The Appeal of Information Transactions

Antonio Cabrales, Olivier Gossner and Roberto Serrano[*]

September 2012

Abstract: An information transaction entails the purchase of information. Formally, it consists of an information structure together with a price. We develop an index of the appeal of information transactions, which is derived as a dual to the agent's preferences for information. The index of information transactions has a simple analytic characterization in terms of the relative entropy from priors to posteriors, and it also connects naturally with a recent index of riskiness.

JEL classification numbers: C00, C43, D00, D80, D81, G00, G11.

Keywords: informativeness, information transactions, Kullback-Leibler divergence, relative entropy, decision under uncertainty, investment, Blackwell ordering.

1 Introduction

Economic agents are constantly receiving new information, yet not all new information is equally valuable. Hence, we often want to measure the informational content of different signals, and both economists and information theorists have devoted much effort to this measurement problem.[1] In addition, new information is often not obtained for free: one has to pay for it. What sets this paper aside from previous contributions is the consideration of this information/price tradeoff, in what we call *information transactions*. An information transaction is defined as a pair consisting of an information structure and a price. If we were sick and wanted a diagnosis, we could hire a famous specialist who also charges high fees for her diagnosis, or we could settle for a cheaper expert who probably would provide a less accurate diagnosis. We aim to find a way to assess "objectively" the relative worth of each of those information transactions.

These questions are heavily inspired by Blackwell (1953), who introduces an ranking of information structures in terms of their informativeness, and shows that an information structure α is more informative than another information structure β if every decision maker prefers α to β in every decision problem. Blackwell's approach has a limited scope for applications, because it does not provide a complete ordering. Researchers have made progress by

---

[1]Veldkamp (2011) describes many of the ways in which economists have measured informativeness and its applications.

focusing on decision makers who have preferences in a particular class.[2] This paper takes a different route that allows us to rank information transactions completely: we develop an index of the appeal of information transactions that is dual to the agent's preferences for information, as we now briefly describe.

Specifically, we say that an agent u_1 *likes information better* than another agent u_2 if for every pair of wealth levels w_1, w_2, and an information transaction a, whenever agent u_2 accepts an information transaction a at wealth w_2, then agent u_1 does so at wealth w_1. This leads to our key definition, which is the requirement that the appeal index satisfy *duality* with respect to preferences for information: when the index deems transaction a more appealing than b, then if the agent who likes information less accepts transaction b, the agent who likes information more must accept transaction a.

In constructing the index, we discovered one important result - which is of interest in its own right - regarding the connection between a decision maker's risk attitudes and the value of information for him. Indeed, we show that an agent u_1 likes information more than agent u_2 if and only if the maximum coefficient of risk aversion of u_1 is lower than the minimum coefficient of risk aversion of u_2. We then use this result to show that the appeal of an information transaction a can be characterized as the risk aversion coeffi-

---

[2]Lehmann (1988), for instance, restricts the analysis to problems that generate monotone decision rules (and hence satisfy single-crossing conditions). Persico (2000), Athey and Levin (2001), and Jewitt (2007) extend Lehmann's analysis to more general classes of monotone problems.

cient of the CARA (constant absolute risk aversion) agent who is indifferent between accepting and rejecting a.

The duality principle described above can only provide an ordinal characterization of the concept of appeal of transactions. It turns out that the appeal index also satisfies the additional property of *price homogeneity of degree -1*. Price homogeneity conveys a sort of separability between information and price in assessing the appeal of an information transaction: when two transactions are ranked in a certain order, multiplying the price of a transaction by a constant divides its appeal by this constant. This implies in particular that multiplying the price of two transactions by the same constant does not alter their ranking. The homogeneity degree is negative because we seek indices that are decreasing with the price μ. Next, we offer an axiomatic characterization of the appeal of information transactions. *Duality* and *homogeneity* define the index uniquely, up to a multiplicative normalization constant.

The results described in the previous paragraphs lead to a simple characterization of the index, given that CARA agents have investment strategies which are easy to express analytically. In words, the index depends negatively on the price of the transaction and positively on the relative entropy from the prior to the posteriors generated by the signals of the information structure. The presence of relative entropy in the index is interesting because this relative entropy, which is also called Kullback-Leibler divergence (after

Kullback and Leibler, 1951), is often used in other sciences as a measure of information gain.

The approach in this paper is very much inspired by that of Aumann and Serrano (2008) for ordering riskiness. The parallels between the indices for informativeness and riskiness are actually rather striking; for example, both indices attribute central importance to the CARA agents, and their respective axiomatic characterizations use identical axioms. This suggests that duality is a powerful principle for ordering multidimensional objects, far beyond the riskiness of one-dimensional random variables.

In Cabrales, Gossner, and Serrano (2012), we have provided another information index. In that paper, the informativeness of an information structure is characterized by the reduction of entropy from the prior. With a uniform prior, and for small amounts of information, that index is close to the index proposed here when the transaction price is kept constant, but they differ significantly when the amount of information in the signals is larger. The difference stems from the assumptions characterizing the index in Cabrales, Gossner, and Serrano (2012): in that paper, all agents have ruin-averse utility functions (i.e., zero wealth leads to negative infinite utility), while in this paper ruin (i.e., negative wealth) is allowed for sufficiently high transaction prices or for certain investment strategies.

Here is how the paper proceeds. Section 2 describes the model. Section 3 defines an ordering of information transactions based on agents' preferences

for information. Section 4 presents our main result, connects the notions of risk aversion and value of information, and relates the notion of appeal to CARA agents. Section 5 provides other characterizations of the index: both a cardinal axiomatization and a perspective based on the total rejection of information transactions. Section 6 presents a number of properties of the index and provides some examples. Relevant literature is discussed in Section 7, and Section 8 concludes. Section 9 contains the proofs.

2 The Model

2.1 The Agent

We consider an investor with initial wealth w and a monetary utility function u defined on $\mathbb{R}$. We assume that u is nondecreasing, concave and twice differentiable. We let $\mathcal{U}$ be the set of such monetary utility functions. We identify agents by their monetary utility functions, thus speaking of agent u to refer to an agent with utility function u.

2.2 Investments

There is a set K of states of nature, about which the agent is uncertain. The agent's prior on K is $p \in \Delta(K)$, assumed to have full support. The set of investment opportunities consists of all no-arbitrage assets given p, that is, assets with a nonpositive expected return: $B^* = \{x \in \mathbb{R}^K, \sum_k p(k) b_k \leq 0\}$.[3]

---

[3]The vector $p = (p_k)_k$ in the definition of no-arbitrage assets corresponds to the price vector of Arrow-Debreu securities, where p_k can be interpreted as the price of an asset

When the agent's initial wealth is w, b is chosen and state k is realized, the agent's final wealth is $w + b_k$.

2.3 Information Transactions

Before choosing an investment, the agent has the opportunity to engage in a (costly) information transaction $a = (\mu, \alpha)$. Here, $\mu > 0$ represents the cost of the information transaction, and α is the information structure representing the information obtained from a. That is, α is given by a finite set of signals S_α, together with probabilities $\alpha_k \in \Delta(S_\alpha)$ for every k. When the state of nature is k, then $\alpha_k(s)$ is the probability that the signal observed by the agent is s. It is standard practice to represent any such information structure by a stochastic matrix, with as many rows as states and as many columns as signals; in the matrix, row k is the probability distribution $(\alpha_k(s))_{s \in S_\alpha}$. Signal s has a total probability $p_\alpha(s) = \sum_k p(k)\alpha_k(s)$, and we assume, without loss of generality, that $p_\alpha(s) > 0$ for every s. For each signal $s \in S_\alpha$, we let $q_k^s \in \Delta(K)$ be the probability of state k conditional on s computed using Bayes's rule.

We say that a is *excluding* if for every signal s, there exists k such that $q_k^s = 0$. It is *nonexcluding* otherwise. Excluding information transactions are such that, for every received signal, there exists a state of nature that the agent can exclude.

---

that pays 1 in state k and 0 in all other states. The fact that this vector coincides with the agent's prior p means that no-arbitrage assets cannot yield a positive expected return. We disentangle the two roles of p, price and priors, in Section 6.2.

2.4 Optimal Investment after Receiving Information

Given a belief q, an agent with wealth w and utility u chooses $b \in B^*$ in order to maximize his expected utility over all states $k \in K$. The maximum expected utility is then $V(u, w, q)$, given by:

$$V(u, w, q) = \sup_{b \in B^*} \sum_k q_k u(w + b_k).$$

2.5 Acceptance of Information Transactions

The agent with utility function u and wealth w *accepts* an information transaction $a = (\mu, \alpha)$ if and only if paying μ upfront to receive information according to α generates an expected utility greater than or equal to staying with wealth w. This is the case if and only if:

$$\sum_s p_\alpha(s) V(u, w - \mu, q_k^s) \geq u(w).$$

3 More Appealing Information Transactions

This section proposes a way to define the "objective" appeal of information transactions. The approach is based on ordering preferences for information.

3.1 Ordering Preferences for Information

Definition 1 *Let $u_1, u_2 \in \mathcal{U}$ represent two agents. We say that agent u_1 uniformly likes (or likes, for short) information better than agent u_2 if, for*

every pair of wealth levels w_1, w_2, and information transaction a, whenever agent u_2 accepts a at wealth w_2, then agent u_1 does so at wealth w_1.

The definition means that independent of their respective wealth levels, agent u_1 is always more prone to accepting information transactions than is agent u_2. This will be the case when there is something intrinsic in agent u_1's preferences that always makes him at least as interested as agent u_2 in purchasing information.

3.2   Ordering information transactions

We move now to define the comparative appeal of two information transactions.

Definition 2 *Let a_1, a_2 be two information transactions. We say that a_1 is more appealing than a_2 if, given two agents u_1, u_2 such that u_1 uniformly likes information better than u_2 and given two wealth levels w_1, w_2, whenever agent u_2 accepts a_2 at wealth level w_2, then agent u_1 accepts a_1 at wealth level w_1.*

Because agent u_1 likes information better than agent u_2, it is clear that, as soon as we know that agent u_2 accepts transaction a_2, so must agent u_1; furthermore, this is true independent of the wealth levels of u_1 and u_2. If appeal is a well-defined objective concept and a_1 is more appealing than a_2, agent u_1 should accept a_1 a fortiori.

4 The Main Result

For two probability distributions p and q, the *relative entropy* from p to q, also called their *Kulback-Leibler divergence*, has been proposed as a nonsymmetric measure of their discrepancy. It is defined as follows:

$$d(p\|q) = \sum_k p_k \ln \frac{p_k}{q_k}.$$

It is always nonnegative, and equals zero if and only if $p = q$. It is finite provided the support of q contains that of p, and we let it take the value $+\infty$ otherwise. Thus, p and q are "maximally different" when q rules out one possibility that p doesn't.[4]

Based on the relative entropy, we define the *appeal* of an information transaction a as this quantity:

$$\mathcal{A}(a) = -\frac{1}{\mu} \ln \left(\sum_s p_\alpha(s) \exp(-d(p\|q_\alpha^s)) \right). \tag{1}$$

In the above formula, and throughout the paper, we use the convention $\exp(-d(p\|q_\alpha^s)) = 0$ by continuity if $d(p\|q_\alpha^s) = +\infty$. The appeal $\mathcal{A}(a)$ of a is thus well-defined and finite if and only if there exists s such that $-d(p\|q_\alpha^s)$ is finite, which is the case if a is nonexcluding. We let $\mathcal{A}(a) = +\infty$ if a is excluding.

The appeal of an information transaction decreases with its price and increases with the relative entropy of the prior to the posteriors. Specifi-

---

[4]If p were the true distribution and q an approximate hypothesis, information theory views the relative entropy from p to q as giving the expected number of extra bits required to code the information if one were to use q instead of p.

cally, the appeal of an information transaction is measured by the inverse of its price multiplied by the natural logarithm of the expected exponentials of the negative of relative entropy from the prior to each of the generated posteriors.[5]

Our central result below asserts that $\mathcal{A}$ properly measures the appeal of information transactions:

Theorem 1 *Let a_1 and a_2 be two information transactions. Then, a_1 is more appealing than a_2 if and only if $\mathcal{A}(a_1) \geq \mathcal{A}(a_2)$.*

To illustrate the appeal ranking measured by Theorem 1, we make the following observations. The maximal elements for the appeal ranking of information transactions are all excluding transactions, since for them the appeal is infinite. On the other hand, the minimal elements are the information transactions that involve completely uninformative information structures (i.e., $\alpha_k = \alpha_{k'}$ for every two states k and k'), in which case the appeal is zero. Also, if $a_1 = (\mu_1, \alpha_1)$ and $a_2 = (\mu_2, \alpha_2)$ are ranked in a certain way in terms of their appeal, then for any $\lambda > 0$, $b_1 = (\lambda\mu_1, \alpha_1)$ and $b_2 = (\lambda\mu_2, \alpha_2)$ are ranked in the same way, which represents a price homogeneity or separability property.

The next subsections pave the way to prove Theorem 1, and the results

---

[5]If we ignore the price μ, this same formula is referred to as "free energy" in theoretical physics (see, e.g., Landau and Lifshitz, 1980), where relative entropy plays the role of the Hamiltonian of the system. Similar formulas appear under the term "stochastic complexity" in machine learning (Hinton and Zemel, 1994)

contained in them have interest in their own right. We begin with connecting preference for information with risk aversion.

4.1 Risk Aversion and Preference for Information

Given $u \in \mathcal{U}$ and $w \in \mathbb{R}$, let $\rho_u(w) = -\frac{u''(w)}{u'(w)}$ be the Arrow-Pratt coefficient of absolute risk aversion of agent u at wealth w. We also let $\overline{R}(u) = \sup_w \rho_u(w)$, and $\underline{R}(u) = \inf_w \rho_u(w)$.

Theorem 2 *Given $u_1, u_2 \in \mathcal{U}$, u_1 likes information better than u_2 if and only if $\overline{R}(u_1) \leq \underline{R}(u_2)$.*

Theorem 2 establishes the connection between preference for information and risk aversion. Lemma 2 in Cabrales, Gossner, and Serrano (2012) shows that an agent with ln utility accepts an information transaction whenever a more risk-averse agent does. Theorem 2 both extends this result to general pairs of utility functions, and shows that a converse result holds, namely, that an agent who likes information better than another is necessarily less risk-averse.

4.2 CARA Agents

Recall the class of CARA (constant absolute risk aversion) utility functions. Given $r > 0$, let u_C^r be the CARA utility function with parameter r given by $u_C^r(w) = -\exp(-rw)$ for every w.

Theorem 3 *Let a be an information transaction and w be any wealth level.*

1. *If $r > \mathcal{A}(a)$, then an agent with utility u_C^r rejects a at wealth w.*

2. *If $r \leq \mathcal{A}(a)$, then an agent with utility u_C^r accepts a at wealth w.*

Theorem 3 shows that $\mathcal{A}(a)$ can be equivalently defined as the level of risk aversion such that every CARA agent who is more risk-averse rejects a, whereas every CARA agent who is less or equally risk-averse accepts it. Theorem 3 is also a key step in the proof of Theorem 1.

It is interesting to see how Theorem 3 particularizes for excluding information transactions and completely noninformative ones. If a is excluding, then $\mathcal{A}(a) = +\infty$, and the theorem shows that all CARA agents accept a. If a is completely noninformative, then $\mathcal{A}(a) = 0$ and the theorem shows that it is rejected by all CARA agents.

5 Other Characterizations of the Appeal

5.1 Axiomatic Characterization

Theorem 1 uses the appeal measured by $\mathcal{A}$ to characterize the ranking of information transactions based on agents' preferences for information. The statement of Theorem 1 would remain unchanged if we replaced $\mathcal{A}$ by any increasing transformation $f(\mathcal{A})$. In this sense, the cardinal measure of appeal is not uniquely characterized by Theorem 1. In this subsection we show that $\mathcal{A}$ is uniquely characterized up to a positive multiplicative constant by two axioms: *duality* and *price homogeneity of degree -1.*

We turn now to the formal definition of the two axioms on the appeal of transactions. Let $\mathcal{B}$ be a map from information transactions to the nonnegative numbers.

Duality: $\mathcal{B}$ satisfies duality if, given two information transactions a_1 and a_2, $\mathcal{B}(a_1) \geq \mathcal{B}(a_2)$ if and only if a_1 is more appealing than a_2.

Price homogeneity: $\mathcal{B}$ is price-homogeneous of degree -1 if, for every $a = (\mu, \alpha)$ and $\lambda > 0$,

$$\mathcal{B}(\lambda\mu, \alpha) = \frac{1}{\lambda}\mathcal{B}(\mu, \alpha).$$

We seek an index that ranks the appeal of information transactions. In this context, conceiving of the appeal of information transactions as a "dual" to the preferences for information embodied in Definition 1 seems rather sensible. That is, whenever an agent who likes information less accepts a less appealing information transaction, an agent who likes information more should accept a more appealing information transaction.[6]

Price homogeneity separates information from price in the basic tradeoff we are trying to capture. Once homogeneity is imposed, one can begin by measuring the appeal of all transactions priced at \$1.[7] Then, homogeneity extends the appeal measure to all transactions. Furthermore, if the appeal index should take positive values, it ought to be decreasing with μ. This is

---

[6]In effect, this was the same reason used in Aumann and Serrano (2008) to justify their duality axiom. (In that case, the "duality" was sought between riskiness and risk aversion.)

[7]It turns out that for this class, the connection of our appeal index and free energy is exact.

14

the reason why the price homogeneity has degree -1 and not 1. Alternatively, we could construct an index having negative values, decreasing with μ, and price-homogeneous of degree 1; this index would be given by $-\frac{1}{\mathcal{B}}$ instead of $\mathcal{B}$.

Theorem 4 *The index $\mathcal{A}$ is the unique index, up to a positive multiplicative constant, that satisfies the axioms of duality and price homogeneity of degree -1.*

It is easy to see that the axioms are logically independent in our characterization. If one drops price homogeneity, any nonlinear monotone transformation of the index $\mathcal{A}$ also satisfies duality. If one drops duality, the index $\frac{1}{\mu}I_e(\alpha)$ satisfies homogeneity, where I_e is the entropy informativeness index proposed in Cabrales, Gossner, and Serrano (2012). Example 2 in Subsection 6.5 argues that the two indices are ordinally different.

5.2   An Approach Based on Total Rejections

Let $\mathcal{U}_{DA}$ be the set of utility functions u that are twice differentiable and such that ρ_u is decreasing (DARA). Following Hart (2011)'s approach (see also Cabrales, Gossner, and Serrano, 2012), we now introduce the definition of uniform wealth-dominance:

Definition 3 *Let a_1 and a_2 be two information transactions. We say that a_1 uniformly wealth-dominates a_2 if any $u \in \mathcal{U}_{DA}$ that rejects a_1 at all wealth levels also rejects a_2 at all wealth levels.*

This definition proposes a uniform or total rejection of transactions within the DARA class of preferences. That is, a_1 uniformly wealth-dominates a_2 because the latter is rejected more often: whenever a_1 is rejected at all wealth levels, so is a_2, but not vice versa. The definition leads to the following result:

Theorem 5 *Let a_1 and a_2 be two information transactions. Then, a_1 uniformly wealth-dominates a_2 if and only if $\mathcal{A}(a_1) \geq \mathcal{A}(a_2)$.*

We observe that the same theorem holds if we restrict the class of functions by imposing IRRA and ruin aversion on top of DARA.[8]

6 Some Properties and Examples

6.1 Some Properties of the Appeal Index

Here we discuss several properties of the appeal index.

6.1.1 Continuity

The appeal index $\mathcal{A}$ is jointly continuous in μ, in p_α, and in $(q_\alpha^s)_s$ on the domain of nonexcluding information transactions. Continuity is a natural and attractive property: small changes in either the price or the conditional probabilities of signals should translate into small changes in the appeal of the transaction. Recall, however, that $\mathcal{A}(a) = +\infty$ when a is excluding.

---

[8]IRRA and ruin aversion are the restrictions on preferences used in Cabrales, Gossner, and Serrano (2012).

6.1.2 Blackwell monotonicity

The appeal index is Blackwell-monotonic, as expressed in the following proposition:

Proposition 1 *If an information structure α_1 is more informative than another information structure α_2 in the sense of Blackwell, then for any price $\mu > 0$, the information transaction (μ, α_1) is more appealing than the information transaction (μ, α_2). Thus we have:*

$$\mathcal{A}(\mu, \alpha_1) \geq \mathcal{A}(\mu, \alpha_2).$$

6.1.3 Mixtures

A third property concerns what happens when an information structure is constructed by randomizing over two other ones. Given information structures α_1, α_2 and $1 > \lambda > 0$, we let $\lambda\alpha_1 \oplus (1 - \lambda)\alpha_2$ be the information structure in which (i) a coin toss determines whether the agent's signal is chosen from α_1 (with probability λ) or α_2 (with probability $1 - \lambda$), and (ii) the agent is informed of both the outcome of the coin toss and the signal drawn from the chosen information structure. Formally, the set of signals in $\lambda\alpha_1 \oplus (1 - \lambda)\alpha_2$ is $S_{\alpha_1} \cup S_{\alpha_2}$ (where we assume that S_{α_1} and S_{α_2} are disjoint), and the probability in state k that the agent receives signal $s \in S_{\alpha_1}$ is $\lambda\alpha_{1,k}(s)$, whereas the probability of a signal $s \in S_{\alpha_2}$ is $(1 - \lambda)\alpha_{2,k}(s)$.

Proposition 2 *Consider $\mu > 0$ and α_1, α_2 such that $\mathcal{A}(\mu, \alpha_1) \geq \mathcal{A}(\mu, \alpha_2)$. For every $1 > \lambda > 0$, we have:*

$$\mathcal{A}(\mu, \alpha_1) \geq \mathcal{A}(\mu, \lambda\alpha_1 \oplus (1-\lambda)\alpha_2) \geq \mathcal{A}(\mu, \alpha_2).$$

6.2 The Role of Prices and Priors

In the model of Section 2, p plays a dual role. Indeed, p is the agent's prior before he receives any information, and it is also a vector of prices for Arrow-Debreu securities that defines the set of no-arbitrage assets B^*. In order to both allow for the agent's prior to be different from the price system, and disentangle the two roles of p, we consider here agents whose prior belief $q \in \Delta(K)$ may differ from the vector p defining the set B^*.

In this more general model, an agent accepts an information transaction $a = (\mu, \alpha)$ at prior q if and only if:

$$\sum_s p_\alpha(s) V(u, w - \mu, q_\alpha^s) \geq V(u, w, q),$$

where q_α^s is the agent's posterior belief after receiving a signal s given the prior q. Note that if $q = p$, then $V(u, w, q)$ equals $u(w)$ so that the definition particularizes to the original one in this case.[9]

Our definition 2 extends as follows: We say that a_1 is more appealing than a_2 at prior q if, given two agents u_1, u_2 such that u_1 uniformly likes information better than u_2 and two wealth levels w_1, w_2, whenever agent u_2

---

[9]It is convenient to write the RHS of this expression this way, given our analysis of sequential transactions in the next subsection.

accepts a_2 at wealth level w_2 and prior q, then agent u_1 accepts a_1 at wealth level w_1 and prior q.

Then, we define the appeal of an information transaction $a = (\mu, \alpha)$ at prior q as:

$$
\begin{aligned}
\mathcal{A}(a, q) &= -\frac{1}{\mu} \ln \left(\sum_s p_\alpha(s) \exp(-d(p \| q_\alpha^s)) \right) - \frac{d(p \| q)}{\mu} \\
&= \mathcal{A}(a) - \frac{d(p \| q)}{\mu}.
\end{aligned}
$$

As a word of caution, we note that in the formula above, as $(q_\alpha^s)_s$ depends on q, so does $\mathcal{A}(a)$. Theorems 1 and 3 can be extended as Theorems 6 and 7 respectively:

Theorem 6 *Let a_1 and a_2 be two information transactions. Then, a_1 is more appealing than a_2 at prior q if and only if $\mathcal{A}(a_1, q) \geq \mathcal{A}(a_2, q)$.*

Theorem 7 *Let a be an information transaction and w be any wealth level.*

1. *If $r > \mathcal{A}(a, q)$, then an agent with utility u_C^r rejects a at wealth w and prior q.*

2. *If $r \leq \mathcal{A}(a, q)$, then an agent with utility u_C^r accepts a at wealth w and prior q.*

6.3 Sequential Transactions

The final property we mention concerns the appeal of an information transaction in which the buyer receives signals sequentially from different information structures. Given an information structure α with a set of signals S_α

and a family $\beta = (\beta_s)_{s \in S_\alpha}$ of information structures, where all the members of β share the same set of signals S_β, we let (α, β) be the information structure in which the agent first receives a signal s from α, then an independently drawn (conditional on k) signal s' from β_s. Formally, the set of signals in (α, β) is $S_\alpha \times S_\beta$, and in state k, the probability of receiving the pair of signals (s, s') is $\alpha_k(s)\beta_{s,k}(s')$. Given an information transaction $a = (\mu, \alpha)$ and a family of information transactions $b = (b_s)_s = (\nu, \beta_s)_s$, where all the members of b have the same price ν, we let $a + b$ denote the information transaction $(\mu + \nu, (\alpha, \beta))$.

Proposition 3 *Given information transactions a and $b = (b_s)_s$, the following hold:*

1. *If for every s, $\mathcal{A}(b_s, q_s) \geq \mathcal{A}(a)$, then $\mathcal{A}(a + b) \geq \mathcal{A}(a)$.*

2. *If for every s, $\mathcal{A}(b_s, q_s) \leq \mathcal{A}(a)$, then $\mathcal{A}(a + b) \leq \mathcal{A}(a)$.*

3. *In particular, if for every s, $\mathcal{A}(b_s, q_s) = \mathcal{A}(a)$, then $\mathcal{A}(a + b) = \mathcal{A}(a)$.*

The proposition relates the appeal of an information transaction involving (α, β) to the appeal of the information transactions involving α and $(\beta_s)_s$. As a result, the appeal of an information transaction involving α has to be measured given the prior p, as in formula (1), but the appeal of an information transaction involving β_s has to be measured given the belief q_α^s of the agent after receiving the signal s. The proposition makes intuitive sense: if the

20

agent faces a sequence of transactions whose individual appeal is increasing, the appeal of the overall transaction is at least that of the appeal of the first-stage transaction, and so on.

6.4   The Role of $\ln$ and $\exp$ in $\mathcal{A}$

So far we have argued that two intuitive properties of the index $\mathcal{A}$ are that it makes the appeal of a transaction (i) a decreasing function of its price and (ii) an increasing function of the relative entropy from the prior to each generated posterior. In this light, one could consider using the following alternative index:

$$\hat{A}(a) = \frac{1}{\mu} \sum_s p_\alpha(s) d(p \| q_\alpha^s).$$

It is apparent that the index $\hat{A}$ retains those two properties, and it also satisfies the desired separability in the form of price homogeneity of degree -1. We know from Theorem 4 that it must violate duality, and indeed, the next example illustrates why it does not rank the appeal of transactions well.

Example 1 *Let $K = \{1, 2, 3\}$ and fix a uniform prior. Consider, for instance, two information structures with each having two signals:*

$$\alpha_1 = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \ \alpha_2 = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ 1/2 & 1/2 \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

Fix an arbitrary $\mu > 0$, and define the transactions $a_1 = (\mu, \alpha_1)$ and $a_2 = (\mu, \alpha_2)$. Note that $\hat{A}(a_1) = +\infty$ because the relative entropy of the prior to the posterior generated by the first signal is infinite. On the other hand, for

21

any $\varepsilon > 0$, $\hat{A}(a_2)$ is finite. We next argue that the appeal of the transactions is not well measured by $\hat{A}$. Indeed, for a small enough $\varepsilon > 0$, the transaction a_2 is almost excluding, and hence, in such a case $r_1 = \mathcal{A}(a_1) < \mathcal{A}(a_2) = r_2$. Here, r_1 and r_2 are the risk-aversion coefficients of the two CARA individuals who define the two corresponding levels of appeal. Let $r = (r_1+r_2)/2$. Clearly, the CARA agent r uniformly likes information more than the CARA r_2 agent; the CARA r_2 agent accepts a_2, which according to the index $\hat{A}$ would be less apealing than a_1; but agent r rejects a_1.

The example makes clear the role of the exponential and its compensating logarithm as a "blow up/shrink down" of relative entropies. The exponential function, being bounded above, avoids the problem of infinite relative entropies attached to a single signal. Only when all relative entropies are infinite does the logarithm restore the infinite value of appeal. This is essential in order to satisfy the duality between uniform preferences for information and the proposed function ranking the appeal of transactions. Less extreme but somewhat more complicated examples can be constructed to make the same point without relying on CARA agents. (To do this, one would need to require a bounded interval within which the agents optimal investment lies.)

6.5 Comparison with Entropy Informativeness

We next present an example, similar to one in Cabrales, Gossner, and Serrano (2012), that illustrates how our framework serves to complete Blackwell's or-

dering. In addition, it shows how the information index in our 2012 paper can sometimes provide a different ranking from the induced index of information structures in the current study (when the price of the transaction is kept constant), while it also shows how both can sometimes point in the same direction.

Example 2 *Let $K = \{1, 2, 3\}$ and fix a uniform prior. Consider two information structures that are not ordered in the sense of Blackwell. For instance, let each of the two information structures have two signals:*

$$\alpha_1 = \begin{bmatrix} 1 - \varepsilon_1 & \varepsilon_1 \\ 1 - \varepsilon_1 & \varepsilon_1 \\ \varepsilon_1 & 1 - \varepsilon_1 \end{bmatrix}, \ \alpha_2 = \begin{bmatrix} 1 - \varepsilon_2 & \varepsilon_2 \\ 0.1 & 0.9 \\ \varepsilon_2 & 1 - \varepsilon_2 \end{bmatrix}$$

For ε_1 and ε_2 small enough, these information structures are not ranked according to Blackwell. To see this, it suffices to consider two decision problems. In Problem 1, the agent must choose one of two actions: action 1 gives a utility of 1 only in the first two states, and 0 otherwise, while action 2 gives a utility of 1 only in the third state, and 0 otherwise. In contrast, Problem 2 has action 1 pay a utility of 1 only in the first state, and 0 otherwise, while action 2 gives a utility of 1 only in states 2 or 3, and 0 otherwise. When facing Problem 1, the decision maker would value α_1 more than α_2: following the first signal in α_1, he would choose the first action and following the second signal in α_1, he would choose the second action, thereby securing a utility of 1. This would be strictly greater than his utility after α_2. On the other hand, when facing Problem 2, he would under α_2 choose action 1 after

23

the first signal and action 2 after the second, yielding a utility close to 29/30,

which is greater than his optimal utility after α_1.

Now let us compute $\mathcal{A}(a_1)$ for $a_1 = (\mu, \alpha_1)$ and $\mathcal{A}(a_2)$ for $a_2 = (\mu, \alpha_2)$.

$$\mathcal{A}(a_i) = -\frac{1}{\mu} \ln \left(\sum_s p_{\alpha_i}(s) \exp(-d(p\|q_{\alpha_i}^s)) \right),$$

and

$$
\begin{aligned}
\sum_s p_{\alpha_1}(s) \exp(-d(p\|q_{\alpha_1}^s)) &= \frac{2-\varepsilon_1}{3} \exp\left(-\frac{1}{3}\left(\ln\left(\frac{\frac{1}{3}}{\frac{1-\varepsilon_1}{2-\varepsilon_1}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{1-\varepsilon_1}{2-\varepsilon_1}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{\varepsilon_1}{2-\varepsilon_1}}\right)\right)\right) \\
&+ \frac{1+\varepsilon_1}{3} \exp\left(-\frac{1}{3}\left(\ln\left(\frac{\frac{1}{3}}{\frac{1-\varepsilon_1}{2-\varepsilon_1}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{\varepsilon_1}{2-\varepsilon_1}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{\varepsilon_1}{2-\varepsilon_1}}\right)\right)\right) \\
&\simeq \frac{2}{3} \exp\left(-\frac{1}{3}\left(\ln\frac{1}{\varepsilon_1}\right)\right) + \frac{1}{3} \exp\left(-\frac{1}{3}\left(\ln\left(\frac{1}{\varepsilon_1}\right)^2\right)\right) \\
&\simeq \frac{2}{3}\varepsilon_1^{1/3} + \frac{1}{3}\varepsilon_1^{2/3}.
\end{aligned}
$$

$$
\begin{aligned}
\sum_s p_{\alpha_2}(s) \exp(-d(p\|q_{\alpha_2}^s)) &= \frac{1.1}{3} \exp\left(-\frac{1}{3}\left(\ln\left(\frac{\frac{1}{3}}{\frac{1-\varepsilon_2}{1.1}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{0.1}{1.1}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{\varepsilon_2}{1.1}}\right)\right)\right) \\
&+ \frac{1.9}{3} \exp\left(-\frac{1}{3}\left(\ln\left(\frac{\frac{1}{3}}{\frac{1-\varepsilon_2}{1.9}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{0.9}{1.9}}\right) + \ln\left(\frac{\frac{1}{3}}{\frac{\varepsilon_2}{1.9}}\right)\right)\right) \\
&\simeq \frac{1.1}{3} \exp\left(\frac{1}{3}\ln\varepsilon_2\right) + \frac{1.9}{3} \exp\left(\frac{1}{3}\ln\varepsilon_2\right) = \varepsilon_2^{1/3}.
\end{aligned}
$$

If $\varepsilon_1 = \varepsilon_2$ and both are small enough, then $\mathcal{A}(a_2) > \mathcal{A}(a_1)$. On the other hand, if $\varepsilon_1 = \varepsilon_2^2$ and both are small, then $\mathcal{A}(a_1) > \mathcal{A}(a_2)$.

Let us now compute the entropy reduction from the uniform prior, which we denote by $I_e(\cdot)$, letting $H(q) = \sum_{k=1}^{3} -q_k \ln(q_k)$.

$$
\begin{aligned}
I_e\left(\alpha_1\right) &= H\left(p\right) - \sum_{s=1}^{2} p_{a_1}^s H\left(q_{a_1}^s\right) \\
&= 3\left(-\frac{1}{3}\ln\left(\frac{1}{3}\right)\right) - \frac{2-\varepsilon_1}{3}\left(-2\frac{1-\varepsilon_1}{2-\varepsilon_1}\ln\left(\frac{1-\varepsilon_1}{2-\varepsilon_1}\right) - \frac{\varepsilon_1}{2-\varepsilon_1}\ln\left(\frac{\varepsilon_1}{2-\varepsilon_1}\right)\right) \\
&\quad -\frac{1+\varepsilon_1}{3}\left(-2\frac{\varepsilon_1}{2-\varepsilon_1}\ln\left(\frac{\varepsilon_1}{2-\varepsilon_1}\right) - \frac{1-\varepsilon_1}{2-\varepsilon_1}\ln\left(\frac{1-\varepsilon_1}{2-\varepsilon_1}\right)\right) \\
&\simeq \ln 3 - \frac{2}{3}\left(\ln 2\right) - \frac{1}{3}\frac{1}{2}\ln 2 = \ln 3 - \frac{5}{6}\ln 2 \simeq \ln 3 - 0.57762265.
\end{aligned}
$$

$$
\begin{aligned}
I_e\left(\alpha_2\right) &= 3\left(-\frac{1}{3}\ln\left(\frac{1}{3}\right)\right) - \frac{1.1}{3}\left(-\frac{1-\varepsilon_2}{1.1}\ln\left(\frac{1-\varepsilon_2}{1.1}\right) - \frac{0.1}{1.1}\ln\left(\frac{0.1}{1.1}\right) - \frac{\varepsilon_2}{1.1}\ln\left(\frac{\varepsilon_2}{1.1}\right)\right) \\
&\quad -\frac{1.9}{3}\left(-\frac{1-\varepsilon_2}{1.9}\ln\left(\frac{1-\varepsilon_2}{1.9}\right) - \frac{0.9}{1.9}\ln\left(\frac{0.9}{1.9}\right) - \frac{\varepsilon_2}{1.9}\ln\left(\frac{\varepsilon_2}{1.9}\right)\right) \\
&\simeq \ln 3 - \frac{1}{3}\left(1.1\ln 1.1 - 0.1\ln 0.1 + 1.9\ln 1.9 - 0.9\ln 0.9\right) \\
&\simeq \ln 3 - 0.549815518.
\end{aligned}
$$

This implies that $I_e(\alpha_1) > I_e(\alpha_2)$ whenever $\varepsilon_1, \varepsilon_2$ are sufficiently close to zero.

The reason for the difference between entropy informativeness and the approach in this paper is the larger sensitivity of the index $\mathcal{A}$ to information concerning low-probability events. In particular, α_1 causes a larger reduction in entropy, being associated with an almost fully informative signal (s_2). In contrast, for equal prices, a purchase of α_2 is more appealing for sequences of small ε_1 and ε_2 where $\varepsilon_1 = \varepsilon_2$. To understand the latter, note that the limits, as ε_1 and ε_2 vanish, of α_1 and α_2 lead to excluding information transactions,

with infinite appeal. The large unbounded appeal of those transactions before going to the limits is explained by the large investments made following each signal. However, in α_1, following signal s_2, two states are becoming extremely unlikely, leading the agent to an optimal investment with large losses in these two states, whereas in α_2 large losses in the optimal investment are confined to only one state. Because of this, when ε_1 and ε_2 go to zero at the same rate, the large negative utility that a CARA agent derives from large negative wealth implies that α_2 is more appealing than α_1. Convergence rates matter, though: this conclusion is overturned if ε_1 goes to zero much faster than ε_2.

To explore somewhat more systematically the difference between the index based on entropy and the one in this paper, we investigate conditions on "small information" that renders them equivalent. Let $a_i = (\alpha_i, \mu)$. We then have the following equation:

$$
\begin{aligned}
\mathcal{A}(a_i) &= -\frac{1}{\mu} \ln \left(\sum_s p_{\alpha_i}(s) \exp(-d(p \| q_{\alpha_i}^s)) \right) \\
&= -\frac{1}{\mu} \ln \left(\sum_s p_{\alpha_i}(s) \exp \left(-\sum_k p(k) \left(\ln p(k) - \ln q_{\alpha_i}^s(k) \right) \right) \right)
\end{aligned}
$$

$$
\begin{aligned}
I_e(\alpha_i) &= -\sum_k p(k) \ln p(k) - \sum_s p_{\alpha_i}(s) \left(-\sum_k q_{\alpha_i}^s(k) \ln q_{\alpha_i}^s(k) \right) \\
&= \sum_s p_{\alpha_i}(s) \left(-\sum_k p(k) \left(\ln p(k) - \ln q_{\alpha_i}^s(k) \right) - \sum_k \left(p(k) - q_{\alpha_i}^s(k) \right) \ln q_{\alpha_i}^s(k) \right).
\end{aligned}
$$

This implies that to a first order approximation when $q_{\alpha_i}^s$ is close to p,

$$\mathcal{A}(a_i) \simeq -\frac{1}{\mu}\ln\left(1 + \sum_s p_{\alpha_i}(s)\left(-\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)\right)\right)$$

$$\simeq -\frac{1}{\mu}\sum_s p_{\alpha_i}(s)\left(-\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)\right)$$

$$= \frac{1}{\mu}\sum_s p_{\alpha_i}(s)\left(\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)\right)$$

and

$$I_e(\alpha_i) = \sum_s p_{\alpha_i}(s)\left(-\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right) - \sum_k \left(p(k) - q_{\alpha_i}^s(k)\right)\ln q_{\alpha_i}^s(k)\right)$$

$$= \sum_s p_{\alpha_i}(s)\left(-\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right) + \left(\sum_k \left(\frac{q_{\alpha_i}^s(k)}{p(k)} - 1\right)p(k)\ln q_{\alpha_i}^s(k)\right)\right)$$

$$\simeq \sum_s p_{\alpha_i}(s)\left(-\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right) - \left(\sum_k \left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)p(k)\ln q_{\alpha_i}^s(k)\right)\right.$$

$$= \sum_s p_{\alpha_i}(s)\left(-\sum_k \left(1 + \ln q_{\alpha_i}^s(k)\right)p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)\right).$$

As a result, it follows that:

$$\mathcal{A}(a_i) \simeq \frac{1}{\mu}\sum_s p_{\alpha_i}(s)\left(\sum_k p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)\right) \qquad (2)$$

and

$$I_e(\alpha_i) \simeq \sum_s p_{\alpha_i}(s)\left(-\sum_k \left(1 + \ln q_{\alpha_i}^s(k)\right)p(k)\left(\ln p(k) - \ln q_{\alpha_i}^s(k)\right)\right). \qquad (3)$$

A comparison of expressions (2) and (3) makes clear that when priors and posteriors are similar, the two indices point in the same direction - as long as it is also true that the $q_{\alpha_i}^s(k)$ vectors are all parallel to the unit

27

vector and that $\ln q_{\alpha_i}^s(k) < -1$, that is, when priors are close to uniform and there are more than two states. Otherwise, cases such as the one provided in Example (2), when posteriors are very informative, are likely to make the indices diverge.

6.6 The Case of the Continuum

We have worked with finitely many states to avoid measure-theoretic technicalities. The results in this paper are easy to extend to distributions with a continuum of states. This is useful because many applications assume such a continuum.

Example 3 *Suppose that each signal drawn generates a distribution with conditional normal density $s \sim N(\eta_s, \Sigma_s)$. Let the prior p be normally distributed with $p \sim N(\eta_p, \Sigma_p)$. Then the relative entropy $(d(p\|q_{\alpha_i}^s))$, also called Kullback-Leibler divergence, can be written as follows:*

$$d(p\|q_{\alpha_i}^s) = \frac{1}{2}\left(\operatorname{tr}\left(\Sigma_s^{-1}\Sigma_p\right) + (\eta_s - \eta_p)^\top \Sigma_s^{-1}(\eta_s - \eta_p) - \ln\left(\frac{|\Sigma_p|}{|\Sigma_s|}\right) - n\right).$$

This implies that $\mathcal{A}(a)$ can be written as follows:

$$\mathcal{A}(a) = -\frac{1}{\mu}\ln\left(\sum_s p_\alpha(s)\exp\left(-\frac{1}{2}\left(\operatorname{tr}\left(\Sigma_s^{-1}\Sigma_p\right) + (\eta_s - \eta_p)^\top \Sigma_s^{-1}(\eta_s - \eta_p) - \ln\left(\frac{|\Sigma_p|}{|\Sigma_s|}\right) - n\right)\right)\right),$$

and also that the expression for I_e is the following:

$$I_e(\alpha) = \frac{1}{2}\ln\left((2\pi e)^n |\Sigma_p|\right) - \sum_s p_\alpha(s)\frac{1}{2}\ln\left((2\pi e)^n |\Sigma_s|\right).$$

Thus, the two indices lend themselves to easy computations, and comparisons as that in Example 2 can be readily drawn.

7 Related literature

As mentioned in the introduction, the natural limitation of Blackwell (1953)'s ordering is that it does not provide a complete ordering of information structures. More recent research has since focused on restricting preferences to a particular class. Lehmann (1988) restricts attention to problems with monotone decision rules, and Persico (2000), Athey and Levin (2001), and Jewitt (2007) do so to some more general classes of monotone problems. The main difference between this line of research and our approach is that we provide a complete order through a duality axiom for problems with a restricted set of investment opportunities.[10]

Relative entropy plays an important role in our index. Kullback and Leibler (1951) showed that relative entropy measures the mean information per sample for distinguishing between two hypotheses when one of them is true.[11] Subsequently, Shore and Johnson (1980) provided an axiomatization of relative entropy with respect to the prior as a criterion for model selection.

The relative-entropy measure of proximity of probability distributions appears prominently in many economic settings. For example, Blume and Easley (1992) and Sandroni (2000) show that, in dynamic exchange economies,

---

[10]Ganuza and Penalva (2010) provides a partial ordering of information structures based on various measures of dispersion of distributions, rather than on decision-theoretic considerations. Many of those measures are presented in Shaked and Shanthikumar (2007). They then study the implications of greater informativeness (in their sense) for auction problems.

[11]A maximum likelihood estimator of a parametric model is also a minimizer of the Kullback-Leibler divergence.

markets favor agents who make the most accurate predictions when accuracy is measured according to relative entropy. Hansen and Sargent (2010, 2001) introduce into the economics literature a multiple priors model that builds on earlier work in the optimal-control literature. In that model, agents take into account the possibility that their beliefs q may not be correct, and they consider other possible alternatives p; the relative likelihood of these alternative distributions is then measured by relative entropy. This model, also called the multiplier preferences model, is in turn a special case of the model of ambiguity aversion in Maccheroni, Marinacci, and Rustichini (2006). Relative entropy has also been used in the reputation model of Gossner (2011) to assess differences between conditional and unconditional predictions about long-run player types. The use of relative entropy in this context allows one to derive explicit bounds on the payoff of the long-run player, thanks to the chain-rule property of relative entropy. However, none of these previous papers in economics uses relative entropy to measure the informativeness of signals or the appeal of information transactions.

Outside economics, relative entropy is widely used to measure both informativeness and differences between distributions. In information theory, the Kraft–McMillan theorem[12] shows that when sending messages to identify a state s in a set S which has a certain probability distribution, the Kullback-Leibler divergence quantifies the expected extra message-length needed if a

[12]From Kraft (1949) and McMillan (1956).

code is optimal for a given (wrong) distribution rather than for one based on the true distribution. Weyl (2007) has used this theorem to propose relative entropy as a good way to learn which is the best scientific theory among those that describe the same data. Soofi and Retzer (2002) provide a summary of the applications of entropy-related indices in statistics and econometrics, and the interrelationships between them. There are numerous applications of relative entropy in a range of disparate fields. Gourieroux, Monfort, and Renault (1987) have proposed the Kullback information criterion as a basis for causality tests in time series analysis. They are used in linguistics, for example, to measure the information content of words (Kuperman, Bertram, and Baayen, 2010; Mishra and Bangalore, 2011); in optics, for measuring the discriminatory power of sensory networks (Ong, Xiaoy, Tham, and Ang, 2009); in hydrology, for assessing data informativeness for risk management (Singh, 1997); in genetics, to measure genetic diversity (Sherwin, 2010); in zoology, where long-run fitness is maximized by the phenotype that adapts itself (best-responds) to the signal which minimizes relative entropy with respect to the true state of the environment (Donaldson-Matasci, Bergstrom, and Lachmann, 2010); and even in archaeology, to infer whether a set of findings is consistent with the known facts of a particular prehistoric period (Justeson, 1973). Although many of these applications use relative entropy to measure informativeness, none of them provides a decision-theoretic microfoundation for such use.

31

8 Conclusion

We propose a novel approach that provides a complete ordering of information transactions. The approach also orders information structures, when the transaction price is kept constant. The complete ordering proposed here is dual to the agent's preferences for information, as stated in Definition 1. We find that only two axioms, duality from preferences for information and homogeneity, yield a cardinal characterization of an index of appeal of information transactions, up to a multiplicative normalization constant. The index has a simple analytic characterization, which is an expectation of the exponential of relative entropy from the prior to the posteriors after signals. This is remarkable since relative entropy has been extensively used as a measure of information gain in computer science and other disciplines. Moreover, the specific appeal formula is actually identical (ignoring the price separable term) to the notion of free energy in theoretical physics (Landau and Lifshitz, 1980). The role of duality in this framework, as well as for measuring riskiness (Aumann and Serrano, 2008), suggests that the approach may be useful in general to index other multidimensional magnitudes of economic interest.

9 Proofs

This section presents the proofs of the results in an appropriate logical order.

9.1 Proof of Theorem 2

We begin by stating and proving several auxiliary lemmas.

Lemma 1 *Fix p and consider a sequence q^n of beliefs such that $q^n \to p$. Let b^n be the optimal investment for an agent with beliefs q^n. Then, it must be true that $b^n \to 0$.*

Proof. If the property does not hold, there exists a sequence $q^n \to p$ and a corresponding sequence of optimal investments b^n together with $\varepsilon > 0$ such that, for every n, $\|b^n\|_\infty \geq \varepsilon$. Since u is strictly concave, there exists $a > 0$ such that for every z with $|z| \geq \varepsilon$,

$$u(w + z) \leq u(w) + zu'(w) - a|z|.$$

We then have for every n:

$$
\begin{aligned}
V(u, w, q^n) &= \sum_k q_k^n u(w + b_k^n) \\
&\leq \; + \sum_{|b_k^n| < \varepsilon} q_k^n (u(w) + u'(w)b_k^n) + \sum_{|b_k^n| \geq \varepsilon} q_k^n (u(w) + u'(w)b_k^n - a|b_k^n|) \\
&= \; u(w) + \sum_{|b_k^n| < \varepsilon} (q_k^n - p_k^n)u'(w)b_k^n + \sum_{|b_k^n| \geq \varepsilon} (q_k^n - p_k^n)u'(w)b_k^n - aq_k^n|b_k^n|,
\end{aligned}
$$

where the last equality uses $\sum_k q_k^n = 1$ and $\sum_k p_k^n b_k^n = 0$. This implies both

$$\lim_{n \to \infty} \sum_{|b_k^n| < \varepsilon} (q_k^n - p_k^n)u'(w)b_k^n = 0$$

33

and

$$\limsup_{n\to\infty} \sum_{|b_k^n|\geq\varepsilon} (q_k^n - p_k^n) u'(w) b_k^n a q_k^n |b_k^n| < 0,$$

since for every n, there exists k such that $|b_n^n| \geq \varepsilon$. This shows that

$$\limsup_{n\to\infty} V(u, w, q^n) < u(w),$$

which is in contradiction with $V(u, w, q) \geq u(w)$ for every q. We conclude that the property holds as claimed. ∎

Lemma 2 *Fix p and consider q close to p. Then, the optimal investment $b(q) = (b_k(q))_{k\in K}$ for an agent with belief $q = (q_k)_{k\in K}$ is*

$$b_k(q) = \frac{1}{p_k \rho(w)}(q_k - p_k) + o(\|q - p\|).$$

Proof. The agent's problem is to maximize $\sum_k q_k u(w + b_k)$ under the constraint $\sum_k p_k b_k = 0$. The solution is uniquely given by the system of first-order conditions:

$$q_k u'(w + b_k) = \lambda p_k,$$

where λ is independent of k. Using a first order Taylor expansion of $u'(w+b_k)$, we obtain:

$$U'(w) + b_k u''(w) = \lambda \frac{p_k}{q_k} + o(b_k). \tag{4}$$

We multiply each equation by p_k and sum over k to get:

$$U'(w) = \lambda \frac{\sum_j p_j^2}{q_j} + o(b_k). \tag{5}$$

34

We replace the value of λ obtained using (5) into equation (4) and get:

$$b_k = \frac{u'(w)}{u''(w)}\left(\frac{p_k}{q_k \sum_j \frac{p_j^2}{q_j}} - 1\right) + o(b_k).$$

In vector form, this can be expressed as:

$$b = F(q) + \gamma(b),$$

where $(F(q))_k = \frac{u'(w)}{u''(w)}\left(\frac{p_k}{q_k \sum_j \frac{p_j^2}{q_j}} - 1\right)$ and $\gamma(b) \in \mathbb{R}^K$ is such that $\frac{\|\gamma(b)\|}{\|b\|} \to 0$ as $\|b\| \to 0$.

We now show that $\|b\| = O(\|q - p\|)$. Assume to the contrary that there exists a sequence $q^n \to p$ and a corresponding sequence b^n such that $\frac{\|b^n\|}{\|q^n - p^n\|} \to \infty$. We would then have:

$$\frac{\|b^n\|}{\|q^n - p^n\|} \leq \frac{\|F(q^n)\|}{\|q^n - p^n\|} + \frac{\gamma(b^n)}{\|b^n\|}\frac{\|b^n\|}{\|q^n - p^n\|}.$$

However, a simple computation shows that $\|F(q^n)\| = O(\|q^n - p^n\|)$, and we know from Lemma 1 that $\|b^n\| \to 0$; hence, $\frac{\gamma(b^n)}{\|b^n\|} \to 0$. This yields a contradiction, and hence the conclusion that $\|b\| = O(\|q - p\|)$.

We thus have $\frac{\gamma(b)}{\|q-p\|} \to 0$ as $\|q - p\| \to 0$. We can therefore write

$$b_k = \frac{u'(w)}{u''(w)}\left(\frac{p_k}{q_k \sum_j \frac{p_j^2}{q_j}} - 1\right) + o(\|q - p\|)$$

$$= \frac{1}{\rho(w)}\left(\frac{q_k \sum_j \frac{p_j^2}{q_j} - p_k}{q_k \sum_j \frac{p_j^2}{q_j}}\right) + o(\|q - p\|)$$

$$= \frac{1}{p_k \rho(w)}(q_k - p_k) + o(\|q - p\|),$$

where the last line uses the fact that $\lim_{q \to p} \sum_j \frac{p_j^2}{q_j} = 1$. ∎

35

Lemma 3 *Fix p and consider q close to p. Then,*

$$V(u, w, q) = u(w) + \frac{1}{2} \sum_k \frac{(q_k - p_k)^2}{\rho(w)p_k} u'(w) + o(\|q - p\|^2).$$

Proof. We have

$$V(u, w, q) = \sum_k q_k u(w + b_k),$$

where $b = (b_k)_{k \in K}$ is defined as in Lemma 2. A second order Taylor expansion gives

$$
\begin{aligned}
V(u, w, q) &= u(w) + \sum_k q_k b_k u'(w) + \frac{1}{2} \sum_k q_k b_k^2 u''(w) + o(\|b\|^2) \\
&= u(w) + \sum_k (q_k - p_k) b_k u'(w) + \frac{1}{2} \sum_k q_k b_k^2 u''(w) + o(\|b\|^2).
\end{aligned}
$$

From Lemma 3 we know that $\|b\| = O(\|q - p\|)$. Hence, we can replace $o(\|b\|^2)$ by $o(\|p - q\|^2)$ in the expression above. By substituting b_k for the expression in Lemma 2 we obtain:

$$
\begin{aligned}
V(u, w, q) &= u(w) + \sum_k \frac{(q_k - p_k)^2}{\rho(w)p_k} u'(w) \\
&\quad + \frac{1}{2} \sum_k \frac{q_k}{\rho(w)^2 p_k^2} (q_k - p_k)^2 u''(w) + o(\|q - p\|^2) \\
&= u(w) + \frac{1}{2} \sum_k \frac{(q_k - p_k)^2}{\rho(w)p_k} u'(w) + o(\|q - p\|^2),
\end{aligned}
$$

which is as claimed. ∎

Fix p, and two states $k, l \in K$. For $\min\{p_k, p_l\} > \varepsilon > 0$, let $q^{\varepsilon,k}$ be given by $q_{k'}^{\varepsilon,k} = p_{k'}$ for $k' \neq k, l$; $q_k^{\varepsilon,k} = p_k + \varepsilon$; and $q_l^{\varepsilon,k} = p_l - \varepsilon$. Similarly, $q^{\varepsilon,l}$ is given by $q_{k'}^{\varepsilon,l} = p_{k'}$ for $k' \neq k, l$; $q_l^{\varepsilon,l} = p_l + \varepsilon$; and $q_k^{\varepsilon,l} = p_k - \varepsilon$. Thus, the

belief $q^{\varepsilon,k}$ gives slightly higher weight to state k and slightly lower weight to state l than p, whereas $q^{\varepsilon,l}$ does the opposite. Now consider an information structure $\alpha(\varepsilon)$ such that with probability $\frac{1}{2}$, the agent's posterior is $q^{\varepsilon,k}$; and with probability $\frac{1}{2}$ it is $q^{\varepsilon,l}$.(Such an information structure exists since $\frac{1}{2}q^{\varepsilon,k} + \frac{1}{2}q^{\varepsilon,l} = p$.)

Lemma 4 *For ε close to 0, the maximal price $\mu(\varepsilon)$ that an agent is willing to pay for $\alpha(\varepsilon)$ is:*

$$\mu(\varepsilon) = \frac{p_k + p_l}{2\rho(w)p_k p_l}\varepsilon^2 + o(\varepsilon^2).$$

Proof. The maximal price $\mu(\varepsilon)$ is such that the informational gains exactly compensate the monetary loss. Such a price satisfies the equation:

$$\frac{1}{2}\left(V(u, w - \mu(\varepsilon), q^{\varepsilon,k}) + V(u, w - \mu(\varepsilon), q^{\varepsilon,l})\right) = u(w).$$

Relying on Lemma 3, we get:

$$u(w) - u(w - \mu(\varepsilon)) = \frac{u'(w - \mu(\varepsilon))}{2\rho(w - \mu(\varepsilon))}\left(\frac{\varepsilon^2}{p_k} + \frac{\varepsilon^2}{p_l}\right) + o(\varepsilon^2).$$

This shows that $\mu(\varepsilon) \to 0$ as $\varepsilon \to 0$, and therefore, by taking a first-order Taylor approximation of $u(w - \mu(\varepsilon))$, we obtain:

$$\mu(\varepsilon)u'(w) + o(\mu(\varepsilon)) = \frac{u'(w)}{2\rho(w)}\frac{p_k + p_l}{p_k p_l}\varepsilon^2 + o(\varepsilon^2).$$

We conclude that:

$$\mu(\varepsilon) = \frac{p_k + p_l}{2\rho(w)p_k p_l}\varepsilon^2 + o(\varepsilon^2),$$

as we wanted to show. ∎

Having proved this series of lemmata, we now proceed to prove Theorem 2.

Proof of Theorem 2. Assume that u_1 likes information better than u_2, and consider wealth levels w_1 for u_1 and w_2 for u_2. The maximal price that u_1 is willing to pay at wealth level w_1 for the information structure $\alpha(\varepsilon)$ is always higher than or equal to the maximal price that u_2 is willing to pay at wealth level w_2 for the same information. From Lemma 4 we obtain that for every w_1, w_2 and for a small enough $\varepsilon > 0$,

$$\frac{p_k + p_l}{2\rho_{u_1}(w_1)p_k p_l}\varepsilon^2 \geq \frac{p_k + p_l}{2\rho_{u_2}(w_2)p_k p_l}\varepsilon^2.$$

Hence, $\rho_{u_1}(w_1) \leq \rho_{u_2}(w_2)$, which implies $\overline{R}(u_1) \leq \underline{R}(u_2)$.

Now assume that $\overline{R}(u_1) \leq \underline{R}(u_1)$. For every x, w_1, and w_2, we have

$$\frac{u_2''(w_2 + z)}{u_2'(w_2 + z)} \leq \frac{u_1''(w_1 + z)}{u_1'(w_1 + z)}.$$

By integration on z, we have:

$$\begin{cases} \ln u_2'(w_2 + z) - \ln u_2'(w_2) \leq \ln u_1'(w_1 + z) - \ln u_1'(w_1) & \text{if } z \geq 0; \\ \ln u_2'(w_2 + z) - \ln u_2'(w_2) \geq \ln u_1'(w_1 + z) - \ln u_1'(w_1) & \text{if } z \leq 0; \end{cases}$$

which is the same as:

$$\begin{cases} \frac{u_2'(w_2+z)}{u_2'(w_2)} \leq \frac{u_1'(w_1+z)}{u_1'(w_1)} & \text{if } z \geq 0; \\ \frac{u_2'(w_2+z)}{u_2'(w_2)} \geq \frac{u_1'(w_1+z)}{u_1'(w_1)} & \text{if } z \leq 0. \end{cases}$$

By a second integration on z, for every z:

$$\frac{u_2(w_2 + z) - u_2(w_2)}{u_2'(w_2)} \leq \frac{u_1(w_1 + z) - u_1(w_1)}{u_1'(w_1)}.$$

38

Thus, for every $q \in \Delta(K)$ and $\mu \geq 0$:

$$\frac{V(u_2, w_2 - \mu, q) - u_2(w_2)}{u_2'(w_2)} \leq \frac{V(u_1, w_1 - \mu, q) - u_1(w_1)}{u_1'(w_1)}.$$

And finally, for every information structure α,

$$\frac{\sum_s p_\alpha(s) V(u_2, w_2 - \mu, q_k^s) - u_2(w_2)}{u_2'(w_2)} \leq \frac{\sum_s p_\alpha(s) V(u_1, w_1 - \mu, q_k^s) - u_1(w_1)}{u_1'(w_1)}.$$

This implies that for every w_1, w_2, if u_1 rejects $a = (\mu, \alpha)$ at wealth w_1, then u_2 also rejects it at wealth w_2. ∎

9.2 Proof of Theorem 3

The first step in the proof of Theorem 3 is to characterize the optimal investment portfolio for a CARA agent as well as the function $V(u_C^r, w, q)$. This is done in the lemmata below.

For a CARA agent with coefficient r of risk aversion and wealth level w, we consider the problem of optimal portfolio choice when the agent's belief is q. The next lemma shows that the solution is interior when q has full support.

Lemma 5 *Let $q \in \Delta(K)$ have full support. The optimal portfolio for the CARA agent with risk-aversion coefficient r and belief q is independent of w, and is given by*

$$b_k = -\frac{1}{r}(-d(p\|q) + \ln \frac{p_k}{q_k}).$$

Proof. The agent's objective is to maximize

$$\sum_k q_k \exp(-r(w + b_k)),$$

subject to the budget constraint $\sum_k p_k b_k = 0$. The first-order condition shows that

$$q_k \exp(-rb_k) = \lambda p_k,$$

where λ is independent of k. We then have, for every k,

$$-rb_k = \ln \lambda + \ln \frac{p_k}{q_k}.$$

Summing over these expressions, after we multiply each of them by p_k, gives

$$0 = \ln(\lambda) + d(p\|q),$$

and hence, the result. ∎

We proceed with a characterization of the function $V(u_C^r, w, q)$. We recall the convention that $\exp(-d(p\|q) - rw) = 0$ by continuity if $d(p\|q) = \infty$, and state:

Lemma 6 *For every r, w, q:*

$$V(u_C^r, w, q) = 1 - \exp(-d(p\|q) - rw).$$

Proof. First, assume that q has full support; hence, $d(p\|q)$ is finite. Using

the optimal-portfolio characterization of Lemma 5, we obtain:

$$
\begin{aligned}
V(u_C^r, w, q) &= 1 - \sum_k q_k \exp(-r(w + b_k)) \\
&= 1 - \exp(-rw) \sum_k q_k \exp(-d(p\|q) + \ln \frac{p_k}{q_k}) \\
&= 1 - \exp(-rw - d(p\|q)) \sum_k q_k \frac{p_k}{q_k} \\
&= 1 - \exp(-rw - d(p\|q)).
\end{aligned}
$$

Now assume that $q_{k_0} = 0$ for some k_0; hence, $d(p\|q) = +\infty$. The investment b^0 given by :

$$
\begin{cases}
b_{k_0}^0 &= -\frac{1 - p_{k_0}}{p_{k_0}}; \\
b_k &= 1 \qquad \text{if } k \neq k_0
\end{cases}
$$

is such that $\lambda b^0 \in B^*$ for every $\lambda \geq 0$. For every such λ, we have

$$
\begin{aligned}
V(u_C^r, w, q) &\geq \sum_k q_k u_C^r(w + \lambda b_k^0) \\
&= u_C^r(w + \lambda b_k^0) \\
&= 1 - \exp(-r(w + \lambda)).
\end{aligned}
$$

Since $\lim_{\lambda \to \infty} \exp(-r(w + \lambda)) = 0$, we have $V(u_C^r, w, q) \leq 1$. On the other hand, $V(u_C^r, w, q) \leq \sup_z u_C^r(z) = 1$. The desired conclusion is therefore that $V(u_C^r, w, q) = 1$. ∎

We now proceed to the proof of Theorem 3.

Proof of Theorem 3. The agent accepts a if and only if

$$
\sum_s p_\alpha(s) V(u_C^r, w - \mu, q_\alpha^s) \geq u_C^r(w).
$$

41

If a is excluding, then the left-hand side of the inequality equals 1, and the inequality is satisfied for all r and w. If a is nonexcluding, then the agent accepts a if and only if

$$\exp(-rw) \geq \exp(-r(w - \mu)) \sum_s p_\alpha(s) \exp(-d(p\|q_\alpha^s)).$$

This is equivalent to

$$\exp(-r\mu) \geq \sum_s p_\alpha(s) \exp(-d(p\|q_\alpha^s)),$$

which in turn is equivalent to $r \leq \mathcal{A}(a)$. Thus, for $r \leq \mathcal{A}(a)$, the agent accepts a at every wealth level, whereas for $r > \mathcal{A}(a)$, the agent rejects a at every wealth level. ∎

9.3 Proof of the Main Result

Equipped with Theorems 2 and 3, we can now prove our main result, Theorem 1.

Proof of Theorem 1. First assume that a_1 is more appealing than a_2, and that $\mathcal{A}(a_2)$ is finite. By Theorem 3, a CARA agent with a coefficient of risk aversion $\mathcal{A}(a_2)$ accepts a_2 at every wealth level. This agent likes information better than itself according to Definition 1 since, by Theorem 3, acceptance or rejection for CARA agents is independent of wealth. Since a_1 is more appealing than a_2, this CARA agent also accepts a_1 at every wealth level, which implies (also by Theorem 3) that $\mathcal{A}(a_1) \geq \mathcal{A}_{(a_2)}$.

The case in which $\mathcal{A}(a_2) = \infty$ is dealt with similarly: by Theorem 3 every CARA agent accepts a_2 at every wealth level, which implies that the same agent also accepts a_1 at every wealth level. By Theorem 3 again, this implies that we also have $\mathcal{A}(a_1) = \infty$.

Now assume that $\mathcal{A}(a_1) \geq \mathcal{A}(a_2)$. Consider two agents u_1 and u_2 such that u_1 likes information better than u_2. Given wealth levels w_1 and w_2, and assuming that u_2 accepts a_2 at w_2, we need to prove that u_1 accepts a_1 at w_1. By Theorem 2 we have $\overline{R}(u_1) \leq \underline{R}(u_2)$. Since $\overline{R}(u_1) > 0$ and $\underline{R}(u_2) < \infty$, $\underline{R}(u_2)$ is positive and finite. Let $r = \underline{R}(u_2)$. Since $\overline{R}(u_r^C) = r$, the agent u_r^C likes information better than agent u_2 does, by Theorem 2; hence the former accepts a_2 at any wealth level. By Theorem 3 this means that $r \leq \mathcal{A}(a_2)$, and hence also $r \leq \mathcal{A}(a_1)$, so that u_r^C also accepts a_1 at any wealth level. Since $\overline{R}(u_1) \leq r = \underline{R}(u_r^C)$ and u_1 likes information better than u_r^C (also by Theorem 2), it follows that u_1 accepts a_1 at wealth level w_1. ■

9.4   Proof of Theorem 4

Proof. The index $\mathcal{A}$ satisfies price homogeneity of degree -1 by its definition, and Theorem 1 shows that it satisfies duality. It only remains to prove uniqueness up to a positive multiplicative constant.

Let $\mathcal{B}$ satisfy both axioms, and fix an information transaction $a_1 = (\mu_1, \alpha_1)$ such that $\mathcal{A}(a_1) \neq 0$. Consider any other information transaction

$a_2 = (\mu_2, \alpha_2)$. First assume that $\mathcal{A}(a_2) \neq 0$. By homogeneity of $\mathcal{A}$, we have:

$$\mathcal{A}(a_2) = \mathcal{A}(\frac{\mathcal{A}(a_1)}{\mathcal{A}(a_2)}\mu_1, \alpha_1).$$

By duality of $\mathcal{A}$, a_2 is more appealing than $(\frac{\mathcal{A}(a_1)}{\mathcal{A}(a_2)}\mu_1, \alpha_1)$. By duality of $\mathcal{B}$, this implies:

$$\mathcal{B}(a_2) \geq \mathcal{B}(\frac{\mathcal{A}(a_1)}{\mathcal{A}(a_2)}\mu_1, \alpha_1).$$

Again, by duality of both $\mathcal{A}$ and $\mathcal{B}$, the converse inequality is also true, so that we have:

$$\mathcal{B}(a_2) = \mathcal{B}(\frac{\mathcal{A}(a_1)}{\mathcal{A}(a_2)}\mu_1, \alpha_1).$$

Finally, by price homogeneity of $\mathcal{B}$, we obtain:

$$\mathcal{B}(a_2) = \frac{\mathcal{B}(a_1)}{\mathcal{A}(a_1)}\mathcal{A}(a_2).$$

Note that $\frac{\mathcal{B}(a_1)}{\mathcal{A}(a_1)}$ is independent of a_2 and is a positive constant because, if $\mathcal{B}(a_1) = 0$, we would also have $\mathcal{B}(2\mu_1, \alpha_1) = 0$. This would imply that $(2\mu_1, \alpha_1)$ is more appealing than (μ_1, α_1), hence that $\mathcal{A}(2\mu_1, \alpha_1) \geq \mathcal{A}(a_1)$. But this is a contradiction with $\mathcal{A}(2\mu_1, \alpha_1) = \frac{1}{2}\mathcal{A}(a_1) > 0$.

Now assume that $\mathcal{A}(a_2) = 0$. This means that for every s, $d(p\|q_{\alpha_2}^s) = 0$; hence, $q_{\alpha_2}^s = p$. Therefore, α_2 has no informational content, and all agents reject a_2 at every price $\mu_2 > 0$. In particular, it is tautologically true that for every $\varepsilon > 0$, the transaction $(\frac{\mu_2}{\varepsilon}, \alpha_1)$ is more appealing than a_2. Thus, $\mathcal{B}(\frac{\mu_2}{\varepsilon}, \alpha_1) \geq \mathcal{B}(a_2)$, and by homogeneity we have $\mathcal{B}(a_2) \leq \varepsilon\mathcal{B}(a_1)$. Since this is true for every $\varepsilon > 0$, we must have $\mathcal{B}(a_2) = 0$.

We have therefore shown the existence of a positive constant $\frac{\mathcal{B}(a_1)}{\mathcal{A}(a_1)}$ such that, for every a_2,

$$\mathcal{B}(a_2) = \frac{\mathcal{B}(a_1)}{\mathcal{A}(a_1)}\mathcal{A}(a_2),$$

which completes the proof. ∎

9.5 Proof of Theorem 5

We begin with an auxiliary lemma:

Lemma 7 *An agent $u \in \mathcal{U}_{DA}$ rejects a at all wealth levels if and only if $\underline{R}(u) \geq \mathcal{A}(a)$.*

Proof. Let $r = \mathcal{A}(a)$. Assume $\underline{R}(u) \geq \mathcal{A}(a)$. Since u is DARA, $\rho_u(w) > \mathcal{A}(a)$ for every w. The same computation as in the proof of Theorem 2 shows that for every z,

$$\frac{u(w+z) - u(w)}{u'(w)} < \frac{u_r^C(w+z) - u_r^C(w)}{u_r^{C'}(w)}.$$

If q has full support, the solution to the maximization problem of $\sum_k q_k u(w + b_k)$ under the constraint $\sum_k p_k b_k \leq 0$ is interior. Let $b(q)$ achieve this maximum. We have:

$$
\begin{aligned}
\frac{V(u, w - \mu, q) - u(w)}{u'(w)} &= \frac{\sum_k q_k u(w - \mu + b_k(q)) - u(w)}{u'(w)} \\
&< \frac{\sum_k q_k u_r^C(w - \mu + b_k(q)) - u_r^C(w)}{u_r^{C'}(w)} \\
&\leq \frac{V(u_r^C, w - \mu, q) - u_r^C(w)}{u_r^{C'}(w)}.
\end{aligned}
$$

45

If q does not have full support, we still have:

$$
\begin{aligned}
\frac{V(u, w - \mu, q) - u(w)}{u'(w)} &= \sup_{b \in B^*} \frac{\sum_k q_k u(w - \mu + b_k) - u(w)}{u'(w)} \\
&\leq \sup_{b \in B^*} \frac{\sum_k q_k u_r^C(w - \mu + b_k(q)) - u_r^C(w)}{u_r^{C\prime}(w)} \\
&\leq \frac{V(u_r^C, w - \mu, q) - u_r^C(w)}{u_r^{C\prime}(w)}.
\end{aligned}
$$

Note that $\mathcal{A}(a) \leq r$ implies that $\mathcal{A}(a)$ is finite, and hence that a is nonexcluding; there therefore exists s such that $p_\alpha(s) > 0$ and q^s has full support. Therefore:

$$
\begin{aligned}
\frac{\sum_s p_\alpha(s) V(u, w - \mu, q^s) - u(w)}{u'(w)} &< \frac{\sum_s p_\alpha(s) V(u_r^C, w - \mu, q^s) - u_r^C(w)}{u_r^{C\prime}(w)} \\
&= 0,
\end{aligned}
$$

where the last equality comes from the fact that the agent u_r^C is indifferent between accepting and rejecting the information transaction a. We conclude that agent u rejects a at wealth level w.

Now, assume that $\underline{R}(u) < r$ and choose r_0 such that $\underline{R}(u) < r_0 < r$. Since an agent u_r^C accepts a at any wealth level, an agent $u_{r_0}^C$ strictly prefers accepting a at wealth level 0, which can be expressed as:

$$
1 - \sum_s p_\alpha(s) \sup_{b^s \in B^*} \sum_k q_k^s \exp(r_0(\mu + b_k^s)) > 0.
$$

Let $(b^s)_s$ then be a family of elements in B^* such that:

$$
1 - \sum_s p_\alpha(s) \sum_k q_k^s \exp(r_0(\mu + b_k^s)) > 0.
$$

46

Let w be such that $\rho(w - \mu + \min_{s,k} q_k^s) < r_0$. We have $\rho(z) < r_0$ for every $z \geq w - \mu + \min_{s,k} q_k^s$. It follows that by the same computation as in the proof of Theorem 2, for every s, k:

$$\frac{u(w - \mu + b_k^s) - u(w)}{u'(w)} \geq \frac{u_{r_0}^C(-\mu + b_k^s) - u_{r_0}^C(0)}{u_{r_0}^{C\prime}(0)}.$$

Therefore:

$$
\begin{aligned}
\frac{\sum_s p_\alpha(s) V(u, w - \mu, q_\alpha^s) - u(w)}{u'(w)} &\geq \frac{\sum_s p_\alpha(s) \sum_k q_k^s u(w - \mu + b_k^s) - u(w)}{u'(w)} \\
&\geq \frac{\sum_s p_\alpha(s) \sum_k q_k^s u_{r_0}^C(-\mu + b_k^s) - u_{r_0}^C(0)}{u_{r_0}^{C\prime}(0)} \\
&> 0.
\end{aligned}
$$

Hence, u accepts a at wealth w. ∎

We now proceed to prove Theorem 5:

Proof of Theorem 5. Assume that a_1 uniformly wealth-dominates a_2. For every $\varepsilon > 0$, Lemma 7 shows that an agent $u_{\mathcal{A}(a_1)+\varepsilon}^C$ rejects a_1 at all wealth levels. Hence such an agent also rejects a_2 at all wealth levels, which implies, again by Lemma 7, that $\mathcal{A}(a_1) + \varepsilon \geq \mathcal{A}(a_2)$. Since this is true for every $\varepsilon > 0$, it follows that $\mathcal{A}(a_1) \geq \mathcal{A}(a_2)$.

For the converse, assume that $\mathcal{A}(a_1) \geq \mathcal{A}(a_2)$, and that $u \in \mathcal{U}_{DA}$ rejects a_1 at all wealth levels. Then by Lemma 7, $\underline{R}(u) \geq \mathcal{A}(a_1) \geq \mathcal{A}(a_2)$, and u also rejects a_2 at all wealth levels. ∎

9.6 Proof of Proposition 1

Proof. Assuming that α_1 is more informative than α_2 in the sense of Blackwell, and fixing any arbitrary wealth level w, then any CARA agent who rejects (μ, α_1) at wealth level w also rejects (μ, α_2) at wealth level w. It follows from the characterization of $\mathcal{A}$ in Theorem 3 that $\mathcal{A}(\mu, \alpha_1) \geq \mathcal{A}(\mu, \alpha_2)$.

∎

9.7 Proof of Proposition 2

Proof. Fix any wealth level. From Theorem 3, a CARA agent with coefficient of risk aversion $\mathcal{A}(\mu, \alpha_2)$ accepts both transactions (μ, α_1) and (μ, α_2) at wealth w; this agent therefore also accepts the transaction $(\mu, \lambda\alpha_1 \oplus (1-\lambda)\alpha_2)$ at that wealth level. This shows that

$$\mathcal{A}(\mu, \lambda\alpha_1 \oplus (1 - \lambda)\alpha_2) \geq \mathcal{A}(\mu, \alpha_2).$$

Now consider $\varepsilon > 0$. Again from Theorem 3, a CARA agent with coefficient of risk aversion $\mathcal{A}(\mu, \alpha_1) + \varepsilon$ rejects both transactions (μ, α_1) and (μ, α_2) at wealth w; this agent therefore also rejects the transaction $(\mu, \lambda\alpha_1 \oplus (1-\lambda)\alpha_2)$ at that wealth level. This shows that

$$\mathcal{A}(\mu, \alpha_1) + \varepsilon \geq \mathcal{A}(\mu, \lambda\alpha_1 \oplus (1 - \lambda)\alpha_2)$$

for every $\varepsilon > 0$, and hence that

$$\mathcal{A}(\mu, \alpha_1) \geq \mathcal{A}(\mu, \lambda\alpha_1 \oplus (1 - \lambda)\alpha_2).$$

∎

9.8 Proofs of Theorems 6 and 7

Proof. We first prove Theorem 7. Agent u_C^r accepts an information transaction $a = (\mu, \alpha)$ at prior q if and only if:

$$\sum_s p_\alpha(s) V(u, w - \mu, q_\alpha^s) \geq V(u, w, q).$$

Once we use the expression of Lemma 6, this becomes equivalent to:

$$\exp(-d(p\|q) \geq \exp(r\mu) \sum_\alpha p_\alpha(s) \exp(-d(p\|q_\alpha^{)}),$$

which is in turn equivalent to

$$r \leq \mathcal{A}(a, q).$$

Theorem 6 follows from Theorem 7 just as Theorem 1 follows from Theorem 3. ∎

9.9 Proof of Proposition 3

Proof. We prove the proposition using the following auxiliary decision problem. In the first stage, the agent can either accept information transaction a or reject it. If the agent accepts a, then a signal s is drawn from α and the agent can either accept the information transaction b_s or reject it. If the agent rejects a, no other information transaction is offered to the agent. Once the agent has acquired some information (or none), any asset in B^* may be purchased; then the state k is realized, and the agent receives the corresponding payoff.

Assume that for every s, $\mathcal{A}(b_s, q_s) \geq \mathcal{A}(a)$, and consider an agent $u_C^{\mathcal{A}(a)}$ at any wealth level and any prior p. In the sequential decision problem, assuming that a is accepted in the first stage by this agent, then b_s is accepted in the second stage for every s. Also, a is accepted in the first stage even if the option of acquiring b_s in the second stage is absent. Therefore, a is also accepted with the option of acquiring b_s in the second stage. Hence, an optimal strategy for the agent is to accept a, and then accept b_s no matter what s is. In particular, this strategy is better for the agent than not acquiring any information transaction. This shows that the agent accepts $a + b$, and hence that $\mathcal{A}(a + b) \geq \mathcal{A}(a)$.

Now assume that for every s, $\mathcal{A}(b_s, q_s) \leq \mathcal{A}(a)$, and consider an agent u_C^ρ with $\rho > \mathcal{A}(a)$ at any wealth level and any prior p. In the sequential decision problem, assuming that a is accepted in the first stage, it is optimal for this agent to reject b_s after every signal s. Hence, the decision to acquire a in the sequential decision problem is equivalent to the decision to acquire a alone, and so this agent rejects a. Hence, the optimal strategy for the agent is to reject information. In particular, not acquiring any information is better than acquiring a, which is itself better than acquiring a and b_s following every s, so that no information is better than $a + b$. Therefore, the agent rejects $a + b$, which shows that $\mathcal{A}(a + b) < \rho$ for every $\rho > \mathcal{A}(a)$. This implies that $\mathcal{A}(a + b) \leq \mathcal{A}(a)$.

The third point follows immediately from the first and second points. ∎

References

ATHEY, S., AND J. LEVIN (2001): "The Value of Information in Monotone Decision Problems," tanford University, Department of Economics Working Papers 01003.

AUMANN, R. J., AND R. SERRANO (2008): "An Economic Index of Riskiness," *Journal of Political Economy*, 116, 810–836.

BLACKWELL, D. (1953): "Equivalent Comparison of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

BLUME, L., AND D. EASLEY (1992): "Evolution and Market Behavior," *Journal of Economic Theory*, 58, 9–40.

CABRALES, A., O. GOSSNER, AND R. SERRANO (2012): "Entropy and the Value of Information for investors," *American Economic Review*.

DONALDSON-MATASCI, M. C., C. T. BERGSTROM, AND M. LACHMANN (2010): "The Fitness Value of Information," *Oikos*, 119, 219–230.

GANUZA, J.-J., AND J. PENALVA (2010): "Signal Orderings Based on Dispersion and the Supply of Private Information in Auctions," *Econometrica*, 78(3), 1007–1030.

GOSSNER, O. (2011): "Simple Bounds on the Value of a Reputation," *Econometrica*, 79.

GOURIEROUX, C., A. MONFORT, AND E. RENAULT (1987): "Kullback Causality Measures," *Annales d'Économie et de Statistique*, 6/7, 370–410.

HANSEN, L., AND T. SARGENT (2001): "Robust Control and Model Uncertainty," *American Economic Review*, 91, 60–66.

HANSEN, L. P., AND T. J. SARGENT (2010): "Wanting Robustness in Macroeconomics," in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3, pp. 1097 – 1157. Elsevier.

HART, S. (2011): "Comparing Risks by Acceptance and Rejection," *Journal of Political Economy*, 119, 617–638.

HINTON, G. E., AND R. S. ZEMEL (1994): "Autoencoders, minimum description length and free energy," in *Advances in neural information processing*, ed. by J. Cowan, G. Tesauro, and J. Alspector, vol. 6. Morgan Kaufmann.

JEWITT, I. (2007): "Information Order in Decision and Agency Problems," Nuffield College.

JUSTESON, J. S. (1973): "Limitations of Archaeological Inference: An Information-Theoretic Approach with Applications in Methodology," *American Antiquity*, 80, 131–149.

KRAFT, L. G. (1949): "A Device for Quantizing, Grouping, and Coding

Amplitude Modulated Pulses," MS Thesis, Electrical Engineering Department, Massachusetts Institute of Technology.

Kullback, S., and R. A. Leibler (1951): "On Information and Sufficiency," *Annals of Mathematical Statistics*, 22, 79–86.

Kuperman, V., R. Bertram, and R. H. Baayen (2010): "Processing Trade-offs in the Reading of Dutch Derived Words," *Journal of Memory and Language*, 62, 83–97.

Landau, L., and E. Lifshitz (1980): *Statistical Physics*, vol. 5 of *Course of Theoretical Physics*. Butterworth-Heinemann, third edn.

Lehmann, E. L. (1988): "Comparing Location Experiments," *The Annals of Statistics*, 16(2), 521–533.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006): "Ambiguity Aversion, Robustness, and the Variational Representation of Preferences," *Econometrica*, 74, 1447–1498.

McMillan, B. (1956): "Two Inequalities Implied by Unique Decipherability," *IEEE Trans. Information Theory*, 2, 115–116.

Mishra, T., and S. Bangalore (2011): "Predicting Relative Prominence in Noun-Noun Compounds," *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers*, 49, 609–613.

ONG, L.-L., W. XIAOY, C.-K. THAM, AND M. H. ANG (2009): "Resource Constrained Particle Filtering for Real-time Multi-Target Tracking in Sensor Networks," *PERCOM '09 Proceedings of the 2009 IEEE International Conference on Pervasive Computing and Communications*, 09, 1–6.

PERSICO, N. (2000): "Information Acquisition in Auctions," *Econometrica*, 68, 135–148.

SANDRONI, A. (2000): "Do Markets Favor Agents Able to Make Accurate Predictions?," *Econometrica*, 68.

SHAKED, M., AND J. G. SHANTHIKUMAR (2007): *Stochastic Orders.* Springer, Berlin.

SHERWIN, W. B. (2010): "Entropy and Information Approaches to Genetic Diversity and its Expression: Genomic Geography," *Entropy*, 12, 1765–1798.

SHORE, J. E., AND R. W. JOHNSON (1980): "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," *IEEE Transactions on Information Theory*, 26, 26–37.

SINGH, V. P. (1997): "The use of Entropy in Hydrology and Water Resources," *Hydrological Processes*, 11, 587–626.

SOOFI, E. S., AND J. J. RETZER (2002): "Information Indices: Unification and Applications," *Journal of Econometrics*, 107, 17–40.

VELDKAMP, L. (2011): *Information Choice in Macroeconomics and Finance.* Princeton University Press, Princeton, NJ.

WEYL, G. E. (2007): "A Simple Theory of Scientific Learning," SSRN: http://ssrn.com/abstract=1324410.