The Implicit Assumptions of Information Cost Functions*

Savitar Sundaresan† Enrico Zanardo‡

March 21, 2022

Abstract

This paper links the literature microfounding information cost functions with the literature that uses them. We introduce six axioms for an information cost function, and prove there is a unique function satisfying them. The function we characterize is increasing in the informativeness of experiments, and additively separable in signal realizations, which is consistent with received wisdom. The solution is simple and intuitive, and importantly, distinct from commonly used alternatives, which do not generally satisfy the properties we assume as axioms. We categorize the implicit assumptions made by other cost functions, and assess the qualitative and quantitative impact on their interpretability.

*We are grateful to Harjoat Bhamra, Navin Kartik, Ramana Nanda, Taraun Ramadorai, and seminar participants at Imperial College and Columbia University for helpful comments and discussions. All errors are our own.

†Business School, Imperial College London; Email: s.sundaresan@imperial.ac.uk

‡Analysis Group; Email: zanardo.enrico@gmail.com

1 Introduction

The economic and financial theoretical literature on information acquisition is massive. Most papers that analyze endogenous information decisions use a cost function for information acquisition, the specifics of which are at the discretion of the modeler. There has been a recent effort to formally justify the functions selected. The work by Sims (2003) and Woodford (2012) has developed functional forms inspired by Shannon (1948), which started the literature on information theory. Works such as Pomatto, Strack, and Tamuz (2018), and Caplin, Dean, and Leahy (2017) have derived classes of cost functions based on axiomatic desiderata. However, for the most part, modelers continue to use a relatively informal basis for selecting cost functions of information acquisition, where the focus is typically on the intuition of the function and its tractability within a model.

But each functional form makes assumptions, both implicit and explicit, about the agent and their environment. What are those assumptions? And how do they qualitatively and quantitatively affect the predictions and conclusions of a model? At present, there does not appear to be a systematic method for determining what properties popular cost functions satisfy, and what they might imply about agents' behavior and beliefs.

This paper provides six axioms which a cost function of information acquisition must satisfy. The first assumption is that reording signals and states does not change the cost of an experiment; the second is that a linear combination of two experiments cannot be more expensive than either experiment separately; the third is that coarsening the set of signals makes experiments cheaper; the fourth is that if any signal become cheaper to observe, the cost of an experiment (i.e., a collection of signals[1])must become cheaper as well; the fifth is that the cost of any experiment can be expressed as the cost of the experiment on two mutually exclusive and exhaustive subsets of the state space, plus the cost of disentangling the two; the sixth is that the cost of running two independent experiments is the sum of the costs of running each experiments separately. The axioms are intuitively justified and allow for a unique solution, whose functional form is consistent with other axiomatic efforts in the literature. In a Gaussian setting, the cost function is linear in the ratio of the variance of prior beliefs to the variance of signals.

We find that for the most common cost functions used in the literature, two of the six axioms are violated frequently. The first violation is of the fourth axiom, where we impose that the cost of

1. For clarity, in this paper "signal" will be used to mean a single realization. An experiment is defined as a collection of signals and a series of conditional probabilities over the set of signals.

information depends on the prior of the agent. Violating this axiom allows modelers to assume that a signal's cost need not depend on the prior beliefs of the agent acquiring the signal. The second violation is of our final axiom, where it is assumed that the cost of acquiring two independent experiments is the sum of the cost of acquiring each experiment. This violation allows modelers to assume economies (or more commonly, diseconomies) of scale in acquiring many signals.

The lack of a function's dependence on the prior is the more pernicious of the two violations, as it has implications on the consistency of agents' beliefs. If we believe that the state of the world matters for the costliness of experiments, then a cost function that is independent of agents' prior beliefs actually makes an implicit assumption: namely, that agents' priors are a uniform distribution.[2] Because models that use such cost functions do not typically assume that agents priors are actually uninformative (that is, most models assume that rational expectations hold, and that agents' priors are consistent with the unconditional data-generating process), the cost function requires a degree of schizophrenia on the part of the agents - they must use one prior for their information acquisition, and then a different prior for their Bayesian updating and utility maximization. Correcting the cost function to contain agents' prior beliefs, or correcting the utility maximization to use an uninformative prior will impact the conclusions of the model: while standard cost functions typically predict that agents tend to pay attention to the most volatile object, correcting for their priors reverses that conclusion.

The violation of the independence axiom is usually implicitly assumed because modelers want a convex cost of information acquisition in signal precision, which allows for unique interior solutions in many settings. However, the assumption of a convex cost of signal precision assumes (not unreasonably) that improving the precision of a signal is more difficult the more precise it already is. Alternatively, the assumption is that running independent experiments successively becomes more and more costly. While the opposite assumption (that running independent experiments become successively easier) may also seem intuitively appealing, our independence assumption fulfills the weak form of both arguments. Violations of the independence assumption can be justified depending on the modeling context. Correcting the cost function to satisfy the independence assumption will

2. As an example, say a cost function can be written as the sum of a function of the state of the world, $C(E) = \sum_{\theta \in \Theta} c(\theta)$, where θ is the state of the world and $c(\cdot)$ is independent of agents' priors. Then, this amounts to assuming the agent has uniform prior, because $C(E)$ is equivalent to assuming the cost of information is $\frac{1}{|\Theta|} \sum_{\theta \in \Theta} c(\theta) = \sum_{\theta} p_\theta c(\theta)$, where p_θ is the uniform distribution (up to a multiplicative constant). We note that some cost functions, as that in Woodford (2012), do not depend on agents' priors and do not make assumptions on agents' priors. Also in that case, the cost of information of an agent with general prior is identical to that of an agent with uniform prior.

change the shape of some model predictions, and can lead to corner solutions as unique interior solutions are no longer guaranteed.

This study provides a framework through which to understand the consequences (both intended and unintended) of different cost functions, and to propose formal and intuitive remedies for potentially undesirable implicit assumptions.

1.1 Literature Review

This paper sits at the nexus of two literatures: the first, a growing literature that drills down into the properties of cost functions of information; the second a massive applied theory literature that takes off-the-shelf cost functions of information and embeds them in theoretical models. Both literatures are too large to summarize completely here, but we highlight some of the key recent contributions that are most closely related to our project.

In the first literature, Caplin, Dean, and Leahy (2017) examine how state-dependent random choice data can impact rational inattention models, and mutual information based cost functions. Chambers, Liu, and Rehbeck (2020) analyzes non-separable models of information acquisition cost functions. Mensch (2018) looks at a characterization of posterior-separable information acquisition cost functions. De Oliveira et al. (2017) and Ellis (2018) look at information acquisition from the perspective of decision theory. Csiszár (2008) and Ebanks, Sahoo, and Sander (1998) provide a survey of axiomatic foundations in information theory. Pomatto, Strack, and Tamuz (2018) provides a general class of cost functions with an axiomatic definition (we discuss this paper in more detail in section 3).

The second literature using off-the-shelf information cost functions for applied theory models stretches back as far as Hayek (1945) and Stigler (1961). In particular, most of the current models bear some resemblance to Grossman and Stiglitz (1980) and Verrecchia (1982) either in the use of noise traders from the former, or the cost functions from the latter. Some papers, such as Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016) use the microfoundation of entropy reduction used in Sims (2003) and Woodford (2012). Others such as Admati and Pfleiderer (1997) or Colombo, Femminis, and Pavan (2014) use less explicit formulations similar to Verrecchia (1982).

The remainder of the paper is as follows. In section 2 we lay out and discuss each of the six axioms, and derive the unique cost function that satisfies them. In section 3 we discuss the implications of the non-axiomatically justified cost functions, and possible remedies. Section 4 provides formalism for the results and the proof of our central theorem. Section 5 concludes.

2 Axioms

In this section we lay out the intuition behind each of the six axioms through the lens of an application. Suppose that a fundamental and unobservable variable (say, GDP next year) has a true, but unobservable value X. An economic agent can conduct an experiment (do some research), and try to acquire information about X. Conducting an experiment and improving their knowledge of X could be beneficial for the agent in making real and financial decisions. We want to understand how to characterize the cost of the experiment. We call the cost of an experiment E, $C(E)$.

2.1 Axiom 1: Anonymity

The first axiom states that reordering the signals and states does not change the cost of an experiment. This axiom constraints the cost of an experiment to depend *only* on the relationship between signals and states of the world. In the context of our example, an experiment that produces a signal about X in US Dollars for the value of GDP in US Dollars will be assumed to be as costly as the same experiment producing a signal in Euros for the value of GDP in Euros (assuming that the USD/EUR exchange rate is known).

2.2 Axiom 2: Quasi-Convexity

The second axiom states that a linear combination of two experiments cannot be more expensive than either experiment separately. In applied terms, suppose that the agent sources research about X from a research company. The research company might delegate the request to an experienced researcher, who costs C_1, or to a novice researcher who costs C_2. Without loss of generality, we can assume that $C_1 > C_2$. If the company has staffing issues, and can only tell the agent the probability that the report will be generated by the experienced researcher versus the novice researcher, the company could not charge a price P such that $P > C_1$.

2.3 Axiom 3: Garbling

The third axiom states that coarsening the signal distribution of an experiment makes it cheaper. In the context of our application, consider again an agent purchasing a signal from a research company about X. There are two types of reports: the first tells the agent whether GDP is "higher" or "lower" than the previous period in expectation; the second gives the agent an estimated value in dollars and cents. The second report must be at least as expensive as the first.

2.4 Axiom 4: Signal Separability

The fourth axiom states that if certain signals become less costly to observe, then the experiment as a whole becomes less costly. Consider an experiment E_1, with cost $C(E_1)$ and then consider an experiment E_2 which is the same as E_1 except that when E_1 would have returned a negative signal, E_2 returns the same negative signal with probability λ and returns a blank signal "" with probability $(1 - \lambda)$. Suppose that E_2 has a cost $C(E_2)$. This axiom states that if λ increases, then $C(E_2)$ increases. Signal separability therefore imposes that the cost of an experiment can be "separated" into the costs of individual signals. Suppose that, in our application, conducting an experiment involves writing a lot of code to process a lot of data. The signal could be multinomial, with each entry coming from a different dataset. Then suppose there is an improvement in one of the datasets such that generating the multinomial signal becomes easier. Then the experiment overall will become less costly to run.

In the mathematical formulation of this axiom (in section 4), we assume that the separability of the cost function in signals depends on the agent's prior over the ex-ante likelihood of each signal subset. By assuming dependence on the prior, we impose that the agent's prior will be a component of the information cost function.

2.5 Axiom 5: State Separability

The fifth axiom states that the cost of an experiment can be expressed as the cost of the experiment on a subset of states, plus the cost of the same experiment on the rest of the states, plus the cost of pairwise disentangling signals between the two subsets. In our example, the cost of an experiment E is equal to the sum of the costs of two separate experiments: the first, one that returns the same distributions of signals as E for all negative values of X (say, the states where the country is in a recession), and one that returns the same distributions of signals as E for all non-negative values of X (the states where the country is not in a recession), plus the cost of disentangling those two separate experiments.

2.6 Axiom 6: Independence

The final axiom states that the cost of running two independent experiments is the sum of the costs of running each experiment separately. In practical terms, flipping a coin twice, independently, should cost twice as much as flipping a coin once. We believe that this requirement is natural, but we discuss common objections in section 3, given that this axiom is frequently violated by off-the-shelf cost function employed in the literature.

2.7 Characterization

The axioms are satisfied by a unique cost of information function that can be written as:

$$C(E) = \int_{\theta,\theta'} D_S(f(\cdot|\theta) \| f(\cdot|\theta')) g(\theta) g(\theta') d\theta d\theta', \tag{1}$$

(see Theorem 1, in Section 4 for a formal mathematical characterization of the cost function and proof) where the variables can be interpreted as follows:

- θ and θ' represent possible states of the world.
- $f(\cdot|\theta)$ represents the distribution of signals conditional on GDP taking the value θ.
- $D_S(f(\cdot|\theta) \| f(\cdot|\theta'))$ represents the 'divergence' between two conditional distributions of signals. It can be interpreted as how different the distributions of signals under the two states of the world are. If an experiment induces distributions that are very different, it will allow to infer the state of the world with large precision and, as such, will be more costly.
- $g(\theta)$, represents the prior probability that the agent assigns to state θ: the more likely a certain state is, the more that state will weight in the calculation of the experiment's cost.

One way to interpret this function is as follows: suppose that there are infinitely many possible worlds, only one of which is the true one. The agent has prior beliefs g over how likely different values of the variable of concern are. If an agent believes that a particular value θ^* is very likely, they will believe that there is a large measure of possible worlds where X takes the value θ^* (that is, $g(\theta^*)$ is large). If an agent believes that a value $\tilde{\theta}$ is unlikely, they will believe that only a few worlds will have X taking the value $\tilde{\theta}$ ($g(\tilde{\theta})$ is small). The experiment involves a pairwise disentangling of worlds (hence the double integral) where disentangling any two possible worlds is equally costly. Discerning two very likely values of X will be costly, because it involves pairwise disentangling *many* possible worlds. Discerning two very unlikely values of X will be cheap, because it involves pairwise disentangling *few* possible worlds. The weights placed on the disentanglement is therefore the agent's prior beliefs.

In the particular case of Gaussian prior beliefs (with variance σ_X^2) and signal distributions (with variance σ_s^2), the cost simplifies to $C(E) = \frac{\sigma_X^2}{\sigma_s^2}$, which can be rewritten in terms of the precision $\tau_i = 1/\sigma_i^2$ as:

$$C(E) = \frac{\tau_s}{\tau_X}.$$

The cost of an experiment increases in the precision of the experiment, and decreases in the precision of the agent's prior beliefs over the fundamental.

2.8 Comparison to Other Microfounded Costs of Information

While, to our knowledge, the axioms laid out in this section represent a novel characterization of cost functions, the solution is similar to other microfounded functions from the literature.

Pomatto, Strack, and Tamuz (2018) characterize a family of information cost functions for discrete state spaces Θ that in our notation would be written as:

$$\sum_{\theta,\theta'} \beta_{\theta,\theta'} D_S(f(\cdot|\theta)\|f(\cdot|\theta')),$$

where $(\beta_{\theta,\theta'})_{\theta\in\Theta,\theta'\in\Theta}$ is a matrix of weakly positive parameters that can be interpreted as the difficulty of discerning between two states.

Therefore, while our information cost function was axiomatized with a different set of properties, it is easy to see that the result of our characterization represents a special case of the information cost functions characterized by Pomatto, Strack, and Tamuz (2018). Specifically, setting their β parameters to be $\beta_{\theta,\theta'} = g(\theta)g(\theta')$, we obtain our information cost function. Given the interpretation of $\beta_{\theta,\theta'}$, our cost function assumes that the "marginal" cost of discriminating between two states is proportional to the ex-ante likelihood of observing the two states. Without this restriction, the cost of an experiment could be very sensitive to the cost of discriminating very unlikely states. From the perspective of an agent who believes a state θ to be extremely unlikely, the cost of information characterized in Pomatto, Strack, and Tamuz (2018) may overweight that particular state.

On the other hand, our case limits the way in which the cost of discriminating two states is related to the a priori probability observing the two states. To illustrate how our version of $\beta_{\theta,\theta'} = g(\theta)g(\theta')$ can be interpreted, suppose that, given an experiment, two states are drawn independently and with probability $g(\cdot)$, and the agent uses the experiment to discern those two states. Our cost of information assumes that the cost of an experiment is proportional to the expected cost of discerning between all possible states.

In other words, the lack of structure on $\beta_{\theta,\theta'}$ in Pomatto, Strack, and Tamuz (2018), which is a direct consequence of their great level of generality, implies the cost of information is unrelated to the agent's prior. We take a different stand by assuming that the a priori belief on the state of the world affects the "perceived" cost of information. This restriction, in our opinion, reflects more closely the trade-off faced by an agent acquiring information. For example, a US trader acquiring information on inflation expectations and their relation with stock prices would not be willing to pay for an analyst to draft a report on trade prices in a 30% inflation rate scenario, if the trader believed that to be an unlikely outcome.

The great level of generality achieved by the characterization of Pomatto, Strack, and Tamuz (2018) also allows to draw comparisons between our cost of information and the family of information costs characterized by Caplin, Dean, and Leahy (2017). Pomatto, Strack, and Tamuz (2018) notes that their information cost functions can be written as posterior-separable information costs through the function F that we rewrite in our notation as:

$$F(g(\cdot|s)) = \sum_{\theta,\theta'} \beta_{\theta,\theta'} \frac{g(\theta|s)}{g(\theta)} \log\left(\frac{g(\theta|s)}{g(\theta'|s)}\right),$$

where $g(\cdot|s)$ represents the posterior associated with signal s. Using this transformation, Pomatto, Strack, and Tamuz (2018) find that the cost of information they derive can be written as:

$$E(F(g(\cdot|s)) - F(g(\cdot)),$$

i.e. the expected difference in F between the prior and the posterior, averaged overall posteriors (using the agent's prior). This reinterpretation provided by function F allows to reinterpret our cost of information as:

$$C(E) = \mathbb{E}\left(F_g(g(\cdot|s)) - F_g(g(\cdot))\right)$$

where F_g represents the adaptation of the general F to our $\beta_{\theta,\theta'}$, specifically:

$$F(g(\cdot|s)) = \sum_{\theta,\theta'} g(\theta)g(\theta')\frac{g(\theta|s)}{g(\theta)}\log\left(\frac{g(\theta|s)}{g(\theta'|s)}\right).$$

Therefore, our cost of information can also be thought of as a special case of the posterior separable information cost functions characterized by Caplin, Dean, and Leahy (2017). As noted in Caplin, Dean, and Leahy (2017), this separability allows for an analytically tractable formulation of the agent's decision to acquire information.

3 Discussion

In this section, we discuss the two properties of our cost function that result in qualitative differences in an agent's incentives to acquire information: dependence on the agent's prior and the independence of multiple experiments.

3.1 Prior-Dependent Cost of Information

As discussed in the previous section, our cost function depends on the agent's prior beliefs. It is helpful to break down the dependence of information cost on an agent's prior into two 'sub-properties' that we discuss in the rest of this section: first, that the cost of acquiring information depends on the state of the world, and second, that given this dependence on the state of the world, each state of the world is weighted according to the agent's prior on the state of the world.

That the cost of information may depend on the state of the world is per se uncontroversial, but consider a simple example: suppose the state of the world is the distance of the Earth from Mars, and the experiment is a photograph of Mars taken with a telescope. The cost of the experiment is likely to depend on the distance of Mars. The same level of precision in the measurement of Mars' distance will be more difficult to achieve if Mars is far (as, for example, one needs more advanced technology, or pictures with higher definition).

The second property has to do with how each state of the world *should* be weighted, when measuring the expected cost of an experiment. In most economic applications it is assumed that the agent's prior enters the calculation of *expected returns* of the decisions she is making. In this sense, it is natural to have the agent's prior enter both the calculation of returns and costs, which in our model, are acquiring information. Both the returns and the costs associated with economic and financial decision making are intrinsically uncertain and, as such, require factoring in the likelihood of different states of the world. Our axiomatic theory suggests that the same prior should be used to weight different states of the world, for both costs and returns.

Essentially, unless one considers only applications in which the cost of the experiment is unrelated to the state of the world, one must assume that an agent conducting an experiment has *a prior* about the state of the world. That prior could be uniform over all states (an agnostic agent), or with a different mean than the 'correct' belief (a biased agent) or with a different variance than the 'correct' belief (an (un)confident agent). But in order for a cost function to be defined, it must be defined relative to *some* distribution. What the cost function we derive here assumes is that the agent's prior is the same as the specified unconditional distribution of the fundamental. What cost functions that neglect the prior variance of the fundamental are doing, is implicitly assuming that the agent is agnostic - and that their prior is uniform.

Although the functional form of the cost function can be justified with different priors (while remaining consistent with our axioms), a problem arises once the information has been collected, and is used to update the agent's beliefs. In our setting, an agent would use the axiomatic cost function, and then update their beliefs using Bayes rule. In settings where the cost of information does not depend on the prior, an agent should update their beliefs using Bayes rule, where the prior is the uniform distribution. Using a cost function that does not depend on the prior and then updating beliefs using Bayes rule with the prior *would* be a violation of our axiom 4.

Suppose we 'correct' the cost functions that ignore the prior by setting their prior beliefs about the fundamental to be the uniform distribution. Such an assumption will allow the modeler to satisfy axiom 4. In this setting, agents' posterior distributions over the fundamental will simply be the distribution of their signals. Therefore, the prior will not show up in the agents' objective functions, as they update their beliefs using their signal and the uniform distribution. As a result, information collection will not vary with the prior at all. Information collection will be independent of the actual variance of the fundamental, which seems undesirable.

To rephrase the two properties, assuming that information cost is *not* dependent on the agent's prior, as done in many previous works, implies the modeler is comfortable with either of the following assumptions:

1. the cost of an experiment is unrelated to the state of the world, i.e. for all states of the world, it is equally costly to achieve a certain precision in the posterior; **or**

2. if the cost of an experiment is related to the state of the world, the maximization problem of the agent uses two, distinct, sets of beliefs: one to calculate expected returns, and a different one to calculate expected costs.[3]

It is important to note that while our cost function does not rely on a Gaussian setting, it is traditional in the literature to consider a setting where the variable of interest is distributed normally (say with precision τ) and the informativeness of an experiment is measured by the precision of the resultant signal, which is also normally distributed (say with precision τ_s). Many papers in these settings assume that the cost of an experiment should depend solely on the precision of the acquired signal, and not on the prior. But this assumption means that the impact of prior beliefs on information collection will be quite different. Under the axiomatic cost function, when agents believe that the fundamental is noisier, maintaining the same level of signal precision becomes more costly, and therefore agents scale down their information acquisition. Conversely, under other cost functions which depend solely on the precision of the signal, agents usually respond to noisier prior beliefs by increasing the precision of their signal, as the change does not affect the cost of doing so, while it increases the benefit of acquiring a more precise signal.

3.2 Independence and Convexity

In a Gaussian setting, assuming strict convexity in the experiment's precision (a common assumption) of the cost of information function plays a crucial role in ensuring that agents always acquire some information (regardless of other parameters of the model). The cost function derived axiomatically in this paper does not satisfy strict convexity, being linear in signal precision. This

3. As mentioned in the introduction, some papers assume that an agent with uniform prior belief has the same cost of information as an agent with a different belief (e.g., Woodford (2012)).

assumption deviates from what is generally accepted in much of the information literature. Therefore, in this section, we discuss the reasons that our cost of information is not strictly convex and how assuming strict convexity is at odds with another popular interpretation of the information acquisition process.

3.2.1 Independence implies the cost of information is not strictly convex

The independence axiom stated above imposes that the cost of running two independent experiments is equal to the sum of the cost of the two experiments. In the context of normally distributed priors and signals, the precision of the two independent experiments is the sum of the precision of the individual experiments.[4] Under the independence axiom, then, a cost of information function expressed in terms of the precision q_1, q_2 of two experiments must satisfy the following equality:

$$\kappa(q_1 + q_2) = \kappa(q_1) + \kappa(q_2). \tag{2}$$

As this condition must hold for any q_1, q_2, under natural smoothness assumptions, the only $\kappa(\cdot)$ satisfying (2) is linear, and therefore, not strictly convex.

3.2.2 Alternative interpretations of information acquisition

A reason for assuming that the cost of information is convex in precision is a parallel between precision and effort. Improving the precision of a very precise experiment could be more costly than improving the precision of an imprecise one. As such, it is natural to assume the cost of information to be convex, which in turn result in the following inequality:

$$\kappa(q_1 + q_2) \geq \kappa(q_1) + \kappa(q_2),$$

with strict inequality for strict convexity.[5]

4. Note that the independence axiom was stated, more generally, for any experiment, for which the concept of "precision" is not unambiguous. For simplicity, in this section, we restrict ourselves to the setting of normal priors and signals, with signal precision being state-independent. In this setting the precision of an experiment is naturally defined as the inverse of the variance of the signal distribution.

5. Notice that this assumption has the additional advantage of an increasing marginal cost of information, which paired with a decreasing marginal value of information results in a unique solution for information acquisition problems. This is of course convenient for comparative statics exercises that require unique equilibria.

Acquiring information, besides being analogous with exerting effort, is also related to learning (so much so that "learning" and "information acquisition" are often used interchangeably). Implicit in the process of learning is the idea that learning becomes "easier," as one's knowledge expands. Under this interpretation, then, there should be decreasing marginal costs to acquiring information which implies that

$$\kappa(q_1 + q_2) \leq \kappa(q_1) + \kappa(q_2),$$

with strict inequality for strict concavity.

It is useful to observe that in both interpretations discussed above, the mathematical inequality assumes the experiments to be independent, while the layman interpretation we provide likely assumes some interdependence between the experiments. In any practical application, it is nigh impossible to obtain exactly independent experiments. As such, the mathematical inequality should be thought of as a theoretical abstraction of the idea of information acquisition as effort or learning.

3.2.3 Independence and interpretation of precision

The two interpretations of information acquisition can be rephrased into commonplace beliefs about the nature of scientific inquiry. Graduate students often find that coming up with original ideas is increasingly difficult. This analogy corresponds to the first interpretation of information acquisition: the more precise the experiment (the larger the academic literature) the more costly it is to improve its precision (the more costly it is to come up with original ideas). Another commonplace belief is that standing "on the shoulder of giants," makes it easier to acquire information. This interpretation translates to a concave information cost.

Both interpretations, besides being inconsistent with each other in their strict form, are also inconsistent with the independence axiom in their strict form, as is seen from the inequalities above. A cost of information function satisfying the independence axiom is *the only* cost of information function that satisfies both inequalities in their weak form.

Alternatively, we can reconcile the two interpretations of information acquisition with the independence axiom by re-stating the "effort" and "learning" interpretation as an assumption on the conditional cost of an experiment. Without providing exact definitions, this solution to the inconsistency above can be formalized as follows. Assume that the cost of running two (potentially correlated) experiments E_1, E_2 can be decomposed as follows:[6]

$$c(E_1, E_2) = c(E_1) + c(E_2|E_1),$$

where the latter expression (not defined formally) stands for the conditional cost of running E_2 after E_1. Now, if the precision of E_1 is q_1, and the precision of (E_1, E_2) is $q_1 + q_2$, then we can rewrite the inequality as:

$$\kappa(q_1 + q_2) = \kappa(q_1) + c(E_2|E_1). \tag{3}$$

Assuming that $c(E_2) = \kappa(q_2)$, observe how this final expression allows to reconcile the three options laid out above:

- *(independence)* Under independence, $c(E_2|E_1) = c(E_2) = \kappa(q_2)$, so (3) yields

$$\kappa(q_1 + q_2) = \kappa(q_1) + \kappa(q_2);$$

- *(information as effort)* If information is thought of as effort, then the cost of E_2 given E_1 is larger than the unconditional cost of E_2, so $c(E_2|E_1) \geq c(E_2) = \kappa(q_2)$. Then, (3) becomes:

$$\kappa(q_1 + q_2) \geq \kappa(q_1) + \kappa(q_2);$$

- *(information as learning)* If information is thought of as learning, then the cost of E_2 given E_1 is smaller than the unconditinoal cost of E_2, so $c(E_2|E_1) \leq c(E_2) = \kappa(q_2)$. Then, (3) becomes:

$$\kappa(q_1 + q_2) \leq \kappa(q_1) + \kappa(q_2);$$

This final interpretation shows informally that the the two interpretations of information acquisition (effort and learning) are compatible with the independence axiom, once the researcher formalizes

6. Note that we denote by $c(E_2|E_1)$ the cost of running E_2 after E_1 and such informal notation makes no assumption on the correlation between experiments (or the lack thereof).

the conditional cost of the second experiment, given the first. The inconsistency we described at the beginning of this section can be reframed as an inaccurate description of the correlation between experiments.

4 Statement of Axioms

Let Θ be the set of states of the world, and $\mathcal{F}$ a $\sigma-$algebra of events. Let $(\Theta, \mathcal{F})$ be a measurable space. A *belief* is a probability measure on $(\Theta, \mathcal{F})$ and we denote by $\mathbb{P}$ the probability measure on $(\Theta, \mathcal{F})$. We assume that the density function of $\mathbb{P}$ exists and is represented by $g(\theta)$.

An experiment E is a distribution on $(\Theta \times S, \mathcal{F} \times \mathcal{S})$, where $(S, \mathcal{S})$ is a measurable space of signals. For a fixed belief $\mathbb{P}$, the experiment can also be defined as a set of conditional probability distributions $f(s|A)$, where $s \in S$ and $A \in \mathcal{F}$, and it will in general be convenient to refer to a generic experiment E as a signal space S and a set of conditional distirbutions $f(s|A)$, for all $A \in \mathcal{F}$. Denote with Π the set of all experiments.

For each experiment E, we want to define the cost of the experiment $C_{\mathbb{P},S}(E) : \Pi \to [0, \infty]$ and we write briefly $C(E)$ whenever unambiguous.

We assume that the cost of information is positive and possibly infinite, $0 \leq C_{\mathbb{P},S}(E) \leq +\infty$. Further, we assume that the cost of information is 0 if and only if the experiment is uninformative, namely if and only if for all $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0 < \mathbb{P}(B)$, $f(s|A) = f(s|B)$. We also impose that the cost of information is infinite only if it allows to reveal a state of the world, in formulas: $C_{\mathbb{P},S}(E) < \infty$ if for all $A, B \in \mathcal{F}$ with $\mathbb{P}(A) > 0 < \mathbb{P}(B)$, $f(s|A)$ and $f(s|B)$ are mutually absolutely continuous.[7]

Define first when two experiments are isomorphic for a given belief $\mathbb{P}$. Let $E = (S, f(s|A))_{A \subset \Theta}$ and $E' = (S', g(s'|A'))_{A' \subset \Theta'}$ where A and A' are measurable. We say that E and E' are $\mathbb{P}$-isomorphic if there exist two bijections $h : S \to S'$ and $i : \Theta \to \Theta'$ such that:

- $f(s|A) = g(s'|A')$ for all $s' = h(s)$ and $A' = i(A)$;
- $\mathbb{P}(A) = \mathbb{P}'(A')$ for all $A' = i(A)$;

Axiom 1 (Anonymity)**.** *If E and E' are $\mathbb{P}-$isomorphic then $C_{\mathbb{P},S}(E) = C_{\mathbb{P}',S'}(E')$.*

7. Two distributions are mutually absolutely continuous if $f(S|A) = 0 \Leftrightarrow f(S|B) = 0$, for all $S \in \mathcal{S}$.

The first axiom requires that S be considered only a "label" without relevant meaning for the cost of the experiment, and analogously for Θ (once accounting for the ex-ante probability of it $\mathbb{P}(\theta)$).

Given two experiments $E = (S, f(s|A))$ and $E' = (S, g(s|A))$, define $\lambda E + (1-\lambda)E' := (s, \lambda f(s|A) + (1-\lambda)g(s|A))$, for any $\lambda \in [0,1]$. Having fixed the signal space S, this experiment is equivalent to observing E with probability λ, and E' with probability $1-\lambda$. The second axiom requires that the linear combination of two experiments be cheaper than the more expensive of the two original experiments.

Axiom 2 (Quasi-Convexity)**.** *For any E, E':*

$$C_{\mathbb{P},S}(\lambda E + (1-\lambda)E') \leq \max\{C_{\mathbb{P},S}(E), C_{\mathbb{P},S}(E')\} \qquad \forall \lambda \in [0,1].$$

Given an experiment $E = (S, f(s|A))$ and a function $h : S \to S'$ define $E \circ h$ by:

$$E \circ h := (S', f(h^{\leftarrow}(s')|A)).$$

In words, the function h coarsens the experiment because signals s_1 and s_2 that are distinguishable under the original experiment E, are not distinguishable if $h(s_1) = h(s_2) = s'$. The third axiom requires that coarsening the signal space cheapens the experiment.

Axiom 3 (Coarsened Experiment)**.** $C(E \circ h) \leq C(E)$ *for any h.*

For any $K \in \mathcal{S}$ subset of signals, define $E(\cdot|K)$ to be the experiment conditional on $s \in K$, formally $E(\cdot|K) = (K, f(s|A,K))_{A \in \mathcal{F}}$. The fourth axiom requires that decreasing the cost of observing a particular signal reduces the cost of the experiment.

Axiom 4 (Signal Separability)**.** *For any K, if $C(E(\cdot|K)) \leq C(E'(\cdot|K))$, then:*

$$C(\mathbb{P}(K)E(\cdot|K) + \mathbb{P}(K^c)E''(\cdot|K^c)) \leq C(\mathbb{P}(K)E'(\cdot|K) + \mathbb{P}(K^c)E''(\cdot|K^c)),$$

for all E''.

For all $B \in \mathcal{F}$ event, let E^B be defined as $(S, f(s|A))_{A \subset B}$. This can be interpreted as the experiment when the state is in B (note that differently from signal separability, $B \subseteq \Theta$). Also,

E^B is equivalent to considering E with the belief $\mathbb{P}(\cdot|B)$ on Θ. The fifth axiom states that the cost of an experiment can be decomposed into the cost of the same experiment on a subset of Θ, the cost of the same experiment on the complement of that subset, and the cost of disentangling the two sub-experiments.

Axiom 5 (State Separability)**.** *For all $B \in F$, we assume that:*

$$C(E) = \mathbb{P}(B)C(E^B) + \mathbb{P}(B^c)C(E^{B^c}) + \int_{K\subset B, K'\subset B^c} C(E^{K\cup K'})\mathbb{P}(K \cup K').$$

For any $E = (S, f(s|A))$ and $E' = (S', g(s'|A))$ experiments, we define:

$$E \otimes E' = (S \otimes S', h((s, s')|A)), \quad h((s, s')|A) := f(s|A)g(s'|A).$$

The sixth axiom requries that the cost of observing independent experiments is the sum of the costs of observing each of them.

Axiom 6 (Independence)**.** *For any E and E' experiments, we assume that:*

$$C(E \otimes E') = C(E) + C(E').$$

Axioms 1 to 6 permit the following characterization.

Theorem 1 (Characterization)**.** *Let $g(\theta)$ be the distribution of the belief $\mathbb{P}$, the only cost function that satisfies Axioms 1-6 is:*

$$\begin{aligned} C(E) = \int_{\theta,\theta'\in\Theta} \left(\int_{s\in S} \log(f(s|\theta)) - \log(f(s|\theta'))(f(s|\theta) - f(s|\theta'))ds \right) g(\theta)g(\theta')d\theta d\theta' \\ = \int_{\theta,\theta'} \left(D_{KL}(f(\cdot|\theta)\|f(\cdot|\theta') + D_{KL}(f(\cdot|\theta')\|f(\cdot|\theta)) \right) g(\theta)g(\theta')d\theta d\theta' \\ = \int_{\theta,\theta'} D_S(f(\cdot|\theta), f(\cdot|\theta'))g(\theta)g(\theta')d\theta d\theta', \end{aligned} \quad (4)$$

where:

- $D_{KL}(h\|i) = \int_{s\in S} i(s) \log\left(\frac{i(s)}{h(s)}\right) ds$ *is the Kullback-Leibler divergence between h and i;*
- $D_S(h, i) = D_{KL}(h\|i) + D_{KL}(i\|h)$ *is the symmetric divergence (or "Jeffreys Divergence").*

Proof of Theorem 1. The proof follows from reinterpreting the mathematical result in Zanardo (2021) related to defining measures of disagreement that satisfy a set of desiderata provided in that paper. The rest of this section is divided in two parts. In the first subsection, we show how the axioms of this paper imply the desiderata of Zanardo (2021), and therefore allow to employ Theorem 1 in Zanardo (2021). After having employed that theorem, one additional step allows to conclude the proof of Theorem 1 of this paper.

Consider the cost of information function axiomatized in this paper, for the particular case of two states of the world θ_1 and θ_2, and with finite signals $s_1, \ldots, s_n$. Any experiment can be described by two probability distributions: $f(\cdot|\theta_1) = (f(s_1|\theta_1), \ldots, f(s_n|\theta_1))$ and $f(\cdot|\theta_2) = (f(s_1|\theta_2), \ldots, f(s_n|\theta_2))$. Re-interpreting these two probability distributions as the "beliefs" of two agents in Zanardo (2021), it is straightforward to show that if C satisfies the axioms of this paper, then C corresponds to a disagreement measure between $f(\cdot|\theta_1)$ and $f(\cdot|\theta_2)$, and given Axiom 1 (isomorphic experiment have identical costs), the disagreement measure has to be symmetric. In formulas, re-interpeting $f(\cdot|\theta_1)$ and $f(\cdot|\theta_2)$ as beliefs, we have that:

$$C(E) = D(f(\cdot|\theta_1), f(\cdot|\theta_2)),$$

where D is a symmetric disagreement function as defined in Zanardo (2021). The interpretation of this analogy between cost functions and disagreement functions is natural. The more an experiment is informative, the more the conditional distribution on signal realizations is different under different states of the world, i.e., the more $f(\cdot|\theta_1)$ and $f(\cdot|\theta_2)$ differ, or "disagree."

Having established this parallel, observe that there are only two symmetric disagreement functions in Zanardo (2021):

- $C(E) = D_1(f(\cdot|\theta_1), f(\cdot|\theta_2)) = \log\left(\sum_s \sqrt{f(s|\theta_1), f(s|\theta_2))}\right)$, or
- $C(E) = D_2(f(\cdot|\theta_1), f(\cdot|\theta_2)) = \sum_s (\log(f(s|\theta_1) - \log(f(s|\theta_2))(f(s|\theta_1) - f(s|\theta_2))$,

where D_1 corresponds to the Bhattacharya divergence, and D_2 corresponds to the symmetric Kullback-Leibler divergence. Extending the state space to include 3 states, θ_1, θ_2, and θ_3, one can apply Axiom 6 (state separability) using sets $B = \theta_1$ and $B^c = \{\theta_2, \theta_3\}$. Axiom 6 implies that:

$$C(E) = \sum_{\theta,\theta'} D(f(\cdot|\theta), f(\cdot|\theta'))g(\theta)g(\theta'), \tag{5}$$

where D can be either D_1 or D_2. One can verify that plugging D_1 into (5), the final cost function fails to be separable in signals, i.e. it doesn't satisfy Axiom 5. On the other hand, plugging in D_2 we obtain a cost function that satisfies all the axioms. So the only cost function satisfying all the axioms on a state space with three states is $C(E) = \sum_{\theta,\theta'} D_2(f(\cdot|\theta), f(\cdot|\theta'))g(\theta)g(\theta')$. Extending to general states of the world and signal spaces, one obtains that only the cost function provided in Theorem 1 satisfies all axioms. □

5 Conclusion

In this paper we exposit and justify six axioms that a cost of information acquisition should satisfy. We then use these axioms as a lens through which to analyze canonical cost functions used in the endogenous information acquisition literature. While most axioms are satisfied by the cost functions we consider, we find that our fourth and sixth axioms are often violated. We show that the six axioms yield a unique solution, which is consistent with a class of functions discussed in the literature. This finding serves to validate the assumptions as reasonable. In a Gaussian setting, the solution has the intuitive formulation of being linear in the ratio of the variance of the prior to the variance of the signal.

In a Gaussian setting, the violations of our fourth and sixth axioms arise because many cost functions are independent of the information collector's prior beliefs, and further are convex in the precision of the acquired signal. We argue that violations of our independence axiom can be justified based on the real-world interpretation of the model design. However, it is harder to intuitively justify cost functions that don't depend on the agents' priors.

References

Admati, Anat R, and Paul Pfleiderer. 1997. "Does it all add up? Benchmarks and the compensation of active portfolio managers." *The Journal of Business* 70 (3): 323–350.

Breugem, Matthijs, and Adrian Buss. 2019. "Institutional investors and information acquisition: Implications for asset prices and informational efficiency." *The Review of Financial Studies* 32 (6): 2260–2301.

Caplin, Andrew, Mark Dean, and John Leahy. 2017. *Rationally inattentive behavior: Characterizing and generalizing Shannon entropy.* Technical report. National Bureau of Economic Research.

Chambers, Christopher P, Ce Liu, and John Rehbeck. 2020. "Costly information acquisition." *Journal of Economic Theory* 186:104979.

Colombo, Luca, Gianluca Femminis, and Alessandro Pavan. 2014. "Information acquisition and welfare." *The Review of Economic Studies* 81 (4): 1438–1483.

Csiszár, Imre. 2008. "Axiomatic characterizations of information measures." *Entropy* 10 (3): 261–273.

De Oliveira, Henrique, Tommaso Denti, Maximilian Mihm, and Kemal Ozbek. 2017. "Rationally inattentive preferences and hidden information costs." *Theoretical Economics* 12 (2): 621–654.

Ebanks, Bruce, Prasanna K Sahoo, and Wolfgang Sander. 1998. *Characterization of information measures.* World Scientific.

Ellis, Andrew. 2018. "Foundations for optimal inattention." *Journal of Economic Theory* 173:56–94.

Grossman, Sanford J, and Joseph E Stiglitz. 1980. "On the impossibility of informationally efficient markets." *The American economic review* 70 (3): 393–408.

Hayek, Friedrich August. 1945. "The use of knowledge in society." *The American economic review* 35 (4): 519–530.

Hébert, Benjamin, and Michael Woodford. 2017. *Rational inattention and sequential information sampling.* Technical report. National Bureau of Economic Research.

Kacperczyk, Marcin, Stijn Van Nieuwerburgh, and Laura Veldkamp. 2016. "A rational theory of mutual funds' attention allocation." *Econometrica* 84 (2): 571–626.

Marschak, Jacob. 1954. *Towards an economic theory of organization and information.* Cowles Commission for Research in Economics.

Mensch, Jeffrey. 2018. "Cardinal representations of information." *Available at SSRN 3148954.*

Nelson, Richard R, and Sidney G Winter Jr. 1964. “A case study in the economics of information and coordination: The weather forecasting system.” *The Quarterly Journal of Economics,* 420–441.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz. 2018. “The cost of information.” *arXiv preprint arXiv:1812.04211.*

Shannon, Claude Elwood. 1948. “A mathematical theory of communication.” *Bell system technical journal* 27 (3): 379–423.

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

Stigler, George J. 1961. “The economics of information.” *Journal of political economy* 69 (3): 213–225.

Verrecchia, Robert E. 1982. “Information acquisition in a noisy rational expectations economy.” *Econometrica: Journal of the Econometric Society,* 1415–1430.

Wilde, Louis L. 1980. “The economics of consumer information acquisition.” *Journal of Business,* S143–S158.

Woodford, Michael. 2012. “Inattentive valuation and reference-dependent choice.”

Zanardo, Enrico. 2017. “Essays in Microeconomics.” PhD diss., Columbia University.