Commonality of Information and Commonality of Beliefs*

Yu Awaya† and Vijay Krishna‡

January 22, 2025

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 may be true. In the long run, conditionally independent signals may achieve greater commonality of beliefs than correlated ones.

Keywords: Common Learning, Correlation

JEL Codes: D83

1 Introduction

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

---

*The research reported here was supported by a grant from the National Science Foundation (SES-2048806). We are grateful to Paulo Barelli, Mira Frick, Kevin He, Tetsuya Hoshino, Ryota Iijima, Yuhta Ishii and Zeming Ma for helpful comments. The referees and editor of this journal also provided valuable suggestions.

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

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

In this paper, we argue that this intuition may be misguided. We identify circumstances in which it is easier for agents to coordinate with *less* correlated signals than with those that are *more* correlated. In fact, it may be that coordination is easier when signals are (conditionally) independent than when they are correlated. We begin with a simple example that illustrates this phenomenon.

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 G ("good") or B ("bad"). The cost of investment is $c > 0$ 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.

Prior to making choices players receive signals that are informative about the state of nature. We will show that for some costs c, efficient coordination can be achieved when these signals are independently distributed but *not* when they are correlated.

First, suppose that the information available to players is generated as follows. Let $\boldsymbol{X} = (X_1, X_2)$ be a pair of binary signals each of which takes on values 0 ("bad news") or 1 ("good news"). The signal $X_i \in \{0, 1\}$ is privately observed by player i. 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. This means that a 1-signal is conclusive evidence that the state is G.

Prior to making decisions, player i sees *two* serially independent realizations of the signal X_i, say X_i^1 and X_i^2 (the state of nature is realized once and remains fixed). It is routine to verify that if $c \leq \frac{24}{25}$, then there is an equilibrium in which player i invests if he gets at least one positive signal, or equivalently, the sum of his private signals, $X_i^1 + X_i^2 \geq 1$. Moreover, if $c > \frac{24}{25}$ the only equilibrium is one in which no investment ever takes place.

Now consider an alternative situation in which players' signals are positively correlated. Specifically, suppose $\boldsymbol{Y} = (Y_1, Y_2)$ are signals that in state G, have the 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, in state G, the players' signals Y_1 and Y_2 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 (Y_1, Y_2). Player i observes Y_i^1 and Y_i^2 prior to making an investment decision. It is routine to verify that if $c \le \frac{47}{50}$, then there is an equilibrium in which player i invests if he gets at least one positive signal, or equivalently, the sum of his private signals, $Y_i^1 + Y_i^2 \ge 1$.

On the other hand, 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 it is dominant to not invest because $\Pr[G \mid Y_i^1 + Y_i^2 = 0] = \frac{1}{26} < c$. Next, if $Y_i^1 + Y_i^2 = 1$, it is iteratively dominant to not invest for $j \ne i$, $\Pr\left[Y_j^1 + Y_j^2 \ge 1 \mid Y_i^1 + Y_i^2 = 1\right] = \frac{47}{50} < c$ as well. Finally, it is then 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 the following.

a. If $c \le \frac{47}{50}$, then with either conditionally independent signals $\boldsymbol{X}$ or correlated ones $\boldsymbol{Y}$, there is an equilibrium with efficient coordination—a player invests if she gets at least one positive signal and so knows that the state is G.

b. If $\frac{47}{50} < c \le \frac{48}{50}$, however, with conditionally independent signals $\boldsymbol{X}$, there is an equilibrium in which both players invest whenever they know G, while with correlated signals $\boldsymbol{Y}$, the unique equilibrium is that no player ever invests.

c. If $c > \frac{48}{50}$, the only equilibrium under either signals $\boldsymbol{X}$ or $\boldsymbol{Y}$ is to not invest.

Why is this? Compared to the case of (conditionally) independent signals, with correlated signals a player that gets good news is more likely to believe that the other player also received good news and so becomes optimistic about the prospects of coordinating on the right outcome. But the opposite is true for a player that gets bad news. With correlated signals, she is more likely to believe that the other player also received bad news and so becomes pessimistic. The second effect dominates—a player with one piece of good news and one piece of bad news is more pessimistic with correlated signals than with independent signals, that is,

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

This type's increased pessimism then spreads to all types.

In the rest of this paper, we explore these ideas in a special case of the common learning setting of Cripps, Ely, Mailath and Samuelson (2008, henceforth CEMS) where both fundamental states of nature and the agents' signals are binary.[1] 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 of nature θ is realized in period 0 and remains fixed. There are T additional periods and in each period t, agents receive private signals $X_i^t \in \{0, 1\}$ that are informative about θ. The signals are independent and identically distributed across time but may be correlated among agents. CEMS showed that in the limit as $T \to \infty$, the true state of nature becomes approximately commonly known with probability approaching one.

We are interested in studying how the commonality of agents' beliefs—that is, how close they are to achieving common knowledge of θ—is affected by the degree of commonality of their information. As in CEMS, "commonality of beliefs" is formalized using the notion of common p-belief introduced by Monderer and Samet (1989). "Commonality of information" is formalized using a multivariate version of "more positively correlated," defined in the next section.

Fix, as in the example, two signals $\boldsymbol{X}$ or $\boldsymbol{Y}$, such that $\boldsymbol{Y}$ exhibits greater positive (but not perfect) correlation than $\boldsymbol{X}$. We show that for any T large enough, there is an interval of p's such that the state of nature can be common p-believed with the less correlated signals $\boldsymbol{X}$ but not with the more correlated signals $\boldsymbol{Y}$. Thus, under the identified conditions, "greater commonality of information is detrimental to commonality of beliefs."[2]

We begin by considering the case when signals are *conclusive,* in the sense that even one piece of "good news" reveals that the state is G (as in the example). This special case is useful because first-order uncertainty—that is, concerning the state of nature θ—is resolved once even a single piece of good news is received. This means that the focus is then solely on higher-order uncertainty—that is, concerning others' knowledge about G, their knowledge about others' knowledge, etc.

We first show that the event that G is common p-believed exhibits a *bang-bang* property: if p is below a threshold, this event is as large as possible and if p is above, it is empty (Proposition 3.1).

[1] A working paper version of this paper (Awaya and Krishna, 2024) studies a more general environment in which the number of signals may exceed two.

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

The relevant threshold is the belief about the event "all agents know G" of the *second-most* pessimistic type—who gets only one piece of good news and $T-1$ pieces of bad news. Only the type that gets only bad news in every period is more pessimistic. We show that whether or not G can be common p-believed depends on whether $p \leq q$ or $p > q$. If $p \leq q$, then G is common p-believed whenever everyone knows G. On the other hand, if $p > q$, then it is *impossible* for G to be common p-believed. Why is this? By definition, the belief of the second-most pessimistic type is too low and so this type cannot believe that all others know G. We show that the pessimism of this type then "infects" all other types so that no one assigns probability greater than p to the event that everyone knows G.

The second step is to show that higher correlation *decreases* the threshold belief q when T is large (Proposition 3.2). As argued above, the second-most pessimistic type is one who receives only one piece of good news. Since this type gets a preponderance of bad news, higher correlation makes her believe that other agents also received a preponderance of bad news, thereby increasing her pessimism. These facts then lead to one of main results (Theorem 1). Consider two kinds of signals, one more correlated than the other. For large enough T, there is an interval of p's (depending on T) such that for all p in that interval, with the more correlated signals, G cannot be common p-believed, but with the less correlated signals, it can be.

In Section 4 we relax the assumption that signals are conclusive. In this more general environment, first-order uncertainty also plays a role. With non-conclusive signals, the "bang-bang" property requires the assumption that this first-order uncertainty is not too large (Proposition 4.1). Because of this, when signals are non-conclusive, the main result, Theorem 2, also requires stronger conditions than Theorem 1.

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 Q is more informative than P, if agent i's signal Y_i from Q is more informative about agent j's signal Y_j than X_i from P is about X_j (see Section 5 for a precise definition). In the same vein as above, it can be shown that in fact, more informative signals can be 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.[3] The literature on common learning asks whether such uncertainty can be made to disappear over time. 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) 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.

Frick, Iijima and Ishii (2023) study how common learning is affected by the underlying signal process. Consider joint distributions over states of nature and signals, P and Q, such that P is more informative about the state θ than is Q. Frick et al. (2023) show that when T is large enough, P results in greater commonality of beliefs than does Q. In particular, how correlated agents' signals are does not matter in the long run. In our work we compare distributions P and Q that are *equally* informative about θ but Q is more correlated than P. We show that when T is large enough, greater correlation may, in fact, be 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). Oyama and Takahashi (2020) study binary-action supermodular games, and as in Example 1, our results on the effects of correlation on common learning have natural counterparts when applied to this class of games.

A paper by Basak, Deb and Kuvalekar (2024) also studies how "commonality of information" can decrease the prospects of coordinated action in regime change games. Unlike our work, the channel by which this results relies on the particular payoff structure of the game.

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

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

There is also work on global games which studies how greater "commonality"—measured by a decrease in the variance of private information relative to that of public information—can, in some circumstances, lead to decreased coordination in equilibrium (see for instance, Iachan and Nenov, 2015). Unlike in our work, in the global games framework, agents' signals are independent conditional on the state of nature θ. The increase in "commonality" of the sort mentioned above affects agents' beliefs about each other only via the change in their beliefs about θ. In our paper, the increase in commonality increases the correlation among agents' signals while keeping their beliefs about the fundamental state θ fixed.

2 Model

A group of agents $i \in \mathcal{I} = \{1, 2, ..., I\}$ 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 remaining periods. At each time $t \geq 1$, each agent i receives a *private* signal that is informative about the state of nature θ.

We assume throughout that signals are *binary* so that $\mathcal{X} = \{0, 1\}$.

The signals are generated as follows.

Let $P \in \Delta\left(\Theta \times \mathcal{X}^I\right)$ be a joint probability distribution over the set of states and signals, one for each agent. We will write a typical element of $\Theta \times \mathcal{X}^I$ as $(\theta, \boldsymbol{x}) = (\theta, x_1, x_2, ..., x_I)$ where x_i is the signal of agent i. Of course, the marginal probability of $\theta = G$ is ρ. To save on notation, we will write $P^\theta \in \Delta\left(\mathcal{X}^I\right)$ as the distribution over signal vectors conditional on the state of nature θ. Thus, $P^\theta(\boldsymbol{x}) = P(\boldsymbol{x} \mid \theta)$.

We will assume that

1. $P^G \neq P^B$ so that the signals carry information about θ.
2. Conditional on θ, the signals are *symmetrically* distributed—that is, $P^\theta(\boldsymbol{x}) = P^\theta(\boldsymbol{x}^\pi)$ for any permutation $\boldsymbol{x}^\pi$ of $\boldsymbol{x}$.

3. P is *affiliated*, that is, for all $(\theta, \boldsymbol{x})$ and $(\theta', \boldsymbol{x}')$,

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

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

Let $\boldsymbol{x}^t \in \mathcal{X}^I$ be the vector of signals, one for each agent, in period t. Conditional on θ, in any period t, the signal vectors $\boldsymbol{x}^t \in \mathcal{X}^I$ are independent draws from the distribution $P^\theta(\cdot) = P(\cdot \mid \theta)$. Thus, in each state of nature θ, the signal vectors are *independently and identically distributed* over time.

It will be convenient to consider the $I+1$ dimensional random vector $(\widetilde{\theta}, \boldsymbol{X})$ which takes values in $\Theta \times \mathcal{X}^I$ and satisfies $\Pr[(\widetilde{\theta}, \boldsymbol{X}) = (\theta, \boldsymbol{x})] = P(\theta, \boldsymbol{x})$.[5] Similarly, for each θ, define the I dimensional random vector $\boldsymbol{X}^\theta$ which takes values in $\mathcal{X}^I$ and satisfies $\Pr\left[\boldsymbol{X}^\theta = \boldsymbol{x}\right] = P^\theta(\boldsymbol{x}) \equiv \Pr[\boldsymbol{X} = \boldsymbol{x} \mid \theta]$.[6]

Now let $Q \in \Delta(\Theta \times \mathcal{X}^I)$ be another distribution such that the marginal probability of G is ρ. Analogously, let $(\widetilde{\theta}, \boldsymbol{Y})$ be the random vector such that $\Pr[(\widetilde{\theta}, \boldsymbol{Y}) = (\theta, \boldsymbol{y})] = Q(\theta, \boldsymbol{y})$. And like $\boldsymbol{X}^\theta$, the random vector $\boldsymbol{Y}^\theta$ also takes values in $\mathcal{X}^I$ and satisfies $\Pr\left[\boldsymbol{Y}^\theta = \boldsymbol{y}\right] = Q^\theta(\boldsymbol{y}) \equiv \Pr[\boldsymbol{Y} = \boldsymbol{y} \mid \theta]$.

Throughout the paper we will assume that $\boldsymbol{X}$ is defined as above from P and $\boldsymbol{Y}$ is defined as above from Q.

We will compare two distributions P and Q such that Q is "more correlated" than P; or equivalently, the signals $\boldsymbol{Y}$ are "more correlated" than signals $\boldsymbol{X}$.

Multivariate correlation When there are more than two variables, there are many ways to measure an increase in correlation (or positive dependence). In what follows, we will use the following notion[7]:

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

$$\Pr[\boldsymbol{X} \leq \boldsymbol{z}] \leq \Pr[\boldsymbol{Y} \leq \boldsymbol{z}] \tag{1}$$

[5] Formally, if $\mathcal{S} = \mathcal{X}^I$, then $(\Theta \times \mathcal{S}, 2^{\Theta \times \mathcal{S}}, P)$ is a finite probability space and $(\widetilde{\theta}, \boldsymbol{X})$ is the identity map from $\Theta \times \mathcal{S}$ to $\Theta \times \mathcal{S}$.

[6] Again, $(\mathcal{S}, 2^{\mathcal{S}}, P^\theta)$ is a probability space and $\boldsymbol{X}^\theta$ is the identity map from $\mathcal{S}$ to $\mathcal{S}$.

[7] This order was first defined by Yanagimoto and Okamoto (1969). It was then developed for $I > 2$ by Joe (1990), who called it the "concordance order".

and

$$\Pr[\boldsymbol{X} \geq \boldsymbol{z}] \leq \Pr[\boldsymbol{Y} \geq \boldsymbol{z}]. \quad (2)$$

If $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$, then for any fixed vector $\boldsymbol{z}$, $\boldsymbol{Y}$ is more likely to take on higher values than $\boldsymbol{z}$ than is $\boldsymbol{X}$ *and* also more likely to take on lower values than $\boldsymbol{z}$. In the bivariate case, this means that a change from P to Q shifts probability weight from the "northwest" and "southeast" quadrants to the "northeast" and "southwest" quadrants. Thus, the values that the variables take are more likely to be closer to each other than before. The PQD order is discussed in detail in Shaked and Shanthikumar (2008) and Meyer and Strulovici (2012).[8] It satisfies the following desirable properties.

First, if $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$, then they have identical univariate marginals, that is, for all $k \in \mathcal{X}$,

$$\Pr[X_i = k] = \Pr[Y_i = k].$$

Second, the PQD order is preserved by monotone transformations of the variables. In other words, if the variables $(Y_1, Y_2, ..., Y_I)$ are more correlated in the PQD order than $(X_1, X_2, ..., X_I)$, then it should be that $(\phi_1(Y_1), \phi_2(Y_2), ..., \phi_I(Y_I))$ are also more correlated than $(\phi_1(X_1), \phi_2(X_2), ..., \phi_I(X_I))$ where each ϕ_i is an increasing function.[9] This is desirable since signals have no inherent cardinal meaning—they only serve to update beliefs.

Third, the PQD order is preserved for marginals over subsets of variables, that is, if the variables $\boldsymbol{Y}$ are more correlated than $\boldsymbol{X}$ then for any non-empty $J \subseteq I$, it should be that the variables $\boldsymbol{Y}_J = (Y_i)_{i \in J}$ are more correlated than $\boldsymbol{X}_J = (X_i)_{i \in J}$. If $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$, then for all i and $j \neq i$, the pairwise covariances satisfy $\text{Cov}(Y_i, Y_j) \geq \text{Cov}(X_i, X_j)$.

Finally, and perhaps most important, the PQD order is *weaker* than all other orders of positive dependence discussed in the references above—for instance, it is weaker than the supermodular order $\succcurlyeq_{SM}$ which requires that $\boldsymbol{Y} \succcurlyeq_{SM} \boldsymbol{X}$ if $E_{\boldsymbol{Y}}[\phi] \geq E_{\boldsymbol{X}}[\phi]$ for all supermodular functions ϕ (see Shaked and Shanthikumar, 2008).

In what follows, we will use the following *strict* version of the PQD order. We will say that $\boldsymbol{Y}$ is *strictly more correlated* than $\boldsymbol{X}$ in the PQD order, and write $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$, if $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$ and (i) the inequality (1) is strict for any $\boldsymbol{z}$ such that for

[8] See Anderson and Smith (2024) for an application of the PQD order in matching problems.

[9] Note that the common (bivariate) notion of greater covariance fails this requirement. It may be that $\text{Cov}(Y_1, Y_2) > \text{Cov}(X_1, X_2)$ but $\text{Cov}(\phi_1(Y_1), \phi_2(Y_2)) < \text{Cov}(\phi_1(X_1), \phi_2(X_2))$. As an example, let $\phi_1(z) = \phi_2(z) = z^2$.

at least two indices i, $z_i = 0$; and (ii) the inequality (2) is strict for any $\boldsymbol{z}$ such that for at least two indices i, $z_i = 1$.[10]

Since the the PQD order $\succcurlyeq_{PQD}$ is implied by other orders, its strict version, $\succ_{PQD}$, will be implied by analogous strict versions of other orders.[11]

Common beliefs A *state of the world*

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

determines the state of nature θ as well as 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. We will refer to a vector $\mathbf{x}_i \in \mathcal{X}^T$ as the *type* of agent i. The set of states of the world is

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

Following Monderer and Samet (1989), given any event $E \subseteq \Omega$ and probability p, the event $B_i^p(E)$ consists of states $\omega \in \Omega$ in which E is *p-believed* by i, 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 E is p-believed by everyone.

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

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

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

$$C^p(E) = \cap_{\ell \geq 1} B^{p,\ell}(E).$$

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

[10] If, for instance, $\boldsymbol{z} = (0, 1, 1, ..., 1)$ then $\Pr[\boldsymbol{X} \leq \boldsymbol{z}] = \Pr[\boldsymbol{Y} \leq \boldsymbol{z}]$ since both equal the marginal probability that $X_1 = 0$.

[11] To see this for the supermodular order, first let $S^-(\boldsymbol{z}) = \{\boldsymbol{x} : \boldsymbol{x} \leq \boldsymbol{z}\}$ be the quadrant below $\boldsymbol{z}$ and $S^+(\boldsymbol{z})$ be the quadrant above $\boldsymbol{z}$. The indicator functions, $I_{S^-(\boldsymbol{z})}$ and $I_{S^+(\boldsymbol{z})}$, are both supermodular. Now say that $\boldsymbol{Y} \succ_{SM} \boldsymbol{X}$, if $\boldsymbol{Y} \succcurlyeq_{SM} \boldsymbol{X}$ and (i) $E_{\boldsymbol{Y}}\left[I_{S^-(\boldsymbol{z})}\right] > E_{\boldsymbol{X}}\left[I_{S^-(\boldsymbol{z})}\right]$ for any $\boldsymbol{z}$ such that for at least two i, $z_i = 0$; and (ii) $E_{\boldsymbol{Y}}\left[I_{S^+(\boldsymbol{z})}\right] > E_{\boldsymbol{X}}\left[I_{S^+(\boldsymbol{z})}\right]$ for any $\boldsymbol{z}$ such that for at least two i, $z_i = 1$. It is now clear that $\boldsymbol{Y} \succ_{SM} \boldsymbol{X}$ implies $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$.

(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.

We are interested in the set $C^p\left(\Omega^G\right)$ after T periods, where $\Omega^G = \{\omega : \theta = G\}$. In other words, we are interested in the set of states of the world in which 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 Conclusive Signals

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

1. a signal $X_i = 1$ is *conclusive* about G—that is, $\Pr[X_i = 1 \mid B] = 0$; and
2. signals have *full support* in state G, for all $\boldsymbol{x}$, $P^G(\boldsymbol{x}) > 0$.

Note that signals are perfectly correlated in state B.

Since signals are binary, the fact that they are independently and identically distributed over time implies that an agent's type can effectively be represented simply by the *total* number of 1-signals received. Thus, a type $\mathbf{x}_i$ can be represented simply as $n_i = \sum_t x_i^t$ and so types can be *linearly* ordered. Let $N_i = \sum_t X_i^t$ denote the random variable which equals the sum of i's signals.

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 of nature 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 of the world in which every agent j received a signal $x_j^t = 1$ at some time t. Since even one positive signal is conclusive about G, at any $\omega \in \Omega^+$ it must be that $\theta = G$. Formally, $\Omega^+ \subseteq \Omega^G = \{\omega : \theta = G\}$. Define

$$q = \Pr\left[\Omega^+ \mid N_i = 1\right] \tag{3}$$

to be the belief of type $N_i = 1$ about the event that everyone else saw at least one positive signal—and so also knows G. Note that Ω^+ and q depend on T although we have suppressed this dependence to reduce the notational burden.

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

$$\Pr\left[\Omega^+ \mid N_i = n\right] \geq \Pr\left[\Omega^+ \mid N_i = 1\right] = q, \tag{4}$$

as established in Lemma A.2 in the Appendix. In other words, among all those that know G, type $N_i = 1$ is most pessimistic about the event that everyone also knows G. Put another way, type $N_i = 1$ is the *second-most* pessimistic type—type $N_i = 0$ is the most pessimistic, of course.

3.1 First result

Consider two signal distributions P and Q with identical univariate marginals. Let $q_{\boldsymbol{X}} = \Pr_{\boldsymbol{X}}[\Omega^+ \mid \sum_t X_i^t = 1]$ as in (3) and let $q_{\boldsymbol{Y}} = \Pr_{\boldsymbol{Y}}[\Omega^+ \mid \sum_t Y_i^t = 1]$ be the analogous belief derived from signals $\mathbf{Y}$.[12]

Define

$$\rho_0 = \Pr\left[\Omega^G \mid N_i = 0\right] \tag{5}$$

to be the belief about G of an agent who receives only 0-signals in each of the T periods. Note that ρ_0 is the same for P and Q as they have the same marginals. As T increases, ρ_0 goes to zero. Note also that Ω, $q_{\boldsymbol{X}}$, $q_{\boldsymbol{Y}}$, as well as ρ_0 all depend on T although we have suppressed this dependence, again to avoid notational clutter.

The main result of this section is[13]:

Theorem 1 *Suppose signals* $\mathbf{X}$ *and* $\mathbf{Y}$ *are conclusive.*

(i) For any T*, if* $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$*, then for* $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$*,*

$$C^p_{\boldsymbol{Y}}\left(\Omega^G\right) = \varnothing \text{ and } C^p_{\boldsymbol{X}}\left(\Omega^G\right) = \Omega^+,$$

that is, G *cannot be common* p*-believed with* $\mathbf{Y}$ *whereas* G *is common* p*-believed with*

[12] The symbol $\Pr_{\boldsymbol{X}}$ indicates that the probability is calculated using P and similarly, $\Pr_{\boldsymbol{Y}}$ is calculated using Q.

[13] $C^p_{\boldsymbol{X}}\left(\Omega^G\right)$ is the set of states of the world in which Ω^G is common p-believed when all the probabilities are calculated using P and $C^p_{\boldsymbol{Y}}\left(\Omega^G\right)$ is the same set when they are calculated using Q. Note also that these depend on T as well.

$\boldsymbol{X}$ *whenever everyone knows* G.

(ii) If $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$*, then for* T *large enough,* $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$.

Theorem 1 says that when T is large enough, there is a non-empty open interval of p's, depending on T, such that for any p in that interval, it is impossible for G to be common p-believed with the more correlated signals $\boldsymbol{Y}$ while it is possible with the less correlated signals $\boldsymbol{X}$.

A few remarks on the theorem are in order.

First, the theorem automatically implies that in the identified circumstances, $\Pr_{\boldsymbol{Y}}[C^p_{\boldsymbol{Y}}\left(\Omega^G\right)] < \Pr_{\boldsymbol{X}}[C^p_{\boldsymbol{X}}\left(\Omega^G\right)]$ since the left-hand probability is zero and the right-hand probability is positive. In this sense, when T is large, greater commonality of information *reduces* the commonality of beliefs.

Second, since we have assumed that Q^G has *full support*, the signals $\boldsymbol{Y}$ are not public—that is, they are not perfectly correlated. If the signals $\boldsymbol{Y}$ were public, then we would have that for all p, $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) = \Omega^+$, which would run counter to (1). But what if $\boldsymbol{Y}$ is "almost" public—that is, for some small ε, for all $k \in \mathcal{X}$, $\Pr\left[\forall j, Y_j = k \mid Y_i = k\right] > 1 - \varepsilon$? 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 almost public, 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 beliefs $q_{\boldsymbol{X}}$ and $q_{\boldsymbol{Y}}$ as functions of T for the two signal distributions in Example 1—the (conditionally) independent signals $\boldsymbol{X}$ and the correlated signals $\boldsymbol{Y}$. For the example, $q_{\boldsymbol{X}} > q_{\boldsymbol{Y}}$, for all $T \geq 2$. Of course, $q_{\boldsymbol{X}}$ and $q_{\boldsymbol{Y}}$ converge to 1 as T increases without bound.

Third, the theorem 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.

Fourth, note also that in Theorem 1 part (1), T must be at least 2—the conclusion cannot hold for $T = 1$. This is because if $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$, then with conclusive signals,

$$\begin{aligned} q_{\boldsymbol{X}} &= \Pr\left[\forall j, X_j = 1 \mid X_i = 1\right] \\ &< \Pr\left[\forall j, Y_j = 1 \mid Y_i = 1\right] \\ &= q_{\boldsymbol{Y}} \end{aligned}$$

and so when $T = 1$, for all p, $C^p_{\boldsymbol{X}}\left(\Omega^G\right) \subseteq C^p_{\boldsymbol{Y}}\left(\Omega^G\right)$.

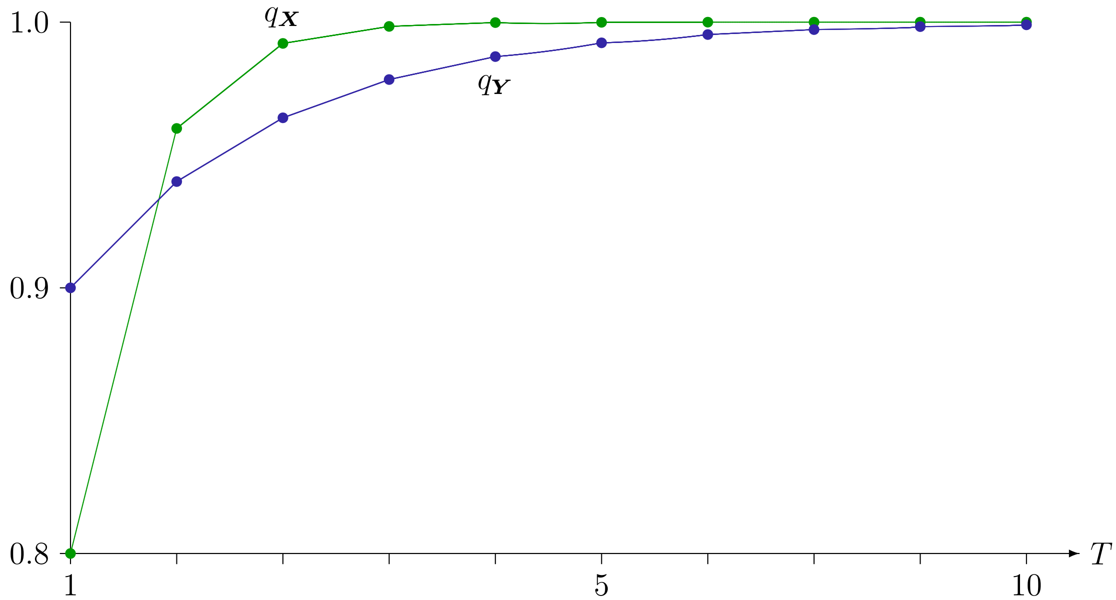

Figure 1: Threshold Beliefs for the Two Signals in Example 1

Finally, if we define T_0 as the smallest T for which $q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$, then T_0 is "relatively small". This is most easily seen when $I = 2$ as the condition that $q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$ is then equivalent to

$$L \equiv \frac{P^G(1,0)}{Q^G(1,0)} < \left(\frac{Q^G(0,0)}{P^G(0,0)}\right)^{T-1} \equiv R^{T-1}.$$

Now $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$ implies that both L and R are greater than one. If $1 < L < R$, then, of course, $T_0 = 2$. And if $1 < R < L$, then since the right-hand side of the inequality above grows exponentially, it will overtake the left-hand side very quickly, that is, for a relatively small T_0. Precisely, when $L > R$, $T_0 = 1 + \lceil \ln L - \ln R \rceil$ where $\lceil z \rceil$ denotes the smallest integer that exceeds z.

3.2 Proof of Theorem 1

The proof of Theorem 1 has two components. We first show that with conclusive signals, for any T, the set $C^p\left(\Omega^G\right)$ has a "bang-bang" property—it is either quite large or empty. Precisely, if $p \leq q$, then $C^p\left(\Omega^G\right)$ is as large as possible—any state of the world in which everyone knows that $\theta = G$ is included. But if $p > q$, $C^p\left(\Omega^G\right)$ is empty. Thus, $C^p\left(\Omega^G\right)$ suddenly goes from being large to being empty as p crosses the threshold q. This is Proposition 3.1 below.

The second step in the proof of Theorem 1 then shows that when T is large enough, an increase in the correlation among agents' signals leads to an increase in the pessimism of the pivotal type who gets only one positive signal. This is Proposition 3.2 below.

3.2.1 Bang-bang property

The important "bang-bang" property of $C^p\left(\Omega^G\right)$, which may be of independent interest, is derived in the following proposition.

Proposition 3.1 *Suppose signals are conclusive. For any T,*

(i) if $\rho_0 < p \leq q$, then

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

(ii) and if $\rho_0 < q < p$, then

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

A formal proof of the proposition is below but the underlying arguments run as follows.

Part (i) is rather intuitive. Consider the type $n_i = 1$ that gets exactly one positive signal. Since signals are conclusive, this type knows G. Moreover, this type assigns probability $q \geq p$ to the event that all others also know G. Because signals are affiliated, *all* types $n_j \geq 1$ also assign probability of at least q to the same event. The fact that G is common p-believed now follows.

Part (ii) says that, in a strong sense, the converse is true as well. Again, consider the type $n_i = 1$ that gets exactly one positive signal. As above, since signals are conclusive, this type knows that G has occurred but assigns only probability $q < p$ to the event that all others also know G. So this type cannot be in $C^p\left(\Omega^G\right)$. Now an infection argument takes over. Consider type $n_i = 2$ with two positive signals. This type is only concerned with the event that all other agents are of type $n_j \geq 2$ since all those with $n_j = 1$ have already been ruled out. We show that type $n_i = 2$ assigns a *lower* probability to the event that all others are of type $n_j \geq 2$, than type $n_i = 1$ assigns to the event that all others are of type $n_j \geq 1$. Why is this? There are two forces at work here. First, the event that all $n_j \geq 2$ is a subset of the event that all $n_j \geq 1$ and, all else being equal, the former has a lower probability than the latter. But on the other hand, affiliation implies that type $n_i = 2$ assigns a higher probability to any event of the sort $n_j \geq n$ than does $n_i = 1$. We show that when signals are serially independent, the first effect is always stronger and so the probability of the event that all $n_j \geq n$ assigned by type $n_i = n$ decreases with n. This now means that the type $n_i = 2$ is also excluded from $C^p\left(\Omega^G\right)$. Once those with $n_i = 2$ are excluded, this argument now carries over to $n_i = 3$ and so on.

Two assumptions are crucial for the argument above. First, since signals X_i are binary, the types N_i can be linearly ordered by the number of positive signals. Second, the types N_i are the result of T identical and *independent* draws of X_i.

Proof of Proposition 3.1 (i) If $\rho_0 < p \leq q$, then the fact that signals are conclusive implies that all types with $n_i \geq 1$ assign probability 1 to the event Ω^G and hence, of course, assign at least probability q to Ω^G. On the other hand, type $n_i = 0$ assigns a probability $\rho_0 < q$ to the event Ω^G. Thus, $B_i^q\left(\Omega^G\right) = \{\omega : n_i \geq 1\}$ and so

$$B^q\left(\Omega^G\right) = \{\omega : \forall j, n_j \geq 1\} = \Omega^+. \tag{6}$$

Moreover, (4) implies that all types 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

$$\Omega^+ \subseteq B^q\left(\Omega^+\right). \tag{7}$$

We will argue by induction that for all $\ell \geq 1$, that $\Omega^+ \subseteq B^{q,\ell}\left(\Omega^G\right)$.

Now (6) implies that the statement is true for $\ell = 1$. So suppose that for some $\ell > 1$, $\Omega^+ \subseteq B^{q,\ell-1}\left(\Omega^G\right)$. Operating on both sides by the monotone operator B^q, we have $B^q\left(\Omega^+\right) \subseteq B^{q,\ell}\left(\Omega^G\right)$. But from (7), $\Omega^+ \subseteq B^{q,\ell}\left(\Omega^G\right)$.

Thus, for all ℓ, $\Omega^+ \subseteq B^{q,\ell}\left(\Omega^G\right)$ and hence $\Omega^+ \subseteq C^q\left(\Omega^G\right)$. Finally, since $p \leq q$, $C^q\left(\Omega^G\right) \subseteq C^p\left(\Omega^G\right)$.

Proof of Proposition 3.1 (ii) Now suppose $\rho_0 < q < p$.

For $n = 0, 1, ..., T+1$, define

$$\Gamma^{(n)} = \{\omega : \forall j, n_j \geq n\}$$

as the set of states of the world ω in which every agent gets *at least* n signals $X_i^t = 1$.

Clearly, for any n, $\Gamma^{(n+1)} \subset \Gamma^{(n)}$ and $\cap_{n=0}^{T+1}\Gamma^{(n)} = \varnothing$ since $\Gamma^{(T+1)} = \varnothing$.

We will argue by induction that for all $n \leq T+1$,

$$C^p\left(\Omega^G\right) \subseteq \Gamma^{(n)}. \tag{8}$$

First, since $\Gamma^{(0)} = \{\omega : \forall j, n_j \geq 0\} = \Omega$, (8) holds for $n = 0$.

Now suppose that $C^p\left(\Omega^G\right) \subseteq \Gamma^{(n)}$. Let $\omega' \in \Gamma^{(n)} \setminus \Gamma^{(n+1)}$. At any such ω', there is an i with $n_i = n$, that is, i gets exactly n positive signals and since $C^p\left(\Omega^G\right) \subseteq \Gamma^{(n)}$,

$$\Pr\left[C^p\left(\Omega^G\right) \mid N_i = n\right] \leq \Pr\left[\Gamma^{(n)} \mid N_i = n\right].$$

Lemma B.1 now implies that

$$\begin{aligned}\Pr\left[C^p\left(\Omega^G\right) \mid N_i = n\right] &\leq \Pr\left[\Gamma^{(1)} \mid N_i = 1\right] \\ &= q\end{aligned}$$

and since $p > q$, $\omega' \notin B_i^p\left(C^p\left(\Omega^G\right)\right)$ and hence $\omega' \notin C^p\left(\Omega^G\right)$. Thus, we have argued that $C^p\left(\Omega^G\right) \subseteq \Gamma^{(n+1)}$ and hence established (8).

Now since $C^p\left(\Omega^G\right) \subseteq \Gamma^{(n)}$ for all n and $\cap_{n=0}^{T+1}\Gamma^{(n)} = \varnothing$, we have that $C^p\left(\Omega^G\right) = \varnothing$.

This completes the proof of Proposition 3.1. ■

3.2.2 Correlation increases pessimism

Proposition 3.1 establishes that with 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 second-most pessimistic agent. In this section, we compare two signal distributions such that $\mathbf{Y}^G \succ_{PQD} \mathbf{X}^G$.[14] We show that a change from $\mathbf{X}^G$ to $\mathbf{Y}^G$ increases the pessimism of type $n_i = 1$.

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

$$q_{\mathbf{Y}} < q_{\mathbf{X}}$$

Proof. Follows from Lemma A.3 and Lemma C.1 in the Appendix. ■

[14] Recall that $\mathbf{X}^\theta$ is a random vector such that $\Pr[\mathbf{X}^\theta = \boldsymbol{x}] = \Pr[\mathbf{X} = \boldsymbol{x} \mid \theta]$. $\mathbf{Y}^\theta$ is similarly defined.

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, with more correlated signals, 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, if signals $\boldsymbol{Y}$ are more correlated than $\boldsymbol{X}$, then

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

while at the same time

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

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 1 holds if $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$ and when T is large enough, $\rho_0 = \Pr\left[\Omega^G \mid N_i = 0\right] < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$.

4 Non-conclusive signals

The sharp result in Theorem 1 was derived for the case of conclusive signals. The sharp result obtains because with conclusive signals, one may 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 of nature θ—also plays a role.

In this section, we assume that conditional on $\theta \in \{G, B\}$, the distribution P has *full support*. This means that a signal $X_i = 1$ does not provide conclusive evidence that the state is G. Recall that since P is affiliated, it is still the case that a signal $X_i = 1$ is more indicative that $\theta = G$ than a signal $X_i = 0$.

Let

$$\mathbf{e}^1 = (1, 0, ..., 0) \in \mathcal{X}^T$$

denote the type that receives a signal of 1 in period 1 and 0's thereafter.[15] Define

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

where, as before, $\Omega^{+} = \{\omega : \forall j, \mathbf{x}_j \neq \mathbf{0}\}$ is the set of states of the world in which everyone gets a signal $X_i = 1$ at least once. Note that because of affiliation, type $\mathbf{e}^1$ is the *second-most* pessimistic type about both Ω^G and Ω^{+}. Only type $\mathbf{0}$ is more pessimistic.

Let $q_{\boldsymbol{Y}}$ be defined in a manner analogous to $q_{\boldsymbol{X}}$.

As in (5), let

$$\rho_0 = \Pr \left[\Omega^G \mid \mathbf{X}_i = \mathbf{0} \right]$$

to be the belief of type $\mathbf{0}$ about G and define

$$\rho_1 \equiv \Pr \left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1 \right]$$

to be the belief of type $\mathbf{e}^1$ about G. Note that if a 1-signal is conclusive, as in last section, then $\rho_1 = 1$. Note that if $\boldsymbol{X}$ and $\boldsymbol{Y}$ are such that conditional on θ, they have the same univariate marginal distribution μ^θ, then both ρ_0 and ρ_1 are the same for $\boldsymbol{X}$ and $\boldsymbol{Y}$. Moreover, the prior probability ρ of G is the same.

4.1 Second result

Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be such that conditional on θ, they have the same univariate marginal distribution μ^θ. Then we have,

Theorem 2

Suppose signals $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are non-conclusive.*

(i) For any T*, if* $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}} < \rho_1$*, then for* $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$*,*

$$C^p_{\boldsymbol{Y}} \left(\Omega^G \right) = \varnothing \text{ and } C^p_{\boldsymbol{X}} \left(\Omega^G \right) = \Omega^{+},$$

that is, G *cannot be common* p*-believed with* $\boldsymbol{Y}$ *whereas* G *is common* p*-believed with* $\boldsymbol{X}$ *whenever everyone gets at least one signal* $X_i = 1$*.*

[15] Since conditional on the state of nature, G or B, signals are serially independent, the beliefs of type $\mathbf{e}^1$ are the same as that of type $\mathbf{e}^2 = (0, 1, 0...0)$ etc. So it enough to consider $\mathbf{e}^1$ as representing all types that got one 1-signal and $T - 1$ signals of 0.

(ii) If $\mathbf{Y}^{\theta} \succ_{PQD} \mathbf{X}^{\theta}$ *for* $\theta = G, B$*, then for* T *large enough,* $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$*.*

Again, Theorem 2 automatically implies that when $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$, $\Pr_{\boldsymbol{Y}}[C^p_{\mathbf{Y}}(\Omega^G)] < \Pr_{\boldsymbol{X}}[C^p_{\boldsymbol{X}}(\Omega^G)]$ since the left-hand probability is zero and the right-hand probability is positive. Like Theorem 1, Theorem 2 says that, under the identified circumstances, greater commonality of information reduces the commonality of beliefs.

With non-conclusive signals, it is possible that even when $q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$, it is the case that $\rho_1 \leq q_{\boldsymbol{Y}}$. This, of course, is impossible in the conclusive-signal model of Section 3 where $\rho_1 = 1$.

4.2 Proof of Theorem 2

Like Theorem 1, the proof of Theorem 2 is in two steps.

We first prove, for non-conclusive signals, an analog of Proposition 3.1.

The second step again shows that when T is large enough, an increase in the correlation among agents' signals again increases the pessimism of the second-most pessimistic type $\mathbf{e}^1$. This is Proposition 4.2 below.

4.2.1 Threshold beliefs

Recall that $\rho_0 = \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{0}\right]$ and $\rho_1 = \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right]$. The following proposition derives the "bang-bang" property when signals are not conclusive. Because now first-order uncertainty also plays a role, an additional condition that ρ_1 is not too small is needed.

Proposition 4.1 *Suppose signals are non-conclusive. For any* T*,*

(i) if $\rho_0 < p \leq q \leq \rho_1$ *then*

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

(ii) and if $\rho_0 < q < p$*, then*

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

Proof. First, in both (i) and (ii), $\rho_0 < p$ and we claim that

$$C^p\left(\Omega^G\right) \subseteq \Omega^+. \tag{9}$$

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

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

and so

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

Part (i) We now argue that if $p \leq q$, $\Omega^+ \subseteq C^p\left(\Omega^G\right)$ and together with (9), this will imply (i),

By assumption, $p \leq q < \rho_1 = \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right]$. 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, for all i,

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

Taking the intersection over i, we have

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

In a similar manner, affiliation implies that for any $\mathbf{x}_i \neq \mathbf{0}$, it is also the case that $\Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right] \leq \Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{x}_i\right]$ 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

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

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

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

Since $\Omega^+ \subseteq B^p\left(\Omega^G\right)$ we have that $C^p\left(\Omega^+\right) \subseteq C^p\left(B^p\left(\Omega^G\right)\right) = C^p\left(\Omega^G\right)$ and so

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

Part (ii) The proof here is identical to that of part (ii) of Proposition 3.1 since the fact that signals were conclusive was not used in proving this. In particular, Lemma B.1 requires only that signals are binary. ■

4.2.2 Correlation increases pessimism

Theorem 1 showed that with conclusive signals, an increase in correlation (as measured by the PQD order) made the second-most pessimistic type even more pessimistic. The same is true with non-conclusive signals, that is, when both P and Q have full-support.

Lemmas A.3 and C.2 in the Appendix imply the following result.

Proposition 4.2 *Suppose signals are non-conclusive. If for $\theta = G, B$, $\mathbf{Y}^\theta \succ_{PQD} \mathbf{X}^\theta$, then for T large enough,*

$$q_{\mathbf{Y}} < q_{\mathbf{X}}.$$

The proof of Theorem 2 is completed by noting that as T increases, ρ_0 goes to zero. Now for large enough T, $\rho_0 < q_{\mathbf{Y}}$ and part (i) of Proposition 4.1 applies to $C^p_{\mathbf{X}}\left(\Omega^G\right)$ and part (ii) to $C^p_{\mathbf{Y}}\left(\Omega^G\right)$.

One may rightly wonder whether condition $\rho_0 < p \leq \rho_1$, required in Theorem 2, holds only when signals are "nearly" conclusive. This is not the case as the following example shows.

Example 2 Suppose that the set of signals $\mathcal{X} = \{0, 1\}$. There are two agents and the prior probability $\rho = \frac{3}{4}$.

Consider signals $\mathbf{Y}$ with the following joint distributions conditional on θ:

$Q^G =$

	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	0.12	0.08
$Y_1 = 1$	0.08	0.72

and $Q^B =$

	$Y_2 = 0$	$Y_2 = 1$
$Y_1 = 0$	0.84	0.075
$Y_1 = 1$	0.075	0.01

.

The two marginal distributions $\mu^G = (0.2, 0.8)$ and $\mu^B = (0.915, 0.085)$.

Let signals $\boldsymbol{X}$ be generated from P such that for each θ, $P^{\theta}(x_1, x_2) = \mu^{\theta}(x_1)\,\mu^{\theta}(x_2)$, that is, P^{θ} is the product of the marginal distributions in each state.

Note that $Q^{B}(0,0) = 0.84 < 1$ and so $(\theta, \boldsymbol{Y})$ is not conclusive (perhaps even "far" from conclusive). It is routine to verify that when $T = 2$, this example satisfies $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}} < \rho_1$ and so for $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$, $C^{p}_{\boldsymbol{Y}}\left(\Omega^{G}\right) = \varnothing$ while $C^{p}_{\boldsymbol{X}}\left(\Omega^{G}\right) = \Omega^{+}$.

5 Blackwell Informativeness

When there are only *two* agents ($I = 2$), 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 conclusive.

In the two-agent case, we first adopt the perspective of agent 1, say. As above, suppose P is a joint distribution over states of nature and signals and let P^{θ} be the joint distribution of signals 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 $P^{\theta}(X_1 \mid X_2)$ 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.[16]

Now consider another distribution Q of states of nature and signals and again let Q^{θ} be the joint distribution of signals conditional on θ. As above, for fixed θ, $Q^{\theta}(Y_1 \mid Y_2)$ is also a Blackwell experiment. When $I = 2$, we will say that

Definition 2 *The signals* $\mathbf{Y}$ *are* mutually more informative *than* $\mathbf{X}$ *if for all* θ *and* $j \neq i$, $Q^{\theta}(Y_j \mid Y_i)$ *is Blackwell more informative than* $P^{\theta}(X_j \mid X_i)$.

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.

[16] 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 the state of nature (X_2, X_3) be the same as the distribution of state of nature (Y_2, Y_3). Together with symmetry, this can hold only if the distribution of $\boldsymbol{Y}$ is the same as the distribution of $\boldsymbol{X}$.

Lemma 5.1 *Suppose that P and Q are both affiliated. If the signals $\boldsymbol{Y}$ are* mutually more informative *than $\boldsymbol{X}$, then*

$$\Pr\left[X_1 = 0, X_2 = 0 \mid \theta\right] \leq \Pr\left[Y_1 = 0, Y_2 = 0 \mid \theta\right]. \tag{10}$$

Proof. Fix θ. From Blackwell (1951), we know that if $Q^\theta\left(Y_1 \mid Y_2\right)$ is more informative than $P^\theta\left(X_1 \mid X_2\right)$, then the posteriors from $\boldsymbol{Y}$ are a mean-preserving spread of those from $\boldsymbol{X}$.

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

$$p_l^k = P^\theta\left(X_2 = l \mid X_1 = k\right),$$

and define

$$\boldsymbol{p}^k = \left(p_0^k, p_1^k\right) \in \Delta\left(\mathcal{X}\right)$$

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

$$\boldsymbol{q}^k \in \Delta\left(\mathcal{X}\right)$$

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,

$$\boldsymbol{p}^k \in \mathrm{co}\{\boldsymbol{q}^0, \boldsymbol{q}^1\},$$

the convex hull of the set of posterior vectors $\boldsymbol{q}^m$ from $\boldsymbol{Y}$.

Moreover, since (X_1, X_2) are affiliated, for any $k > 0$, the distribution $\boldsymbol{p}^1 \in \Delta\left(\mathcal{X}\right)$ (first-order) *stochastically dominates* the distribution $\boldsymbol{p}^0 \in \Delta\left(\mathcal{X}\right)$. Similarly, the distribution $\boldsymbol{q}^1 \in \Delta\left(\mathcal{X}\right)$ *stochastically dominates* $\boldsymbol{q}^0 \in \Delta\left(\mathcal{X}\right)$.

Since $\boldsymbol{p}^0 \in \mathrm{co}\{\boldsymbol{q}^0, \boldsymbol{q}^1\}$ we can write

$$\boldsymbol{p}^0 = \alpha_0 \boldsymbol{q}^0 + \left(1 - \alpha_0\right) \boldsymbol{q}^1,$$

where $\alpha_0 \in [0, 1]$.

Now note that

$$\begin{aligned} p_0^0 &= \alpha_0 q_0^0 + (1-\alpha_0)\, q_0^1 \\ &\leq q_0^0 \end{aligned}$$

because the distribution $\boldsymbol{q}^1$ stochastically dominates $\boldsymbol{q}^0$, that is, $q_0^1 \leq q_0^0$.

By definition, the inequality $p_0^0 \leq q_0^0$ is equivalent to

$$P^\theta (X_2 = 0 \mid X_1 = 0) \leq Q^\theta (Y_2 = 0 \mid Y_1 = 0)$$

and since $P^\theta (X_1 = 0) = Q^\theta (Y_1 = 0)$, the result follows. ∎

Lemma 5.1 implies that when there are *two* agents, in *all* of the results of the earlier sections, the condition "$\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$" can be replaced with "$\boldsymbol{Y}$ is mutually more informative than $\boldsymbol{X}$," provided that the inequality in (10) is strict. This is because Lemmas C.1 and C.2 only require (the strict version) of the inequality.

A Appendix: Affiliation and the PQD Order

Recall that the probability distribution $P \in \Delta\left(\mathcal{X}^I\right)$ 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 $\boldsymbol{x}^t = (x_i^t)_{i \in I}$ (slanted bold) is the I-vector of all I signal realizations in period t, while $\mathbf{x}_i = (x_i^t)_{t \in T}$ (upright bold) is the T-vector of i's signals over the T periods.

Lemma A.1 *Suppose that the I variables $\boldsymbol{X} = (X_1, X_2, ..., X_I)$ are affiliated with distribution P. If $\boldsymbol{X}^1, \boldsymbol{X}^2, ..., \boldsymbol{X}^T$ are independently and identically distributed according to P, then the $I \times T$ variables $(\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_I)$ also have an affiliated joint distribution.*

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 the $\boldsymbol{X}^t$'s are independently distributed over time,

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

Thus,

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

■

Lemma A.2 *Let* $\mathbf{e}^1 = (1, 0, ..., 0) \in \mathcal{X}^T$. *Suppose that the variables* $\mathbf{X}$ *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 $\mathcal{I}_{\Omega^+} : \left(\mathcal{X}^T\right)^I \to \{0,1\}$ 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,

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

The inequality in the third line is the result of the following argument. First, since the 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 Shanthikumar (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* $\mathbf{Y}^\theta \succ_{PQD} \mathbf{X}^\theta$. *Then*

$$\Pr[X_i = 0, X_j = 0 \mid \theta] < \Pr[Y_i = 0, Y_j = 0 \mid \theta].$$

Proof. Recall that $\boldsymbol{Y}^\theta \succ_{PQD} \boldsymbol{X}^\theta$ implies that for any $\boldsymbol{z}$ such that for at least two indices l, $z_l = 0$, then

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

If $\boldsymbol{z}$ such that $z_i = z_j = 0$ and $z_l = 1$, for all $l \neq i, j$, then the conclusion follows. ■

B Appendix: Posterior Monotonicity

Consider an agent with n signals $X_i = 1$. His belief that all other agents also received at least n signals $X_j = 1$ decreases with n.

Lemma B.1 *For any $n \geq 1$,*

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

Proof. Without loss of generality, suppose that the conditioning events are such that $\sum_{t=1}^{T-1} X_i^t = n$ 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. This is without loss of generality because the signals X_i^t are serially independent.

For $j = 1, 2, ..., I$, define $M_j = \sum_{t=1}^{T-1} X_j^t$ to be the sum of the first $T-1$ signals received by j and let $\boldsymbol{M}_{-i} = (M_j)_{j \neq i}$ denote the vector of sums of the first $T-1$ signals received by agents other than i. Then $N_j = M_j + X_j^T$.

We will argue that for all $\boldsymbol{m}_{-i}$,

$$\begin{aligned} & \Pr\left[\forall N_j \geq n+1, \boldsymbol{M}_{-i} = \boldsymbol{m}_{-i} \mid M_i = n, X_i^T = 1\right] \\ \leq\ & \Pr\left[\forall N_j \geq n, \boldsymbol{M}_{-i} = \boldsymbol{m}_{-i} \mid M_i = n, X_i^T = 0\right]. \end{aligned} \tag{11}$$

This is because if the left-hand side of (11) is positive, then it must be that after $T-1$ periods everyone has already received at least n positive signals, that is, for all j, $m_j \geq n$. But then the right-hand side of (11) is 1.

Thus, for all $\boldsymbol{m}_{-i}$, the probability that $N_j \geq n+1$ occurs conditional on $M_i = n$ and $X_i^T = 1$ is no greater than the probability that $N_j \geq n$ occurs conditional on $M_i = n$ and $X_i^T = 0$.

Finally, since, conditional on θ, the random variable $\boldsymbol{M}_{-i} = \sum_{t=1}^{T-1} \mathbf{X}_{-i}^{t}$ is independent of X_i^T, summing both sides of the inequality over all the $\boldsymbol{m}_{-i}$, we have

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

which establishes the result. ∎

C Appendix: Effect of Correlation

How does correlation affect the probability $\Pr[\Omega^+ \mid \mathbf{X}_1 = \mathbf{e}^1]$ that type $\mathbf{e}^1 = (1, 0, ..., 0) \in \mathcal{X}^T$ assigns to the event that all j get at least one $X_j = 1$?

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

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

If we define $A_j = \{\omega : \mathbf{x}_1 = \mathbf{e}^1, \mathbf{x}_j = \mathbf{0}\}$ as the set of states of the world 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 $P \in \Delta\left(\Theta \times \mathcal{X}^I\right)$ is the joint distribution of states of nature and signals.

By the inclusion-exclusion principle,

$$P\left(\cup_{j \neq 1} A_j\right) = \sum_{1<j} P\left(A_j\right) - \sum_{1<j<k} P\left(A_j \cap A_k\right) + \sum_{1<j<k<l} P\left(A_j \cap A_k \cap A_l\right) - ...$$

But since agents are symmetric, we have

$$\begin{aligned} P\left[\cup_{j \neq 1} A_j\right] &= \tbinom{I-1}{1} P\left(A_2\right) - \tbinom{I-1}{2} P\left(A_2 \cap A_3\right) + \tbinom{I-1}{3} P\left(A_2 \cap A_3 \cap A_4\right) - ... \\ &= \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} P\left(A_2 \cap A_3 \cap ... \cap A_l\right). \end{aligned} \tag{12}$$

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

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

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

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

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

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

and

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

So we can rewrite (12) more compactly as

$$P[\cup_{j\neq 1} A_j] = \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} \left(\rho \alpha_l^G \left(\beta_l^G\right)^{T-1} + (1-\rho)\, \alpha_l^B \left(\beta_l^B\right)^{T-1}\right). \tag{13}$$

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 Q, then we have

$$Q[\cup_{j\neq 1} A_j] = \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} \left(\rho \overline{\alpha}_l^G (\overline{\beta}_l^G)^{T-1} + (1-\rho)\, \overline{\alpha}_l^B (\overline{\beta}_l^B)^{T-1}\right), \tag{14}$$

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 Q of $\boldsymbol{Y}$. As above, both $\overline{\alpha}_l^\theta$ and $\overline{\beta}_l^\theta$ are non-increasing sequences.

Lemma C.1 *Suppose that both signals* $\mathbf{X}$ *and* $\mathbf{Y}$ *are conclusive. If*

$$\Pr\left[Y_i = 0, Y_j = 0 \mid G\right] > \Pr\left[X_i = 0, X_j = 0 \mid G\right], \tag{15}$$

then there exists a T_0 *such that for all* $T > T_0$*,*

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

Proof. First, since the signals $\boldsymbol{X}$ and $\boldsymbol{Y}$ are conclusive, then for all l,

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

and $\overline{\alpha}_l^B = 0$ as well. Then from (13) and (14) we have that the ratio

$$\frac{P\left(\cup_{j\neq 1} A_j\right)}{Q\left(\cup_{j\neq 1} A_j\right)} = \frac{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \alpha_l^G \left(\beta_l^G\right)^{T-1}}{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \overline{\alpha}_l^G (\overline{\beta}_l^G)^{T-1}}.$$

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

$$\frac{P\left(\cup_{j\neq 1} A_j\right)}{Q\left(\cup_{j\neq 1} A_j\right)} = \frac{(I-1)\,\alpha_2^G \left(\frac{\beta_2^G}{\overline{\beta}_2^G}\right)^{T-1} + \sum_{l=3}^{I} (-1)^l \binom{I-1}{l-1} \alpha_l^G \left(\frac{\beta_l^G}{\overline{\beta}_2^G}\right)^{T-1}}{(I-1)\,\overline{\alpha}_2^G + \sum_{l=3}^{I} (-1)^l \binom{I-1}{l-1} \overline{\alpha}_l^G (\frac{\overline{\beta}_l^G}{\overline{\beta}_2^G})^{T-1}}.$$

Now note that since $\overline{\beta}_l^G$ is a strictly decreasing sequence, each of the terms of the form $\left(\overline{\beta}_l^G / \overline{\beta}_2^G\right)$ is less than one. Moreover, (15) is the same as $\beta_2^G < \overline{\beta}_2^G$,

$$\frac{\beta_l^G}{\overline{\beta}_2^G} < \frac{\beta_2^G}{\overline{\beta}_2^G} < 1$$

and so we have that when T is large enough,

$$\frac{\Pr\left[\mathbf{X}_1 = \mathbf{e}^1, \exists j, \mathbf{X}_j = \mathbf{0}\right]}{\Pr\left[\mathbf{Y}_1 = \mathbf{e}^1, \exists j, \mathbf{Y}_j = \mathbf{0}\right]} = \frac{P\left(\cup_{j\neq 1} A_j\right)}{Q\left(\cup_{j\neq 1} A_j\right)} < 1. \tag{16}$$

Now since $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same univariate marginals, $\Pr[\mathbf{X}_1 = \mathbf{e}^1] = \Pr[\mathbf{Y}_1 = \mathbf{e}^1]$ and so from (16)

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

■

Lemma C.2 *Suppose* $\mathbf{X}$ *and* $\mathbf{Y}$ *have full-support distributions. If for* $\theta = G, B,$ *and* $i \neq j$,

$$\Pr[Y_i = 0, Y_j = 0 \mid \theta] > \Pr[X_i = 0, X_j = 0 \mid \theta], \tag{17}$$

then there exists a T_0 *such that for all* $T > T_0$,

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

Proof. From (13) and (14) we have that the ratio

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

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

$$\frac{P\left(\cup_{j\neq 1} A_j\right)}{Q\left(\cup_{j\neq 1} A_j\right)} = \frac{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho \alpha_l^G \left(\frac{\beta_l^G}{\overline{\beta}_2^B}\right)^{T-1} + (1-\rho)\, \alpha_l^B \left(\frac{\beta_l^B}{\overline{\beta}_2^B}\right)^{T-1}\right)}{\sum_{l=2}^{I} (-1)^l \binom{I-1}{l-1} \left(\rho \overline{\alpha}_l^G (\frac{\overline{\beta}_l^G}{\overline{\beta}_2^B})^{T-1} + (1-\rho)\, \overline{\alpha}_l^B (\frac{\overline{\beta}_l^B}{\overline{\beta}_2^B})^{T-1}\right)}. \tag{18}$$

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

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

Moreover, (17) implies that

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

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 (18), 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 $l > 2$

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

and so as $T \to \infty$, all the terms with $l > 2$ in the denominator of the right-hand side of (18) go to zero. The $l = 2$ term in the denominator, however, stays positive (it is at least $(1 - \rho)\,\overline{\alpha}_l^B > 0$).

So we have that when T is large enough,

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

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 (16),

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

■

References

[1] Anderson, Axel and Lones Smith (2024): "The Comparative Statics of Sorting," *American Economic Review*, 114, 709–751.

[2] Awaya, Yu and Vijay Krishna (2024): "Commonality of Information and Commonality of Beliefs: General Signals," SSRN Working Paper Series 4114760.

[3] Basak, Deepal, Joyee Deb and Aditya Kuvalekar (2024): "Similarity of Information and Collective Action," Working paper.

[4] 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.

[5] Cripps, Martin W., Jeffrey C. Ely, George J. Mailath and Larry Samuelson (2008): "Common Learning," *Econometrica*, 76, 909–933.

[6] Cripps, Martin W., Jeffrey C. Ely, George J. Mailath and Larry Samuelson (2013): "Common Learning with Intertemporal Dependence," *International Journal of Game Theory*, 42, 55–98.

[7] Frick, Mira, Ryota Iijima and Yuhta Ishii (2023): "Learning Efficiency of Multiagent Information Structures," *Journal of Political Economy*, 131, 3377–3414.

[8] Iachan, Felipe S., and Plamen T. Nenov (2015): "Information Quality and Crises in Regime-Change Games," *Journal of Economic Theory,* 158, 739–768.

[9] Joe, Harry (1990): "Multivariate Concordance," *Journal of Multivariate Analysis,* 35, 12–30.

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

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

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

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

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

[15] Shaked, Moshe and J. George Shanthikumar (2008): *Stochastic Orders*, Springer, 2008.

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

[17] Yanagimoto, Takemi and Masashi Okamoto (1969): "Partial Orderings of Permutations and Monotonicity of a Rank Correlation Statistic, *Annals of the Institute of Statistical Mathematics*, 21, 489–506.