Commonality of Information and Commonality of Beliefs*

Yu Awaya† and Vijay Krishna‡

May 19, 2022

Abstract

A group of agents with a common prior receive informative signals about an unknown state repeatedly over time. If these signals were public, agents' beliefs would be identical and commonly known. This suggests that if signals were private, then the more correlated these are, the greater the commonality of beliefs. We show that, in fact, the opposite is true. In the long run, conditionally independent signals achieve greater commonality of beliefs than correlated ones. We then apply this result to binary-action, supermodular games.

1 Introduction

What kind of information increases the possibility of coordination? If a group of agents with a common prior receive *public* signals about an unknown state, they will have identical beliefs and moreover, these beliefs will be commonly known, thereby facilitating coordination. This suggests that if agents' signals are *private*, then the more correlated these are, the easier it will be for agents to coordinate.

Here we examine this intuition in the context of the common learning framework of Cripps, Ely, Mailath and Samuelson (2008, henceforth CEMS), where informative signals comes repeatedly over time. CEMS (2008) showed that if agents' signals were independently and identically distributed over time, then regardless of the degree of correlation among agents' signals, the realized state would, in the limit, become (approximately) commonly known. Frick, Iijima and Ishii (2022) have recently shown that when the number of signals each individual sees is large enough, the rate of

*The research reported here was supported by a grant from the National Science Foundation (SES-2048806). We are grateful to Paulo Barelli, Mira Frick, Tetsuya Hoshino, Ryota Iijima, Yuhta Ishii and Zeming Ma for helpful comments.

†Department of Economics, University of Rochester, E-mail: yuawaya@gmail.com.

‡Department of Economics, Penn State University, E-mail: vkrishna@psu.edu.

common learning is the same regardless of the degree of correlation among agents' signals.

Does correlation have any role to play in determining the commonality of beliefs? We begin by examining this question in the context of a canonical game where a high degree of common belief is needed for coordination.

Example 1 Two players simultaneously choose whether to invest or not in the face of uncertainty. Specifically, there are two equally-likely states of nature "good" (G) or "bad" (B). The cost of investment is c and a player's investment is successful and yields a gross return of 1 if and only if the state is G and the other player also invests. If a player invests and the other does not, then the gross return is 0.

The information available to players is generated as follows. Let $\boldsymbol{X} = (X_1, X_2) \in \{0,1\}^2$ be a pair of binary random variables. In state G, X_1 and X_2 are symmetrically and *independently* distributed with $\Pr[X_i = 0 \mid G] = \frac{1}{5}$. In state B, the joint distribution of the signals is degenerate—with probability 1, both players receive a signal of 0.

Prior to making decisions, player i sees *two* serially independent realizations of X_i, say X_i^1 and X_i^2. It is routine to verify that if $c < \frac{24}{25}$, then there is an equilibrium in which player i invests if $X_i^1 + X_i^2 \geq 1$.

Now consider an alternative situation in which the players' signals are correlated. Specifically, suppose $\boldsymbol{Y} = (Y_1, Y_2) \in \{0,1\}^2$ is a pair of random variables that in state G, have the joint distribution

	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	$\frac{3}{25}$	$\frac{2}{25}$
$Y_1 = 1$	$\frac{2}{25}$	$\frac{18}{25}$

Notice that while the marginal distributions of Y_i and X_i are the same, the signals $\boldsymbol{Y}$ are positively *correlated*. In state B, the joint distribution of (Y_1, Y_2) is again degenerate, with $\Pr[(Y_1, Y_2) = (0,0) \mid B] = 1$.

Again, there are two serially independent realizations of $\boldsymbol{Y}$, say $\boldsymbol{Y}^1$ and $\boldsymbol{Y}^2$. Player i observes Y_i^1 and Y_i^2 prior to making an investment decision. Now we claim that if $c > \frac{47}{50}$, then the unique equilibrium is for neither player to invest regardless of her information. This follows from a standard infection argument. First, if $Y_i^1 + Y_i^2 = 0$ then, given the cost, it is dominant to not invest because $\Pr[G \mid Y_i^1 + Y_i^2 = 0] = \frac{1}{6} < c$. Next, if $Y_i^1 + Y_i^2 = 1$, it is iteratively dominant to not invest either because if $j \neq i$, $\Pr\left[Y_j^1 + Y_j^2 \geq 1 \mid Y_i^1 + Y_i^2 = 1\right] = \frac{47}{50} < c$ as well. Finally, given the behavior of those with $Y_j^1 + Y_j^2 \leq 1$, it is optimal even for a player with $Y_i^1 + Y_i^2 = 2$ to not invest because $\Pr\left[Y_j^1 + Y_j^2 = 2 \mid Y_i^1 + Y_i^2 = 2\right] = \frac{81}{100} < c$.

So we obtain that if $c \in \left(\frac{47}{50}, \frac{48}{50}\right)$, under two independent signals $\boldsymbol{X}$, there is an equilibrium in which both players invest when they know the state is G while under two correlated signals $\boldsymbol{Y}$, the unique equilibrium is that no player ever invests. Thus, for these costs, *correlated information signals hinders coordination*!

Why is this? Compared to the case of (conditionally) independent signals $\boldsymbol{X}$, with correlated signals $\boldsymbol{Y}$, a player that gets a positive signal assigns a higher probability that the other player also received a positive signal and becomes optimistic about the prospects of coordination. But the opposite is true for a player that gets a zero signal. With correlated signals, she assigns a higher probability that the other player also received a zero signal and so becomes pessimistic. Here the second effect dominates. In fact, a player with one positive signal and one zero is more pessimistic when signals are correlated than when they are independent, that is,

$$\Pr\left[Y_j^1 + Y_j^2 \geq 1 \mid Y_i^1 + Y_i^2 = 1\right] < \Pr\left[X_j^1 + X_j^2 \geq 1 \mid X_i^1 + X_i^2 = 1\right]$$

Now suppose that players receive T serially independent signals X_i^t and Y_i^t (in the example above $T = 2$) prior to making decisions. It is easy to see that for all $T \geq 2$,

$$\Pr\left[\sum_t Y_j^t \geq 1 \mid \sum_t Y_i^t = 1\right] < \Pr\left[\sum_t X_j^t \geq 1 \mid \sum_t X_i^t = 1\right] \tag{1}$$

and it can be argued in a manner similar to that above, that for any cost c in between the two sides of (1), coordination is possible with the independent signals X but not with the correlated signals Y.

While the common learning result of CEMS (2008) implies that both sides of (1) tend to 1 as $T \to \infty$, away from the limit, correlation reduces the prospects for coordination.

In the remainder of this paper, we explore this phenomenon in the common learning setting of CEMS (2008). There is an unknown fundamental state of nature $\theta \in \{G, B\}$ that is of concern to a group of $I \geq 2$ agents. The state is realized in period 0 and remains fixed. There are T additional periods and in each period t, agents receive private signals X_i^t that are informative about θ. We are interested in the degree of commonality of agents beliefs—that is, how close are the agents to achieving common knowledge about the state of nature θ.

In this paper, we show that the phenomenon demonstrated in the example above is general. Informally stated, our main result is[1]:

Theorem *Commonality of information is detrimental to commonality of beliefs.*

In what follows, "commonality of information" means the degree of correlation among agents' information. Precisely, correlation is measured using a multivariate positive dependence order. "Commonality of beliefs" is formalized using the notion of common p-belief introduced by Monderer and Samet (1989). An event E is common p-believed if (1) everyone assigns at least probability p to E, and (2) also assigns at least probability p to the event that everyone assigns at least probability p to E, and also (3) assigns at least probability p to the event that everyone assigns at least probability p to the event that everyone assigns at least probability p to E and so on.

[1] This is formalized in various settings as Theorems 1, 2 and 3.

Binary and Conclusive Signals We begin by considering a case where agents' signals are (i) *binary* (either 0 or 1); and (ii) *conclusive* in the sense that even one 1-signal reveals that the state is G. This special case is useful because first-order uncertainty—that is, concerning the state θ—is resolved once a 1-signal is received. This means that the focus is then solely on higher-order uncertainty—that is, concerning others' knowledge about θ, etc.[2]

We first show that in our model the type that is most pessimistic about the event that everyone else knows G is one that gets exactly one positive signal. This one positive signal is conclusive evidence that the state is G but the other $T-1$ signals of 0 make this type pessimistic that others know G as well. Fix a T and let q denote the belief of this most pessimistic type about the event that everyone knows G. Now for any $p \leq q$, since the most pessimistic type assigns a probability of at least p to the event that everyone knows G, all other types do so as well. It is then simple to see that if everyone gets at least one positive signal—and so everyone knows G—then this event is common p-believed (Proposition 3.1).

We then show that the converse is true as well: for any $p > q$, it is impossible for G to be common p-believed. The reasoning here is more delicate. Clearly, the most pessimistic type cannot assign a probability higher than p to the event that everyone knows G. But what about more optimistic types? We show that in fact, the pessimism of the type with only one positive signal "infects" the beliefs of all other types and so G cannot be common p-believed (Proposition 3.1).

The final step is to show that higher correlation *decreases* the threshold belief q when T is large (Proposition 3.2). As argued above, the most pessimistic type is the one who receives only one positive signal. Since this type sees a preponderance of 0 signals, higher correlation makes her believe that other agents also received a preponderance of 0 signals, thereby increasing her pessimism. These results then lead to Theorem 1.

Kajii and Morris (1997) have shown a close connection between common p-beliefs and p-dominant equilibria of games, under assumptions equivalent to ours that signals are conclusive.[3] Thus, our result on how commonality of information affects common p-beliefs has implications for information design in games. Specifically, consider a planner who wishes agents to play a particular action profile a^* which is a p-dominant equilibrium of the game in state G, in an equilibrium of the incomplete information game. Our result then suggests that the planner should choose a less-correlated information structure over a more correlated one.

In Section 4 we consider the class of *binary-action, supermodular* games where a tight connection between a generalized version of common beliefs, due to Morris and Shin (2007), and equilibria of games can be made (Oyama and Takahashi, 2020). This class includes, of course, the pure coordination game considered in the introduction

[2] The signals in Rubinstein's E-Mail game are also binary and conclusive.

[3] An action profile a^* is p-dominant if every player i wishes to play a_i^* whenever she assigns a probability p or greater to the event that others will play a_{-i}^*.

but also allows for more general situations of interest—for instance, games of regime change and currency attacks. As an example, consider an investment game in which investment is successful if at least $2/3$ of the players invest and the state is G. In such situations, common p-belief is too strong a requirement because now it is not necessary that *all* players are optimistic enough, only that sufficiently many are. Theorem 2 extends the reach of Theorem 1 to such games and shows that correlation can have a detrimental effect even in situations where common p-belief is not the right notion.

General Signals In Section 5 we relax the assumption that signals are binary and conclusive. In this more general environment, the results are similar but not as sharp. The reason is that first-order uncertainty also plays a role now. We show that one part of Theorem 1 remains almost unchanged but the infection argument underlying the other part no longer goes through. But it is still the case that higher correlation now weakly decreases the threshold beliefs (Theorem 3).

Informativeness Finally, for the case of *two* agents and general signals, we show that our results can be recast in the language of Blackwell informativeness. Say that $\boldsymbol{Y}$ is more informative than $\boldsymbol{X}$, if agent i's signal Y_i is more informative about agent j's signal Y_j than X_i is about X_j (see Section 6 for a precise definition). In the same vein as above, it can be shown that in fact, more informative signals are detrimental to common learning.

Related literature The importance of higher-order uncertainty in game theory was brought to the fore by Rubinstein's (1989) E-Mail game. The literature on common learning asks whether such uncertainty can be made to disappear over time. As mentioned above, Cripps, Ely, Mailath and Samuelson (2008) show that if the set of signals is finite and these are independent over time, then common learning occurs in the limit.[4]

In a subsequent paper, Cripps et al. (2013) the same authors show that common learning may fail if signals are not serially independent and find some more general sufficient conditions for common learning. Steiner and Stewart (2011) consider a version of the common learning model in which signals—which are binary and conclusive—arrive at random times. They ask how communication between agents affects common learning and show that under certain conditions it prevents common learning. In our model, common learning always occurs in the limit. We are interested in examining agents' beliefs away from the limit and how these are affected by correlation.

In the CEMS framework, Frick, Iijima and Ishii (2021) study how common learning is affected by the underlying information structure. Consider two information

[4] They also show that if the set of signals is infinite then common learning may fail if agents' signals are correlated.

structures $\boldsymbol{X}$ and $\boldsymbol{Y}$ such that $\boldsymbol{X}$ is more informative about the state θ than is $\boldsymbol{Y}$. Frick et al. (2021) show that when T is large enough, $\boldsymbol{X}$ results in greater commonality of beliefs than does $\boldsymbol{Y}$. In particular, how correlated either information structure is does not matter in the long run. In our work we compare information structures $\boldsymbol{X}$ and $\boldsymbol{Y}$ that are *equally* informative about θ and show that when T is large enough, correlation is, in fact, detrimental to commonality of beliefs.

There is, of course, a close connection between common beliefs and equilibria of games. This connection has been explored in various manners by Monderer and Samet (1989), Kajii and Morris (1997) and more recently by Oyama and Takahashi (2020). Along these lines, in Section 4 we study the effect of correlation on equilibria of a class of games studied by Oyama and Takahashi (2020).

A related strand of work concerns the problem of information design in games. Mathevet, Perego and Taneva (2020) outline a general framework for this problem and study some interesting examples. Hoshino (2022) considers games that have a **p**-dominant equilibrium—where different players may have different p_i's and where the sum of the p_i's is less than one. He shows that for such games, a designer can always choose an information structure to implement the **p**-dominant equilibrium outcome as the unique rationalizable outcome in a suitable incomplete information version of the original game. Of course, the existence of a **p**-dominant equilibrium satisfying the condition is not guaranteed. Morris, Oyama and Takahashi (2022) study information design in binary-action supermodular games as do Li, Song and Zhao (2019). Our results from Section 4 indicate that a designer who wishes players to coordinate on a **p**-dominant/high-action equilibrium via private signals should choose information structures that are (conditionally) independent. On the other hand, a designer who wishes players to not coordinate, thereby resulting in a unique equilibrium with low actions (as in Li, Song and Zhao, 2019), should choose information structures that are highly correlated.

2 Model

A group of agents $i \in I = \{1, 2, ..., I\}$ face an uncertain fundamental state of nature $\theta \in \Theta$ that can take on two possible values G and B with commonly known prior probabilities $\rho \in (0, 1)$ and $1 - \rho$, respectively. We will suppose that G and B take on numerical values such that $G > B$, say $G = 1$ and $B = 0$.

Time is discrete and there is a finite number of periods, denoted by $t = 0, 1, 2, ...T$. At time $t = 0$, nature chooses $\theta \in \Theta = \{G, B\}$ and this choice remains fixed for all the remaining periods.

At time $t \geq 1$, each agent i receives private information about the state of nature θ.

Specifically, let $\boldsymbol{X} = (X_1, X_2, ..., X_I)$ be a random vector of signals where each X_i takes values in an ordered, finite set $\mathcal{X} = \{0, 1, 2, ..., K\}$. Let $P \in \Delta\left(\Theta \times \mathcal{X}^I\right)$ denote the joint distribution of the state of nature θ and the signals $\boldsymbol{X}$. We will assume that $P(\cdot \mid \theta = G) \neq P(\cdot \mid \theta = B)$ so that the signals carry information about θ. Moreover, conditional on θ, the signals $\boldsymbol{X}$ are *symmetrically* distributed—that is, $P(\boldsymbol{x} \mid \theta) = P(\boldsymbol{x}^{\pi} \mid \theta)$ for any permutation $\boldsymbol{x}^{\pi}$ of $\boldsymbol{x}$.

Let $\boldsymbol{X}^{\theta}$ denote the random vector $\boldsymbol{X}$ in state θ, that is, conditional on θ. We will also suppose that $\boldsymbol{X}^{G}$ has *full support*, that is, for all $\boldsymbol{x} \in \mathcal{X}^{I}$,

$$P(\boldsymbol{x} \mid \theta = G) > 0$$

Note that the full support assumption guarantees that signals are not *public*.

Conditional on θ, the signals $\boldsymbol{X}^{t}$ are independently and identically distributed over time according to the distribution $P(\cdot \mid \theta)$. At any time t, agent i privately sees the realization of X_i^t.

Thus, the signals X_i^t are independent across time but may be correlated across agents.

In what follows, we will assume that

Condition 1 *The random variables* $(\theta, \boldsymbol{X})$ *are* affiliated*, that is,*

$$P(\theta, \boldsymbol{x}) \times P(\theta', \boldsymbol{x}') \leq P(\theta \vee \theta', \boldsymbol{x} \vee \boldsymbol{x}') \times P(\theta \wedge \theta', \boldsymbol{x} \wedge \boldsymbol{x}')$$

where $\boldsymbol{x} \vee \boldsymbol{x}'$ *is the component-wise maximum of* $\boldsymbol{x}$ *and* $\boldsymbol{x}'$ *and* $\boldsymbol{x} \wedge \boldsymbol{x}'$ *is the component-wise minimum.*

Correlation In what follows, we will compare two *information structures* $(\theta, \boldsymbol{X})$ and $(\theta, \boldsymbol{Y})$, say, such that the signals $\boldsymbol{Y} \in \mathcal{X}^{I}$ are "more correlated" than $\boldsymbol{X}$. We will only compare signals with the property that for all $k \in \mathcal{X}$ and $\theta \in \Theta$,

$$\Pr[X_i = k \mid \theta] = \Pr[Y_i = k \mid \theta] \tag{2}$$

This condition guarantees that

$$\Pr[\theta = G \mid X_i = k] = \Pr[\theta = G \mid Y_i = k]$$

as well, so that agents' beliefs about θ are the *same* in $(\theta, \boldsymbol{X})$ as in $(\theta, \boldsymbol{Y})$.

Since correlation itself is a bivariate concept, we will use the following multivariate generalization of positive correlation (see Shaked and Shantikumar, 2008).

Definition 1 $\boldsymbol{Y}$ *is greater than* $\boldsymbol{X}$ *in the* positive quadrant dependence *(PQD) order, written* $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$*, if for any* $\boldsymbol{z} \in \mathcal{X}^{I}$,

$$\Pr[\boldsymbol{X} \leq \boldsymbol{z}] \leq \Pr[\boldsymbol{Y} \leq \boldsymbol{z}] \tag{3}$$

and

$$\Pr[\boldsymbol{X} \geq \boldsymbol{z}] \leq \Pr[\boldsymbol{Y} \geq \boldsymbol{z}] \tag{4}$$

The PQD order is weaker than all other orders of positive dependence discussed in Shaked and Shantikumar (2008) and Meyer and Strulovici (2012). For instance, the PQD order is weaker than the supermodular order or the weak associative order.

Note that it follows from (3) and (4) that if $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$, then $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same univariate marginals, that is, for all z, $\Pr[X_i \leq z] = \Pr[Y_i \leq z]$.

In what follows, we will use the following strict version of the PQD order. We will say that $\boldsymbol{Y}$ is *strictly* greater than $\boldsymbol{X}$ in the PQD order, and write $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$, if the inequality (3) is strict for $\boldsymbol{z} \neq (K, K, ..., K)$ and the inequality (4) is strict for $\boldsymbol{z} \neq 0$.

Common beliefs A state of the world ω, hereafter simply *state*,

$$\omega = \left(\theta, \boldsymbol{x}^1, \boldsymbol{x}^2, ..., \boldsymbol{x}^T\right)$$

determines the state of nature θ as well as a list of the agents' signal realizations $\boldsymbol{x}^t \in \mathcal{X}^I$ (slanted bold $\boldsymbol{x}$) in each period. Alternatively, we can write $\omega = (\theta, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_I)$ where $\mathbf{x}_i \in \mathcal{X}^T$ (upright bold $\mathbf{x}$) is a list of the T signals received by i. The set of states is denoted by

$$\Omega = \Theta \times \mathcal{X}^I \times ... \times \mathcal{X}^I$$

Following Monderer and Samet (1989), given any event $E \subseteq \Omega$ and probability p, the event $B_i^p(E)$ consists of the set of states $\omega \in \Omega$ in which i *p-believes* E, that is, i assigns probability exceeding p to the event E given her information $\mathbf{x}_i$. Next, write $B^p(E) = \cap_i B_i^p(E)$ as the set of states in which everyone p-believes E.

Now for $r = 1, 2, ...$ define the operator $B^{p,r}$ recursively by

$$B^{p,r}(E) = B^p\left(B^{p,r-1}(E)\right)$$

where $B^{p,0}(E) = E$ and finally,

$$C^p(E) = \cap_r B^{p,r}(E)$$

Thus, $C^p(E)$ is the set of states in which E is *common p-believed*. In other words, (i) everyone assigns probability exceeding p to the event E, and also (ii) assigns probability exceeding p to the event that everyone assigns probability exceeding p to the event E, and also (iii) assigns probability exceeding p to the event that everyone assigns probability exceeding p to the event that everyone assigns probability exceeding p to the event E, and so on.

In what follows, we will be interested in the set $C^p\left(\Omega^G\right)$ after T periods, where $\Omega^G = \{\omega : \theta = G\}$. In other words, we will be interested in the set of states where G is common p-believed.

The common learning result of CEMS (2008) implies that for any $p < 1$,

$$\lim_{T\to\infty} \Pr\left[C^p\left(\Omega^G\right) \mid \theta = G\right] = 1$$

3 Binary and Conclusive Signals

We begin by considering a special case of the general model in which

1. signals are *binary*, so that $\mathcal{X} = \{0, 1\}$; and
2. a signal $X_i = 1$ is *conclusive* about G—that is, $\Pr[X_i = 0 \mid B] = 1$.

Because signals are independently and identically distributed over time, an agent's type can effectively be represented by the *total* number of 1-signals received, that is, with binary signals, a type can be represented simply as $n_i = \sum_t x_i^t$. This means that with binary signals the set of types of can be linearly ordered.

The assumption of conclusive signals allows us to focus solely on *higher-order* uncertainty—an agent who gets even one signal $x_i^t = 1$ knows for sure that the state is G but remains unsure about whether others know G, whether others know that she knows G, etc.

This higher-order uncertainty is captured via agents' beliefs about the set

$$\Omega^+ = \{\omega : \forall j, n_j \geq 1\}$$

that is, the set of states in which every agent j received a signal of $x_j^t = 1$ at some time t. Since even one signal is conclusive about G, at any state in Ω^+ everyone knows that the state is G. Formally, $\Omega^+ \subseteq \Omega^G$.

3.1 Main result

The main result of this section, for binary and conclusive signals is

Theorem 1 *If $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$, then for T large enough,*
(i) for all p,

$$C^p_{\mathbf{Y}}\left(\Omega^G\right) \subseteq C^p_{\mathbf{X}}\left(\Omega^G\right)$$

(ii) for p in a non-empty open interval, under $\mathbf{Y}$, G cannot be common p-believed:

$$C^p_{\mathbf{Y}}\left(\Omega^G\right) = \varnothing$$

whereas under $\mathbf{X}$, G is common p-believed whenever everyone knows G:

$$C^p_{\mathbf{X}}\left(\Omega^G\right) = \Omega^+$$

Theorem 1 says that for large T, greater commonality of information actually *reduces* the commonality of beliefs. Before proving the theorem, a few remarks are in order.

First, since we have assumed that $\mathbf{Y}^G$ has *full support*, $(\theta, \mathbf{Y})$ is *not* a public information structure—that is, the signals are not perfectly correlated. If the signals

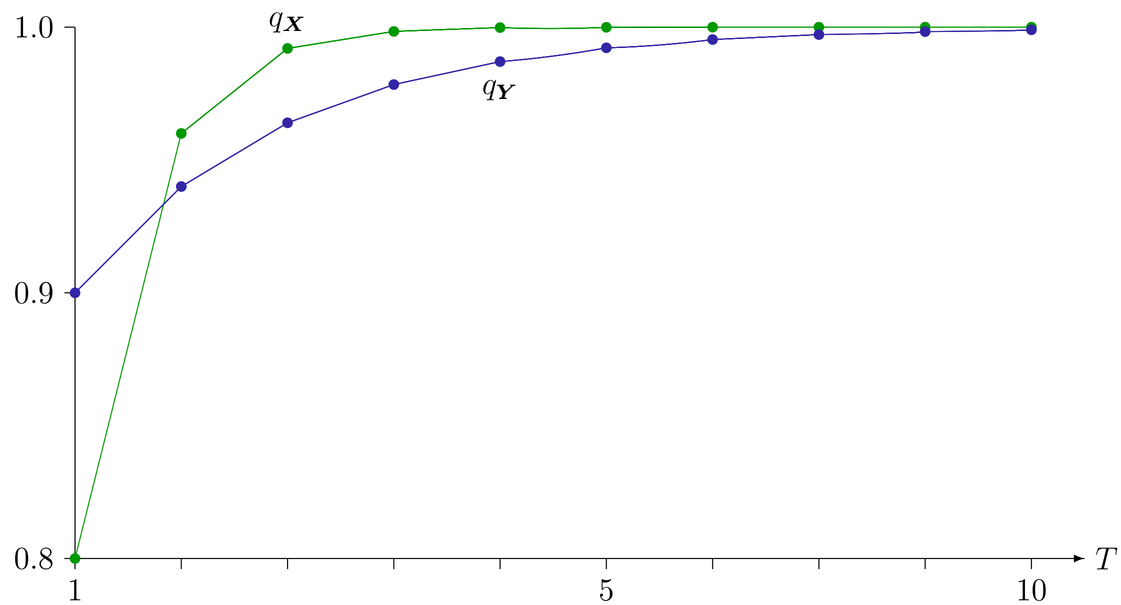

Figure 1: Threshold Beliefs for the Information Structures in Example 1

$\boldsymbol{Y}$ were perfectly correlated—that is, for all $k \in \mathcal{X}$, $\Pr[Y_j = k \mid Y_i = k] = 1$—then we would have that for all p, $C^p_{\boldsymbol{Y}}(\Omega^G) = \Omega^+$, which would run counter to (ii). But what if $\boldsymbol{Y}$ is "nearly" perfectly correlated—that is, for some small ε, for all $k \in \mathcal{X}$, $\Pr[X_j = k \mid X_i = k] > 1 - \varepsilon$? Is there a discontinuity at $\varepsilon = 0$? Here the order of quantifiers in the theorem is important. For a fixed T, it may be that if $\boldsymbol{Y}$ is nearly perfectly correlated, it leads to greater commonality of beliefs than $\boldsymbol{X}$. What the theorem says is that this cannot persist once T is large enough. Figure 1 depicts the threshold beliefs $q_{\boldsymbol{X}}$ and $q_{\boldsymbol{Y}}$ as functions of T for the two information structures in Example 1 from the Introduction—the (conditionally) independent structure $(\theta, \boldsymbol{X})$ and the correlated $(\theta, \boldsymbol{Y})$.

Second, part (i) of the theorem does not imply that the probability of the event that G is common p-believed under $\boldsymbol{Y}$ is smaller than that under $\boldsymbol{X}$. This is because $\boldsymbol{X}$ and $\boldsymbol{Y}$ have different probability distributions. Of course, in part (ii) under $\boldsymbol{Y}$ the probability of common p-belief is zero while under $\boldsymbol{X}$ it is positive.

Third, the result does not conflict with the CEMS (2008) result that common learning occurs in the limit regardless of the commonality of signals. Theorem 1 requires T to be large enough but not infinite. Note also that T must be at least 2—the conclusion of the theorem cannot hold for $T = 1$.

3.2 Proof of Theorem 1

The proof of Theorem 1 has two components. We first show that with binary, conclusive signals the set $C^p(\Omega^G)$ has a "bang-bang" property. When p is above a certain threshold, $C^p(\Omega^G)$ is empty—that is, $\theta = G$ cannot be p-believed—and when p is below the threshold, $\Omega^+ \subseteq C^p\left(\Omega^G\right)$—that is, whenever everyone knows $\theta = G$, it is also common p-believed. This is Proposition 3.1 below.

The second step in the proof of Theorem 1 then shows that when T is large enough, the threshold decreases with an increase in the "correlation" among agents' signals—specifically, an increase in positive quadrant dependence. This is Proposition 3.2 below.

As above, let $N_i = \sum_{t=1}^{T} X_i^t$ denote the random variable which is the number of positive signals received by agent i and define

$$\begin{aligned} q &= \Pr\left[\forall j, N_j \geq 1 \mid N_i = 1\right] \\ &= \Pr\left[\Omega^+ \mid N_i = 1\right] \end{aligned} \tag{5}$$

as the belief of type $N_i = 1$ about the event that everyone else saw at least one positive signal—and so also knows that $\theta = G$.

Since signals are affiliated, for all $k \geq 1$,

$$\Pr\left[\Omega^+ \mid N_i = k\right] \geq \Pr\left[\Omega^+ \mid N_i = 1\right] = q \tag{6}$$

as established in Lemma A.2 in the Appendix. In other words, among all those that know that $\theta = G$, the type $N_i = 1$ is most pessimistic about the event that everyone knows $\theta = G$.

3.2.1 Threshold beliefs

Our first result is that q is the "threshold" belief such that if $p \leq q$, then in all states in which everyone gets at least one signal $x_i^t = 1$, the event $\theta = G$ is common p-believed. This is rather intuitive—if the belief about $\Omega^+ = \{\omega : \forall j, n_j \geq 1\}$ of the most pessimistic type exceeds p, then the beliefs of *all* types exceed p and in fact $\theta = G$ is common p-believed.

The result below says that, in a strong sense, the converse is true as well—if the belief about Ω^+ of the most pessimistic type is smaller than p, then it is *impossible* that the event $\theta = G$ is common p-believed. The assumption that signals are binary and conclusive is important for the converse.

Define

$$\rho_0 = \Pr\left[\Omega^G \mid N_i = 0\right]$$

be the belief about G of an agent who only receives 0-signals in each of the T periods. Note that as T increases, ρ_0 goes to zero.

Proposition 3.1 *(i) If $p \leq q$, then*

$$\Omega^+ \subseteq C^p\left(\Omega^G\right)$$

(ii) If $\rho_0 < q < p$, then

$$C^p\left(\Omega^G\right) = \varnothing$$

Proof of Proposition 3.1 (i) Since the right-hand side of (6) is just q, all types of agent i with $n_i \geq 1$ assign at least probability q to the event Ω^+ that everyone got at least one positive signal. Formally,

$$\{\omega : n_i \geq 1\} \subseteq B_i^q\left(\Omega^+\right)$$

and since $\Omega^+ = \{\omega : \forall j, n_j \geq 1\} \subset \{\omega : n_i \geq 1\}$, we have

$$\Omega^+ \subseteq B_i^q\left(\Omega^+\right)$$

In the language of Monderer and Samet (1989) this says that Ω^+ is *evident q-believed* (or is *q-evident,* for short). Proposition 3 in Monderer and Samet (1989) now implies that Ω^+ is common q-believed at any $\omega \in \Omega^+$. Formally,

$$\Omega^+ \subseteq C^q\left(\Omega^+\right)$$

Since signals are conclusive $\Omega^+ \subset \Omega^G$. Moreover, since the C^q operator is monotone and $p \leq q$, we have

$$\Omega^+ \subseteq C^p\left(\Omega^G\right)$$

This completes the proof of part (i) of Proposition 3.1.

Proof of Proposition 3.1 (ii) Now suppose $p > q$. We will argue that now $C^p\left(\Omega^+\right) = \varnothing$ and then that $C^p\left(\Omega^G\right) = \varnothing$ as well.

As a first step, we show that every additional signal makes an agent more *pessimistic* about the event that all other agents got an *additional* signal as well.

Lemma 3.1 *Suppose signals are binary. For any $k \geq 1$,*

$$\Pr\left[\forall j, N_j \geq k+1 \mid N_i = k+1\right] \leq \Pr\left[\forall j, N_j \geq k \mid N_i = k\right]$$

Proof. Again, since signals are serially independent, without loss of generality, suppose that the conditioning events are such that $\sum_{t=1}^{T-1} X_i^t = k$ and then on the left-hand side $X_i^T = 1$ whereas on the right-hand side $X_i^T = 0$. In other words, the additional 1-signal received by i occurs in period T.

For $j \neq i$,define $M_j = \sum_{t=1}^{T-1} X_j^t$ to be the random variable that is the sum of the first $T-1$ signals received by j and let $\boldsymbol{M}_{-i} = \sum_{t=1}^{T-1} \boldsymbol{X}_{-i}^t$ denote the corresponding vector random variable of the sum of the signals of agents other than i. Then $N_j = M_j + X_j^T$.

If $M_j < k-1$, then both $N_j = k$ and $N_j = k+1$ are impossible.

If $M_j = k-1$, then $N_j = k$ is possible while $N_j = k+1$ is impossible.

If $M_j = k$, then $N_j = k$ occurs with probability 1 while $N_j = k+1$ occurs with probability less than one.

If $M_j > k$, then both $N_j = k$ and $N_j = k+1$ occur with probability one.

Thus, in all cases the probability that $N_j \geq k$ occurs is at least as large as the probability that $N_j \geq k+1$ occurs and so for all $\boldsymbol{m}_{-i}$,

$$\begin{aligned} & \Pr\left[\forall j, m_j + X_j^T \geq k+1 \mid \sum_{t=1}^{T-1} X_i^t = k, X_i^T = 1\right] \\ \leq \ & \Pr\left[\forall j, m_j + X_j^T \geq k \mid \sum_{t=1}^{T-1} X_i^t = k, X_i^T = 0\right] \end{aligned}$$

Finally, since the probability distribution of $\boldsymbol{M}_{-i} = \sum_{t=1}^{T-1} \boldsymbol{X}_{-i}^t$ is independent of X_i^T, integrating both sides of the inequality over the m_j, we have

$$\begin{aligned} & \Pr\left[\forall j, M_j + X_j^T \geq k+1 \mid \sum_{t=1}^{T-1} X_i^t = k, X_i^T = 1\right] \\ \leq \ & \Pr\left[\forall j, M_j + X_j^T \geq k \mid \sum_{t=1}^{T-1} X_i^t = k, X_i^T = 0\right] \end{aligned}$$

which establishes the result. ■

Lemma 3.2 *Suppose signals are binary and conclusive and $\rho_0 < q$. For any $p > q$,*

$$C^p\left(\Omega^+\right) = \varnothing$$

Proof. Define

$$\Omega^{(k)} = \{\omega : \exists j, n_j < k\}$$

as the set of states in which at least one agent gets fewer than k signals. Clearly, for any k, $\Omega^{(k)} \subset \Omega^{(k+1)}$ and $\cup_{k=1}^{T+1} \Omega^{(k)} = \Omega$.

We will argue by induction that if $p > q$, then for all k,

$$\Omega^{(k)} \cap C^p\left(\Omega^+\right) = \varnothing \tag{7}$$

First, since $\Omega^{(1)} = \{\omega : \exists j, n_j = 0\}$, in any state $\omega \in \Omega^{(1)}$, there is an i who never gets a positive signal and so assigns probability ρ_0 to Ω^G. Since signals are conclusive, $\Omega^+ \subset \Omega^G$, and so i assigns a probability no more than ρ_0 to Ω^+. Since $\rho_0 < q < p$, this implies that any such $\omega \notin B_i^p(\Omega^+)$ and so $\omega \notin C^p(\Omega^+)$. Thus, $\Omega^{(1)} \cap C^p(\Omega^+) = \varnothing$

Next, $\Omega^{(2)} = \{\omega : \exists j, n_j < 2\}$. Let $\omega \in \Omega^{(2)} \setminus \Omega^{(1)}$. At any such ω, there is an i such that $n_i = 1$. But by definition, $\Pr[\Omega^+ \mid N_i = 1] = q$ and so for any $p > q$, $\omega \notin B_i^p(\Omega^+)$ and hence $\omega \notin C^p(\Omega^+)$ as well. Thus, we have shown that $\Omega^{(2)} \cap C^p(\Omega^+) = \varnothing$.

Now suppose that $\Omega^{(k)} \cap C^p(\Omega^+) = \varnothing$. This implies that

$$C^p\left(\Omega^+\right) \subset \Omega \diagdown \Omega^{(k)} = \{\omega : \forall j,\ N_j \geq k\}$$

Let $\omega \in \Omega^{(k+1)} \setminus \Omega^{(k)}$. At any such ω, there is an i with $n_i = k$, that is, i gets exactly k 1-signals. By the induction hypothesis

$$\Pr\left[C^p\left(\Omega^+\right) \mid N_i = k\right] \leq \Pr\left[\forall j,\ N_j \geq k \mid N_i = k\right]$$

Lemma 3.1 now implies that

$$\Pr\left[C^p\left(\Omega^+\right) \mid N_i = k\right] \leq \Pr\left[\forall j,\ N_j \geq 1 \mid N_i = 1\right] = q$$

and so $\omega \notin B_j^p\left(C^p\left(\Omega^+\right)\right)$ and hence $\omega \notin C^p\left(\Omega^+\right)$. Thus, we have argued that $\Omega^{(k+1)} \cap C^p\left(\Omega^+\right) = \varnothing$ and hence established (7).

Now since $\Omega^{(k)} \cap C^p\left(\Omega^+\right) = \varnothing$ for all k and $\cup_{k=1}^{\infty}\Omega^{(k)} = \Omega$, we have that $C^p\left(\Omega^+\right) = \varnothing$. ■

To complete the proof of Proposition 3.1 (ii), first note that if $p > q > \rho_0$, then for all i,

$$B_i^p\left(\Omega^G\right) = \{\omega : n_i \geq 1\} \tag{8}$$

that is, i assigns probability exceeding p to G if and only if i gets at least one positive signal. To see this, observe that if i gets a signal, then she knows for sure that $\theta = G$ and so assigns probability 1 to $\theta = G$. On the other hand, if i did not get a positive signal, then the posterior probability of $\theta = G$ is $\rho_0 < p$ and so $B_i^p\left(\Omega^G\right) \subseteq \{\omega : n_i \geq 1\}$.

Now using (8),

$$\begin{aligned} B^p\left(\Omega^G\right) &= \cap_i B_i^p\left(\Omega^G\right) \\ &= \{\omega : \forall i, n_i \geq 1\} \\ &= \Omega^+ \end{aligned}$$

and since $C^p\left(\Omega^G\right) = \cap_r B^{p,r}\left(\Omega^G\right)$ we have from Lemma 3.2,

$$C^p\left(\Omega^G\right) = C^p\left(\Omega^+\right) = \varnothing$$

This completes the proof of Proposition 3.1.

3.2.2 Correlation increases pessimism

Proposition 3.1 establishes that with binary and conclusive signals, the maximum commonality of beliefs—that is, the highest p for which Ω^G can be common p-believed—is exactly q, the belief of the most pessimistic type among all those who know that $\theta = G$. In this section, we compare two information structures $(\theta, \boldsymbol{X})$ and $(\theta, \mathbf{Y})$ such that $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$. We show that a change from $\boldsymbol{X}$ to $\boldsymbol{Y}$ increases the pessimism of the most pessimistic type $N_i = 1$.

For binary and conclusive signals, we then have

Proposition 3.2 *If $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$, then for T large enough,*

$$q_{\boldsymbol{X}} = \Pr_{\boldsymbol{X}}\left[\Omega^+ \mid N_i^X = 1\right] > \Pr_{\boldsymbol{Y}}\left[\Omega^+ \mid N_i^Y = 1\right] = q_{\boldsymbol{Y}}$$

where $N_i^X = \sum_t X_i^t$ and $N_i^Y = \sum_t Y_i^t$.

Proof. Follows from Lemma A.3 and Lemma B.1 (see (13)) in the Appendix. ■

The result is rather intuitive. Consider a type $N_i = 1$ who gets one 1-signal in period 1 and in every subsequent period $t > 1$ gets signal 0. What happens if signals become more correlated? At the end of period 1, under $\boldsymbol{Y}$, this type is *more* optimistic about the event that other agents also know G. However, when T is large this initial optimism is overwhelmed by the increased pessimism resulting from a string of $T-1$ zeros. Formally, while

$$\Pr[X_j = 1 \mid X_i = 1] < \Pr[Y_j = 1 \mid Y_i = 1]$$

at the same time

$$\Pr[X_j = 1 \mid X_i = 0] > \Pr[Y_j = 1 \mid Y_i = 0]$$

and for large enough T, the second inequality dictates the effect of greater "correlation" on the beliefs of type $N_i = 1$.

Propositions 3.1 and 3.2 together prove Theorem 1 since part (ii) of the theorem holds if $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$ and when T is large enough, $\rho_0 = \Pr\left[\Omega^G \mid N_i = 0\right] < q_{\boldsymbol{Y}}$.

4 Binary-action Supermodular Games

Consider any *binary* action, *symmetric, supermodular* game Γ where each player i chooses an action $a_i \in \{0, 1\}$. Because of symmetry we can write the payoff function,

$$u_i(a_1, a_2, ..., a_I) = u_i(a_i, s_{-i})$$

where s_{-i} is the number of other players playing $a_j = 1$. It will be convenient to write

$$u_i(1, s_{-i}) - u_i(0, s_{-i}) = h(s_{-i}) - c$$

where c is a parameter. Supermodularity ensures that $h(\cdot)$ is increasing and we suppose that

$$h(I-1) > c$$

In other words, everyone playing $a_i = 1$ is an equilibrium of Γ.

Following Kajii and Morris (1997), let $\mathcal{G}$ be an *elaboration* of the complete information game where there are two states $\theta = \{G, B\}$ and where in state G, the game is Γ and in state B, the game is Γ' in which action $a_i = 0$ is strictly dominant. Oyama and Takahashi (2020) have studied the relationship between a generalized version of common beliefs and equilibria of binary-action supermodular games.

As in the previous section, suppose signals are *binary* and *conclusive*. Player types are again $N_i = \sum X_i^t$.

Consider the strategies that specify: play $a_i = 1$ if $N_i \geq 1$. If such strategies constitute an equilibrium of $\mathcal{G}$, then we will call such an equilibrium *maximal.* This is because in such an equilibrium, the *only* types that may possibly play $a_i = 0$ are those with $N_i = 0$.

Theorem 2 *If $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$, then for T large enough,*
(i) for all c, if there is a maximal equilibrium under $\mathbf{Y}$, then there is also such an equilibrium under $\mathbf{X}$;
(ii) there is a non-empty, open set of c's such that there is a maximal equilibrium under $\mathbf{X}$, while under $\mathbf{Y}$, the only rationalizable outcome is $a_i = 0$.

Theorem 2 extends the reach of Theorem 1 to the game context. For games of pure coordination, like Example 1 in the introduction, Theorem 1 provides conditions under which there is a maximal equilibrium. Precisely, in the example such an equilibrium exists if and only if G is common q-believed for $q \geq c$. Moreover, correlation reduces the threshold belief q. Theorem 2 shows that correlation can have a detrimental effect any binary-action, symmetric, supermodular game.

4.1 Proof of Theorem 2

It is useful to define the set

$$S_{-i}^k = \{j \neq i : N_j \geq k\} \tag{9}$$

as the set of players other than i, who receive at least k positive signals and let $\#S_{-i}^k$ denote the cardinality of S_{-i}^k.

Proposition 4.1 *(i) If $E[h\left(\#S_{-i}^1\right) \mid N_i = 1] \geq c$, then there is a maximal equilibrium of $\mathcal{G}$.*
(ii) If $E[h\left(\#S_{-i}^1\right) \mid N_i = 1] < c$ and $a_i = 0$ is dominant for a player with $N_i = 0$, then the unique rationalizable outcome of $\mathcal{G}$ is that all play $a_i = 0$.

Proof. (i) Suppose $E[h\left(\#S_{-i}^1\right) \mid N_i = 1] \geq c$. Define $H\left(\mathbf{x}_{-i}\right) = h\left(\#S_{-i}^1\right)$ and then note that H is non-decreasing. Since $(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_I)$ are affiliated (Lemma A.1), we have that for any $\mathbf{x}_i \geq \mathbf{e}^1$,

$$E\left[H\left(\mathbf{X}_{-i}\right) \mid \mathbf{X}_i = \mathbf{x}_i\right] \geq E\left[H\left(\mathbf{X}_{-i}\right) \mid \mathbf{X}_i = \mathbf{e}^1\right]$$

which is equivalent to: for any $k \geq 1$,

$$E[h\left(\#S_{-i}^1\right) \mid N_i = k] \geq E[h\left(\#S_{-i}^1\right) \mid N_i = 1]$$

and so if all players $j \neq i$ follow the strategy of playing $a_j = 1$ whenever they receive at least one positive signal, it is a best response for player i to do so as well.

(ii) By assumption, it is dominant for any player with $N_i = 0$ to play $a_i = 0$. Thus only those who receive at least one positive signal may play $a_i = 1$. But since $E[h(\#S^1_{-i}) \mid N_i = 1] < c$, any player that gets exactly one positive signal will also play $a_i = 0$.

Now Lemma B.2 in the Appendix guarantees that $E[h(\#S^2_{-i}) \mid N_i = 2] < E[h(\#S^1_{-i}) \mid N_i = 1]$ and so if those with $N_i = 0$ and $N_i = 1$ play $a_i = 0$, then it is in the interests of those with $N_i = 2$ to do so as well.

Proceeding in this way, repeated application of Lemma B.2 implies that for all k, all those with $N_i = k$ will play $a_i = 0$. ■

Proposition 4.2 *Suppose* $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$. *For T large enough,*

$$E_{\boldsymbol{Y}}[h(\#S^1_{-i}) \mid N_i^Y = 1] < E_{\boldsymbol{X}}[h(\#S^1_{-i}) \mid N_i^X = 1]$$

where $N_i^X = \sum_t X_i^t$ *and* $N_i^Y = \sum_t Y_i^t$.

Proof. From Lemmas A.3 and B.1 in the Appendix we know that if for $s = 0, 1, ..., I-1$, $q_{\boldsymbol{X}}(s) = \Pr[\#S^1_{-i} = s \mid N_i = 1]$ and $q_{\boldsymbol{Y}}(s)$ is similarly defined, then the distribution $q_{\boldsymbol{X}}(\cdot)$ (strictly) stochastically dominates $q_{\boldsymbol{Y}}(\cdot)$. Since $h(\cdot)$ is a non-decreasing, but not constant, function, the result follows. ■

Propositions 4.1 and 4.2, together with the fact that when T is large $\rho_0 = \Pr[G \mid N_i = 0]$ goes to zero, complete the proof of Theorem 2.

5 General Model

The sharp result in Theorem 1 was derived for the case of binary and conclusive signals. With conclusive signals, we were able to focus solely on higher-order uncertainty—that is, agents' beliefs about the beliefs of other agents etc. When signals are not conclusive, first order uncertainty—that is, agents' beliefs about the state θ—also plays a role. We now turn to consider the general case where the set of signals $\mathcal{X} = \{0, 1, 2, ..., K\}$. Here we will assume that conditional on $\theta \in \{G, B\}$, the signals have *full support*.

Let $\mathbf{e}^1 = (1, 0, ..., 0) \in \mathcal{X}^T$ be the unit vector and define

$$q_{\boldsymbol{X}} = \min\left\{\Pr_{\boldsymbol{X}}\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right], \ \Pr_{\boldsymbol{X}}\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right]\right\} \tag{10}$$

where, as before, $\Omega^+ = \{\omega : \forall j, \mathbf{x}_j \neq \mathbf{0}\}$ is the set of states in which everyone got at least one non-zero signal. Thus, $q_{\boldsymbol{X}}$ is the smaller of the belief of type $\mathbf{X}_i = \mathbf{e}^1$ about Ω^+ *and* her belief that $\theta = G$. The latter represents first-order uncertainty.

If signals were conclusive, as in the previous section, then $\Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right] = 1$ and so in that case, the definition above reduces to the one in (5).

Once again we will compare two information structures $(\theta, \boldsymbol{X})$ and $(\theta, \boldsymbol{Y})$ where $\boldsymbol{Y}$ is "more correlated" than $\boldsymbol{X}$. Let $q_{\boldsymbol{X}}$ be defined as above and let $q_{\boldsymbol{Y}}$ be analogously defined for $(\theta, \mathbf{Y})$. Recall that because we only compare information structures where $\Pr[X_i \mid \theta] = \Pr[Y_i \mid \theta]$ (see (2)),

$$\Pr_{\boldsymbol{X}}\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right] = \Pr_{\boldsymbol{Y}}\left[\Omega^G \mid \mathbf{Y}_i = \mathbf{e}^1\right]$$

and note that this probability goes to zero as $T \to \infty$.

5.1 Main result

We will say that $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$ if for each $\theta \in \{B, G\}$, $\mathbf{Y}^\theta \succ_{PQD} \mathbf{X}^\theta$

For general information structures, we have the following result:

Theorem 3 *If $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$, then for T large enough, $q_{\boldsymbol{Y}} \leq q_{\boldsymbol{X}}$ and for $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$,*

$$C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subsetneq C^p_{\boldsymbol{X}}\left(\Omega^G\right) = \Omega^+$$

Comparing Theorem 3 to Theorem 1, we see that the latter reaches a much stronger conclusion. First, part (i) of Theorem 1 has no counterpart in the general model—it is no longer the case that for all $p, C^p_{\mathbf{Y}}\left(\Omega^G\right) \subseteq C^p_{\mathbf{X}}\left(\Omega^G\right)$. For instance, in Example 2 below, for p close to 1, $C^p_{\mathbf{X}}\left(\Omega^G\right) = \varnothing$ while $C^p_{\mathbf{Y}}\left(\Omega^G\right) \neq \varnothing$. Moreover, without the assumption of binary signals, there is no analog of the infection argument underlying part (ii) of Theorem 1 that leads to the conclusion that for $p > q_{\mathbf{Y}}$, $C^p_{\mathbf{Y}}\left(\Omega^G\right) = \varnothing$.

Note however, that the strong conclusion of Theorem 1 would continue to hold if signals were binary and "nearly conclusive". Precisely, suppose $(\theta, \boldsymbol{X})$ and $(\theta, \boldsymbol{Y})$ are binary and conclusive information structures satisfying the conditions of Theorem 1 and that T is such that the conclusion holds. Now if we perturb both information structures so that they are "nearly conclusive"—for a small ε, $\Pr[X_i = 1 \mid B] \leq \varepsilon$—then the conclusion of Theorem 1 would continue to hold.

5.2 Proof of Theorem 3

Like Theorem 1, the proof of Theorem 3 is in two parts. We first prove, for general signals, an analog of Proposition .3.1. With general signals, the conclusion reached is weaker.

Recall that $\rho_0 = \Pr\left[\Omega^G \mid N_i = 0\right]$,

Proposition 5.1 *(i) If $p \leq q$, then*

$$\Omega^+ \subseteq C^p\left(\Omega^G\right)$$

that is, if everyone got at least one positive signal, then G is common p-believed.
(ii) if $\rho_0 < q < p$, then

$$C^p\left(\Omega^G\right) \subsetneq \Omega^+$$

that is, if G is common p-believed, then everyone got at least one positive signal (but there are states where everyone gets at least one positive signal but G is not common p-believed).

Proof. (Part i) Since $p \leq q$ (defined in (10)) we have that $p \leq \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right]$. Since $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_I$ are affiliated (Lemma A.1), this implies that for any $\mathbf{x}_i \neq \mathbf{0}$, $\Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right] \leq \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{x}_i\right]$ and so for any $\mathbf{x}_i \neq \mathbf{0}$, $p \leq \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{x}_i\right]$ as well. Thus,

$$\{\omega : \mathbf{x}_i \neq \mathbf{0}\} \subseteq B_i^p\left(\Omega^G\right)$$

Taking the intersection over i, we have

$$\Omega^+ \subseteq B^p\left(\Omega^G\right)$$

and operating by B_i^p,

$$B_i^p\left(\Omega^+\right) \subseteq B_i^p\left(B^p\left(\Omega^G\right)\right)$$

Moreover, as above, affiliation implies that for any $\mathbf{x}_i \neq \mathbf{0}$, $\Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right] \leq \Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{x}_i\right]$ and so $p \leq \Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{x}_i\right]$ as well. Thus,

$$\{\omega : \mathbf{x}_i \neq \mathbf{0}\} \subseteq B_i^p\left(\Omega^+\right)$$

Taking intersections over i, we have that

$$\Omega^+ \subseteq B^p\left(\Omega^+\right)$$

that is, the event Ω^+ is p-evident. Since $\Omega^+ \subseteq B^p\left(\Omega^G\right)$ it follows that

$$\Omega^+ \subseteq C^p\left(\Omega^G\right)$$

(Part ii) First, note that when $\rho_0 < p$, then

$$C^p\left(\Omega^G\right) \subseteq \Omega^+$$

To see this, note that if $\omega \notin \Omega^+$, then there exists an agent, say 1, such that $\mathbf{x}_1 = \mathbf{0}$ and since $\Pr\left[G \mid \mathbf{X}_1 = \mathbf{0}\right] = \rho_0 < p$,

$$\omega \notin B_1^p\left(\Omega^G\right)$$

and so

$$\omega \notin C^p\left(\Omega^G\right)$$

Thus, we have that $C^p\left(\Omega^G\right) \subseteq \Omega^+$.

Next we argue that the inclusion is strict. In particular, if $\omega' \in \Omega^+$ is such that $\mathbf{x}_1 = \mathbf{e}^1$, then $\omega' \notin C^p\left(\Omega^G\right)$.

There are two cases to consider. Since $p > q$, either (a) $p > \Pr\left[\Omega^G \mid \mathbf{X}_1 = \mathbf{e}^1\right]$ or (b) $p > \Pr\left[\Omega^+ \mid \mathbf{X}_1 = \mathbf{e}^1\right]$ or both.

If (a), then $\omega' \notin B_1^p\left(\Omega^G\right)$ and so $\omega' \notin C^p\left(\Omega^G\right)$.

If (b), then $\omega' \notin B_1^p\left(\Omega^+\right)$ and so $\omega' \notin C^p\left(\Omega^+\right)$. But since

$$C^p\left(\Omega^G\right) \subseteq \Omega^+$$

operating on both sides by C^p and using the fact that $C^p\left(\Omega^G\right)$ is a fixed point of the operator C^P,

$$C^p\left(\Omega^G\right) \subseteq C^p\left(\Omega^+\right)$$

and so $\omega' \notin C^p\left(\Omega^G\right)$. ■

5.2.1 Correlation increases pessimism

Theorem 1 showed that with conclusive signals, an increase in correlation (as measured by the PQD order) made the most pessimistic type even more pessimistic. Here we show that modulo some minor qualifications, the same is true in general—that is, even when signals are not conclusive.

As before we will compare the information structure $(\theta, \boldsymbol{X})$ with another information structure $(\theta, \boldsymbol{Y})$ with the same level of *individual* learning. Lemmas A.3 and C.1 in the Appendix imply the result that if $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$, then there exists a $\overline{T}$ such that for all $T > \overline{T}$,

$$\Pr_{\boldsymbol{X}}\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right] > \Pr_{\boldsymbol{Y}}\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right]$$

Recall that

$$q_{\boldsymbol{X}} = \min\left\{\Pr_{\boldsymbol{X}}\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right], \Pr_{\boldsymbol{X}}\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right]\right\}$$

and similarly,

$$q_{\boldsymbol{Y}} = \min\left\{\Pr_{\boldsymbol{Y}}\left[\Omega^+ \mid \mathbf{Y}_i = \mathbf{e}^1\right], \Pr_{\boldsymbol{Y}}\left[\Omega^G \mid \mathbf{Y}_i = \mathbf{e}^1\right]\right\}$$

and the assumption that $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same extent of individual learning implies that

$$\Pr_{\boldsymbol{X}}\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right] = \Pr_{\boldsymbol{Y}}\left[\Omega^G \mid \mathbf{Y}_i = \mathbf{e}^1\right]$$

Thus, we obtain

Proposition 5.2 *Suppose that* $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$. *Then there exists a* $\overline{T}$ *such that for all* $T > \overline{T}$,

$$q_{\boldsymbol{X}} \geq q_{\boldsymbol{Y}} \tag{11}$$

Unlike in the case of conclusive signals, the inequality (11) is weak.

Example 2 We now consider an example in which there are more than two signals and these are non-conclusive—every signal occurs with positive probability in each state of nature θ. The purpose of the example is to show that the conclusion of Theorem 3 can hold for a non-empty, open set of p's.

Suppose that the set of signals $\mathcal{X} = \{0, 1, 2\}$. There are two agents and the two states B and G are equally likely.

Consider a correlated information structure $(\theta, \mathbf{Y})$ with the following joint distributions conditional on the state θ, where $\varepsilon > 0$ is a small number.

$\overline{P}(\cdot \mid G) =$

	$Y_2 = 0$	$Y_2 = 1$	$Y_2 = 2$
$Y_1 = 0$	0.1	0.08	ε^3
$Y_1 = 1$	0.08	$0.7 - 2\varepsilon - 2\varepsilon^3$	ε
$Y_1 = 2$	ε^3	ε	0.04

$\overline{P}(\cdot \mid B) =$

	$Y_2 = 0$	$Y_2 = 1$	$Y_2 = 2$
$Y_1 = 0$	$0.997 - 3\varepsilon^3 - 2\varepsilon^4$	0.001	ε^4
$Y_1 = 1$	0.001	0.001	ε^3
$Y_1 = 2$	ε^4	ε^3	ε^3

When ε is small enough, the random variables $(\theta, \mathbf{Y})$ are affiliated. In fact, all the (non-trivial) affiliation inequalities are strict.

Consider an alternative information structure $(\theta, \mathbf{X})$ where for each θ, the conditional distribution $\Pr[X_i \mid \theta] = \Pr[Y_i \mid \theta]$ and

$$\Pr[X_i = k, X_j = l \mid \theta] = \Pr[X_i = k \mid \theta] \times \Pr[X_j = l \mid \theta]$$

In other words, conditional on θ, the signals $\mathbf{X}$ are independently distributed.

Suppose that $T = 2$, so that signals are generated twice. Now we have that when ε close to zero,

$$\rho_0 = \Pr\left[\Omega^G \mid \mathbf{X} = \mathbf{0}\right] = \Pr\left[\Omega^G \mid \mathbf{Y} = \mathbf{0}\right] \approx 3.12 \times 10^{-2}$$

whereas

$$\begin{aligned} q_{\boldsymbol{Y}} &\approx 0.936 \\ q_{\boldsymbol{X}} &\approx 0.954 \end{aligned}$$

so that $q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$.

If $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$, then from Theorem 3, we know that

$$C^p_{\boldsymbol{X}}\left(\Omega^G\right) = \Omega^+ = \{\omega : \forall i,\ \mathbf{x}_i \neq \mathbf{0}\}$$

and it may be verified that

$$C^p_{\boldsymbol{Y}}\left(\Omega^G\right) = \left\{\omega : \forall i,\ \max_t \mathbf{y}_i^t = 2\right\}$$

that is, $C^p_{\boldsymbol{Y}}\left(\Omega^G\right)$ consists of those states in which both players receive at least one signal $k = 2$.

Finally, when $\varepsilon \approx 0$,

$$\Pr_{\boldsymbol{Y}}\left[C^p_{\boldsymbol{Y}}\left(\Omega^G\right)\right] \approx 0.039$$

whereas

$$\Pr_{\boldsymbol{X}}\left[C^p_X\left(\Omega_G\right)\right] \approx 0.468$$

Thus, in the example, we not only have

$$\varnothing \neq C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subsetneq C^p_{\boldsymbol{X}}\left(\Omega^G\right)$$

but

$$0 < \Pr_{\boldsymbol{Y}}\left[C^p_{\boldsymbol{Y}}\left(\Omega^G\right)\right] < \Pr_{\boldsymbol{X}}\left[C^p_{\boldsymbol{X}}\left(\Omega_G\right)\right]$$

as well.

But there are large p's for which $C^p_{\boldsymbol{X}}\left(\Omega_G\right) = \varnothing$ while $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \neq \varnothing$. Thus, with non-conclusive signals it is not the case that for all p, $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subseteq C^p_{\boldsymbol{X}}\left(\Omega_G\right)$.

6 Blackwell Informativeness

When there only *two* agents, our main result can be reinterpreted in the language of Blackwell's (1951) informativeness notion. Blackwell's setting, of course, is that of a *single* agent facing a decision whose payoff is influenced by an unknown state of nature. In what follows, signals need not be binary nor need they be conclusive.

In the two-agent case, we first adopt the perspective of agent 1, say. Suppose P is the joint distribution of (θ, X_1, X_2). Let P^θ be the joint distribution of (X_1, X_2) conditional on θ. For fixed θ, from agent 1's perspective, the signal X_2 of agent 2 can be interpreted as a "state of nature" and X_1 as agent 1's informative signal about X_2. The conditional distribution $P^\theta\left(X_1 \mid X_2\right)$ is then a Blackwell experiment. The same is true if we adopt the perspective of agent 2 and treat X_1 as a "state of nature" and X_2 as agent 2's signal about X_1.[5]

Now consider another pair of signals (Y_1, Y_2) and suppose $\overline{P}$ joint distribution (θ, Y_1, Y_2). Again let $\overline{P}^\theta$ be the joint distribution of (Y_1, Y_2) conditional on θ. As above, for fixed θ, $\overline{P}^\theta\left(Y_1 \mid Y_2\right)$ is also a Blackwell experiment.

We will say that

Definition 2 *Suppose $I = 2$.The information structure $(\theta, \mathbf{Y})$ is* mutually more informative *than $(\theta, \mathbf{X})$ if for all θ, $\overline{P}^\theta\left(Y_j \mid Y_i\right)$ is Blackwell more informative than $P^\theta\left(X_j \mid X_i\right)$.*

[5] This reinterpretation cannot work when there are more than two agents. For instance, suppose signals are binary and $I = 3$. Now from agent 1's perspective the state of nature is (X_2, X_3). Blackwell's informativeness criterion would require that if $\boldsymbol{Y}$ is another signal structure, then for all i, the distribution of (X_2, X_3) be the same as the distribution of (Y_2, Y_3). Together with symmetry, this can hold only if the distribution of $\boldsymbol{Y}$ is the same as the distribution of $\boldsymbol{X}$.

Note that this definition focuses on how informative one agent's signals are about the other agent's signals. Also, this guarantees that conditional on θ, $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same univariate marginal distributions.

Lemma 6.1 *Suppose that conditional on θ, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both affiliated. If $(\theta, \boldsymbol{Y})$ is mutually more informative than $(\theta, \boldsymbol{X})$, then*

$$\Pr\left[X_1 = 0, X_2 = 0 \mid \theta\right] \leq \Pr\left[Y_1 = 0, Y_2 = 0 \mid \theta\right] \tag{12}$$

Proof. Fix θ. From Blackwell, we know that if $\overline{P}^{\theta}(Y_1 \mid Y_2)$ is more informative than $P^{\theta}(X_1 \mid X_2)$, then the posteriors from $\boldsymbol{Y}$ are a mean-preserving spread of the posteriors from $\boldsymbol{X}$.

Formally, if we define for every k and l in $\mathcal{X}$,

$$p_l^k = P^{\theta}\left[X_2 = l \mid X_1 = k\right]$$

and

$$\boldsymbol{p}^k = \left(p_l^k\right)_{l \in \mathcal{X}} \in \Delta(\mathcal{X})$$

to be the vector of posterior beliefs of agent 1 with signal $X_1 = k$ about the signals X_2 of agent 2. Similarly, define

$$\overline{\boldsymbol{p}}^k \in \Delta(\mathcal{X})$$

to be the vector of posterior beliefs of agent 1 with signal $Y_1 = k$ about the signals Y_2 of agent 2.

Now Blackwell's Theorem implies that for all k,

$$\boldsymbol{p}^k \in \mathrm{co}\{\overline{\boldsymbol{p}}^m : m \in \mathcal{X}\}$$

the convex hull of the set of posterior vectors from Y.

Moreover, since (X_1, X_2) are affiliated, for any $k > 0$, the distribution $\boldsymbol{p}^k \in \Delta(\mathcal{X})$ *stochastically dominates* the distribution $\boldsymbol{p}^0 \in \Delta(\mathcal{X})$. Similarly, for any $k > 0$, the distribution $\overline{\boldsymbol{p}}^k \in \Delta(\mathcal{X})$ *stochastically dominates* the distribution $\overline{\boldsymbol{p}}^0 \in \Delta(\mathcal{X})$.

Since $\boldsymbol{p}^0 \in \mathrm{co}\{\overline{\boldsymbol{p}}^m : m \in \mathcal{X}\}$ we can write

$$\boldsymbol{p}^0 = \sum\nolimits_{m=0}^{K} \alpha_m \overline{\boldsymbol{p}}^m$$

where $\alpha_m \in [0,1]$ and $\sum_{m=0}^{K} \alpha_m = 1$.

We claim that the distribution $\boldsymbol{p}^0 \in \Delta(\mathcal{X})$ stochastically dominates $\overline{\boldsymbol{p}}^0 \in \Delta(\mathcal{X})$. This is the same as, for any $L \in \mathcal{X}$,

$$\begin{aligned}
\sum\nolimits_{l=0}^{L} p_l^0 &= \sum\nolimits_{l=0}^{L} \sum\nolimits_{m=0}^{K} \alpha_m \overline{p}_l^m \\
&= \sum\nolimits_{m=0}^{K} \alpha_m \left(\sum\nolimits_{l=0}^{L} \overline{p}_l^m\right) \\
&\leq \sum\nolimits_{m=0}^{K} \alpha_m \left(\sum\nolimits_{l=0}^{L} \overline{p}_l^0\right) \\
&= \sum\nolimits_{l=0}^{L} \overline{p}_l^0
\end{aligned}$$

where the inequality in the third line follows from the fact that the distribution for all $m > 0$, $\overline{\boldsymbol{p}}^m$ stochastically dominates $\overline{\boldsymbol{p}}^0$.

In particular, for $L = 0$, this implies that

$$p_0^0 \leq \overline{p}_0^0$$

which is equivalent to

$$P^\theta [X_2 = 0 \mid X_1 = 0] \leq \overline{P}^\theta [Y_2 = 0 \mid Y_1 = 0]$$

and since $P^\theta [X_1 = 0] = \overline{P}^\theta [Y_1 = 0]$, the result follows. ■

Lemma 6.1 implies that when there are *two* agents, in *all* of the results of the earlier sections, the condition that "$\boldsymbol{Y} \succ_{PQD} \mathbf{X}$" can be replaced with the condition "$\boldsymbol{Y}$ is mutually more informative than $\mathbf{X}$," provided that the inequality in (12) is strict. This is because Lemmas B.1 and C.1 only require (the strict version) of the inequality.

A Appendix: Affiliation and PQD Order

Recall that a vector of random variables $\mathbf{X} \in \mathcal{X}^I$ with joint probability distribution P is said to be *affiliated* if for all $\boldsymbol{x}$ and $\boldsymbol{x}'$ in $\mathcal{X}^I$

$$P[\boldsymbol{x}] \times P[\boldsymbol{x}'] \leq P(\boldsymbol{x} \vee \boldsymbol{x}') \times P(\boldsymbol{x} \wedge \boldsymbol{x}')$$

Also recall the notation that if $\mathbf{x} = (x_i^t)_{i \in I, t \in T}$ is a realization of all I signals in all T periods, then

$$\boldsymbol{x}^t = \left(x_i^t\right)_{i \in I}$$

(slanted bold) is the I-vector of all I signal realizations in period t, while

$$\mathbf{x}_i = \left(x_i^t\right)_{t \in T}$$

(upright bold) is the T-vector of i's signals over the T periods.

Lemma A.1 *Suppose that the variables $\mathbf{X} \in \mathcal{X}^I$ are affiliated with distribution P. If $\boldsymbol{X}^1, \boldsymbol{X}^2, ..., \boldsymbol{X}^T$ are independently and identically distributed according to P, then $(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_I) \in \left(\mathcal{X}^I\right)^T$ are also affiliated.*

Proof. Suppose $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_I)$ and $\mathbf{x}' = (\mathbf{x}_1', \mathbf{x}_2', ..., \mathbf{x}_I')$ are both in $\left(\mathcal{X}^I\right)^T$. Because of independence

$$\Pr[\mathbf{x}] = \prod_{t=1}^T P\left(\boldsymbol{x}^t\right)$$

and

$$\Pr[\mathbf{x}'] = \prod_{t=1}^T P\left(\boldsymbol{x}'^t\right)$$

and so

$$\begin{aligned}
\Pr[\mathbf{x}]\Pr[\mathbf{x}'] &= \textstyle\prod_{t=1}^{T} P\left(\boldsymbol{x}^t\right)\prod_{t=1}^{T} P\left(\boldsymbol{x}'^t\right) \\
&= \textstyle\prod_{t=1}^{T} P\left(\boldsymbol{x}^t\right) P\left(\boldsymbol{x}'^t\right) \\
&\leq \textstyle\prod_{t=1}^{T} P\left(\boldsymbol{x}^t \vee \boldsymbol{x}'^t\right) P\left(\boldsymbol{x}^t \wedge \boldsymbol{x}'^t\right) \\
&= \textstyle\prod_{t=1}^{T} P\left(\boldsymbol{x}^t \vee \boldsymbol{x}'^t\right)\prod_{t=1}^{T} P\left(\boldsymbol{x}^t \wedge \boldsymbol{x}'^t\right) \\
&= \Pr[\mathbf{x}\vee\mathbf{x}']\Pr[\mathbf{x}\wedge\mathbf{x}']
\end{aligned}$$

■

Lemma A.2 *Suppose that the variables* $\boldsymbol{X} \in \mathcal{X}^I$ *are affiliated. For any* $\mathbf{x}_i \neq \mathbf{0}$,

$$\Pr\left[\Omega^+ \middle| \mathbf{X}_i = \mathbf{x}_i\right] \geq \Pr\left[\Omega^+ \middle| \mathbf{X}_i = \mathbf{e}^1\right]$$

Proof. Clearly, the indicator function $I_{\Omega^+} : \left(\mathcal{X}^T\right)^I \to \{0,1\}$ of the set $\Omega^+ = \{\omega : \forall j, \mathbf{x}_j \neq \mathbf{0}\}$ is non-decreasing. For any $\mathbf{x}_i \neq \mathbf{0}$ there is a permutation $\mathbf{x}_i^\pi$ of $\mathbf{x}_i$ such that $\mathbf{x}_i^\pi \geq \mathbf{e}^1$. Since the set Ω^+ is permutation invariant

$$\begin{aligned}
\Pr\left[\Omega^+ \middle| \mathbf{X}_i = \mathbf{x}_i\right] &= \Pr\left[\Omega^+ \middle| \mathbf{X}_i = \mathbf{x}_i^\pi\right] \\
&= E\left[I_{\Omega^+}(\mathbf{X}) \mid \mathbf{X}_i = \mathbf{x}_i^\pi\right] \\
&\geq E\left[I_{\Omega^+}(\mathbf{X}) \mid \mathbf{X}_i = \mathbf{e}^1\right] \\
&= \Pr\left[\Omega^+ \middle| \mathbf{X}_i = \mathbf{e}^1\right]
\end{aligned}$$

The inequality in the third line is the result of the following argument. First, since the random variables $\mathbf{X} = (X_i^t)$ are affiliated (Lemma A.1), the probability distribution of $\mathbf{X}_{-i}$ conditional on $\mathbf{X}_i = \mathbf{x}_i^\pi$ dominates the distribution of $\mathbf{X}_{-i}$ conditional on $\mathbf{X}_i = \mathbf{e}^1$ in the *multivariate likelihood order*, as defined in Section 6.E of Shaked and Shantikumar (2008). Their Theorem 6.E.8 now implies that the two distributions are also ranked by the usual *stochastic order*. ■

Lemma A.3 *Suppose that* $\mathbf{Y}^\theta \succ_{PQD} \boldsymbol{X}^\theta$*. Then, for any subset* $S \subset I$

$$\Pr[\boldsymbol{X}_S = \mathbf{0} \mid \theta] < \Pr[\boldsymbol{Y}_S = \mathbf{0} \mid \theta]$$

where $\boldsymbol{X}_S = (X_i)_{i \in S}$.

Proof. Recall that if $\mathbf{Y}^\theta \succ_{PQD} \boldsymbol{X}^\theta$ then for $\boldsymbol{z} \neq (K, K, ..., K)$,

$$\Pr[\boldsymbol{X} \leq \boldsymbol{z} \mid \theta] < \Pr[\boldsymbol{Y} \leq \boldsymbol{z} \mid \theta]$$

If we choose $\boldsymbol{z}$ such that for all $i \in S$, $z_i = 0$ and $z_j = K$, for all $j \notin S$, then the conclusion follows. ■

B Appendix: Binary and Conclusive Signals

Recall that from (9) that

$$S_{-i}^{1} = \{j \neq i : N_j \geq 1\}$$

is the set of agents other than i, who receive at least one positive signal and let $\#A$ denote the cardinality of A.

Now define for $s = 0, 1, ..., I-1$

$$q_{\boldsymbol{X}}(s) = \Pr\left[\#S_{-i}^{1} = s \mid N_i = 1\right]$$

as the probability assigned by a type who saw only one positive signal to the event that s other agents saw positive signals so that

$$\sum_{s=0}^{I-1} q_{\boldsymbol{X}}(s) = 1$$

Note that $q_{\boldsymbol{X}}(I-1)$ is the same as $q_{\boldsymbol{X}}$, as defined in (5) in Section 3.

Lemma B.1 *Suppose that for any subset $S \subset I$*

$$\Pr[\boldsymbol{X}_S = \mathbf{0} \mid G] < \Pr[\mathbf{Y}_S = \mathbf{0} \mid G]$$

For T large enough, for all $m < I-1$,

$$\sum_{s=0}^{m} q_{\boldsymbol{X}}(s) < \sum_{s=0}^{m} q_{\boldsymbol{Y}}(s)$$

and so

$$q_{\boldsymbol{X}}(I-1) > q_{\boldsymbol{Y}}(I-1) \tag{13}$$

In other words, the distribution $q_{\boldsymbol{X}}(\cdot)$ strictly stochastically dominates $q_{\boldsymbol{Y}}(\cdot)$.

Proof. We will, in fact, prove the stronger statement that when T is large enough, for all $s < I-1$,

$$q_{\boldsymbol{X}}(s) < q_{\boldsymbol{Y}}(s)$$

Without loss of generality, we will suppose that $i = 1$ and so for any $n < I-1$,

$$\begin{aligned} q_{\boldsymbol{X}}(s) &= \tbinom{I-1}{s} P\left[\mathbf{x}_{s+2} = \mathbf{0}, ..., \mathbf{x}_I = \mathbf{0}\right] \\ &\quad - \tbinom{I-1}{s} \sum_{m=0}^{s-1} \tbinom{s}{m} P\left[\mathbf{x}_2 \neq \mathbf{0}, ..., \mathbf{x}_{m+1} \neq \mathbf{0}, \mathbf{x}_{m+2} = \mathbf{0}, .., \mathbf{x}_I = \mathbf{0}\right] \end{aligned}$$

where for any event A,

$$P[A] = \Pr[A \mid N_1 = 1]$$

The formula for $q_{\mathbf{X}}(s)$ then just uses the fact that the event $\{\omega : \mathbf{x}_{s+2} = \mathbf{0},..., \mathbf{x}_I = \mathbf{0}\}$ is the union of mutually exclusive events in which some subset consisting of m agents from the set $\{2, 3, ..., s+1\}$ get non-zero signals. In the expression above, we have used symmetry to write the probability of each of these events in a way that for the s players $i = 2, 3, ..., m+1$, $\mathbf{x}_i \neq \mathbf{0}$ whereas for the $I-1-m$ agents $j > m+1$, $\mathbf{x}_j = \mathbf{0}$.

Similarly,

$$\begin{aligned} q_{\mathbf{Y}}(s) &= \tbinom{I-1}{s}\overline{P}\left[\mathbf{y}_{s+2} = \mathbf{0},..., \mathbf{y}_I = \mathbf{0}\right] \\ &\quad -\tbinom{I-1}{s}\sum_{m=0}^{s-1}\tbinom{s}{m}\overline{P}\left[\mathbf{y}_2 \neq \mathbf{0}, ..., \mathbf{y}_{m+1} \neq \mathbf{0}, \mathbf{y}_{m+2} = \mathbf{0}, .., \mathbf{y}_I = \mathbf{0}\right] \end{aligned}$$

where for any event A,

$$\overline{P}[A] = \Pr\left[A \mid \mathbf{Y}_1 = \mathbf{e}^1\right]$$

The ratio of the two is then

$$\begin{aligned} &\frac{q_{\mathbf{X}}(s)}{q_{\mathbf{Y}}(s)} \\ =\ &\frac{P\left[\mathbf{x}_{s+2} = \mathbf{0},..., \mathbf{x}_I = \mathbf{0}\right] - \sum_{m=0}^{s-1}\binom{s}{m}P\left[\mathbf{x}_2 \neq \mathbf{0}, ..., \mathbf{x}_{m+1} \neq \mathbf{0}, \mathbf{x}_{m+2} = \mathbf{0}, .., \mathbf{x}_I = \mathbf{0}\right]}{\overline{P}\left[\mathbf{y}_{s+2} = \mathbf{0},..., \mathbf{y}_I = \mathbf{0}\right] - \sum_{m=0}^{s-1}\binom{s}{m}\overline{P}\left[\mathbf{y}_2 \neq \mathbf{0}, ..., \mathbf{y}_{m+1} \neq \mathbf{0}, \mathbf{y}_{m+2} = \mathbf{0}, .., \mathbf{y}_I = \mathbf{0}\right]} \\ =\ &\frac{\alpha - \beta}{\overline{\alpha} - \overline{\beta}} \end{aligned}$$

where

$$\begin{aligned} \alpha &= P\left[\mathbf{x}_{s+2} = \mathbf{0},..., \mathbf{x}_I = \mathbf{0}\right] \\ \beta &= \sum_{m=0}^{s-1}\tbinom{s}{m}P\left[\mathbf{x}_2 \neq \mathbf{0}, ..., \mathbf{x}_{m+1} \neq \mathbf{0}, \mathbf{x}_{m+2} = \mathbf{0}, .., \mathbf{x}_I = \mathbf{0}\right] \end{aligned}$$

and $\overline{\alpha}$ and $\overline{\beta}$ are similarly defined but for $\overline{P}$.

First, observe that

$$\begin{aligned} \frac{\alpha}{\overline{\alpha}} &= \frac{\Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \mathbf{X}_{s+2} = \mathbf{0},..., \mathbf{X}_I = \mathbf{0}\right]}{\Pr\left[\mathbf{Y}_1 = \mathbf{e}^1, \mathbf{Y}_{s+2} = \mathbf{0},..., \mathbf{Y}_I = \mathbf{0}\right]} \\ &= \frac{\Pr\left[X_1 = 1, X_{s+2} = ... = X_I = 0\right]}{\Pr\left[Y_1 = 1, Y_{s+2} = ... = Y_I = 0\right]} \times \left(\frac{\Pr\left[X_1 = 0, X_{s+2} = ... = X_I = 0\right]}{\Pr\left[Y_1 = 0, Y_{s+2} = ... = Y_I = 0\right]}\right)^{T-1} \end{aligned}$$

and by hypothesis, the term raised to the power of $T-1$ is less than one. Thus as T increases, the ratio above goes to zero.

Second, since we have assumed that $\alpha > \beta$, we have that $\frac{\beta}{\overline{\alpha}} = \frac{\alpha}{\overline{\alpha}} \times \frac{\beta}{\alpha}$ goes to zero as well.

Finally, for any $m \leq s-1$,

$$\begin{aligned}
& \frac{\overline{P}\left[\mathbf{y}_2 \neq \mathbf{0}, ..., \mathbf{y}_{m+1} \neq \mathbf{0}, \mathbf{y}_{m+2} = \mathbf{0}, .., \mathbf{y}_I = \mathbf{0}\right]}{\overline{P}\left[\mathbf{y}_{s+2} = \mathbf{0}, ..., \mathbf{y}_I = \mathbf{0}\right]} \\
= \ & \Pr\left[\mathbf{Y}_2 \neq \mathbf{0}, ..., \mathbf{Y}_{m+1} \neq \mathbf{0}, \mathbf{Y}_{m+2} = \mathbf{0}, ..., \mathbf{Y}_{s+1} = \mathbf{0} \mid \mathbf{Y}_1 = \mathbf{e}^1, \mathbf{Y}_{s+2} = \mathbf{0}, ..., \mathbf{Y}_I = \mathbf{0}\right] \\
< \ & \Pr\left[\mathbf{Y}_{s+1} = \mathbf{0} \mid \mathbf{Y}_1 = \mathbf{e}^1, \mathbf{Y}_{s+2} = \mathbf{0}, ..., \mathbf{Y}_I = \mathbf{0}\right] \\
= \ & \Pr\left[\mathbf{Y}_{s+1} = \mathbf{0} \mid \mathbf{Y}_1 = \mathbf{e}^1, \mathbf{Y}_{s+2} = \mathbf{0}, ..., \mathbf{Y}_I = \mathbf{0}\right]
\end{aligned}$$

which also goes to zero since conditional on $\mathbf{Y}_1 = \mathbf{e}^1$, and hence also conditional on $\theta = G$, the probability than $\mathbf{Y}_{s+1} = \mathbf{0}$ goes to zero. Thus, we also have that $\frac{\overline{\beta}}{\alpha}$ goes to zero as T increases.

Combining all these and using the fact $s \leq I-1$, we obtain for T large enough, for all $s < I-1$

$$\frac{q_{\boldsymbol{X}}(s)}{q_{\boldsymbol{Y}}(s)} < 1$$

■

We now establish a generalization of Lemma 3.1. Recall from (9) that

$$S_{-i}^k = \{j \neq i : N_j \geq k\}$$

is the set of players other than i, who receive at least k positive signals where $\#S_{-i}^k$ denotes the cardinality of S_{-i}^k. Define

$$q_{\mathbf{X}}^{(k)}(s) = \Pr\left[\#S_{-i}^k = s \mid N_i = k\right]$$

be the probability assigned by an agent with k positive signals to the event that s other agents received at least k positive signals. Note that $\sum_{s=0}^{I-1} q_{\mathbf{X}}^{(k)}(s) = 1$.

Lemma B.2 *The distribution $q_{\mathbf{X}}^{(k)}(\cdot)$ stochastically dominates $q_{\mathbf{X}}^{(k+1)}(\cdot)$.*

Proof. Stochastic dominance is the same as

$$1 - \sum_{s=m}^{I-1} q_{\mathbf{X}}^{(k)}(s) \geq 1 - \sum_{s=m}^{I-1} q_{\mathbf{X}}^{(k+1)}(s)$$

or equivalently

$$\Pr\left[\#\{j \neq i : N_j \geq k\} \geq m \mid N_i = k\right] \geq \Pr\left[\#\{j \neq i : N_j \geq k+1\} \geq m \mid N_i = k+1\right]$$

Let $M_j = \sum_{t=1}^{T-1} X_j^t$ be a random variable that counts the number of 1-signals received by agent j in the first $T-1$ periods. Thus, $N_j = M_j + X_j$.

If $M_j < k-1$, then both $N_j = k$ and $N_j = k+1$ are impossible.

If $M_j = k-1$, then $N_j = k$ is possible while $N_j = k+1$ is impossible.

If $M_j = k$, then $N_j = k$ occurs with probability 1 while $N_j = k+1$ occurs with probability less than one.

If $M_j > k$, then both $N_j = k$ and $N_j = k+1$ occur with probability one.

Thus, in all cases the probability that $N_j \geq k$ occurs is at least as large as the probability that $N_j \geq k+1$ occurs. ■

C Appendix: Non-conclusive signals

We begin by developing a formula for the *joint* probability

$$\begin{aligned} \Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \Omega^+\right] &= \Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \forall j, \mathbf{X}_j \neq \mathbf{0}\right] \\ &= \Pr\left[\mathbf{X}_1 = \mathbf{e}^1\right] - \Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \exists j, \mathbf{X}_j = \mathbf{0}\right] \end{aligned} \tag{14}$$

If we define $A_j = \{\omega : \mathbf{x}_1 = \mathbf{e}^1, \mathbf{x}_j = \mathbf{0}\}$ as the set of states in which 1's type is $\mathbf{e}^1$ and j's type is $\mathbf{0}$, then

$$\Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \exists j, \mathbf{X}_j = \mathbf{0}\right] = P\left[\cup_{j\neq 1} A_j\right]$$

where P is joint distribution of $(\theta, \mathbf{X})$.

By the inclusion-exclusion principle,

$$P\left[\cup_{j\neq 1} A_j\right] = \sum_{1<j} P\left[A_j\right] - \sum_{1<j<k} P\left[A_j \cap A_k\right] + \sum_{1<j<k<l} P\left[A_j \cap A_k \cap A_l\right] - ... \tag{15}$$

But since agents are symmetric, we have

$$\begin{aligned} P\left[\cup_{j\neq 1} A_j\right] &= \tbinom{I-1}{1} P\left[A_2\right] - P\left[A_2 \cap A_3\right] + \tbinom{I-1}{3} P\left[A_2 \cap A_3 \cap A_4\right] - ... \\ &= \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} P\left[A_2 \cap A_3 \cap ... \cap A_l\right] \end{aligned}$$

Now, since conditional on θ, the signals are independent over time

$$\begin{aligned} P\left[A_2\right] &= P\left[\mathbf{X}_1 = \mathbf{e}^1, \mathbf{X}_2 = \mathbf{0}\right] \\ &= \rho\left(P\left[(X_1, X_2) = (1,0) \mid G\right] \times \left(P\left[(X_1, X_2) = (0,0) \mid G\right]\right)^{T-1}\right) \\ &\quad + (1-\rho)\left(P\left[(X_1, X_2) = (1,0) \mid B\right] \times \left(P\left[(X_1, X_2) = (0,0) \mid B\right]\right)^{T-1}\right) \end{aligned}$$

In general, for all $l = 2, 3, ..., I$

$$\begin{aligned} P\left[A_2 \cap A_3 \cap ... \cap A_l\right] &= P\left[\mathbf{X}_1 = \mathbf{e}^1, \mathbf{X}_2 = \mathbf{X}_3 = ... = \mathbf{X}_l = \mathbf{0}\right] \\ &= \rho\left(P\left[(X_1, X_2, , ..., X_l) = (1, 0, ...0) \mid G\right]\right. \\ &\quad \left.\times \left(P\left[(X_1, X_2, , ..., X_l) = (0, 0, ...0) \mid G\right]\right)^{T-1}\right) \\ &\quad + (1-\rho)\left(P\left[(X_1, X_2, , ..., X_l) = (1, 0, ...0) \mid B\right]\right. \\ &\quad \left.\times \left(P\left[(X_1, X_2, , ..., X_l) = (0, 0, ...0) \mid B\right]\right)^{T-1}\right) \end{aligned}$$

It will be convenient to define, for $l = 2, 3, ..., I$ and $\theta = G, B$,

$$\alpha_l^\theta = P\left[(X_1, X_2, , ..., X_l) = (1, 0, ...0) \mid \theta\right]$$

and

$$\beta_l^\theta = P\left[(X_1, X_2, , ..., X_l) = (0, 0, ...0) \mid \theta\right]$$

and so we can rewrite (15) more compactly as

$$P\left[\cup_{j\neq 1} A_j\right] = \sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho\alpha_l^G \left(\beta_l^G\right)^{T-1} + (1-\rho)\,\alpha_l^B \left(\beta_l^B\right)^{T-1}\right) \tag{16}$$

Note that for $\theta = G, B$, both α_l^θ and β_l^θ are non-increasing sequences since the event that $X_2 = X_2 = ... = X_l = 0$ includes the event that $X_2 = X_2 = ... = X_l = X_{l+1} = 0$. Moreover, if conditional on θ, signals have full support, then α_l^θ and β_l^θ are *strictly* decreasing.

Analogously, if $(\theta, \mathbf{Y})$ are distributed according to $\overline{P}$, then we have

$$\overline{P}\left[\cup_{j\neq 1} A_j\right] = \sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho\overline{\alpha}_l^G (\overline{\beta}_l^G)^{T-1} + (1-\rho)\,\overline{\alpha}_l^B (\overline{\beta}_l^B)^{T-1}\right) \tag{17}$$

where $\overline{\alpha}_l^\theta$ and $\overline{\beta}_l^\theta$ are defined in the same manner as α_l^θ and β_l^θ but for the probability distribution $\overline{P}$ of $\mathbf{Y}$. As above, both $\overline{\alpha}_l^\theta$ and $\overline{\beta}_l^\theta$ are decreasing sequences.

Lemma C.1 *Suppose* $(\theta, \mathbf{X})$ *and* $(\theta, \mathbf{Y})$ *are two non-conclusive, full-support information structures such that for* $\theta = G, B$

$$\Pr\left[X_i = 0, X_j = 0 \mid \theta\right] < \Pr\left[Y_i = 0, Y_j = 0 \mid \theta\right] \tag{18}$$

Then there exists a $\overline{T}$ *such that for all* $T > \overline{T}$,

$$\Pr_{\mathbf{X}}\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right] > \Pr_{\mathbf{Y}}\left[\Omega^+ \mid \mathbf{Y}_i = \mathbf{e}^1\right]$$

Proof. From (16) and (17) we have that the ratio

$$\frac{P\left[\cup_{j\neq 1} A_j\right]}{\overline{P}\left[\cup_{j\neq 1} A_j\right]} = \frac{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho\alpha_l^G \left(\beta_l^G\right)^{T-1} + (1-\rho)\,\alpha_l^B \left(\beta_l^B\right)^{T-1}\right)}{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho\overline{\alpha}_l^G \left(\overline{\beta}_l^G\right)^{T-1} + (1-\rho)\,\overline{\alpha}_l^B \left(\overline{\beta}_l^B\right)^{T-1}\right)}$$

Dividing the numerator and denominator by $\left(\overline{\beta}_2^B\right)^{T-1} > 0$, we obtain

$$\frac{P\left[\cup_{j\neq 1} A_j\right]}{\overline{P}\left[\cup_{j\neq 1} A_j\right]} = \frac{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho\alpha_l^G \left(\frac{\beta_l^G}{\overline{\beta}_2^B}\right)^{T-1} + (1-\rho)\,\alpha_l^B \left(\frac{\beta_l^B}{\overline{\beta}_2^B}\right)^{T-1}\right)}{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho\overline{\alpha}_l^G (\frac{\overline{\beta}_l^G}{\overline{\beta}_2^B})^{T-1} + (1-\rho)\,\overline{\alpha}_l^B (\frac{\overline{\beta}_l^B}{\overline{\beta}_2^B})^{T-1}\right)} \tag{19}$$

and observe that since both $(\theta, \boldsymbol{X})$ and $(\theta, \boldsymbol{Y})$ are affiliated,

$$\begin{aligned} \beta_2^G &= \Pr\left[(X_1, X_2) = (0,0) \mid G\right] \leq \Pr\left[(X_1, X_2) = (0,0) \mid B\right] = \beta_2^B \\ \overline{\beta}_2^G &= \Pr\left[(Y_1, Y_2) = (0,0) \mid G\right] \leq \Pr\left[(Y_1, Y_2) = (0,0) \mid B\right] = \overline{\beta}_2^B \end{aligned}$$

Moreover, (18) implies that

$$\begin{aligned} \beta_2^B &= \Pr\left[(X_1, X_2) = (0,0) \mid B\right] < \Pr\left[(Y_1, Y_2) = (0,0) \mid B\right] = \overline{\beta}_2^B \\ \beta_2^G &= \Pr\left[(X_1, X_2) = (0,0) \mid G\right] < \Pr\left[(Y_1, Y_2) = (0,0) \mid G\right] = \overline{\beta}_2^G \end{aligned}$$

Thus, for all l,

$$\beta_l^G \leq \beta_2^G < \overline{\beta}_2^G \leq \overline{\beta}_2^B$$

and since β_l^B is a strictly decreasing sequence, for $l > 2$,

$$\beta_l^B < \beta_2^B < \overline{\beta}_2^B$$

These inequalities in turn imply that in the numerator of (19), for all l

$$\frac{\beta_l^G}{\overline{\beta}_2^B} < 1 \text{ and } \frac{\beta_l^B}{\overline{\beta}_2^B} < 1$$

and so as $T \to \infty$, the numerator goes to zero.

Moreover, for all $l > 2$

$$\frac{\overline{\beta}_l^G}{\overline{\beta}_2^B} < \frac{\overline{\beta}_2^G}{\overline{\beta}_2^B} \leq 1$$

and for $l > 2$,

$$\frac{\overline{\beta}_l^B}{\overline{\beta}_2^B} < 1$$

and so as $T \to \infty$, all the terms with $l > 2$ in the denominator of the right-hand side of (19) go to zero. The $l = 2$ term in the denominator, however, stays positive (the $l = 2$ term in the denominator is at least $(1 - \rho)\,\overline{\alpha}_l^B > 0$).

So we have that when T is large enough,

$$\frac{\Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \exists j, \mathbf{X}_j = \mathbf{0}\right]}{\Pr\left[\mathbf{Y}_1 = \mathbf{e}^1, \exists j, \mathbf{Y}_j = \mathbf{0}\right]} = \frac{P\left[\cup_{j \neq 1} A_j\right]}{\overline{P}\left[\cup_{j \neq 1} A_j\right]} < 1$$

Now since $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same univariate marginals, $\Pr\left[\mathbf{X}_1 = \mathbf{e}^1\right] = \Pr\left[\mathbf{Y}_1 = \mathbf{e}^1\right]$ and so from (14)

$$\Pr\left[\forall j, \mathbf{X}_j \neq \mathbf{0} \mid \mathbf{X}_1 = \mathbf{e}^1\right] > \Pr\left[\forall j, \mathbf{Y}_j \neq \mathbf{0} \mid \mathbf{Y}_1 = \mathbf{e}^1\right]$$

■

References

[1] Blackwell, David (1951): "Comparison of Experiments," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, Volume 1, University of California Press, 93–102.

[2] Cripps, M., J. Ely, G. Mailath and L. Samuelson (2008): "Common Learning," *Econometrica*, 76, 909–933.

[3] Cripps, M., J. Ely, G. Mailath and L. Samuelson (2013): "Common Learning with Intertemporal Dependence," *International Journal of Game Theory*, 42, 55–98.

[4] Frick, M., R. Iijima and Y. Ishii (2022): "Learning Efficiency of Multi-agent Information Structures," CEPR Discussion Paper No. DP16877, Available at SSRN: https://ssrn.com/abstract=4026761.

[5] Hoshino, T. (2022): "Multi-agent Persuasion: Leveraging Strategic Uncertainty," *International Economic Review* (forthcoming).

[6] Kajii, A. and S. Morris (1997): "The Robustness of Equilibria to Incomplete Information," *Econometrica*, 65, 1283–1309.

[7] Li, F., Y. Song and M. Zhao (2019): "Global Manipulation by Local Obfuscation," Available at http://dx.doi.org/10.2139/ssrn.3471491.

[8] Mathevet, L., J. Perego and I. Taneva (2020): "On Information Design in Games," *Journal of Political Economy*, 128, 1370–1404.

[9] Meyer, M. and B. Strulovici (2012): "Increasing Interdependence of Multivariate Distributions," *Journal of Economic Theory*, 147, 1460–1489.

[10] Monderer, D. and D. Samet (1989): "Approximating Common Knowledge with Common Beliefs," *Games and Economic Behavior*, 1, 170-190.

[11] Morris, S., and H. Shin (2007). Common belief foundations of global games. Working Paper, Princeton University, https://scholar.princeton.edu /sites/default/files/smorris/files/cbf.pdf.

[12] Morris, S., Oyama, D. and S. Takahashi (2020): "Implementation via Information Design in Binary-Action Supermodular Games," Available at SSRN: http://dx.doi.org/10.2139/ssrn.3697335

[13] Oyama, D. and S. Takahashi (2020): "Generalized Belief Operator and Robustness in Binary-Action Supermodular Games," *Econometrica*, 88, 693–726.

[14] Rubinstein, A. (1989): "The Electronic Mail Game: Strategic Behavior under Almost Common Knowledge," *American Economic Review*, 79, 385–391.

[15] Shaked, M. and G. Shantikumar (2008): *Stochastic Orders*, Springer, 2008.

[16] Steiner, J. and C. Stewart (2011): "Communication, Timing and Common Learning," *Journal of Economic Theory*, 146, 230–247.