A Note on Complementary and Choice of Information

Hershdeep Chopra [*]

October 11, 2025

1 Introduction

In decision-making, individuals often allocate limited resources to gather information, while anticipating the arrival of future information. This note how the anticipation of future data influences decision makers' immediate incentives to acquire information. The decision maker seeks to learn an uncertain state using two signals: the anticipated future information and the signal chosen today. These dynamics manifest in various contexts, such as:

1. An investor deciding on resource allocation among different economic sectors while awaiting an upcoming Federal Reserve announcement.

2. A doctor chooses between diagnostic tests while still waiting for results from a previously conducted test.

This note aims to highlight some interesting properties associated with the incentives to acquire information in such choice problems and how it is related to the notion of complementarity between information. By restricting the choice of information between Blackwell equivalent signals, we isolate the effect of anticipated information on the present-day incentive to acquire information.

The value of a signal today depends on all the information a decision maker expects to learn before making a decision in the future. Complementary information can be described by the positive dependence of the value of the signal and the presence of additional information. In order to study this dependence, we need an appropriate concept for the value of information while anticipating future information. We provide a new definition of complementarity based on this, in terms of the joint correlation structure of the signals and the uncertain state. We show how existing notions of complementarity of information might overlook information that otherwise (and intuitively) seems complementary. Most of the paper considers comparisons that are robust to preferences; thus, the results and insights apply to many different economic environments.

2 Literature Review

Some previous works on the complementarity of information and the effect of anticipated information on present-day incentives:

Complementarity between signals has been studied in Börgers et al. (2013). They study how having a signal improves the marginal value of having another signal. Similar to our setting, they also consider choices that are robust to the preferences and beliefs of the decision-maker. We provide a new definition of complementarity that is neither implied by nor implies the definition in Börgers et al. (2013) (henceforth

[*]hershdeepchopra2026@u.northwestern.edu

referred to as BHK13). We discuss this in detail in section 6.1. Liang and Mu (2020) extends the BHK13's notion of complementarity to a dynamic environment. They study how access to a signal improves the marginal value of having access to another signal.

Our temporal choice of information resembles the setting of Brooks et al. (2022). They also consider the choice between signals in the presence of an additional signal. Their setting is robust to the preferences and beliefs of the decision-maker as well as the nature of additional information that the DM possesses (or will possess). In contrast, we fix the structure of the decision maker's additional information; this allows us to study complementarity, which otherwise would not be possible in the model of Brooks et al. (2022).

3 Motivation

In this section, we provide motivation for the type of information acquisition problems considered in this paper.

3.1 Medical Testing

Let the uncertain state represent presence (P) or absence (A) of a certain disease ($\mathcal{X} = \{P, A\}$). There are two decision makers Bob and Kevin. Bob has no history of the disease and assigns a prior $q_B(P) = \frac{1}{2}$. On the other hand, Kevin has a family history of the disease and assigns prior $q_K(P) = \alpha \in (\frac{1}{2}, 1)$. Consider the situation when Bob gets a free medical screening Y, whereas Kevin does not get this screening. The outcome of that screening is $Y = y$ and it increases Bob's belief about the likelihood of him having the disease from $\frac{1}{2}$ to α. Additionally, Bob and Kevin have the choice to get tested for the disease using a test Z at some cost $c > 0$.

The outcome $Y = y$ increases the value of the additional signal Z if Bob is willing to pay at least as much as Kevin for the test Z. For example, if the medical screening Y detects the disease for Bob and there is a test Z available such that whenever Y incorrectly detects the disease, the test Z is very unlikely to also incorrectly detect the disease. Then Bob is willing to pay more for the test Z compared to Kevin who has the same beliefs as Bob about the likelihood of having the disease. Bob values the signal Z more due to its correlation with the signal Y.

We say that the screening Y strongly complements Z if for all outcomes of Y, the value of test Z for Bob is greater than the value of test Z for Kenvin. We say that the screening Y complements the test Z if on average a DM who observes the realization of Y values Z more than another DM who doesn't observe Y but has the same beliefs about the state as if he saw Y. A consequence of this is that if Y is a cheap screening and Z is an expensive medical test, then complementarity would imply providing free screenings Y increases the demand for the more expensive medical test Z, leading to higher profit. These ideas are formalized in example 1.

3.2 Delay and Investment

Ann is an investment manager who is interested in the optimal portfolio X. She is waiting for an important announcement from the Federal Reserve about interest rates, (Y). She is deciding whether to buy a detailed financial report (Z) today or delay the purchase until after the announcement. Having access to the report before the announcement lets her react fast to the news. If she delays, she might miss out on some good investment chances because markets can change fast after such announcements. Hence, she incurs some delay costs. The report is priced today based on the market's expectations of the announcement, but if she waits until after the announcement, the report is priced according to the market's reaction to the announcement.

So, Ann has two options. She could buy the report today, expecting it to be more valuable and thus more expensive after the announcement. Or choose to delay the purchase, if she thinks the delay cost is smaller relative to the flexibility she gains by making her purchase decision contingent on the announcement.

In making her decision, Ann will think about the losses she could face. If she buys today, her loss is the cost of the report. If she waits, her loss could be the higher price for the report plus any good investment chances she missed. The announcement is a substitute for the report if no matter what the announcement turns out to be, the market price of the report is greater than her willingness to pay for it. In this case, even for arbitrarily small delay costs, Ann is willing to acquire the information today. We come back to this in section 9.1.

4 Model

Fix a probability space $(\Omega, \mathcal{F}, Pr)$. On this space define $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}_1$ $\mathcal{Z}_2$ valued random variables X, Y, Z_1, Z_2 respectively[1]. We refer to Y, Z_1, Z_2 as signals. We will use $\Delta(A)$ to represent the space of all probability distributions on some finite set A. The decision maker's preferences are represented by bounded positive $u : \mathcal{X} \times \mathcal{D} \to [0, \infty)$, where $d \in \mathcal{D}$ represents a decision[2]. In particular, we assume $\arg\max_{d \in \mathcal{D}} u(x, d)$ exists for each $x \in \mathcal{X}$. The choice problem is described by the following sequence (figure 1)

1. Choice between random variables Z_1 and Z_2.
2. If Z_i was chosen then observe realization of bivariate random variable (Y, Z_i).
3. Let[3] $\mathcal{D}$ be a finite decision space, after observing $(Y = y,\ Z_i = z)$ the decision maker chooses $d \in \arg\max_{\delta \in \mathcal{D}} \mathbb{E}[u(X, \delta) | Y = y,\ Z_i = z]$.

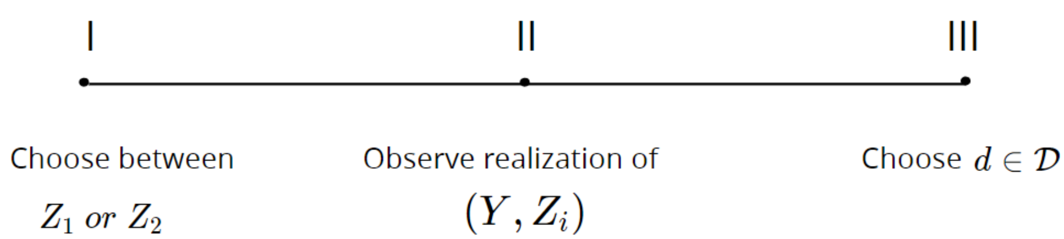

Figure 1: Timing

Let p represent the joint pmf[4] of (X, Y, Z_1, Z_2). We refer to the marginal distribution of X, p_X as the prior. In later parts, we use q to represent this prior. A decision problem is defined by the triple $(p_X, u, \mathcal{D})$. Often we will suppress dependence on the decision space and write decision problems in terms of (p_X, u). With finite $\mathcal{X}$, most results hold if we restrict attention[5] to the uniform prior $p_X = \frac{1}{|\mathcal{X}|}$. We define the value of a collection of signals relative to a decision problem, formally value function assigns a real number to the family of conditional pmf $\{p(.|X = x) \in \Delta(\mathcal{S}_1 \times ... \times \mathcal{S}_k); x \in \mathcal{X}\}$ for each decision problem $(q, u, \mathcal{D})$.

[1] We assume $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}_i$ are finite subset of $\mathbb{R}$ (unless stated otherwise).

[2] We assume the decision space $\mathcal{D}$ is convex and compact. Without loss of generality, we assume that this space only contains undominated decisions.

[3] We restrict attention to pure strategies for clarity.

[4] We restrict attention to joint distributions where each marginal has full support.

[5] We use Blackwell's notion of comparison in Blackwell (1953). It is shown that a signal S is more informative than S' if and only if under the uniform prior, the distribution of posteriors generated by S is a mean preserving spread of the distribution of posteriors generated by S'.

Definition 1. For the decision problem $(q, u, \mathcal{D})$ the ex-ante value of a collection of random variables $(S_1, ..., S_k)$ is given by

$$V_{q,u}(S_1, ..., S_k) := \mathbb{E}_p \left[\max_{\{d \in \mathcal{D}\}} \mathbb{E}_p[u(X, d)|S_1, ..., S_k] \right] - \max_{\{d \in \mathcal{D}\}} \mathbb{E}_p[u(X, d)]$$

Where $p \in \Delta(\mathcal{X} \times \mathcal{S}_1 \times ... \times \mathcal{S}_k)$ represents the joint distribution of $(X, S_1, ..., S_k)$, with the marginal probability $p_X = q$ being determined by the decision problem.

The value can be interpreted as the maximum amount a decision maker is willing to pay to learn the realization of $S_1, .., S_k$.

Definition 2. A collection $S_1, S_2, ..., S_k$ of random variables is more valuable than another collection $S'_1, S'_2, ..., S'_m$ if for any decision problem (q, u) we have:

$$V_{q,u}(S_1, ..., S_k) \geq V_{q,u}(S'_1, ..., S'_m)$$

The ordering of being more valuable in the above sense is from Blackwell (1953). In particular, it is shown that for $S_i : \Omega \to \mathcal{S}_i$ and $S'_i : \Omega \to \mathcal{S}'_i$. The collection $S_1, S_2, ..., S_k$ is more valuable than $S'_1, S'_2, ..., S'_m$ if and only if there is a mapping $M : S_1 \times ..., \times S_k \to \Delta(S'_1 \times S'_2 \times ... \times S'_m)$ such that for all $x \in \mathcal{X}$ we have $\sum_{S_1 \times ..., \times S_k} Pr(s_1, ..., s_k|X = x)M(s'_1, ..., s'_m|s_1, ..., s_k) = Pr(s'_1, ..., s'_m|X = x)$. We will write this as $(S_1, S_2, ..., S_k) \succeq_B (S_1, S_2, ..., S_m)$, and refer to this partial order as the Blackwell order. There are many characterizations and generalizations of this order [6].

In what follows we will restrict the choice between signals Z_1, Z_2 that are Blackwell equivalent ($Z_1 \sim_B Z_2$). We make this restriction to isolate the effect of the signal Y on the choice between Z_1 and Z_2. As Blackwell equivalence will imply that a decision maker with arbitrary utility (only dependent on X and d) and prior (i.e. the marginal probability of X, p_X) will have the same value $V_{p_X,u}(Z_1) = V_{p_X,u}(Z_2)$. The indifference in value of Z_1 and Z_2 in the absence of Y implies that the preference of signal Z_i over Z_{-i} is due to anticipation of learning Y in the future. Assuming that $Z_1 \sim_B Z_2$ is the same as assuming equality of distribution of posteriors induced by signals Z_1 and Z_2. Formally consider the random variables X, S_1, S_2. Define the following:

$$t_1 : \mathcal{S}_1 \to \Delta(\mathcal{X});\ s \mapsto \left\{ \frac{Pr(S_1 = s|X = x)}{\sum_{\mathcal{X}} Pr(S_1 = s|X = x')}; x \in \mathcal{X} \right\}$$

$$t_2 : \mathcal{S}_2 \to \Delta(\mathcal{X});\ s \mapsto \left\{ \frac{Pr(S_2 = s|X = x)}{\sum_{\mathcal{X}} Pr(S_2 = s|X = x')}; x \in \mathcal{X} \right\}$$

Then the posterior belief on X (assuming uniform prior) after observing signals S_1 and S_2 are given by random variables $t_1(S_1)$ and $t_2(S_2)$. From theorem 25.4 and corollary 25.5 in Strasser (1985)[7] we have the following conclusion:

Fact 1. *The following are equivalent:*

1. $S_1 \sim_B S_2$
2. $t_1(S_1) =_d t_2(S_2)$

The following lemma will be useful in later sections

Lemma 1. *For random variable S_1, S_2 and full support prior $q \in \Delta(\mathcal{X})$ if $V_{q,u}(S_1) \geq V_{q,u}(S_2)$ for all bounded positive utility u then $S_1 \succeq_B S_2$.*

[6]The equivalence stated above relates having less value of information to being reproducible from the more valuable signals by some randomization procedure. For other characterizations see Blackwell and Girshick (1979),Strasser (1985)

[7]The result was originally proved in Blackwell (1951).

Proof. The result follows from noting that transforming the utility function state-wise has the same effect as re-weighting the prior belief. Fix two full support priors $q, q' \in \Delta(\mathcal{X})$. Define a map γ from the set of positive bounded utilities to itself such that $\gamma(u) : (x,d) \mapsto \frac{q(x)}{q'(x)} u(x,d)$. It is easy to check that γ is a bijection. Now, note that for any strategy $\varphi_i : \mathcal{S}_i \to \mathcal{D}$ and utility $u : \mathcal{X} \times \mathcal{D} \to [0,\infty)$ we have

$$\sum_{\mathcal{S}_i} \sum_{x \in \mathcal{X}} u(x, \varphi_i(s)) Pr(S_i = s|X = x) q(x) = \sum_{\mathcal{S}_i} \sum_{x \in \mathcal{X}} \gamma(u(x, \varphi_i(s))) Pr(S_i = s|X = x) q'(x)$$

Thus $V_{q,u}(S_1) \geq V_{q,u}(S_2)$ for all u implies that $V_{q',\gamma(u)}(S_1) \geq V_{q',\gamma(u)}(S_2)$ for all u. As gamma is a bijection we get that $V_{q',u}(S_1) \geq V_{q',u}(S_2)$ for all u. This holds for any full support prior q'.

When q' dosen't have full support on $\Delta(\mathcal{X})$, we define γ' from the set of positive bounded utilities to itelf such that $\gamma'(u) : (x,d) \mapsto \frac{q(x)}{q'(x)} u(x,d)$ when $x \in \text{supp}(q')$ and $\gamma'(u) : (x,d) \mapsto 0$ when $x \notin \text{supp}(q')$. Thus γ' is a bijection from the set of positive bounded utilities that are zero outside support of q' to itself. Now, note that for any strategy $\varphi_i : \mathcal{S}_i \to \mathcal{D}$ and utility $u : \mathcal{X} \times \mathcal{D} \to [0,\infty)$ such that $u(x,.) = 0$ whenever $x \notin \text{supp}(q')$ we have

$$\sum_{\mathcal{S}_i} \sum_{x \in \mathcal{X}} u(x, \varphi_i(s)) Pr(S_i = s|X = x) q(x) = \sum_{\mathcal{S}_i} \sum_{x \in \mathcal{X}} \gamma'(u(x, \varphi_i(s))) Pr(S_i = s|X = x) q'(x)$$

Thus $V_{q,u}(S_1) \geq V_{q,u}(S_2)$ for all u implies that $V_{q',\gamma'(u)}(S_1) \geq V_{q',\gamma'(u)}(S_2)$ for all u which are zero outside the support of q'. As γ' is bijective on the restricted set of utilities this implies $V_{q',u}(S_1) \geq V_{q',u}(S_2)$ for all u which are zero outside the support of q'. As the decisions (and hence the value of a signal) don't depend on the states outside the support of the prior we have $V_{q',u}(S_1) \geq V_{q',u}(S_2)$ for all u. This concludes the proof.

□

Note that the above proof extends to the case when the comparison is restricted to some class of utilities $\mathcal{U}$, as long as the bijection γ and γ' can be appropriately defined for $\mathcal{U}$. In particular the proof extends to the class of IDO preferences defined in 8.

5 Definitions

Based on the choice problem described in section 4 Z_i is chosen over another signal Z_{-i} if the value of having signals (Y, Z_i) is more than the value of having joint signal (Y, Z_{-i}). Let $\mathcal{U}$ represent some arbitrary class of bounded positive utility $u : \mathcal{X} \times \mathcal{D} \to [0,\infty)$. Let $p_X \in \Delta(\mathcal{X})$ be the marginal distribution of X. A $\mathcal{U}-$decision problem is a triple $(p_X, u, \mathcal{D})$ where utility $u \in \mathcal{U}$. Whenever we refer to the set of all possible bounded utilities, we drop the dependence on $\mathcal{U}$. For the following definitions fix the conditional distribution[8] $p_{|X} \in \Delta(\mathcal{Y} \times \mathcal{Z})$ of Y, Z given X. Then the joint pmf $p \in \Delta(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ for prior q is given by $p(x,y,z) = q(x) p(y,z|X = x)$.

Definition 3. *Y ($\mathcal{U}$ -) complements Z_i more than Z_j if for all ($\mathcal{U}-$) decision problem the ex-ante value of signals (Y, Z_i) is greater than ex-ante value of signals (Y, Z_j). Formally for all decision problems (q, u)*

$$V_{q,u}(Y, Z_i) \geq V_{q,u}(Y, Z_j)$$

Written as $Z_i \succeq_Y (\succeq_Y^{\mathcal{U}}) Z_j$.

Remark 1. The $Z_i \succeq_Y Z_j$ if and only if $(Y, Z_i) \succeq_B (Y, Z_j)$.

[8]For any y, z, $p(y,z|X) = \mathbb{E}[1_{y,z}(Y,Z)|X]$

Two signals complement (substitute) each other if the presence of one signal increases (decreases) the incentive to acquire another signal. In order to study complementarity between information due to anticipation of future information we need to define an appropriate notion for value of information when the decision maker anticipates to learn another piece of information before making his decision. Once the decision maker learns some signal Y his beliefs about the state X change[9]. Thus in order to define the value of signal Z we need a reference point, that accounts for this change in beliefs, against which we measure the change in utility. To this end we require that the change in the value of the second signal be determined with respect to the posterior induced by the realization of the first signal. Formally define the following:

For random variables X, Y, Z with the joint pmf $p \in \Delta(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$, where the marginal probability of X is given by $p_X = q$. Given $y \in \text{supp}(Y)$, let $p_y \in \Delta(X)$ be the pmf of X conditional on $Y = y$, i.e. $p_y(x) = Pr(X = x|Y = y) = \frac{p(y|x)q(x)}{p(y)}$. Define a random variable Z_y such that the joint distribution $r \in \Delta(\mathcal{X} \times \mathcal{Z})$ of (X, Z_y) is given by $r(x,z) = Pr(Z = z, X = x|Y = y) = p_y(x)p(z|x,y)$. Then the value of signal Z after observing the realization $Y = y$ is given by (definition 1):

$$V_{p_y,u}(Z_y) = \sum_{\mathcal{Z}} \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)r(x|z)r(z) - \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)q(x)$$

$$= \sum_{\mathcal{Z}} \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)p(z|x,y)p(x|y) - \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)p(x|y)$$

Similarly, the value of signal Z can be considered without taking into account the effect of observing $Y = y$ on the joint distribution of (X, Z). To this end, we consider the ex-ante value of signal Z when the prior (marginal distribution of X) is given by $p(.|Y = y)$ i.e. the distribution of X after observing $Y = y$. Define the following joint pmf $m \in \Delta(\mathcal{X} \times \mathcal{Z})$ of (X, Z) by $m(x,z) = Pr(X = x|Y = y)Pr(Z = z|X = x) = p_y(x)p(z|x)$. Following definition 1:

$$V_{p_y,u}(Z) = \sum_{\mathcal{Z}} \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)m(x|z)m(z) - \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)m(x)$$

$$= \sum_{\mathcal{Z}} \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)p(z|x)p(x|y) - \max_{\{d \in \mathcal{D}\}} \sum_{x \in \mathcal{X}} u(x,d)p(x|y)$$

This represents the value of the signal Z to a decision maker who doesn't observe the realization of Y, but whose prior belief about X is as if the decision maker saw the realization $Y = y$. The change in the value of signal Z after observing the realization $Y = y$ relative to the value of signal Z for a decision maker with the same beliefs about X is given by:

$$I_{q,u}(y; Y, Z) = V_{p_y,u}(Z_y) - V_{p_y,u}(Z)$$

Using the ideas developed above, we describe two notions of complementarity between signals. First, complementarity in a strong sense would require that no matter what realization of a signal is observed the decision maker's value of the second signal increases.

Definition 4. For two random variables Y, Z. Y is $[\mathcal{U}-]$ strong complement (strongly substitutes) of Z if for all $y \in \text{supp}(Y)$, prior on $q \in \Delta(\mathcal{X})$, and utility u $[\mathcal{U}-]$ we have $I_{q,u}(y; Y, Z) \geq (\leq) 0$.

Remark 2. The above condition describes complementarity in a strong sense. Knowing any realization of one of the signals makes the other signal more valuable relative to a decision maker with the same beliefs about the state.

Our definition of strong complements requires the inequality to hold pointwise for all realizations of the signal Y, this can be seen as an interim notion of complementarity. As we are interested in ex-ante choices, we require an ex-ante notion of complements. To this end, we can relax the pointwise inequality in definition 4 to an average.

[9]Similar ways of defining the value of information are discussed in Degroot (1994)

Definition 5. Two random variables Y, Z are said to be $[\mathcal{U}-]$ complements (substitutes) if for all $y \in \text{supp}(Y)$, prior on $p_X \in \Delta(\mathcal{X})$, and utility u $[\in \mathcal{U}]$ we have $\sum_{\mathcal{Y}} I_{q,u}(y; Y, Z)p(y) \geq (\leq)0$.

An obvious conclusion of the definition is the following observation:

Observation 1. For two random variables (Y, Z), Y is $[\mathcal{U}-]$ strong complement (substitute) of Z then (Y, Z) are $[\mathcal{U}-]$ complements (substitutes).

5.1 Binary example

Example 1. (Binary state and signals)

To fix the ideas, consider an example with binary state and signals. Let $\mathcal{X} = \{P, A\}$, $\mathcal{Y} = \{0, 1\}$, $\mathcal{Z} = \{0, 1\}$. The state X represents the presence (P) or absence (A) of a disease. The variables Y and Z_i represent diagnostic tests that detect the presence (1) or absence (0) of the disease. For simplicity, we assume that Y and Z_i have the same true positive and true negative rate p (with $p \geq \frac{1}{2}$). The decision maker is expected to learn the results of the test Y at some future date. While awaiting the results of Y the decision makers choose to conduct another test from Z_1, Z_2 or $Z_\varnothing$. The distribution of the joint experiments (Y, Z_i) is shown in figure 2.

$$\begin{array}{c} \begin{array}{cc} & \begin{array}{cccc} 11 & 10 & 01 & 00 \end{array} \\ \begin{array}{c} P \\ A \end{array} & \begin{pmatrix} 2p-1 & 1-p & 1-p & 0 \\ 0 & 1-p & 1-p & 2p-1 \end{pmatrix} \end{array} \\ (Y, Z_1) \end{array} \quad \begin{array}{c} \begin{array}{cc} & \begin{array}{cccc} 11 & 10 & 01 & 00 \end{array} \\ \begin{array}{c} P \\ A \end{array} & \begin{pmatrix} p^2 & p(1-p) & p(1-p) & (1-p)^2 \\ (1-p)^2 & p(1-p) & p(1-p) & p^2 \end{pmatrix} \end{array} \\ (Y, Z_\varnothing) \end{array} \quad \begin{array}{c} \begin{array}{cc} & \begin{array}{cccc} 11 & 10 & 01 & 00 \end{array} \\ \begin{array}{c} P \\ A \end{array} & \begin{pmatrix} p & 0 & 0 & 1-p \\ 1-p & 0 & 0 & p \end{pmatrix} \end{array} \\ (Y, Z_2) \end{array}$$

Figure 2: Binary example

It is easy to check that $Z_1 \succeq_Y Z_\varnothing \succeq_Y Z_2$, by constructing appropriate garbling matrices[10]. Intuitively, Z_2 represents a test that will always give the same result as test Y, as the information in Z_2 is redundant it has no value if Y is known. Z_1 represents a test, such that both Z_1 and Y can never have false positives (and false negatives) simultaneously. In this sense, the test Y adds complementary information to Z_1 which is not present in the tests individually. Finally, the tests $(Y, Z_\varnothing)$ can be interpreted as the same test being repeated twice and each time has (unconditionally) independent diagnostic errors.

Remark 3. From the Blackwell ordering we can conclude for any prior q and utility u the following inequalities hold: for any $y \in \mathcal{Y}$ we have $I_{q,u}(y; Y, Z_1) \geq I_{q,u}(y; Y, Z_\varnothing) = 0 \geq I_{q,u}(y; Y, Z_2)$. In fact, the inequalities can be strict, to see this consider the decision problem with a uniform prior $q = (1/2,\ 1/2)$ on $\mathcal{X} = \{P, A\}$ and utility given by $u(P, d_1) = u$ and $u(A, d_2) = 1 - u$ and zero otherwise. Let $p = \frac{2}{3}$ and $u = \frac{2}{5}$. For $Y = y = 1$ the posterior belief on X is given by $(p, 1-p)$. Then

$$I_{q,u}(1; Y, Z_1) = \frac{2}{3}u + \frac{1}{3}(2p-1)(1-u) - \frac{2}{3}pu - \frac{1}{3}p(1-u) = \frac{1}{3}(1-p)(3u-1) > 0$$

and

$$I_{q,u}(1; Y, Z_2) = \frac{2}{3}u - \frac{2}{3}pu - \frac{1}{3}p(1-u) = \frac{-2}{45} < 0$$

[10]The garbeling matrices $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ & $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & p & 1-p \\ 0 & 0 & 0 & 1 \end{pmatrix}$ suffice

Remark 4. For the bivariate signal (Y, Z_1) we observe that the conditional distributions $p(Y, Z|X = P)$ and $P(Y, Z|X = A)$ don't have the same support. This is not necessary for the ordering, to see this consider $p = \frac{2}{3}$ and observe that

$$\begin{pmatrix} \frac{2}{5} & \frac{4}{15} & \frac{4}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{4}{15} & \frac{4}{15} & \frac{2}{5} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{6} & \frac{5}{6} & 0 & 0 \\ 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{4}{9} & \frac{2}{9} & \frac{2}{9} & \frac{1}{9} \\ \frac{1}{9} & \frac{2}{9} & \frac{2}{9} & \frac{4}{9} \end{pmatrix}$$

Later we use Z_1 to discuss our definitions in the context of Börgers et al. (2013), If we replace Y, Z_1 with the above example the main insights are the same.

Remark 5. From Z_2 we see that in some situations it's possible to have signals that provide no additional information even when individually the signal is informative. More specifically when is it possible that for a bivariate signal $(Y \sim_B (Y, Z))$ or $(Z \sim_B (Y, Z))$? Our proposition 3 shows that this is possible only if $Y \succeq_B Z$ or $Z \succeq_B Y$.

Similar to the minimal signal discussed above, another extreme is bivariate signal (Y, Z) such that $(Y, Z) \sim_B X$, meaning that the combined signal (Y, Z) is fully informative. Apart from the trivial case when one of the signals Y or Z is fully informative, we characterize these maximal signals in remark **??**.

Remark 6. If Y and Z are conditionally independent given X, then knowing one doesn't improve understanding or change beliefs about the other once the state X is fixed. The signal $(Y, Z_\varnothing)$ represents this situation.

This example fits in with the intuition that positive correlation decreases and negative correlation increases the informativeness of the joint signal. A natural question is whether this intuition holds more generally. As we show in example 2 this is in fact false.

5.2 Linear Normal Model

Example 2. (Linear normal signals with unknown mean)[11]

This example will look at linear normal experiments. Let X represent the unknown mean of normally distributed random variables Y and Z with variance σ_Y^2 and σ_Z^2 respectively. Let the joint signal $(Y, Z)_\rho$ be given by $(Y, Z|X)_\rho \sim \mathcal{N}\left(\begin{pmatrix} X \\ X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_Z \\ \rho\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix} \right)$, here the correlation coefficient ρ is used for indexing. Let Σ_ρ represent the correlation matrix of $(Y, Z)_\rho$. Following Hansen and Torgersen (1974) [12] [13] [14] we get that $(Y, Z)_{\rho_1} \succeq_B (Y, Z)_{\rho_2}$ iff $1^T\Sigma_{\rho_1}^{-1}1 \geq 1^T\Sigma_{\rho_2}^{-1}1$.

From figure 3 we can see that when Z and Y are Blackwell equivalent (i.e. have the same variance) then the value is decreasing in ρ. Note that $1^T\Sigma_\rho^{-1}1 = \frac{\sigma_Y^2+\sigma_Z^2-2\rho\sigma_Y\sigma_Z}{\sigma_Y^2\sigma_Z^2(1-\rho^2)}$. This indicates for the same variance higher correlation leads to a lower complementarity of information. On the other hand, when Z and Y have different variances, the value first decreases with the correlation coefficient and then increases.

[11] For this example we may relax the finitness assumption on $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$

[12] Hansen and Torgersen (1974) show that for unknown parameter β signal $S_A \sim \mathcal{N}(A\beta, I)$ Blackwell dominates $S_B \sim \mathcal{N}(B\beta, I)$ if and only if $A^TA - B^TB$ is positive semi definite. For positive definite Σ_A and Σ_B, the signals $S_A \sim \mathcal{N}(A\beta, \Sigma_A)$ and $S_B \sim \mathcal{N}(B\beta, \Sigma_B)$ are Blackwell equivalent to $\widetilde{S}_A := (U_A^T)^{-1}S_A \sim \mathcal{N}((U_A^T)^{-1}A\beta, I)$ and $\widetilde{S}_B := (U_B^T)^{-1}S_B \sim \mathcal{N}((U_B^T)^{-1}B\beta, I)$ respectively. Where Σ_A and Σ_B uniquely (up to unitary transformations) decompose into $U_A^TU_A$ and $U_B^TU_B$ respectively. Then S_A Blackwell dominates S_B if and only if $A^T\Sigma_A^{-1}A - B^T\Sigma_B^{-1}B$ is positive semi-definite.

[13] The above requires the linear normal experiments to be regular (i.e. $ker(\Sigma) \subset ker(A^T)$), which holds in our case when $\rho \in (-1, 1)$ and the variances are non-zero.

[14] The Blackwell comparison of the linear normal model is equivalent to the comparison between covariances of the best linear unbiased estimators. For linear normal experiments A and B. A Blackwell dominates B if and only if $Cov(\hat{\beta}_B) - Cov(\hat{\beta}_A)$ is positive semidefinite. Interested readers can see ch 8 of Torgersen (1991).

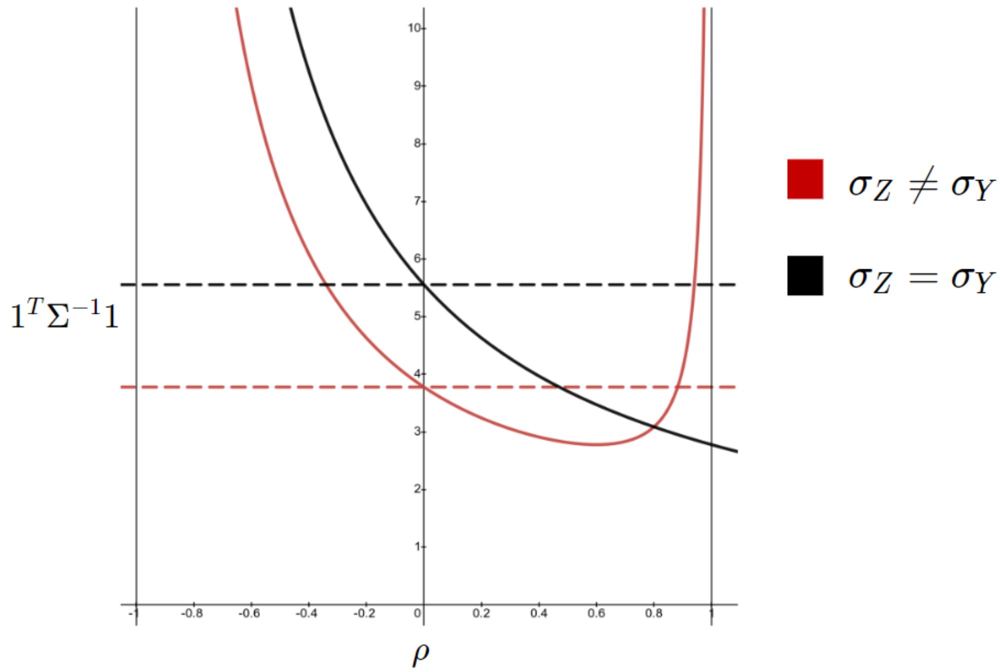

Figure 3: Gaussian example

Remark 7. Similar behavior at the extreme correlation (i.e. $|\rho| = 1$) is true for more general linear signals. In particular consider signals Y, Z to be given by:

$$Y = X + \varepsilon$$
$$Z = X + \eta$$

Where η and ε are zero mean noise terms with variance σ_Y^2 and σ_Z^2 respectively. Let ρ represent the correlation coefficient between Y, Z. Then $|\rho| = 1$ corresponds to $\eta = \rho\frac{\sigma_Z}{\sigma_Y}\varepsilon$ almost surely[15]. Thus whenever $\rho = -1$ the above linear equations give unique solution $X = \frac{Y+Z}{2}$. Whereas when $\rho = 1$ the system of equation determine X only if $\sigma_Y \neq \sigma_Z$.

From the example, we see that the intuition that a positive correlation between signals leads to less information isn't always correct. When considering joint signals, such as Y and Z, about an unknown variable X, the information they offer can be divided into two channels. Firstly, the direct insight that each signal individually provides about X. Secondly, there's the indirect insight due to the relationship between the signals. In the next section, we study the relation between the correlation structure of the signals and their informativeness.

6 Complements

When two pieces of information are conditionally independent, they are effectively separate sources of information about X. This separation is what leads to the lack of complementarity. Example 1 demonstrates this, in this section, we want to generalize the example. To this end, we can construct the random variable

[15] $\mathbb{E}[(\eta - a\varepsilon)^2] = 0$ when $|\rho| = 1$ and $a = \rho\frac{\sigma_Z}{\sigma_Y}$.

$Z_\varnothing$ in the following way. For any joint distribution p of (X, Y, Z) we can define the joint distribution $\widetilde{p}$ of $(X, Y, Z_\varnothing)$ such that $\widetilde{p}(x, y, z) = p(x, y)p(z|x) = p(x)p(y|x)p(z|x)$. By construction, $Z_\varnothing$ is Blackwell equivalent to Z and $Z_\varnothing \perp Y|X$.

Proposition 1. *The pair of random variables (Y, Z) are complements (substitutes) as per definition 5 if and only if $Z \succeq_Y (\preceq_Y) Z_\varnothing$.*

Proof. This follows from noting that for any decision problem (q, u) we have $V_{q,u}(Y, Z) - V_{q,u}(Y, Z_\varnothing) = \sum_{\mathcal{Y}} I_{q,u}(y; Y, Z)p(y)$. □

Remark 8. This shows that our definition of complementarity is symmetric, meaning Y is a complement of Z if and only if Z is a complement of Y. To see this observe that for any decision problem (q, u) we have $V_{q,u}(Y, Z_\varnothing) = V_{q,u}(Y_\varnothing, Z)$ as $(Y, Z_\varnothing|X) =_d (Y_\varnothing, Z|X)$ by construction[16]. Thus $V_{q,u}(Y, Z) \geq V_{q,u}(Y, Z_\varnothing)$ if and only if $V_{q,u}(Y, Z) \geq V_{q,u}(Y_\varnothing, Z)$.

Remark 9. Referring back to example 2 we see that for fixed variances $\rho = 0$ corresponds to $Z_\varnothing$, in figure 3 value for which the curve lies above (below) the dotted line are complements (substitutes).

6.1 Relation to BHK(2013)

Börgers et al. (2013), defines two signals to be complementary if the marginal value of a signal is higher when the other signal is present. In this sense, information is treated like a physical good.

Definition 6. Signals Y and Z are complements (substitutes) if for all preference u and priors $p_X(= q)$.

$$V_{q,u}(Y, Z) - V_{q,u}(Y) \geq (\leq) V_{q,u}(Z) - V_{q,u}(\varnothing)$$

In contrast, our definition of complements is defined in terms of the choice between signals relative to another signal. By restricting choice to Blackwell equivalent experiments we have

$$V_{q,u}(Y, Z_1) - V_{q,u}(Z_1) \geq V_{q,u}(Y, Z_2) - V_{q,u}(Z_2) \overset{Z_1 \sim_B Z_2}{\iff} V_{q,u}(Y, Z_1) \geq V_{q,u}(Y, Z_2)$$

Remark 10. The notion of complementarity introduced in definition 5 is substantively different than the notion of complementarity introduced in Börgers et al. (2013). To see this we recall example 1 the signal Z_1 was preferred (in $\succeq_Y$ order) over $Z_\varnothing$. This implies under definition 5, Z_1 and Y are complements. From proposition 3 of Börgers et al. (2013), the signals Y and Z_1 are not complements, as there is no "meaning reversal"[17][18] which is necessary (and also sufficient in binary states) for the condition in definition 6 to hold. More directly, Börgers et al. (2013) observe that the requirement of definition 6 is equivalent to the ordering of auxiliary signals $\widetilde{S}_S$ and $\widetilde{S}_C$. Where $\widetilde{S}_S$ represents a signals which takes value of Y with probability $\frac{1}{2}$ and Z with probability $\frac{1}{2}$. The signal $\widetilde{S}_C$ represents a signal that takes value of (Y, Z) with probability $\frac{1}{2}$ and is uninformative signal $\varnothing$ with probability $\frac{1}{2}$. Then definition 6 is equivalent to $\widetilde{S}_C \succeq_B \widetilde{S}_S$. This mean for example 1 with $Z = Z_1$ we get:

$$\widetilde{S}_S : \begin{pmatrix} \frac{1}{2}p & \frac{1}{2}(1-p) & \frac{1}{2}p & \frac{1}{2}(1-p) \\ \frac{1}{2}(1-p) & \frac{1}{2}p & \frac{1}{2}(1-p) & \frac{1}{2}p \end{pmatrix}$$

[16] For any joint distribution p of X, Y, Z. The joint distribution p' of $(X, Y, Z_\varnothing)$ is such that $p'(x, y, z) = p(x)p(y|x)p(z|x)$. Similarly, the joint distribution p'' of $(X, Y_\varnothing, Z)$ is such that $p''(x, y, z) = p(x)p(y|x)p(z|x)$. Thus the signals $(Y, Z_\varnothing)$ and $(Y_\varnothing, Z)$ are Blackwell equivalent

[17] For any prior p_X, $p(P|Y = 1, Z = 1) > p_X(P)$ and $p(P|Y = 0, Z = 0) < p_X(P)$.

[18] Roughly speaking, for binary states, meaning reversal requires signals that individually move the decision maker's beliefs into one direction, if received together move the decision maker's beliefs into the opposite direction. See Börgers et al. (2013) for the general definition.

$$\widetilde{S}_C : \begin{pmatrix} \frac{1}{2}(2p-1) & \frac{1}{2}(1-p) & \frac{1}{2}(1-p) & 0 & \frac{1}{2} \\ 0 & \frac{1}{2}(1-p) & \frac{1}{2}(1-p) & \frac{1}{2}(2p-1) & \frac{1}{2} \end{pmatrix}$$

Under the following decision problem, the value of $\widetilde{S}_S$ is greater than $\widetilde{S}_C$. Let $\mathcal{D} = \{d_1, d_2\}$, and $u(d_1, P) = \frac{1}{3}$, $u(d_2, A) = \frac{2}{3}$ and zero utility otherwise. Then for uniform prior on $\mathcal{X} = \{P, A\}$, and $p = \frac{5}{6}$ the value of signal $\widetilde{S}_S$ is greater than the value of $\widetilde{S}_C$. Similar conclusions hold for remarks 4 and **??**.

Remark 11. From example 2 we have $(Y, Z|X)_\rho \sim \mathcal{N}\left(\begin{pmatrix} X \\ X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_Z \\ \rho\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix} \right)$. In the linear normal model for ρ approaching -1, the value of the combined signal is increasing. Consider for instance the limiting case when Y, Z are together fully informative about X (i.e. $(Y, Z) \sim_B X$). Let $\mathcal{X} = \{x_1 = 1, x_2 = -1\}$. Consider the decision problem with a uniform prior and utility such that $u(x_1, d_1) = u(x_2, d_2) = 1$ and 0 otherwise. In this case the value of signal $\widetilde{S}_c$ is 0.25. If we let $\sigma_Z = \sigma_Y < 1.48$, then the value of $\widetilde{S}_S > 0.25$. Thus certain perfectly correlated linear normal signals are not complements under definition 6.

The above remarks show that studying the complementarity of information based on the marginal increase in value of a signal in the presence of another signal isn't able to account for complementarity induced by anticipation of information. In particular, definition 6 is too narrow to capture behaviourally relevant ways in which two signals can "add to" each others' information.

6.2 Strong Complements

In this section[19], we will present an equivalent condition for signals to strongly complement each other as per definition 4. Let p be the joint distribution of (X, Y, Z), then define a family of random variables $\{Z_y\}_{y\in\mathcal{Y}}$ taking value in $\mathcal{Z}$ such that Z_y is defined as in section 5.

Proposition 2. *For two random variables Y, Z. If for all $y \in \mathcal{Y}$ we have $Z_y \succeq_B Z$ then Y is a strong complement of Z as per definition 4. The converse holds when $t_{\mathcal{Y}}(y)$ has full support in $\mathcal{X}$ for every $y \in \mathcal{Y}$.*

Proof. (Sketch)

Fix arbitrary $y \in \mathcal{Y}$.

The first statement follows by noting that $Z_y \succeq_B Z$ then for any decision problem the value of Z_y is greater than Z. Formally $V_{q,u}(Z_y) \geq V_{q,u}(Z)$ for every decision problem $(q, u, \mathcal{D})$. Consider the case when $q = p_y$, then $V_{p_y,u}(Z_y) \geq V_{p_y,u}(Z)$ for every $(u, \mathcal{D})$. Finally recall that $I_{q,u}(y; Y, Z) = V_{p_y,u}(Z_y) - V_{p_y,u}(Z)$.

For the second statement let Y be a strong complement of Z. Thus for all decision problems (q, u) we have $I_{q,u}(y; Y, Z) \geq 0$. In particular, for uniform prior $q = \frac{1}{|\mathcal{X}|}$ we have $I_{\frac{1}{|\mathcal{X}|},u}(y; Y, Z) = V_{t_{\mathcal{Y}}(y),u}(Z_y) - V_{t_{\mathcal{Y}}(y),u}(Z) \geq 0$ for every positive bounded utility. By assumption $t_{\mathcal{Y}}(y)$ has full support on $|\mathcal{X}|$ thus by lemma 1 we have that $Z_y \succeq_B Z$. As the choice of y was arbitrary we have proved the result. $\square$

Remark 12. For the linear normal models (see example 2), definition 5 and 4 are equivalent[20]. To see this consider the signals $(Y, Z|X)_\rho \sim \mathcal{N}\left(\begin{pmatrix} X \\ X \end{pmatrix}, \begin{pmatrix} \sigma_Y^2 & \rho\sigma_Y\sigma_Z \\ \rho\sigma_Y\sigma_Z & \sigma_Z^2 \end{pmatrix} \right)$. Then the distribution of Z_y is given by $\mathcal{N}\left(X\left(1 - \rho\frac{\sigma_Z}{\sigma_Y}\right) + y\rho\frac{\sigma_Z}{\sigma_Y},\ \sigma_Z^2(1-\rho^2) \right)$. We can define a random variable $\widetilde{Z}_y$ with distribution $\mathcal{N}\left(X,\ \frac{\sigma_Y^2\sigma_Z^2(1-\rho^2)}{(\sigma_Y - \rho\sigma_Z)^2} \right)$, by construction $Z_y \sim_B \widetilde{Z}_y$. Now note $Z_y \succeq_B Z$ if and only if $\frac{\sigma_Y^2\sigma_Z^2(1-\rho^2)}{(\sigma_Y-\rho\sigma_Z)^2} \geq \sigma_Z^2$. This equates to $\rho \geq \frac{2\sigma_Y\sigma_Z}{\sigma_Y^2+\sigma_Z^2}$ when $\rho > 0$ and $\rho \leq \frac{2\sigma_Y\sigma_Z}{\sigma_Y^2+\sigma_Z^2}$ when $\rho < 0$ (this always holds as the expression on right is positive). From example 2 we see that $(Y, Z) \succeq_B (Y, Z_\varnothing)$ if and only if $\frac{\sigma_Y^2+\sigma_Z^2-2\rho\sigma_Y\sigma_Z}{\sigma_Y^2\sigma_Z^2(1-\rho^2)} \geq \frac{\sigma_Y^2+\sigma_Z^2}{\sigma_Y^2\sigma_Z^2}$. After some

[19]Here we restrict attention to full support signals such that $\text{supp}(p) = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$.

[20]Here we assume that $1 - \rho\frac{\sigma_Z}{\sigma_Y} \neq 0$

algebra, this equates to $\rho \geq \frac{2\sigma_Y\sigma_Z}{\sigma_Y^2+\sigma_Z^2}$ when $\rho > 0$ and $\rho \leq \frac{2\sigma_Y\sigma_Z}{\sigma_Y^2+\sigma_Z^2}$ when $\rho < 0$. Thus by proposition 1 for the linear normal model complementarity (definition 5) is equivalent to strong complementarity (definition 4). Combining the above with remark 11 we see that even complementarity in the stronger sense of proposition 2 isn't sufficient for complementarity as per definition 6.

7 Redundant Information

As seen in example 1, we can construct signal Z_2 such that it has no value conditional on seeing the signal Y. Thus once Y is known there is no value in knowing Z_2, in this sense Y acts like a substitute for Z_2. This happens as the signal Z_2 contains redundant information that is already contained in Y. In this section, we expand on these ideas. In this section, the results are stated for full support priors; $\text{supp}(p_X) = \mathcal{X}$.

Definition 7. A signal Z is redundant with respect to Y, written $Y \rhd Z$ if the following holds for all preference u and priors q:

$$V_{q,u}(Y, Z) = V_{q,u}(Y)$$

Proposition 3. *The following conditions are equivalent:*

(a) Redundancy; $Y \rhd Z$

(b) Sufficiency; $Z \perp X|Y$

(c) Minimality; $Y \succeq_B Z$ and $Z \preceq_Y Z'$ for all $Z' \sim_B Z$

Condition (b) highlights the role of correlation in information redundancy, it requires that once Y is known, knowing Z doesn't provide additional information about X. More precisely, the signal Y is a sufficient statistic for Z, meaning the distribution of Z depends on X only through Y. Condition (c) relates the notion of "higher" redundancy to being less desirable (substitutability) in the information acquisition problem described above. The requirement that Y Blackwell dominated Z, says that in order for Z to be redundant with respect to Y it is necessary that Y is at least as informative as Z. In particular proposition 3 shows the equivalence between the extreme case of information redundancy (conditional independence) and minimal incentive to acquire information.

Proof. **(Sketch)**

It is easy to check that $(b) \implies (a)$ by noting that $\mathbb{E}[u(X,d)|Y = y, Z = z] = \mathbb{E}[u(X,d)|Y = y]$ for all u, $d \in \mathcal{D}, y \in \mathcal{Y}$ and $z \in \mathcal{Z}$.
For $(a) \implies (c)$ note that for any signal Z', $V_{q,u}(Z'), V_{q,u}(Y) \leq V_{q,u}(Y, Z')$ as the value of information is non negative. Now , by redundancy we get $V_{q,u}(Z) \leq V_{q,u}(Y, Z) = V_{q,u}(Y) \leq V_{q,u}(Y, Z')$. This holds for any preference and any prior on X, thus (c) follows.
Finally, we show $(c) \implies (b)$. Let p represent the joint distribution of (X, Y, Z). By (c) we know $Y \succeq_B Z$, thus there exist a map $\kappa : \mathcal{Y} \to \Delta(\mathcal{Z})$ such that $p(z|x) = \sum_{\mathcal{Y}} \kappa(z|y)p(y|x)$. Define a random variable[21] Z' such that the joint distribution r of (X, Y, Z') is given by $r(x, y, z') = p(x)p(y|x)\kappa(z'|y)$, then $Z \sim_B Z'$ and $Z' \perp X|Y$. In order to show (b), we assume for contradiction $Z \not\perp X|Y$ and then show that $V_{q,u}(Y, Z') < V_{q,u}(Y, Z)$ for some prior q and preference u which contradicts $Z \preceq_Y Z'$.
Let u, q be some arbitrary preference and prior on X.

$$\delta^*_{q,u}(y) \in argmax_{d \in \mathcal{D}} \mathbb{E}[u(X, d)|Y = y]$$

[21] Le (1964) shows a similar statement for a more general setting.

$$\delta'_{q,u}(y,z) \in argmax_{d\in\mathcal{D}}\mathbb{E}[u(X,d)|Y=y,Z'=z]$$

$$\delta_{q,u}(y,z) \in argmax_{d\in\mathcal{D}}\mathbb{E}[u(X,d)|Y=y,Z=z]$$

As $Z' \perp X|Y$, for all $z \in \mathcal{Z}$ we have $\delta'_{q,u}(y,z) = \delta^*_{q,u}(y)$. For contradiction we assume that $Z \not\perp X|Y$ which implies that there exists (x',y',z') such that $p(x',y',z') > 0$ and $p(x'|y',z') \neq p(x'|y')$. Without loss of generality[22] let $p(x'|y',z') > p(x'|y')$. Define the strategy $\sigma : \mathcal{Y}\times\mathcal{Z} \to \mathcal{D}$ such that forall $(y,z) \neq (y',z')$ we have $\sigma_{q,u}(y,z) = \delta^*_{q,u}(y)$ and $\sigma_{q,u}(y',z') = \delta'_{q,u}(y,z)$. By definition payoff from $\delta_{q,u}$ is greater than payoff from $\sigma_{q,u}$ conditional on seeing (Y,Z). Thus it suffices to show that there is a decision problem under which the payoff from strategy $\delta^*(Y)$ is strictly worse than strategy $\sigma(Y,Z)$. From the definition, it follows that (we omit the dependence of decision strategy on u,q for clarity)

$$\mathbb{E}\left[\mathbb{E}[u(X,\sigma(Y,Z))|Y,Z]\right] - \mathbb{E}\left[\mathbb{E}[u(X,\delta^*(Y))|Y]\right]$$

$$\begin{aligned}
&= \sum_{\mathcal{Y}}\sum_{\mathcal{Z}}\left[\sum_{\mathcal{X}}(u(x,\sigma(y,z)) - u(x,\delta^*(y))p(x|y,z)\right]p(z|y)p(y)\\
&= \sum_{\mathcal{Y}-y'}\sum_{\mathcal{Z}-z'}\left[\sum_{\mathcal{X}}(u(x,\sigma(y,z)) - u(x,\delta^*(y))p(x|y,z)\right]p(z|y)p(y)\\
&\quad + \left[\sum_{\mathcal{X}}(u(x,\sigma(y',z')) - u(x,\delta^*(y'))p(x|y',z')\right]p(z'|y')p(y')\\
&= \sum_{\mathcal{Y}-y'}\sum_{\mathcal{Z}-z'}\left[\sum_{\mathcal{X}}(u(x,\delta^*(y)) - u(x,\delta^*(y))p(x|y,z)\right]p(z|y)p(y)\\
&\quad + \left[\sum_{\mathcal{X}}(u(x,\delta(y',z')) - u(x,\delta^*(y'))p(x|y',z')\right]p(z'|y')p(y')\\
&= \left[\sum_{\mathcal{X}}u(x,\delta(y',z'))) - u(x,\delta^*(y'))\right]p(x|y',z')p(z'|y')p(y')
\end{aligned}$$

Let $k \in \mathbb{R}$ be such that $p(x'|y') < \frac{1}{1+k} < p(x'|y',z')$, $\mathcal{D} = \{d_1,d_2\}$.The utility is given by $u(d_1,x') = k$, $u(d_2,x) = 1$ for $x \neq x'$ and zero otherwise. This gives us:

$$\sum_{\mathcal{X}}\left[u(x,\delta(y',z')) - u(x,\delta^*(y'))\right]p(x|y',z')$$

$$= (k+1)p(x'|y',z') - 1 > 0$$

As $p(y',z') > 0$ the above implies that $Z \not\succeq_Y Z'$. Thus the result follows by contradiction. □

Remark 13. Blackwell (1951), Blackwell (1953) showed that if Y is more valuable than Z for all decision problems then we can generate the signals Z by randomizing the signals Y post-realization. But in general, the joint distribution of (X,Y,Z) isn't uniquely determined by Blackwell dominance. Proposition 3 says that if in addition to $Y \succeq_B Z$, the signal Z is the minimal element (under the order $\succeq_Y$) in the equivalent class of signals as informative as Z, then the additional restriction of conditional independence (c) holds for the joint distribution of (X,Y,Z).

[22] If $p(x'|y',z') < p(x'|y)$ then as $p(x'|y')$ is convex combination of $p(x'|y',z)$ with positive weight for $z = z'$ there exists a z'' such that $p(z''|y') > 0$ and $p(x'|y',z'') > p(x'|y')$.

Remark 14. Within the linear Gaussian setting of example 2 we can consider the following:

$$Y = X + \varepsilon$$

$$Z = X + \varepsilon + \eta$$

Where η, ε are mean zero normal variables with variance δ^2 and σ^2 respectively. We let X, η, ε to be jointly independent. This gives us the conditional distribution $(Y, Z|X) \sim \mathcal{N}\left(\begin{pmatrix} X \\ X \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \delta^2 \end{pmatrix}\right)$. Note that $1^T \Sigma_Z^{-1} 1 = \frac{1}{\sigma^2}$. Proposition 3 implies among linear Gaussian signals (Y, Z') such that $(Y, Z'|X) \sim \mathcal{N}\left(\begin{pmatrix} X \\ X \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma(\sigma^2+\delta^2)^{\frac{1}{2}} \\ \rho\sigma(\sigma^2+\delta^2)^{\frac{1}{2}} & \sigma^2 + \delta^2 \end{pmatrix}\right)$, the signal (Y, Z) is least preferred (in Blackwell order). More directly for correlation coefficient ρ, $1^T \Sigma_\rho^{-1} 1 = \frac{2\sigma^2 + \delta^2 - 2\rho\sigma(\sigma^2+\delta^2)^{\frac{1}{2}}}{\sigma^2(\sigma^2+\delta^2)(1-\rho^2)}$, the minimum value[23] of this expression is $\frac{1}{\sigma^2}$.

8 Affiliated information

So far we have made comparisons over all possible utilities and over all possible correlation structures. In this section, we will restrict attention to a smaller class of utilities and distribution to get a better understanding of the complementarity between signals.
Positive dependence between the state and information is common in economics and has been studied in Milgrom and Weber (1982), Persico (2000), de Castro (2009). Following that we restrict attention to the information acquisition environment with positive dependence. First, we restrict to the state $(\mathcal{X})$, sample $(\mathcal{Y}, \mathcal{Z})$ and decision space $(\mathcal{D})$ to be subsets of $\mathbb{R}$ inhereting the usual order from $\mathbb{R}$. Just like before we require $\mathcal{X}$, $\mathcal{Z}$, $\mathcal{Y}$ to be finite, and require $\mathcal{D}$ to be a closed interval [24]. We look at signals such that the joint pmf p of (X, Y, Z) is affiliated and has full support. When studying the comparison of positive dependent signals it is useful to restrict preferences to the ones under which optimal decisions are increasing in the signal value. Such preferences arise naturally in statistical and economic problems (Karlin and Rubin (1956), Lehmann (1988), Athey (2002), Quah and Strulovici (2009)). With this in mind, we provide sufficient conditions for choosing a signal over another in a restricted set of decision problems.

Definition 8. The random vector $(S_1, ..., S_k)$ with pmf (distribution function for continuous variables) p is affiliated if for any $s, s' \in \mathcal{S}_1 \times \mathcal{S}_2 \times ...\mathcal{S}_k$ we have the following[25]:

$$p(s \vee s')p(s \wedge s') \geq p(s)p(s')$$

Definition 9. If $r, m \in \Delta(\mathcal{S})$ are two probability mass functions then we say r likelihood ratio dominates m if $\frac{r(s)}{m(s)}$ is increasing in s. Moreover, a family $\{r_i\}_{i \in \mathcal{I}}$ of mass functions is MLR ordered if for any $i \geq j \in \mathcal{I}$ we have that r_i likelihood ratio dominates r_j.

Note, that the signal space $\mathcal{S}$ and the indexing set $\mathcal{I}$ in the above definition can be partially ordered. In particular $\mathcal{S}$ and $\mathcal{I}$ can be a subset of $\mathbb{R}^2$ with the usual partial order.

Proposition 4. *For random variables (X, Y, Z) with pmf p. If $\{Pr(Y|X = x)\}_{x \in \mathcal{X}}$ and $\{Pr(Z|X = x)\}_{x \in \mathcal{X}}$ are MLR ordered then $\{Pr(Y, Z_\varnothing|X = x)\}_{x \in \mathcal{X}}$ is MLR ordered.*

[23] $\rho^* = \underset{(-1,1)}{\arg\min}\, 1^T \Sigma_\rho^{-1} 1$, then $1^T \Sigma_{\rho^*}^{-1} 1 = \frac{1}{\sigma^2}$

[24] Whenever we refer to convergence of elements in $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ it's with respect to relative euclidean topology induced by the subsets.

[25] $s \vee s' = (\max\{s_1, s_1', \}..., \max\{s_k, s_k'\})$ and $s \wedge s' = (min\{s_1, s_1', \}..., \min\{s_k, s_k'\})$

Proof. (Sketch)

To see this let $r \in \Delta(\mathcal{X} \times \mathcal{Y} \times \mathcal{Z})$ represent the pmf of $(X, Y, Z_\varnothing)$, then for any $x' \geq x$, $y' \geq y$, $z' \geq z$ we get that

$$\begin{aligned} r(x', y', z')r(x, y, z) &= p(y'|x')p(z'|x')p(y|x)p(z|x)p(x')p(x) \\ &\geq p(y|x')p(z|x')p(y'|x)p(z'|x)p(x')p(x) \quad \text{(by MLR property)} \\ &= r(x', y, z)r(x, y', z') \end{aligned}$$

This proves the desired MLR property. □

Corollary 1. *For random variables (X, Y, Z) if $\{Pr(Y|X = x)\}_{x\in\mathcal{X}}$ and $\{Pr(Z|X = x)\}_{x\in\mathcal{X}}$ are MLR ordered then $\{Pr(X|Y = y, Z_\varnothing = z)\}_{(y,z)\in\mathcal{Y}\times\mathcal{Z}}$ is MLR ordered.*

Proof. For some random vector (S_1, S_2) taking values in $\mathcal{S}_2 \times \mathcal{S}_2$. If $\{Pr(S_1|S_2 = s)\}_{s\in\mathcal{S}_2}$ is a MLR ordered family then $\{Pr(S_2|S_1 = s)\}_{s\in\mathcal{S}_1}$ is a MLR ordered family. To see this note that Bayes rule gives us that for any $s_i, s_i' \in \mathcal{S}_i$

$$\frac{Pr(S_1 = s_1|S_2 = s_2')Pr(S_2 = s_2)}{Pr(S_1 = s_1|S_2 = s_2)Pr(S_2 = s_2')} = \frac{Pr(S_2 = s_2'|S_1 = s_1)}{Pr(S_2 = s_2|S_1 = s_1)}$$

If $s_i' > s_i$ then the MLR property we get that $\frac{Pr(S_1=s_1|S_2=s_2')}{Pr(S_1=s_1|S_2=s_2)} \leq \frac{Pr(S_1=s_1'|S_2=s_2')}{Pr(S_1=s_1'|S_2=s_2)}$

$$\implies \frac{Pr(S_1 = s_1|S_2 = s_2')Pr(S_2 = s_2)}{Pr(S_1 = s_1|S_2 = s_2)Pr(S_2 = s_2')} \leq \frac{Pr(S_1 = s_1'|S_2 = s_2')Pr(S_2 = s_2)}{Pr(S_1 = s_1'|S_2 = s_2)Pr(S_2 = s_2')}$$

$$\implies \frac{Pr(S_2 = s_2'|S_1 = s_1)}{Pr(S_2 = s_2|S_1 = s_1)} \leq \frac{Pr(S_2 = s_2'|S_1 = s_1')}{Pr(S_2 = s_2|S_1 = s_1')}$$

This gives us that $\{Pr(S_2|S_1 = s)\}_{s\in\mathcal{S}_1}$ is a MLR ordered family. The result then follows from proposition 4. □

Lemma 2. *If the random vector (X, Y, Z) is affiliated then $\{Pr(Y|X = x)\}_{x\in\mathcal{X}}$ and $\{Pr(Z|X = x)\}_{x\in\mathcal{X}}$ are MLR ordered.*

Proof. Let p be the distribution of (X, Y, Z), then by affiliation it follows that for $x' \geq x \in \mathcal{X}$, $z' \geq z \in \mathcal{Z}$ and some $y' \geq y \in \mathcal{Y}$:

$$p(x', z', y')p(x, z, y) \geq p(x', z, y)p(x, z', y')$$

We prove the result for the case of $\{Pr(Z|X = x)\}_{x\in\mathcal{X}}$, the result for $\{Pr(Y|X = x)\}_{x\in\mathcal{X}}$ follows analogusly. To show this statement we will use the four function theorem of[26] Ahlswede and Daykin (1978). Let $x' > x \in \mathcal{X}$, $z' > z \in \mathcal{Z}$ then define $f_1(y) = p(x', z, y)$, $f_2(y) = p(x, z', y)$, $f_3(y) = p(x, z, y)$, $f_4(y) = p(x', z', y)$. By affiliation, we get that for $y, y' \in \mathcal{Y}$, $f_1(y)f_2(y') \leq f_3(y \wedge y')f_4(y \vee y')$. The desired result then follows from the four-function theorem after summing over $\mathcal{Y}$:

$$\sum_{y\in\mathcal{Y}} p(x', z, y) \sum_{y\in\mathcal{Y}} p(x, z', y) \leq \sum_{y\in\mathcal{Y}} p(x, z, y) \sum_{y\in\mathcal{Y}} p(x', z', y)$$

$$\implies p(x', z)p(x, z') \leq p(x, z)p(x', z')$$

□

[26] From Eaton (1982) section 3: For four non-negative functions f_1, f_2, f_3 and f_4 on S such that for all $s, s' \in S$, $f_1(s)f_2(s') \leq f_3(s \wedge s')f_4(s \vee s')$ then $\sum f_1(s)\mu(s) \sum f_2(s)\mu(s) \leq \sum f_3(s)\mu(s) \sum f_4(s)\mu(s)$.

Let $u : \mathcal{X} \times \mathcal{D} \to [0, \infty)$ represent the preference of the decision maker. Then $\{u(x,.)\}_{x\in\mathcal{X}}$ defines a family of functions parameterized by x. For each $x \in \mathcal{X}$ let $D^u(x) := \{d \in \mathcal{D} | d \in \arg\max_{d'\in\mathcal{D}} u(x, d')\}$. Similarly, for each $q \in \Delta(\mathcal{X})$ we define $D^u(q) := \{d \in \mathcal{D} | d \in \arg\max_{d'\in\mathcal{D}} \sum_{x\in\mathcal{X}} u(x, d')q(x)\}$. We restrict attention to utilities[27] such that $D^u(x)$ is non-empty for each $x \in \mathcal{X}$. Additionally, we restrict attention to preferences such that $\{u(x,.)\}_{x\in\mathcal{X}} \in \mathcal{U}_{IDO}$, where $\mathcal{U}_{IDO}$ represents the collection of interval dominance ordered family of functions (defined in Quah and Strulovici (2009)).

Definition 10. A utility function $u \in \mathcal{U}_{IDO}$ if and only if for any $x' \geq x$ and $d'' > d'$

$$u(x, d'') - u(x, d') \geq (>) 0 \text{ and } u(x, d'') \geq u(x, d) \text{ for all } d \in [d', d''] \implies u(x', d'') - u(x', d') \geq (>) 0$$

We say $u(x',.)$ interval order dominates $u(x,.)$; written $u(x',.) \geq_I u(x,.)$.

The following lemma 3 is a restatement of theorem 1 in Quah and Strulovici (2009).

Lemma 3. *If $u \in \mathcal{U}_{IDO}$ then $D^u(x)$ is increasing in strong set order*[28].

Proof. (Sketch)

Fix some arbitrary $u \in \mathcal{U}_{IDO}$. We prove the above by contradiction, to this end let $x' > x \in \mathcal{X}$ be such that $D^u(x') \not\geq D^u(x)$. Thus there exist $d' \in D^u(x')$ and $d \in D^u(x)$ such that $d > d'$ and either $d \notin D^u(x')$ or $d' \notin D^u(x)$.

First let $d \notin D^u(x')$. As $d \in D^u(x') \implies d \in \arg\max_{\delta\in\mathcal{D}} u(x'\delta)$, thus $u(x', d') > u(x', d)$. Moreover, as $d \in D^u(x)$ we have $u(x, d) \geq u(x, \delta)$ for all $\delta \in [d', d]$. By definition of $\mathcal{U}_{IDO}$ we get that $u(x', d) \geq u(x', \delta)$ for all $\delta \in [d', d]$. In particular, $u(x', d) \geq u(x', d')$ which is a contradiction.

Now let $d' \notin D^u(x)$ and $d \in D^u(x')$. In this case $u(x', \delta)$ is constant for $\delta \in [d', d]$. As $d' \notin D^u(x)$ we have $u(x, d) > u(x, d')$ and $u(x, d) \geq u(x, \delta)$ for $\delta \in [d', d]$. By definition of $\mathcal{U}_{IDO}$ we get that $u(x', d) > u(x', d')$ which is a contradiction.

Thus $D^u(x') \geq D^u(x)$. □

The above lemma can be interpreted as a monotone comparative static result stating that higher states lead to higher optimal actions.

Remark 15. For some $u \in \mathcal{U}_{IDO}$ and prior $q \in \Delta(\mathcal{X})$ define the the function $u^q : d \mapsto \sum_{x\in\mathcal{X}} u(x, d)q(x)$. From Theorem 2 of Quah and Strulovici (2009) we get that if $r, m \in \Delta(X)$ such that r likelihood ratio dominates m then $u^r \geq_I u^m$. Combining this with lemma 3 we get that $D^u(r) \geq D^u(m)$.

The following proposition follows from a slight modification to lemma 7 in Quah and Strulovici (2009).

Proposition 5. *For any $u \in \mathcal{U}_{IDO}$. If a decision rule $\delta : \mathcal{X} \to \mathcal{D}$ is decreasing in x then there is a decision $d^* \in \mathcal{D}$ such that $u(x, d^*) \geq u(x, \delta(x))$ for all $x \in \mathcal{X}$.*

Proof. Fix some $u \in \mathcal{U}_{IDO}$. By lemma 3 we get that $D^u(x') \geq D^u(x)$ for any $x' \geq x$, this allows us to construct an increasing decision rule $\delta^* : \mathcal{X} \to \mathcal{D}$ such that $\delta^*(x) \in D^u(x)$.

We construct δ^* in the following way. Enumerate $\mathcal{X} = \{x_1, ..., x_M\}$ where $x_i < x_{i+1}$. Choose some $d_1 \in D^u(x_1)$ and let $\delta^*(x_1) = d_1$. For the induction hypothesis assume for $1 < i \leq k < M$ we have $\delta^*(x_i) \in D^u(x_i)$ and $\delta^*(x_{i-1}) \leq \delta^*(x_i)$. For the induction step consider x_{k+1} then if $D^u(x_{k+1}) - \cup_{i=1}^k D^u(x_i) \neq \varnothing$, let $\delta^*(x_{k+1}) = d_{k+1}$ for some $d_{k+1} \in D^u(x_{k+1}) - \cup_{i=1}^k D^u(x_i)$. As $D^u(x_{k+1}) \geq D^u(x_i)$ for $i \leq k$ and by the inductive hypothesis $\delta^*(x_i) \in D^u(x_i)$, we get that $\delta^*(x_{k+1}) \geq \delta^*(x_i)$ for $i \leq k$. If on the other hand $D^u(x_{k+1}) - \cup_{i=1}^k D^u(x_i) = \varnothing$ we let $\delta^*(x_{k+1}) = \delta^*(x_k)$, it then follows that $\delta^*(x_{k+1}) \in D^u(x_{k+1})$. Thus by induction, we have an increasing decision rule δ^*.

[27] In fact we require that the arg max exists for every sub-interval of $\mathcal{D}$.

[28] Let A and B be two subsets of $\mathbb{R}$. We say that A larger than B in the strong set order (written $A \geq B$) if for any for $a \in A$ and $b \in B$, we have $\max\{a, b\} \in A$ and $\min\{a, b\} \in B$. See Topkis (1998) for details

In the following, we have two cases. First, if $\delta(x) \geq \delta^*(x)$ for all $x \in \mathcal{X}$. In this case we let $d^* = \delta^*(x_M)$. We claim that $u(x, d^*) = u(x, \delta^*(x_M)) \geq u(x, d)$ for all $d \geq \delta^*(x_M)$ and $x \in \mathcal{X}$. To see this we proceed by contradiction, assume that there exists $x < x_M$ such that $u(x, \delta^*(x_M)) < u(x, d)$ for some $d > \delta^*(x_M)$. Let $\widetilde{d} \in \underset{d' \in [\delta^*(x_M),\, d]}{\arg\max}\; u(x, d')$. We get that $u(x, \widetilde{d}\,) \geq u(x, d') > u(x, \delta^*(x_M))$ and $u(x, \widetilde{d}\,) \geq u(x, d')$ for all $d' \in [\delta^*(x_M),\, \widetilde{d}]$. Thus by IDO property $u(x_M, \widetilde{d}\,) > u(x_M, \delta^*(x_M))$, this is a contradiction as $\delta^*(x_M) \in D^u(x_M)$. We conclude that $u(x, d^*) = u(x, \delta^*(x_M)) \geq u(x, d)$ for all $d \geq \delta^*(x_M)$ and $x \in \mathcal{X}$. From $\delta(x) \geq \delta^*(x)$ and $\delta^*(x)$ is increasing in x, we get $\delta(x) \geq \delta^*(x_M)$ for all $x \in \mathcal{X}$. Thus, $u(x, d^*) \geq u(x, \delta(x))$ for all $x \in \mathcal{X}$.

Now we show the second case. As δ is decreasing and δ^* is increasing, if $\delta \ngeq \delta^*$, there is some $x^* \in \mathcal{X}$ such that $x' < x^*$ implies $\delta(x') > \delta^*(x^*)$ and $x' \geq x^*$ implies $\delta(x') \leq \delta^*(x^*)$. Let $d^* = \delta^*(x^*)$. As $\delta^*(x^*) \in D^u(x^*)$ we get that $u(x, \delta^*(x^*)) \geq u(x, d)$ for all $d \leq \delta^*(x^*)$. By IDO property, $u(x', \delta^*(x^*)) \geq u(x', d)$ for all $x' \geq x^*$ and $d \leq \delta^*(x^*)$. In particular, $u(x', \delta^*(x^*)) \geq u(x', \delta(x'))$ for all $x' \geq x^*$. We are just left with showing $u(x', \delta^*(x^*)) \geq u(x', \delta(x'))$ for all $x' < x^*$. To do this we will repeat the argument in the first case with $x_M = x^*$. □

The following theorem gives a sufficient condition for complementarity with affiliated information and IDO preferences.

Theorem 1. *For random variables Y, Z such that $\{Pr(Y|X = x)\}_{x \in \mathcal{X}}$ and $\{Pr(Z|X = x)\}_{x \in \mathcal{X}}$ are MLR ordered, if there exists a function $T : \mathcal{X} \times \mathcal{Y} \times \mathcal{Z} \to \mathcal{Y} \times \mathcal{Z}$ such that*

1. *$T(x, y, z)$ is decreasing[29] in x for each $(y, z) \in \mathcal{Y} \times \mathcal{Z}$; and*
2. *For $Z_\varnothing$ as defined in section 6, we have $(X, Y, Z_\varnothing) =_d (X, T(X, Y, Z))$; where $=_d$ refers to equality in distribution.*

Then Y and Z are $\mathcal{U}_{IDO}-$ complements, i.e. $Z \succeq_Y^{\mathcal{U}_{IDO}} Z_\varnothing$.

Proof. [30] Fix some utility $u \in \mathcal{U}_{IDO}$ and full support prior $q \in \Delta(\mathcal{X})$. First, observe that by corollary 1 and remark 15, there is an optimal decision rule $\varphi : \mathcal{Y} \times \mathcal{Z} \to \mathcal{D}$ that solves the equation below and is increasing in (y, z).

$$\mathbb{E}\left[u(X, \varphi(Y, Z_\varnothing))\right] = \mathbb{E}\left[\max_{d \in \mathcal{D}} \mathbb{E}\left[u(X, d)|Y, Z_\varnothing\right]\right]$$

Composing φ with T gives us that $\varphi \circ T(x, y, z)$ is decreasing in x for each y, z. Fix some $(y, z) \in \mathcal{Y} \times \mathcal{Z}$, then $\varphi(T(., y, z)) : \mathcal{X} \to \mathcal{D}$ is a decreasing decision rule and by proposition 5 there exists a decision $d^*_{y,z} \in \mathcal{D}$ such that $u(x, d^*_{y,z}) \geq u(x, \varphi(x, y, z))$ for all $x \in \mathcal{X}$. Define a new decision rule $\psi : \mathcal{Y} \times \mathcal{Z} \to \mathcal{D}$ such that $\psi(y, z) = d^*_{y,z}$. For each $x \in \mathcal{X}$ we have the following:

$$\mathbb{E}\left[u(X, \varphi(Y, Z_\varnothing))|X = x\right] = \mathbb{E}\left[u(x, \varphi(T(x, Y, Z))|X = x\right]$$

$$\leq \mathbb{E}\left[u(x, \psi(Y, Z))|X = x\right] = \mathbb{E}\left[u(X, \psi(Y, Z))|X = x\right]$$

Summing over $\mathcal{X}$ with respect to q we get that

$$\sum_{x \in \mathcal{X}} \mathbb{E}\left[u(X, \varphi(Y, Z_\varnothing)|X = x\right] q(x) \leq \sum_{x \in \mathcal{X}} \mathbb{E}\left[u(X, \psi(Y, Z))|X = x\right] q(x)$$

Let $\psi^* : \mathcal{Y} \times \mathcal{Z} \to \mathcal{D}$ be such that

$$\mathbb{E}\left[u(X, \psi^*(Y, Z))\right] = \mathbb{E}\left[\max_{d \in \mathcal{D}} \mathbb{E}\left[u(X, d)|Y, Z\right]\right]$$

[29] By decreasing we mean decreasing in the usual partial order on $\mathbb{R}^2$, i.e. $(y, z) \geq (y', z')$ if and only if $y \geq y'$ and $z \geq z'$, and strict inequality if at least one of the coordinates is strictly greater.

[30] The proof follows from a slight modification of theorem 3 in Quah and Strulovici (2009) to the case of two-dimensional signals.

Thus we have

$$\sum_{x\in\mathcal{X}} \mathbb{E}\left[u(X,\varphi(Y,Z_\varnothing))|X=x\right]q(x) \leq \sum_{x\in\mathcal{X}} \mathbb{E}\left[u(X,\psi(Y,Z))|X=x\right]q(x) \leq \sum_{x\in\mathcal{X}} \mathbb{E}\left[u(X,\psi^*(Y,Z))|X=x\right]q(x)$$

$$V_{u,q}(Y,Z_\varnothing) \leq V_{u,q}(Y,Z)$$
$$\implies \sum_{\mathcal{Y}} I_{q,u}(y;Y,Z)p(y) \geq 0$$

This establishes the required statement as the choice of decision problem was arbitrary.

□

8.1 Strong complements

In this section, we characterize strong complementarity for affiliated information and IDO preferences. We also extend this characterization to the case of continuous random variables in order to provide a more intuitive result.

Proposition 6. *For random variables (X,Y,Z) such that $\{Pr(Z|X=x)\}_{x\in\mathcal{X}}$ is MLR ordered. If the function T in theorem 1 is such that $T(x,y,z)=(y,\tau(x,y,z))$ for some $\tau:\mathcal{X}\times\mathcal{Z}\to\mathcal{Z}$, then Y in a strong $\mathcal{U}_{IDO}-$ complement of Z.*

Proof. Note that $\{Pr(Z_\varnothing|X=x,Y=y)\}_{x\in\mathcal{X}} = \{Pr(Z|X=x)\}_{x\in\mathcal{X}}$ is MLR ordered. Fix some $y\in\mathcal{Y}$, there is an increasing decision rule $\varphi_y:\mathcal{Z}\to\mathcal{D}$ such that

$$\mathbb{E}\left[u(X,\varphi_y(Z_\varnothing))|Y=y\right] = \mathbb{E}\left[\max_{d\in\mathcal{D}}\mathbb{E}\left[u(X,d)|Z_\varnothing\right] \mid Y=y\right]$$

Define $\tau_y:\mathcal{X}\times\mathcal{Z}\to\mathcal{D}$ such that $\tau_y(x,z)=\tau(x,y,z)$. Composing φ_y with τ_y gives us that $\varphi_y\circ\tau_y(x,z)$ is decreasing in x for each z. Fix some $z\in\mathcal{Z}$, then $\varphi(\tau_y(.,z)):\mathcal{X}\to\mathcal{D}$ is a decreasing decision rule and by proposition 5 there exists a decision $d^*_{y,z}\in\mathcal{D}$ such that $u(x,d^*_{y,z})\geq u(x,\varphi_y\circ\tau_y(x,z))$ for all $x\in\mathcal{X}$. Define a new decision rule $\psi_y:\mathcal{Z}\to\mathcal{D}$ such that $\psi_y(z)=d^*_{y,z}$. For each $x\in\mathcal{X}$ we have the following:

$$\mathbb{E}\left[u(X,\varphi_y(Z))|X=x\right] = \mathbb{E}\left[u(X,\varphi_y(Z_\varnothing))|X=x,Y=y\right] = \mathbb{E}\left[u(x,\varphi_y(\tau_y(x,Z))|X=x,Y=y\right]$$
$$\leq \mathbb{E}\left[u(x,\psi_y(Z))|X=x,Y=y\right] = \mathbb{E}\left[u(X,\psi_y(Z))|X=x,Y=y\right]$$

Summing over $\mathcal{X}$ with respect to $p_y\in\Delta(\mathcal{X})$ where $p_y(x)=p(x|Y=y)=\frac{Pr(Y=y|X=x)q(x)}{\sum_{x'\in\mathcal{X}}Pr(Y=y|X=x')q(x')}$ we get that

$$\sum_{x\in\mathcal{X}}\mathbb{E}\left[u(X,\varphi_y(Z_\varnothing))|X=x,Y=y\right]p(x|Y=y) \leq \sum_{x\in\mathcal{X}}\mathbb{E}\left[u(X,\psi_y(Z))|X=x,Y=y\right]p(x|Y=y)$$

Let $\psi^*_y:\mathcal{Z}\to\mathcal{D}$ be such that

$$\mathbb{E}\left[u(X,\psi^*_y(Z))|Y=y\right] = \mathbb{E}\left[\max_{d\in\mathcal{D}}\mathbb{E}\left[u(X,d)|Z\right]|Y=y\right]$$

Thus we have

$$\sum_{x\in\mathcal{X}}\mathbb{E}\left[u(X,\varphi_y(Z_\varnothing)|X=x\right]p_y(x) \leq \sum_{x\in\mathcal{X}}\mathbb{E}\left[u(X,\psi_y(Z))|X=x,Y=y\right]p_y(x) \leq \sum_{x\in\mathcal{X}}\mathbb{E}\left[u(X,\psi^*_y(Z))|X=x,Y=y\right]p_y(x)$$
$$\implies V_{u,p_y}(Z) \leq V_{u,p_y}(Z_y)$$
$$\implies I_{q,u}(y;Y,Z) \geq 0$$

This establishes the required statement as the choice of decision problem was arbitrary. □

8.1.1 Continuous Case

In this section we will assume that $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are compact intervals in $\mathbb{R}$. Additionally, we make the following assumptions:

Assumption 1. The conditional distribution function given $X = x$ and $Y = y$, $Pr(Z \leq z|X = x)$ and $Pr(Z \leq z|X = x, Y = y)$ are continuous in z for all x, y are continuous is x for all z, y.

Let $F(y|X = x) = Pr(Y \leq y|X = x)$ and $G(z|x) = Pr(Z \leq z|X = x)$ represent the conditional cdf of Y and Z given $X = x$ respectively and let $H_y(z|x) = Pr(Z \leq z|X = x, Y = y)$ be the conditional cdf of Z given $X = x$ and $Y = y$. Let $f(y|x)$, $g(z|x)$ and $h_y(z|x)$ represent the corresponding probability density functions. Extending definition 9 to continuous random variables we say that the family of distributions $\{G(Z|x)\}_{x\in\mathcal{X}}$ is MLR ordered if $\frac{g(z|x')}{g(z|x)}$ is increasing in z for each $x' \geq x$.

Assumption 2. The family of functions $\{u(x,.)\}_{x\in\mathcal{X}}$ is equicontinous.

The above assumption guarantees the existence of an optimal decision rule mapping $\mathcal{Z}$ to $\mathcal{D}$ for each $y \in \mathcal{Y}$

Proposition 7. *For random variables (X, Y, Z) such that $\{G(Z|x)\}_{x\in\mathcal{X}}$ is MLR ordered, if for all $y \in \mathcal{Y}, z, z' \in \mathcal{Z}$, $x \leq x' \in \mathcal{X}$:*

$$Pr(Z \leq z|X = x) \geq (\leq) Pr(Z \leq z'|X = x, Y = y) \implies Pr(Z \leq z|X = x') \geq (\leq) Pr(Z \leq z'|X = x', Y = y)$$

Then Y is $\mathcal{U}_{IDO}-$ strong complement (substitute) of Z, i.e. for all $y \in \mathcal{Y}$ we have $Z \succeq_Y^{\mathcal{U}_{IDO}} (\preceq_Y^{\mathcal{U}_{IDO}}) Z_y$. Additionally when (X, Y, Z) are affiliated then the converse holds.

Proof. (Sketch)
Define[31] $\tau_y(x, z) := H_y^{-1}(G(z|x)|x)$.
Assume for all $y \in \mathcal{Y}, z, z' \in \mathcal{Z}, x \leq x' \in \mathcal{X}$:

$$G(z|x) \geq (\leq) H_y(z'|x) \implies G(z|x') \geq (\leq) H_y(z'|x') \quad (*)$$

Fix some $y \in \mathcal{Y}$. Assume that $(*)$ holds for all z, z' and $x < x'$. Then we can show the function τ_y defined in step 1 is increasing in x. To this end fix $x < x' \in \mathcal{X}$ and some $z \in \mathcal{Z}$. Let z' be the largest value[32] for which $G(z|x) = H_y(z'|x)$. By $(*)$ we get that $G(z|x') \geq H_y(z'|x')$. The desired statement follows from:

$$\tau_y(x', z) = H_y^{-1}(G(z|x')|x') \geq H_y^{-1}(H_y(z'|x')) \geq z'$$

and

$$z' = H_y^{-1}(H_y(z'|x)) = H_y^{-1}(G(z|x)) = \tau_y(x, z)$$

Fix some $y \in \mathcal{Y}$. Let $\tau_y(x, z)$ to be increasing in x. Assume that for some x, z, z' we have $G(z|x) \geq H_y(z'|x)$, it follows that $\tau_y(x, z) \geq z'$. By definition for any $x' \geq x$ we have $G(z|x') = H_y(\tau_y(x', z)|x')$, thus:

$$G(z|x') = H_y(\tau_y(x', z)|x') \implies G(z|x') \geq H_y(\tau_y(x, z)|x') \implies G(z|x') \geq H_y(z'|x') \text{ for all } x' \geq x$$

The choice of y was arbitrary so we get that $(*)$ holds all $y \in \mathcal{Y}, z, z' \in \mathcal{Z}, x \leq x' \in \mathcal{X}$ if and only if for every $y \in \mathcal{Y}, z \in \mathcal{Z}, \tau_y(x, z)$ is increasing in x. In particular, we have shown that for every $y \in \mathcal{Y}$, H_y is Lehmann-more accurate[33] than G if and only if $(*)$ holds for for every $y \in \mathcal{Y}, z \in \mathcal{Z}$.

[31] $H_y^{-1}(r|x) = \inf\{z : H_y(z|x) > r\}, 0 < r < 1$

[32] Such a value exists as $G(|x), H_y(|x)$ are continuous cdf.

[33] We follow the terminology of Persico (2000) and Quah and Strulovici (2009). The concept has been studied in Lehmann (1988) and Yanagimoto and Okamoto (1969).

As the $G\{(Z|X=x)\}_{x\in\mathcal{X}}$ is MLR-ordered[34] from Theorem 3 of Quah and Strulovici (2009) we conclude that

$$\mathbb{E}\left[\max_{d\in\mathcal{D}}\mathbb{E}\left[u(X,d)|Z_{\varnothing}\right]|Y=y\right]\leq\mathbb{E}\left[\max_{d\in\mathcal{D}}\mathbb{E}\left[u(X,d)|Z\right]|Y=y\right]$$
$$\implies V_{u,p_y}(Z_y)\geq V_{u,p_y}(Z)$$

For the converse note that affiliation implies that $\frac{h_y(z|x')}{h_y(z|x)}$ is increasing in z for each $x'\geq x\in\mathcal{X}$ and $y\in\mathcal{Y}$. The result then follows from proposition 11 in Quah and Strulovici (2009).

□

Remark 16. Roughly speaking the condition in proposition 7 says that Y strongly complements Z for IDO preferences if knowing any realization $Y=y$ makes Z and X more positively dependent compared to not knowing Y. Where positive dependence captures the idea that higher values of X and Z are more likely to occur together. More concretely, consider the following stronger condition: [35] for all $y\in\mathcal{Y},z,z'\in\mathcal{Z},x\leq x'\in\mathcal{X}$

$$P(Z\leq z|X=x)-P(Z\leq z'|X=x,Y=y)\geq P(Z\leq z|X=x')-P(Z\leq z'|X=x',Y=y)$$

As (X,Y,Z) are affiliated we get $P(Z|X=x')\succeq_{FOSD}P(Z|X=x)$ and $P(Z|X=x',Y=y)\succeq_{FOSD}P(Z|X=x,Y=y)$. Thus the above inequality can be interpreted as for all values of $Y=y$, Z conditional on $X=x,Y=y$ increases faster in FOSD sense than Z conditional on $X=x$ as x increases.

9 Applications

9.1 Delayed Acquisition

In this section, we look at an application similar to section 3.2. Consider the following information acquisition problem:

1. In the first period, the DM chooses between acquiring a signal Z at some cost $c(q,Y)\geq 0$ or delaying the acquisition at a cost $\varepsilon>0$.
2. In the second period
   (a) If Z was acquired then the decision maker observes realization (y,z) of the bivariate random variable (Y,Z).
   (b) If the acquisition is delayed, the agent first observes the realization y of Y and then decides to acquire signal Z at cost $c(q,y)\geq 0$
3. Let[36] $\mathcal{D}$ be a finite decision space. In the third period:
   (a) If the agent acquires Z, then after observing $(Y=y,\ Z=z)$ the decision maker chooses $d\in\arg\max_{\delta\in\mathcal{D}}\mathbb{E}[u(X,\delta)|Y=y,\ Z_i=z]$.
   (b) If the agent chooses not to acquire Z, then after observing $Y=y$ the decision maker chooses $d\in\arg\max_{\delta\in\mathcal{D}}\mathbb{E}[u(X,\delta)|Y=y]$.

The cost $c(q,Y)=V_{q,u}(Y,Z)-V_{q,u}(Y)$ represents the price of signal Z in period 1. The interpretation of the cost is that if the prior, utility and expectation of learning Y are common knowledge among the seller

[34] Note the same conclusion also follows by adapting the proof of proposition 5 to continuous $\mathcal{X}$.
[35] The condition in proposition 7 is a single crossing condition on the difference of cdfs whereas the condition in the remark is an increasing difference condition on the same difference of cdfs.
[36] We restrict attention to pure strategies for clarity.

and buyer of the information then $V_{p,u}(Y,Z) - V_{q,u}(Y)$ is the highest amount the buyer is willing to pay to learn Z on top of Y.

After the realization of information $Y = y$, the new beliefs about the state p_y and the utility u are commonly known among the buyer and the seller. The cost $c(q,y) = V_{p_y,u}(Z)$ represents the price of signal Z in period 2 after observing the realization $Y = y$. Here, we make the implicit assumption that the seller doesn't know if the buyer actually saw $Y = y$ or not, more specifically the seller cannot elicit the buyer's knowledge of the correlation between Y and Z. Thus the seller can't use this correlation to price Z.

If there is no delay cost the decision maker always weakly prefers to delay the decision to acquire information. With the presence of delay cost $\varepsilon > 0$, the decision maker might have the incentive to forgo a contingent decision in favor of taking advantage of mispricing by the seller.

The following propositions clarify the relationship between complementarity and the decision to delay information acquisition.

Proposition 8. *If Y and Z are complements then there exists a small enough delay cost $\varepsilon > 0$ such that the decision maker prefers to delay his acquisition decision.*

Proposition 9. *If Y is a strong substitute of Z then for all delay costs $\varepsilon > 0$, the decision maker prefers to acquire Z today rather than delaying.*

Proof. If no delay: the expected utility of the agent is $V_{u,q}(Y)$. If delay: $\sum_{\mathcal{Y}_+} I_{u,q}(y;Y,Z)p(y) + V_{u,q}(Y) - \varepsilon$; where $\mathcal{Y}_+ := \{y \in \mathcal{Y} | I(y;Y,Z) \geq 0\}$. Strong substitute implies that $\mathcal{Y}_+$ has zero probability, and complementarity implies that $\mathcal{Y}_+$ has strictly positive probability. □

References

Ahlswede, Rudolf and David E Daykin, "An inequality for the weights of two families of sets, their unions and intersections," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 1978, *43*, 183–185.

Athey, Susan, "Monotone comparative statics under uncertainty," *The Quarterly Journal of Economics*, 2002, *117* (1), 187–223.

Blackwell, David, "Comparison of experiments," in "Proceedings of the second Berkeley symposium on mathematical statistics and probability," Vol. 2 University of California Press 1951, pp. 93–103.

_ , "Equivalent comparisons of experiments," *The annals of mathematical statistics*, 1953, pp. 265–272.

Blackwell, David A and Meyer A Girshick, *Theory of games and statistical decisions*, Courier Corporation, 1979.

Börgers, Tilman, Angel Hernando-Veciana, and Daniel Krähmer, "When are signals complements or substitutes?," *Journal of Economic Theory*, 2013, *148* (1), 165–195.

Brooks, Benjamin, Alexander Frankel, and Emir Kamenica, "Comparisons of Signals," Technical Report, Working paper.[2198] 2022.

de Castro, Luciano I, "Affiliation and dependence in economic models," Technical Report, Discussion paper 2009.

Degroot, Morris H, "Changes in utility as information," *Theory and Decision*, 1994, *17*, 287–303.

Eaton, Morris L, "A review of selected topics in multivariate probability inequalities," *The Annals of Statistics*, 1982, pp. 11–43.

Hansen, Ole Havard and Erik N Torgersen, "Comparison of linear normal experiments," *The Annals of Statistics*, 1974, pp. 367–373.

Karlin, Samuel and Herman Rubin, "The theory of decision procedures for distributions with monotone likelihood ratio," *The Annals of Mathematical Statistics*, 1956, pp. 272–299.

Le, L, "Sufficiency and approximate sufficiency," *The Annals of Mathematical Statistics*, 1964, pp. 1419–1455.

Lehmann, EL, "Comparing location experiments," *Annals of statistics*, 1988, *16* (2), 521–533.

Liang, Annie and Xiaosheng Mu, "Complementary information and learning traps," *The Quarterly Journal of Economics*, 2020, *135* (1), 389–448.

Milgrom, Paul R and Robert J Weber, "A theory of auctions and competitive bidding," *Econometrica: Journal of the Econometric Society*, 1982, pp. 1089–1122.

Persico, Nicola, "Information acquisition in auctions," *Econometrica*, 2000, *68* (1), 135–148.

Quah, John K-H and Bruno Strulovici, "Comparative statics, informativeness, and the interval dominance order," *Econometrica*, 2009, *77* (6), 1949–1992.

Strasser, Helmut, *Mathematical Theory of Statistics*, Berlin, New York: De Gruyter, 1985.

Topkis, Donald M, *Supermodularity and complementarity*, Princeton university press, 1998.

Torgersen, Erik, *Comparison of statistical experiments*, Vol. 36, Cambridge University Press, 1991.

Yanagimoto, Takemi and Masashi Okamoto, "Partial orderings of permutations and monotonicity of a rank correlation statistic," *Annals of the Institute of Statistical Mathematics*, 1969, *21*, 489–506.