correlations between information goods together in understanding optimal information acquisition.

Some insights from the applications show how it is possible that as information from resume screens and other sources becomes cheaper, firms may actually interview candidates more. The application also shows how resume screens may be used by firms even though they are not directly payoff-relevant if they are correlated with payoff-relevant variables and are cheap. In the education testing setup, the model shows how school districts may assess creativity from students more as it becomes easier to administer tests and why tests may be valuable even if they are very weakly correlated with a student's ability or future outcomes.

There is other related work that has made great progress on the topic of modeling costs of information acquisition with an eye towards applications. The rational inattention literature (e.g. Sims, 2003; Caplin and Dean, 2013; Matějka and McKay, 2015) has shown the implications of costly information acquisition for consumer behavior under entropy-based cost functions. Pomatto et al. (2019) prove the existence of a unique cost function that satisfies desirable axioms. This paper provides another way for applied researchers to think of the demand for information. Compared to these literatures, this paper reduces generality with respect to the information acquisition space but allows for cost functions to be relatively unrestricted, focusing more on using basic economic tools to derive predictions.

Probably the most related paper to this work is Van Neiuwerburgh and Veldkamp (2010) who consider a very similar problem. Similar to my applications, they consider a world with normal priors and study a particular application in portfolio acquisition. They only allow normal signals which means that their class of signals is indexed by the posterior variance. Importantly they assume that the consumer has a budget constraint and costs and variance reduction is costly. They look at the implications of two standard cost function representations: additive precision and entropy. Their focus on specific predictions as opposed to general characterizations makes their set of results comparable to my own. Verrecchia (1982) also studies the implications of the posterior variance selection problem in a normal world in the context of asset trading, and derives comparative statics that are related to the ones derived in this paper.

The organization of this paper is as follows: first I exposit the most general form of my model and compare it to standard models (Section 2), then I analyze the model's insights to the firm recruiting application (Section 3) and the education application (Section 4). I discuss further insights into media and dating markets in Section 5. Finally I conclude with discussion of the contributions and limitations of the paper (Section 6). Most of the proofs and longer derivations are left for the Appendix.

2 Model

2.1 Setup and Comparison to Standard Models

There is a single decision-maker (DM) that has a utility function over actions and states of the world $u(a, \omega)$. Suppose that $a \in \mathcal{A}$ is some finite set of possible actions and the state space satisfies $\omega \in \Omega = \times_{i=1}^{n} \Omega_i \subseteq \mathbb{R}^n$.

The value in having Ω be multidimensional instead of a finite set as is common in the literature is that it allows for a more intuitive breakdown of the potential scarcity in resources decision-makers must allocate across sources of information. This will become particularly important when I examine the impact of correlation in the next section.

The key difference between this model and standard models of information acquisition is the set of information structures that the DM has access to. In particular the DM has access to signals indexed by $q \in [0, 1]^n$ where the signal realization space is $S = \times_{i=1}^{n} (\Omega_i \cup \{\emptyset\})$ and the marginal signal realization distribution on dimension i, π_i, satisfies

$$\pi_i(\omega_i|\omega_i, \omega_{-i}) = q_i, \ \pi_i(\emptyset|\omega_i, \omega_{-i}) = 1 - q_i$$

for all ω_i and ω_{-i}. π is then the product of these marginal distributions.

This signal space is best interpreted in the following way: the DM looks over the n dimensions of the state-space, and selects a probability q_i of fully-revealing the information on dimension i for each dimension. After selecting such a bundle of probabilities, the DM sees each dimension with probability q_i where the event of seeing a dimension i is independent of seeing another dimension j. Importantly, instead of interpreting signal selection as selecting a posterior belief over the state as is common in the literature, this framework allows for the interpretation that the DM's information acquisition problem is instead an n dimensional choice problem. The interpretation of each dimension is as a separate "information good", where q_i is a measure of the "consumption" of that good for the DM.

One of the major benefits of such a stylized signal structure is that costs can now be modeled flexibly and more analogously to prices for these information goods. I assume that the cost of a signal indexed by q is $c(q)$ and I assume that the payoff of the DM is quasi-linear in the sense that her payoff is the difference between her expected utility and this signal cost. Assume that $c(q) = \sum_i c_i(q_i)$ so that the cost is additive and also assume that these costs are smooth, convex and satisfy $c_i(0) = 0$ and $\lim_{q_i \to 1} c_i(q_i) = +\infty$.[1] The interpretation of the cost $c_i(q_i)$ is as the price for information good i with added appropriate convexity assumptions since $q_i \in [0, 1]$ and so consumption is necessarily bounded. By assuming that $c_i(0) = 0$ I have assumed that there is no fixed cost for consuming information good i.

As in the standard models, the DM has some prior $\mu_0 \in \Delta(\Omega)$ with arbitrary correlation over the dimensions of the state. The existence of non-trivial correlation between the different dimensions of the state will be important in examples and applications.

As emphasized, the starkest simplification in this model compared to most models in the literature is the reduction of feasible signals that the DM has access to. I now discuss what this restriction looks like. Since it is simpler and w.l.o.g. to work in posterior belief space I now discuss the restriction within this space.

To gain some intuition first consider the case where the dimensions of the state are independent. In this case, the support of the distribution of posterior beliefs is greatly reduced. In particular, the support of the posterior beliefs consists of the combinations of point masses at each individual point in each dimension of the state space and the support of the prior in each dimension of the state space. More simply put, in the model with independent dimensions of the state, the DM can only have either exact knowledge or the same as prior knowledge on each dimension. As Kamenica and Gentzkow (2011) show, any distribution of posterior beliefs can be achieved by some signal in the most general framework, so this encompasses a substantial restriction.

[1]It is possible to assume that costs go to some very large scalar instead, and I will do this in the application, but the math and logic is simpler under this assumption in the more general model.

When the dimensions of a state are correlated, the revelation of only one dimension can alter the support away from the two extremes of point masses and the prior. However, the restriction is still strong and highly dependent on the prior.

I now aim to provide some intuition on why this restriction may be valuable. What this restriction does is shift more of the economic burden towards the costs and prices of information $c(q) = (c_1(q_1), \dots c_n(q_n))$ and the demand decision $q = (q_1, \dots, q_n)$ in attempting to explain the DM's information acquisition behavior. In particular, this allows for the interpretation of different signals as different levels of consumption of the information goods as is common in classical demand theory, leading to natural definitions of comparative statics and notions of "consuming more of good i". As shown in the examples and applications below, this generates predictions on how firms and policymakers may change their information acquisition decisions in the face of changing costs and technology.

The main results of this paper focus on a two-dimensional state space model. In particular, suppose that $n = 2$ and index the first dimension of the state space by $\omega_1 = A$ and the second dimension by $\omega_2 = B$. Then the set of feasible signals is indexed by $(q_A, q_B) \in [0, 1]^2$.

First I provide some examples to ground ideas and illustrate how the model may fit with some important economic scenarios. These examples and applications are returned to after presenting the general results.

2.2 Some Examples

To solidify the motivation for my modeling choices I examine three examples in which I find the dimensionality of Ω and the simplified signal space to be intuitively appealing. I will pursue the firm recruiting example further in Section 3 and the education testing example further in Section 4.

Example 1: Firm Recruiting

Suppose that a firm is screening an applicant. There are two skills in the world: technical skills and social skills. Thus A is technical skills of the applicant and B is social skills of the applicant. The utility function of the firm is their profit function when hiring the worker and the only action they take is to hire or not hire the worker.

As the model points out, the costs of information can now be directly mapped to the outcomes produced in producing information on that dimension. So, one can think of the acquiring information on A as the costs of testing the applicant, asking for technical references, etc. while one can think of the costs of acquiring information on B relating to the costs of interviewing applicants, getting to know them better and asking for social references. These costs are flexibly incorporated into c_A and c_B.

Example 2: Education Testing

Suppose that a school district is testing students at different schools to assess their skills. There are two skills in the world: test-taking skills A and creativity B. The utility function of the school maps how these skills into a labor market valuation function. The school district is deciding whether or not to reward or punish a school through funding for their ability to give labor-market relevant skills to their students.

Test-taking skills can be examined by administering tests. Creativity is much harder to assess - it may be assessed by interviewing the children for extensive periods and asking them targeted questions, or by having them write stories or essays as part of the tests.

Example 3: Demand for Media

Suppose that a consumer is interested in deciding whether to support a certain policy or candidate and is interested in gaining information from the media. There are two news outlets in the world A and B. The consumer values both the information that A and B provide in making the decision whether to support the policy or candidate, but also has (dis)utility from viewing news on A and B modeled by c_A and c_B.

Example 4: Dating Markets

Suppose a woman is interested in finding a marriage partner and values two characteristics of mates: A is the "looks" of the potential partner and B is how well the woman gets along with the potential partner. Then the cost of inferring A, potentially through looking at a dating profile, is given by c_A while the cost of inferring B, which may consist of actually going on dates and spending time together, is captured by c_B.

2.3 Results of the Two Dimensional Model

The information structure of the model makes the solution a simple dynamic program since the DM can only transition to 4 possible states: seeing nothing (0), seeing only A (A), seeing only B (B), and seeing both A and B (AB).

Let the ex-ante value of each of these transitions to $X \in \{0, A, B, AB\}$ be written as V_X. In terms of the model

$$V_X = \mathbb{E}\Big[\max_a \mathbb{E}[u(a, \omega)|X]\Big].$$

For example, when $X = 0$ this is simply $\mathbb{E}\max_a \mathbb{E}[u(a,\omega)]$ since no information is revealed and when $X = AB$ it is $\mathbb{E}\max_a u(a,\omega)$ since $(A,B) = \omega$ characterizes the full state.

Because there are only 4 states to transition to in the dynamic program, and the probabilities of transitions are independent, it is easy to write down the objective function of the DM in closed form using the V_X terms and the probability terms q_A and q_B. When picking how to acquire information, the DM faces the following objective function:

$$\pi(q) = q_A q_B V_{AB} + q_A(1-q_B)V_A + q_B(1-q_A)V_B + (1-q_A)(1-q_B)V_0 - c_A(q_A) - c_B(q_B). \quad (1)$$

and solves the problem

$$\max_{q\in[0,1]^2} \pi(q). \quad (2)$$

Equation (1) and the problem (2) are the most important equations in the methodology of the theory. It shows how one can express this special case of the information acquisition problem as a simple maximization problem. (1) shows the analogy to the standard consumer problems (with quasi-linear utility) and firm problems in classical economic theory.

Before moving onto understanding the maximization problem and general properties of the solution, some algebraic manipulation makes the objective function a bit more manageable:

$$\pi(q) - V_0 = q_A q_B \tilde{V}_{AB} + q_A \tilde{V}_A + q_B \tilde{V}_B - c_A(q_A) - c_B(q_B) \quad (3)$$

where $\tilde{V}_{AB} = V_{AB} + V_0 - V_A - V_B$ and $\tilde{V}_A = V_A - V_0$, $\tilde{V}_B = V_B - V_0$.

This expression is already suggestive of some important features of the problem: the marginal benefit of good A will be related to $\tilde{V}_A$ (similar for B) and the complementarities between the information goods will be measured by $\tilde{V}_{AB}$. Moreover, $\tilde{V}_{AB}$ is exactly analogous to the Börgers et al. (2013) definition of substitutes and complements.

The value of (1)-(3) is that, because of the technical assumptions on the cost functions, one can use the first-order conditions to pin down the solutions.[2] Because the cost of full revealing is infinite, the maximizing choices of q_A and q_B satisfy

$$\begin{aligned} q_B^* \tilde{V}_{AB} + \tilde{V}_A &\le c_A'(q_A^*), \text{ with equality if } q_A^* > 0, \\ q_A^* \tilde{V}_{AB} + \tilde{V}_B &\le c_B'(q_B^*), \text{ with equality if } q_B^* > 0. \end{aligned} \quad (4)$$

Each of these conditions exhibit the marginal benefit of each information good on the left of each, which is the value to the DM of finding the true value of that dimension of the state. The

[2]Existence of a solution follows because the choice set $q \in [0,1]^2$ is compact and the objective is continuous due to the well-behaved cost functions.

right hand side is the marginal cost of the goods and is a flexible function of the probability of revelation. The marginal benefit includes both the standard marginal "value of information" $\tilde{V}_A$ and $\tilde{V}_B$ which measures the value of realizing the dimension A and B over having no information along with a demand-weighted term on $\tilde{V}_{AB}$. It is in this sense that $\tilde{V}_{AB}$ suggests substitutes and complements: as q_B^* increases $\tilde{V}_{AB}$ shows that the marginal benefit of A is increasing or decreasing depending on the sign of $\tilde{V}_{AB}$.

This first-order condition alone can be adapted to many applications of information acquisition. With an appropriate selection of utility, state space, prior and cost function, one can easily pin down the optimal information acquisition problem as I show in Section 3.

I now seek two questions with analogies to the standard demand problems. The first is what Becker (1962) calls the "fundamental theorem" of traditional economic theory. What is the cost (or price) effect of information demand for a single information good? The second is important for understanding the dependencies of demand over different information goods. When are information goods complements and/or substitutes?

To answer these questions it is necessary to develop a meaningful sense of comparative statics on the cost functions. To do this I introduce dummy parameters γ_A and γ_B s.t.

$$\frac{\partial c'_A}{\partial \gamma_A} > 0, \frac{\partial c'_B}{\partial \gamma_B} > 0$$

so that these are marginal cost shifters. They are analogous to prices.

The first result shows that there are no Giffen goods in this model.

Proposition 1. $\frac{\partial q_A^*}{\partial \gamma_A} \leq 0$ and $\frac{\partial q_B^*}{\partial \gamma_B} \leq 0$

The proof of this result follows quite easily from the first order conditions and the intuition comes from looking at the first order conditions (4) and noticing that utility has a quasi-linear form. Because of quasi-linear utility, there are no income effects and so price effects must be negative. The technical innovation is in generalizing to non-linear information costs instead of prices.

The importance of this result is that this provides a uniform and robust prediction of how demand for an information good changes as its marginal cost increases. As motivated by the examples above, this prediction could be tested in empirical settings.

The next goal is to think about complements and substitutes. To do this, I develop the following definition of substitutes and complements using the dummy parameters.

Definition 1. A and B are *substitutes* if $\frac{\partial q_A^*}{\partial \gamma_B} \geq 0$ and $\frac{\partial q_B^*}{\partial \gamma_A} \geq 0$, and A and B are *complements* if $\frac{\partial q_A^*}{\partial \gamma_B} \leq 0$ and $\frac{\partial q_B^*}{\partial \gamma_A} \leq 0$.

These definitions are natural in the sense that as the marginal cost or price of realizing information on one dimension increases, you either substitute to using another dimension or react similarly on another dimension.

The next result provides a necessary and sufficient condition for substitutes.

Proposition 2. A and B are substitutes if and only if $\tilde{V}_{AB} \leq 0$ and they are complements if and only if $\tilde{V}_{AB} \geq 0$.

The proof of this result is similar to the proof of Proposition 1 and heavily utilizes the first-order conditions. The intuition is very clear from these conditions. Suppose that the marginal cost of dimension B is reduced. By Proposition 1, this will induce an increase in q_B^*. Now the left-hand-side of the first-order condition for q_A^* will increase if $\tilde{V}_{AB}$ is positive and decrease if $\tilde{V}_{AB}$ is negative. If the condition holds at equality, since $c_A'(q_A)$ is increasing (by convexity) this will lead to a respective increase or decrease.

These conditions are the same as in Börgers et al. (2013). That paper provides an excellent intuition for understanding the $\tilde{V}_{AB}$ object from an information experiment perspective. Consider two possible experiments I present to the DM: (1) I toss a coin, if it lands heads then I tell the DM the full state ω and if it lands tails I tell the firm nothing; (2) I toss a coin, if it lands heads then I tell the firm A only and if it lands tails then I tell the firm B only. The value of experiment (1) is $(1/2)V_{AB} + (1/2)V_0$ and the value of experiment (2) is $(1/2)V_A + (1/2)V_B$ and so the difference in values of these experiments properly explains this object.

Since the result is so strikingly similar to the definition in Börgers et al. (2013) it is worth devoting some time and analysis to the comparisons between the two. Börgers et al. (2013) consider a more classical Blackwell setup of information acquisition. They posit a definition of signal complements and substitutes, emphasizing the quantifier that the definition must hold *for all* decision problems in the Blackwell spirit. Their definition is meant to capture a "willingness to pay" for a signal based on the availability of the other signal. However, Börgers et al. (2013) has no notion of prices or costs in their model. My definition of substitutes and complements is motivated more directly by economic theories of firm and consumer choice.[3] Since traditional notions of substitutes and complements do not hold for all utility functions, my definition does not have this requirement.[4]

[3]Of course, in more general classical theory, there are multiple definitions of substitutes and complements with regards to Hicksian and Marshallian demand. Because of the quasi-linear form of utility in this model, I do not need to worry about such distinctions.

[4]Indeed to demand such a property of traditional notions of substitutes and complements would rob them of their information completely. A utility function of the form $u(x_1, x_2) = x_1 + x_2 + c \cdot x_1 x_2$ can have substitutes or complements depending on the sign of c and so basic examples show no goods would be substitutes and complements. In this sense, it is actually quite remarkable that any two signals are substitutes and complements in Börgers et al. (2013).

The intuition that Börgers et al. (2013) have for complements and substitutes is comparable to the intuition I gain in my parametric analysis below. However, I argue that my analysis unifies and simplifies the separate intuitions from Börgers et al. (2013) since in my quasi-linear model the signal classes must always be substitutes or complements due to the looser requirements on the definition holding for all decision problems and the quasi-linear assumptions. In particular, their definition of conditional uninformativeness can be simplified in my analysis to "with high probability seeing the other signal will not change the DM's decision", whereas their definition of meaning reversal can be simplified to "with high enough probability seeing the other signal will change the DM's decision". As well, due to the fact that I drop the "for all decision problem" quantifier in my analysis these intuitions come from simple probability theory arguments as opposed to more abstract Blackwell arguments.

Propositions 1 and 2 guarantee that this model has well-behaved price effects that can be tested in the data, and also has a clear condition characterizing if multi-dimensional information acquisition displays substitutes and complements.

Equipped with these propositions and the model, it is much easier to map specific empirical applications to the model and then make economic predictions. As shown in the examples above, this requires picking the dimensions of the state space to correspond to some empirical proxy, eliciting beliefs of the prior or estimating the distribution of the state space from previous data, specifying a utility function, and then specifying a cost function. The cost function could come naturally from the selection of the state space, as discussed in the examples above. Thus I see a contribution of this paper and this model to be suggesting one way to moving the theory of information acquisition closer to the realm of empirical and applied work.

To exposit the usefulness and practical insights of the model, the rest of the paper focuses on applications. These applications show that the model has two important and novel insights for understanding these information acquisition problems: understanding better the roots of information substitutes and complements, and showing the importance of correlation in the prior and costs of acquisition for optimal information acquisition demand.

3 Application: Firm Recruiting

Firm recruiting has an extensive information demand component and is an important problem currently being heavily affected by changing technology.[5] In particular, firms now face the challenge of deciding how much time to devote to technology driven assessments better suited towards

[5] Another important theory paper on firm recruiting and screening studying different parts of the problem include Frankel (2019) who emphasizes potential bias of interviewers and information design problems in the recruiting problem.

assessing candidates technical skills or investing in more traditional interviews. What will they choose to do?

The set-up of the firm recruiting application is as follows. Let A be technical skills and B be social skills and suppose that the firm is deciding whether to hire ($a = 1$) or not hire ($a = 0$) a worker with unknown skills so that the action space is simply $\{0, 1\}$. A worker with skills A and B produces profit

$$f(A, B) = \lambda A + (1 - \lambda)B$$

where $\lambda \in [0, 1]$ so that there is a job or firm-specific weighting of skills as in Gibbons and Waldman (2004) and Lazear (2009).

It is costly for the firm to hire the worker independent of gathering information. Suppose that the profit is normalized so that $A \in \mathbb{R}$ and $B \in \mathbb{R}$ so that workers can have negative skills relative to this cost.[6] The firm's utility function is their profit from hiring which is

$$u(a, \omega) = \mathbf{1}\{a = 1\}f(A, B).$$

Clearly the firm wants to hire only if $f(A, B) \geq 0$.

Finally suppose that the firm's prior over the worker's skill (or the market distribution of skills) is normal with mean 0, unit variance and correlation ρ:[7]

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Note that this implies that $\mathbb{E}[f(A, B)] = 0$ so that the firm is indifferent between hiring the worker and not hiring the worker with no extra information.

This application is of particular interest when we think about the sources of gaining information on technical and social skills. Suppose for simplicity that we group technical skill information sources as being advanced technology (resume screening, AI, etc.) and we group social skill information sources as being human based screening (interviews, informal talking etc.). Then this question speaks to a large literature on labor economics which examines the impacts of AI and robots on labor demand.[8] In this setting, one particularly interesting question is whether HR and interviews will be entirely replaced by AI and robots. This model examines this problem through

[6]This cost might be dealing with legal requirements of hiring, or other fixed costs of hiring.

[7]Variances act to "scale" the importance of the relative dimensions and play very similar roles to the weights λ so I normalize them to reduce the number of parameters. The mean 0 assumption on both the marginals and the value of the worker is not w.l.o.g. but my focus is on the profit function parameters and the correlation.

[8]E.g. Autor (2015), Acemoglu and Restrepo (2018). Hoffman et al. (2018) is directly related as it looks at the value of manager interviews and decisions in hiring. This paper provides a particularly interesting contrast to the insights from the model below.

the lens of firm recruiting under the stylized assumption that all technical skill screening is done by robots and all social skill screening is done by humans.

3.1 Resources devoted to evaluating technical skills

Proposition 1 states that as the information in any of the dimensions becomes cheaper, the firm will use that demand that information good more. This is independent of any assumptions on the cost functions other than the convexity and smoothness assumptions.

What seems most likely if we take the view expressed earlier about the partitioning of the dimensions is that the cost of finding out technical skills has been reduced over time: firms are learning better how to use AI and technology to find out this information. Proposition 1 then tells us that, holding fixed their profit functions and the structure of skills and jobs, firms should be using more AI and technology, even through basic testing and resume screening, to gather information on their potential hires. Anecdotally, AI-based recruiting firms has seen a rise in recent history.[9]

Is the cost of receiving information about social skills also decreasing? This is less clear. While technology allows for easier communication, these types of information retrievals still require human face-to-face interactions under the stylized interpretation that only humans can learn about social skills. If we proxy the costs of these methods by the cost of the time of those conducting the interviews it is not clear which direction these costs have gone.

Thus while it seems a relatively sensible prediction within this model that the use of technology for technical screening has increased over time, it is less clear how the use of interviews has been impacted. To gain more insight into this requires an understanding of the cross-price interactions - whether these information goods are substitutes or complements.

3.2 How will interviews be impacted?

Proposition 2 shows that $\tilde{V}_{AB}$ is the object that fully characterizes substitutes and complements. Under the assumptions made above I show in the Appendix that

$$\begin{aligned}
V_0 &= 0 \\
V_A &= \phi(0)|\lambda + \rho(1-\lambda)| \\
V_B &= \phi(0)|(1-\lambda) + \rho\lambda| \\
V_{AB} &= \phi(0)\sqrt{\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda)\rho}
\end{aligned}$$

where $\phi(\cdot)$ is the density for the standard normal distribution.

[9] See this company for example which claims that the current theme in recruiting is AI.

The derivation of these values is actually quite informative in that it provides useful intuition on the structure of recruiting and examples of the derivations required of the more general model in the normal prior case. Other algebraic details are left to the Appendix.

There are two cases that are less interesting - these are the cases in which no information is revealed and where full information is revealed. If no information is revealed the firm is indifferent about hiring the worker receiving an expected payoff of 0. If all the information is revealed the firm can make a perfectly informed decision. Thus using the fact that $f(A, B) \sim N(0, \lambda^2 + (1 - \lambda)^2 + 2\lambda(1 - \lambda)\rho)$ and taking a conditional expectation of $f(A, B)$ given that $f(A, B) > 0$ gives us V_{AB}.

The interesting case is when information on only one dimension is revealed by a good. Suppose for example that dimension A is revealed to be a. The conditional expectation of $f(A, B)$ is

$$\mathbb{E}[f(A, B)|A = a] = \lambda a + (1 - \lambda)\rho a = a \cdot d_A$$

using properties of the conditional distribution when variables are distributed as a bivariate distribution.

Since the firm will hire after seeing some a if and only if this conditional expectation is non-negative, this expression shows that for any realized value $a \in \mathbb{R}$ whether or not the firm hires the worker depends on the sign and size of a and the sign of $d_A = \lambda + (1 - \lambda)\rho$. In particular if $d_A > 0$ then the firm will hire if and only if $a \geq \underline{a}$ while if $d_A < 0$ the firm will hire if and only if $a \leq \bar{a}$ for some thresholds $\underline{a}$ and $\bar{a}$.

These two scenarios represent distinct (second-stage) hiring schemes: in one scenario if the firm realizes a worker is particularly good at technical tasks ($a \geq \underline{a}$) the firm will hire the worker. In another, if the firm realizes a worker is particularly good at technical tasks the firm will not hire the worker; the firm will hire the worker if and only if the worker is particularly bad at technical tasks ($a \leq \bar{a}$) due to the negative correlation. Thus these represent distinct "good news is good news" and "good news is bad news" scenarios.

The analysis is symmetric when realizing the value of dimension B. The analogous parameterized expression is $d_B = 1 - \lambda + \lambda\rho$. Note the hiring mechanism can be asymmetric in the following sense - the firm may hire workers who are technically skilled and hire workers who are socially inept (or vice versa). This is true because the signs of d_A and d_B do not need to match. In fact, analogous to the parameterized Roy model, the firm will not adopt the "good news is bad news" philosophy in *both* directions: it is impossible for d_A and d_B to both be negative.

The importance of the expressions is that they fully characterize $\tilde{V}_{AB}$. In Figure 1 I plot the complements and substitutes in (ρ, λ) space.

This plot shows a few important things. First, there exists a non-trivial region wherein infor-

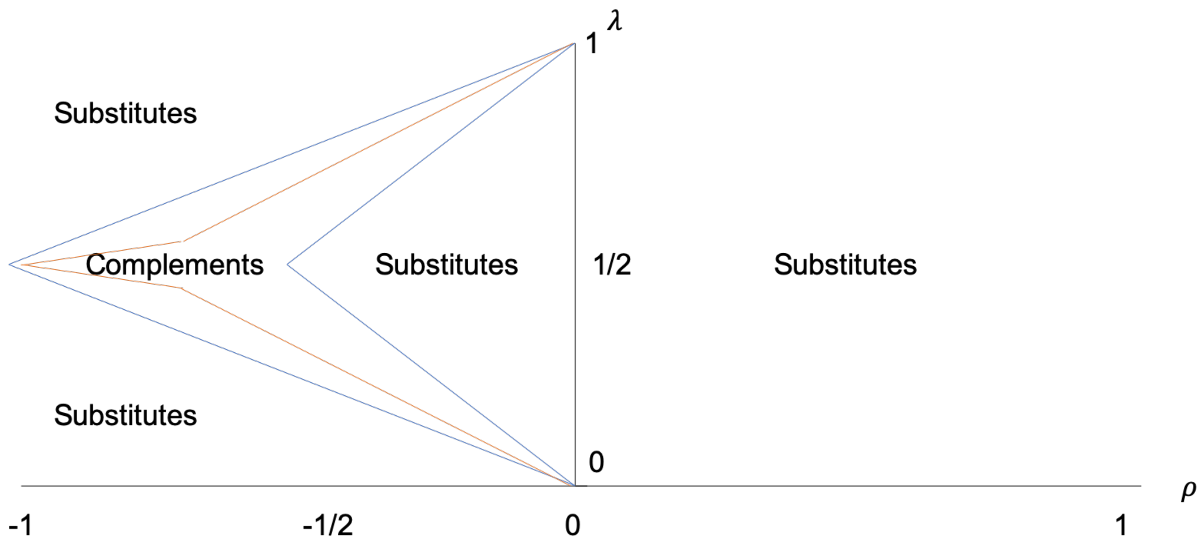

Figure 1: Substitutes and Complements in the Firm Recruiting Problem
Notes: This is a plot of the regions where information goods are substitutes and complements in (ρ, λ) space in the firm recruiting application example. Substitutes indicates that $\tilde{V}_{AB} \leq 0$ and Complements indicates that $\tilde{V}_{AB} \geq 0$.

mation sources are complements. This insight has important implications for the applied question of interest: under the parametric assumptions, there are cases where a decrease in the cost of AI induces more use of AI, but also induces more use of interviews in recruiting. Thus, the worry of AI completely replacing humans is not unconditionally true within this model of information demand treating AI and interviews as different information goods. Within the model, the distinction between substitutes and complements depends crucially on the state of the world with regards to the specific job technology (λ) and the correlation of skills (ρ).

A key feature in Figure 1 is that negative correlation is necessary to induce complementarities. As well, the measure of correlations ρ such that complements exist is increasing as λ approaches 1/2. The interpretation of this is that as jobs or firms become more balanced and skills become more negatively correlated in the job market, it is more likely that technology based information goods and human based information goods are complementary. This is a testable implication.

To gain mathematical intuition for generating complements, consider the set of (ρ, λ) s.t. $d_A = 0$. In Figure 1 this is given by the bottom orange line. When $d_A = 0$, the expectation of f given $A = a$ is independent of a. Thus, $V_A = 0$ the prior mean. When $V_A = 0$ we know that $\tilde{V}_{AB} = V_{AB} - V_B > 0$ if λ and ρ are not at their extreme values.[10] Since $\tilde{V}_{AB}$ is continuous in

[10]This can be seen by easy computation:

$$\sqrt{\lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda)\rho} = \sqrt{(1-\lambda+\rho\lambda)^2 + (1-\rho)\lambda^2} > |1-\lambda+\rho\lambda|.$$

(ρ, λ) there must be some neighborhood around each point of this (ρ, λ) in which $\tilde{V}_{AB}$ is positive in this neighborhood. That neighborhood clearly depends on the original (ρ, λ) and shrinks as the dimensions approach independence. The same discussion applies for (ρ, λ) where $d_B = 0$.

These exercises also make clear the value of the general model. Using only simple probability theory arguments and computations, one can derive important predictions on the structure of the demand for information.

3.3 Closed form demand and the irrelevance of assessing technical skills

The model also yields useful insights in this application as it allows for closed-form solutions. To illustrate this and take advantage of the flexible cost modeling I examine an example with a quadratic parametric assumption on the cost function.

Suppose that

$$c_A(q_A) = \gamma_A \frac{q_A^2}{2},\ c_B(q_B) = \gamma_B \frac{q_B^2}{2}.$$

To ensure that there is no corner solution at $q_A = 1$ or $q_B = 1$ I assume that γ_A and γ_B are sufficiently large. In this case $\min\{\gamma_A, \gamma_B\} \geq \phi(0)$ is sufficient.

Doing some simple algebra it is easy to derive from the first-order conditions (4) at equality that:[11]

$$\begin{aligned} q_A^* &= \frac{\gamma_B \tilde{V}_A + \tilde{V}_B \cdot \tilde{V}_{AB}}{\gamma_A \gamma_B - \tilde{V}_{AB}^2} \\ q_B^* &= \frac{\gamma_A \tilde{V}_B + \tilde{V}_A \cdot \tilde{V}_{AB}}{\gamma_A \gamma_B - \tilde{V}_{AB}^2} \end{aligned} \tag{5}$$

These are the information acquisition demand functions.

First we can verify the own price effects: clearly each demand is decreasing in their respective marginal cost shifters γ_A and γ_B. Verifying substitutes and complements is a little less obvious from (5) but worth the exercise to show the value of these closed form expressions.

Taking the derivative of q_A^* with respect to γ_B yields

$$\frac{\partial q_A^*}{\partial \gamma_B} = -\frac{\tilde{V}_{AB} \cdot (\gamma_A \tilde{V}_B + \tilde{V}_A \cdot \tilde{V}_{AB})}{(\gamma_A \gamma_B - \tilde{V}_{AB}^2)^2}$$

Ignoring the denominator which is always positive we see that if $\tilde{V}_{AB} \geq 0$ since all terms inside the parentheses are positive, this must be negative (thus complements). If $\tilde{V}_{AB} \leq 0$ then, by

[11]Equality follows here because of the quadratic costs. One can show that if $q_A = 0$ or $q_B = 0$, one can strictly improve profits by raising q_A or q_B by some small enough amount.

assumption, we have that $\gamma_A \geq \tilde{V}_A$ and also it is always true that $\tilde{V}_B + \tilde{V}_{AB} \geq 0$.[12] Thus this is positive (thus substitutes).

I now aim to convey an important insight that presents itself clearly within this model related to the correlation of information sources and how that impacts demand. Consider the case when $\lambda = 0$ so that the profit and output of the worker *only* depends on B the social skills. Standard intuition might suggest that there is no value to screening on technical skills A as they are not directly payoff relevant, and so the firm will have a 0 demand for information good A. However, this intuition is wrong when there is non-trivial correlation.

To make this computation simpler rescale the marginal costs so that $\phi(0)$ is the new unit of scale (which implies that both γ_A and γ_B are larger than 1 under the assumptions). When $\lambda = 0$ the values above simplify to $V_A = |\rho|$, $V_B = 1$ and $V_{AB} = 1 \Rightarrow \tilde{V}_{AB} = -|\rho|$. Already we see that if $\rho \neq 0$, $V_A > 0$. This is precisely because correlation is informative on the payoff-relevant dimension of the state. We also see that A and B are information substitutes.

We can write the demand functions as

$$q_A^* = \frac{|\rho|(\gamma_B - 1)}{\gamma_A \gamma_B - \rho^2}$$
$$q_B^* = \frac{\gamma_A - |\rho|}{\gamma_A \gamma_B - \rho^2}.$$

This shows us two things: first as long as $\rho \neq 0$ and $\gamma_B > 1$, the firm will gather information on the worker's technical skills *even though the worker's technical skills are irrelevant for their profit*. The interior solution for information good A is due to the fact that (1) the marginal costs are close to 0 for small amounts of information gathering and (2) A is correlated with B. The strong assumption of (1) can be dispensed with by modifying the costs and assuming that the cost of A is small enough. But in any case the demand for information A is increasing in $|\rho|$, the informativeness of A about B.

At an extreme case suppose that $|\rho| = 1$. Then $q_A^* \geq q_B^* \Leftrightarrow \gamma_B \geq \gamma_A$ so that costs completely determine relative demand. This is intuitive: when $|\rho| = 1$ A is just as informative as B and so only the costs of each dimension should be considered in the demand.

At another extreme case, suppose that $\gamma_B \to \infty$. Then $q_B^* \to 0$ and $q_A^* \to |\rho|/\gamma_A$. When the payoff relevant dimension of the state becomes infinitely costly the firm completely gives up on gathering information on that dimension and focuses its efforts on the less costly dimension *as long as there is non-trivial correlation*. Even for very small correlations, it is worth it for the firm to invest in the other dimension.

In this firm recruiting application these results gives another insight into the evolution of firm

[12] To see why, note that $\tilde{V}_B + \tilde{V}_{AB} = V_{AB} - V_A \geq 0$ using the ex-ante value of information theorem (AB provides more information than A).

recruiting. Consider jobs where almost all of the job is based on social skills so that $\lambda = 0$. What we have seen is that as long as the information provided by AI and technology is somewhat correlated with the payoff relevant dimension of the state space, and the marginal costs of AI and technology are small enough for a small enough amount of information revelation, these firms will use AI and cutting-edge technology in their recruiting. As stated, marginal costs are assumed to be small and an important model feature that is left out is the fixed cost of information acquisition - there are non-trivial costs for setting up advanced applicant tracking systems that incorporate advanced technologies. These fixed costs will drive down the demand for information good A at these firms.

4 Application: Education Testing

There is a large debate about the value of test-taking for assessing children's skills for the labor market.[13] A concrete economic component of this debate is how schools are funded or rewarded based off of their performance in certain areas. Importantly, school districts and policymakers want to reward schools that achieve standards to improve their students well-being.

In this case suppose that information sources exist in the form of A, the ability to pass a standardized mathematics test, and B, student creativity. Similar to the recruiting example, the labor market returns to the student population for these skills is

$$f(A, B) = \lambda A + (1 - \lambda)B.$$

A policymaker can choose how to assess schools according to the model so that with some probability that can reveal A through assigning tests and with some probability they can reveal B. The policymaker can reward the school ($a = 1$) or not and wants to only reward schools that meet a certain threshold normalized to 0. The policymaker gets a payoff of 1 for rewarding correctly ($a = 1$ and $f(A, B) \geq 0$), and for not rewarding correctly ($a = 0$ and $f(A, B) < 0$) and otherwise gets 0. Then her payoff is

$$u(a, \omega) = \mathbf{1}\{a = \text{sign}(f(A, B))\}$$

As before make the joint-normality assumption on how A and B are distributed with unrestricted correlation:

$$\begin{pmatrix} A \\ B \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

[13] See for example in the popular press: The Washington Post. There is also a large academic literature on the subject. (e.g. Amrein and Berliner (2002))

4.1 Test-taking on the rise

With the advent of technology, the cost of administering large scale tests and processing or summarizing the data should be decreasing over time. The model then predicts through Proposition 1 that, because information on A is a standard good, the policymaker will use more test-taking in their optimal information acquisition strategy. But what are the implications for assessing creativity in this changing world?

4.2 Should creativity be assessed more?

Within the model, how the policy-maker responds to assessing creativity more depends on whether or not the information goods of test-taking and creativity assessment are substitutes or complements. Using Proposition 2, this depends on the sign of $\tilde{V}_{AB}$.

In this application problem, it is easy to derive the value of V_0 and V_{AB}. If the policymaker guesses, she has a 1/4 chance of making the right decision regardless, and so $V_0 = 1/4$. If the policymaker knows the true labor market outcomes of the students, she will always make the right decision so that $V_{AB} = 1$.

Suppose now the policymaker observes $A = a$. Then $f(A, B)$ is distributed as $N((\lambda + (1 - \lambda)\rho)a, (1 - \lambda)^2(1 - \rho^2))$ so the probability that $f(A, B)$ is positive, the probability that the school has given students sufficient labor market skills, is

$$1 - \Phi\left(\frac{-(\lambda + (1 - \lambda)\rho)a}{(1 - \lambda)\sqrt{1 - \rho^2}}\right) = \Phi\left(\frac{(\lambda + (1 - \lambda)\rho)a}{(1 - \lambda)\sqrt{1 - \rho^2}}\right) \tag{6}$$

The complementary probability is the probability it is negative. The optimal strategy is to clearly pick $a = 1$ when (6) is larger than 1/2, and otherwise pick $a = 0$. The probability (6) passes through 1/2 at $a = 0$.

Then, integrating over A, the potential realizations of test scores according to the prior distribution, the ex-ante expected payoff to seeing testing A is

$$V_A = \int_{-\infty}^{0} (1 - \Phi\left(\frac{(\lambda + (1 - \lambda)\rho)}{(1 - \lambda)\sqrt{1 - \rho^2}}a\right))\phi(a)da + \int_{0}^{\infty} \Phi\left(\frac{(\lambda + (1 - \lambda)\rho)}{(1 - \lambda)\sqrt{1 - \rho^2}}a\right)\phi(a)da$$

and this can be simplified using normal distribution rules to

$$V_A = 2\int_{0}^{\infty} \Phi(d_A a)\phi(a)da$$

where $d_A = \frac{\lambda+(1-\lambda)\rho}{(1-\lambda)\sqrt{1-\rho^2}}$. By symmetry,

$$V_B = 2\int_0^\infty \Phi(d_B b)\phi(b)db$$

where $d_B = \frac{1-\lambda+\lambda\rho}{\lambda\sqrt{1-\rho^2}}$.

Now, whether creativity is assessed more as it becomes easier to administer tests depends on whether $\tilde{V}_{AB} \geq 0$. In this case

$$\tilde{V}_{AB} = \frac{5}{4} - 2\int_0^\infty \Phi(d_A a)\phi(a)da - 2\int_0^\infty \Phi(d_B b)\phi(b)db.$$

To gain some intuition for when this term is positive, consider $\lambda = 1/2$. In this case

$$d_A(\rho) = d_B(\rho) = \frac{1+\rho}{\sqrt{1-\rho^2}}$$

where these are decreasing in $\rho \in (-1, 1)$. Note that at $\rho = 0$, so that math skills and creativity are completely independent, we have $\int_0^\infty \Phi(a)\phi(a)da = 3/8$.[14] and so $\tilde{V}_{AB} = \frac{5}{4} - \frac{6}{4} = -1/4 < 0$ thus more test taking assessments can replace assessing creativity. However, if ρ is close enough to -1, then $d_A = d_B$ is small and the integrals are smaller, thus leading to the possibility of complements.

This intuition can be extended: creativity should and will be assessed more by the policymaker if λ is close to $1/2$, so that passing math tests and creativity skills are approximately equally weighted in labor market returns, and $\rho < 0$ so that these skills are negatively correlated. In particular, if students that tend to be better at math are also more creative, or the weights on math or creativity are close to 1, then the optimal strategy would invest more in using cheaper tests instead of other methods to assess creativity.

4.3 Using testing, even if it's irrelevant

Consider the case of $\lambda = 0$: passing a math test has no direct impact on the labor market returns of students. Instead, labor market returns are entirely determined by creativity. Should the policymaker still administer tests to students?

In this case, c_B is undefined as defined above. Instead, when $\lambda = 0$ we know that $V_B = V_{AB}$ because the policymaker knows all the relevant skills for labor market returns. Consider V_A it is

$$V_A = 2\int_0^\infty \Phi(\rho a/\sqrt{1-\rho^2})\phi(a)da$$

[14] This can be found by integrating by parts.

which is strictly positive if $\rho \neq 0$. Thus, using the cost structure used above, it is easy to show that in this case, test taking is still valuable as long as it is not too costly. In particular, as above, if test taking is cheap enough relative to assessing creativity in students and $\rho \neq 0$, it may be optimal to invest more in revealing test scores accurately than revealing creativity.

5 Further Applications

5.1 Dating Markets

With the proliferation of dating apps and online dating, there are more information sources available about prospective partners. The model makes two predictions in this setting. First, even if information on a dating profile is not directly relevant to the payoff of dating someone, since it is viewable with a low cost and likely correlated with the true payoff to dating that person it is likely to be heavily used in gathering information on prospective partners. Second, if superficial attributes on dating profiles that reveal positive information about a prospective partner are negatively correlated with other important attributes, such as how well you get along with them, then more effort will be spent on discovering those attributes as the cost of seeing dating profiles decreases. A concrete example could be putting in more effort to have face-to-face interactions.

To see how these predictions come out of the model, consider again the setup of the firm recruiting problem but now interpret the firm as a person and the "applicant" as a prospective dating or marriage partner. Suppose A is the "looks" and "superficial" parts of a partner. These are things you can see on someone's dating profile. B is how well you get along with the partner, which may require more costly investments such as going on dates and meeting up in-person. $f(A, B) = \lambda A + (1 - \lambda)B$ describes the marriage or long-term dating payoff.

Then, the first prediction comes from the fact that even if $\lambda = 0$ as long as $\rho \neq 0$ and γ_A is small, q_A^* could be large relative to q_B^*. The second prediction comes from the fact that if $\rho < 0$ and $\lambda \approx 1/2$ then $\tilde{V}_{AB} > 0$ so that according to Proposition 2 the information sources are complements.

5.2 Media Consumption

Political polarization and how consumers consume information from media sources is important in today's combative political environment. Consider the following view of media consumption: consumers use the Left (A) or Right (B) media to make political decisions, such as choosing who to vote for in a Presidential election. The true state of the world is that either the Democratic candidate is better for the country ($f(A, B) > 0$) or the Republican candidate is better ($f(A, B) < 0$). The consumer weights the opinions of each media source according to λ so that $f(A, B) = \lambda A + (1 -$

$\lambda)B$ and wants to elect the "correct" candidate. The model makes two predictions in this setting. Since information from these sources is likely negatively correlated, impartial consumers (those with $\lambda \approx 1/2$) will utilize more of both media consumption as it becomes cheaper to consume media. Second, if people get more utility from gathering information from one media source then they will acquire more information from that source holding all else constant. To see how these predictions are generated by the model, notice that the setup is identical to the education application and utilizes Proposition 2. The second prediction simply comes from the own-cost effect result Proposition 1.

6 Conclusion

This paper presented a model of costly information acquisition. The setup interprets information revelation on the dimensions of the state space as information goods each with their own cost or price. The optimal information acquisition decision is pinned down by a first-order condition. The model exhibits some predictions with respect to marginal costs independent of cost structures and any other assumptions on utility or action spaces. First, as the marginal cost of one information good increases, the DM demands less of it. Second, the model gives a necessary and sufficient conditions for when goods are substitutes and complements.

The model is applied to two applications. In the context of labor recruiting, the model predicts that with an exogenous decrease in the marginal cost of using advanced technologies to screen candidates (1) advanced technology is used more in recruiting (2) interviews may decrease or increase in use in recruiting but it is necessary that technical and social skills are negatively correlated for interviews to increase. The model also shows that companies hiring for a position requiring none of a certain skill may still screen for this skill if there is non-trivial correlation and marginal costs are small enough at small consumption. They may even screen for the irrelevant skill more if the correlation is high enough and the marginal costs of screening for the relevant skill are high enough.

In the context of education, the model predicts that technology making testing cheaper will cause testing to be used more in evaluating school performance. However, policymakers may also want to invest more in assessing creative skills if creative skills and analytical skills assessed by testing are approximately equally weighted in the labor market wage function. Also, even if test taking skills are not directly applicable to labor market returns, if they have some correlation and are less costly, policymakers will use them to assess student performance.

It is clear that the model is extremely stylized with respect to the information structure assumptions - in particular the case of partial information revelation, a huge case in the literature, is not considered at all by the model. Adapting the model to noisy signals is not straightforward. One potential avenue for adapting the methodology to noisy signals is to choose a vector of precisions over the different goods which would be analogous to information goods. This has proved to be tractable in the normal-normal world where the prior and signals are noisy, but it becomes slightly more difficult to characterize the ex-ante values as a function of the precisions without making specific assumptions on utility and actions as Van Neiuwerburgh and Veldkamp (2010) do.

Another potential way to approximate noisy signals through the current setup of the model is to instead enlarge the dimension of the state space and think about every possible signal received as being binary on each dimension. Because of the number of possible transitions in the dynamic program, state spaces with more than 2 dimensions are difficult to characterize with the simple independent probability draw model.[15] However, some perturbations of the signaling technology here could yield valuable insights.

Overall, the goal of this paper was to provide a way for economists to conceptualize information acquisition as a standard consumption or investment problem, and also illustrate some subtle insights from such a conceptualization that can be taken to the data.

References

Acemoglu, D. and Restrepo, P. (2018). Robots and Jobs: Evidence from US Labor Markets. *Manuscript*, page 98.

Amrein, A. L. and Berliner, D. (2002). High-Stakes Testing, Uncertainty, and Student Learning. *Education Policy Analysis Archives*, 10(18):74.

Autor, D. H. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives*, 29(3):3–30.

Becker, G. S. (1962). Irrational Behavior and Economic Theory. *Journal of Political Economy*, 70(1):1–13.

Börgers, T., Hernando-Veciana, A., and Krhmer, D. (2013). When are signals complements or substitutes? *Journal of Economic Theory*, 148(1):165–195.

Caplin, A. and Dean, M. (2013). Behavioral Implications of Rational Inattention with Shannon Entropy. Technical Report w19318, National Bureau of Economic Research, Cambridge, MA.

Frankel, A. (2019). Selecting Applicants. *Manuscript*.

[15] One needs to compute 2^K values for K dimensions.

Gibbons, R. and Waldman, M. (2004). Task-Specific Human Capital. *American Economic Review*, 94(2):203–207.

Hoffman, M., Kahn, L. B., and Li, D. (2018). Discretion in Hiring. *The Quarterly Journal of Economics*, 133(2):765–800.

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

Lazear, E. (2009). FirmSpecific Human Capital: A SkillWeights Approach. *Journal of Political Economy*, 117(5).

Matějka, F. and McKay, A. (2015). Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model. *American Economic Review*, 105(1):272–298.

Pomatto, L., Stack, P., and Tamuz, O. (2019). The Cost of Information. *Manuscript*.

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

Van Neiuwerburgh, S. and Veldkamp, L. (2010). Information Acquisition and Under-Diversification. *The Review of Economic Studies*, 77(2):779–805.

Verrecchia, R. E. (1982). Information Acquisition in a Noisy Rational Expectations Economy. *Econometrica*, 50(6):1415–1430.

7 Appendix Proofs and Derivations

7.1 Proofs of Propositions 1 and 2

The proofs of Propositions 1 and 2 are established by the following two Lemmas.

Lemma 1. Let

$$\eta_X(p;\gamma) := -\frac{\frac{\partial c'_X(p_X;\gamma_X)}{\partial \gamma_X}}{\tilde{V}_{AB}^2 - c''_A(p_A;\gamma_A)c''_B(p_B;\gamma_B)}.$$

Then

$$\frac{\partial p_X^*}{\partial \gamma_X} = c''_X(p_X;\gamma_X)\cdot \eta_X(p^*;\gamma)$$
$$\frac{\partial p_X^*}{\partial \gamma_{-X}} = \tilde{V}_{AB}\cdot \eta_X(p^*;\gamma)$$

The proof of this lemma simply uses the implicit function theorem and algebraic manipulations to rearrange the first-order conditions (4) at equality. I omit the algebra for simplicity. It is clear from this lemma that we should be interested in the sign of η_X.

Lemma 2. $\eta_X(p^*;\gamma) \geq 0$. If $\frac{\partial c'_X(p_X;\gamma_X)}{\partial \gamma_X} > 0$ then $\eta_X(p^*;\gamma) > 0$.

Proof. It is sufficient to prove that $\tilde{V}_{AB}^2 < c''_A c''_B$ since the numerator of η_X is always non-positive by assumption. Note that if these expressions are equal then we cannot apply the Implicit Function theorem and perform these calculations (the relevant determinant is 0).

p^* is a local maximum so the Hessian of the objective function must be negative semi-definite. The Hessian is given by

$$\begin{bmatrix} -c''_A(p^*;\gamma) & \tilde{V}_{AB} \\ \tilde{V}_{AB} & -c''_B(p^*;\gamma) \end{bmatrix}.$$

This is negative semi-definite iff its eigenvalues are non-positive. The eigenvalues are given by solving

$$(c''_A + \lambda)(c''_B + \lambda) - \tilde{V}_{AB}^2 = 0 \Leftrightarrow \lambda^2 + (c''_A + c''_B)\lambda + (c''_A c''_B - \tilde{V}_{AB}^2) = 0.$$

Since the matrix is symmetric, the eigenvalues are real and thus characterized by the quadratic formula

$$\lambda = \frac{-(c''_A + c''_B) \pm \sqrt{c''^2_A - 2c''_A c''_B + c''^2_B + 4\tilde{V}_{AB}^2}}{2}$$

Now suppose that $\tilde{V}_{AB} > c''_A c''_B$. Then

$$\sqrt{c''^2_A - 2c''_A c''_B + c''^2_B + 4\tilde{V}_{AB}^2} > \sqrt{c''^2_A + 2c''_A c''_B + c''^2_B} = c''_A + c''_B$$

and so we have that one of the solutions λ is strictly positive, contradicting that the Hessian was negative semi-definite. Thus, we must have that $\tilde{V}_{AB}^2 < c''_A c''_B$ (at p^*) and so $\eta_X(p^*;\gamma) \geq 0$ as required.

If the numerator is negative (i.e. marginal costs strictly shift) then we have shown that $\eta_X > 0$ as required. □

Lemmas 1 and 2 and the fact that $c''_X \geq 0$ due to convex costs imply Propositions 1 and 2.

7.2 Derivation of Value Formulas in Section 3

Next I derive the values in the Application in Section 3.

Derivation of value formulas. A lot of the work is done in the main body of the text. This is to iron out the details more clearly.

I will use the following facts about the normal distribution: Let

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

Then

$$\begin{aligned}
Y|_{X=x} &\sim N(\rho x, (1-\rho^2)) \\
aX + bY &\sim N(0, a^2 + b^2 + 2ab\rho) \\
\mathbb{E}[X|X \geq 0] &= \frac{\phi(0)}{1-\Phi(0)} \\
\mathbb{E}[X|X \leq 0] &= -\frac{\phi(0)}{\Phi(0)}
\end{aligned}$$

where $\phi(\cdot)$ is the standard normal pdf and $\Phi(\cdot)$ is the standard normal cdf.

The first two facts are used in the main text to get $\mathbb{E}[f|X = x]$ and the second and third facts are used to compute V_{AB}.

Now consider computing V_A. As shown in the main text we have that

$$\mathbb{E}[f|A = a] = a \cdot d_A.$$

Suppose that $d_A > 0$. Then this is clearly non-negative if and only if $a \geq 0 = \underline{a}$ and thus the acceptance region is $[0, +\infty)$ as required.

Thus

$$\begin{aligned}
V_A &= \mathbb{E}[f|A \geq \underline{a}]\mathbb{P}\left(A \geq \underline{a}\right) \\
&= \mathbb{E}[\mathbb{E}[f|A = a]|A \geq \underline{a}](1 - \Phi(0)) \\
&= \mathbb{E}[ad_A|A \geq \underline{a}](1 - \Phi(0)) \\
&= \left(d_A \frac{\phi(0))}{1-\Phi(0)}\right)(1 - \Phi(0)) \\
&= \phi(0)(\lambda + (1-\lambda)\rho)
\end{aligned}$$

where the second equality uses the law of iterated expectations.

Suppose that $d_A < 0$, then $\mathbb{E}[f|A = a]$ is non-negative if and only if $a \leq 0 = \bar{a}$. Then we get that, by a similar calculation, and using the fact that $-X \geq 0 \Rightarrow X \leq 0$ and the last fact above,

$$V_A = -\phi(0)(\lambda + (1-\lambda)\rho).$$

If $d_A = 0$ then $V_A = \mathbb{E}[f] = 0$ (it is invariant to the hiring choice).

Thus

$$V_A = \phi(0)|\lambda + (1-\lambda)\rho|.$$

The computation for V_B is symmetric. $\square$