# Commonality of Information and Commonality of Beliefs*

Yu Awaya[†] and Vijay Krishna[‡]

April 5, 2023

**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.

## 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.

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 come 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 (2023) have recently shown that when the number of signals each individual sees is large enough, the rate of common learning is not affected by the degree of correlation among agents' signals.

---

*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.

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

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

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 efficient 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 $G$ ("good") or $B$ ("bad"). 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.

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"). 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 even a single 1-signal tells a player 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 $X_i^2$. It is routine to verify that if $c \leq \frac{24}{25}$, then there is an equilibrium in which player $i$ invests if $X_i^1 + X_i^2 \geq 1$. 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. 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 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 $j \neq i$, $\Pr\left[Y_j^1 + Y_j^2 \geq 1 \mid Y_i^1 + Y_i^2 = 1\right] = \frac{47}{50} < c$ 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 the following.

a. If $c \leq \frac{47}{50}$, then with either conditionally independent signals $\boldsymbol{X}$ or correlated ones $Y_i$, 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 $\frac{47}{50} < c \leq \frac{48}{50}$, however, with conditionally independent signals $\boldsymbol{X}$, there is an equilibrium in which the players both players invest whenever they know $G$, while with correlated signals $\boldsymbol{Y}$, the unique equilibrium is that no player ever invests. Thus, for these costs, *correlated information hinders efficient coordination*!

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 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,

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

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

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

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

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 but not with the correlated signals.

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 rest 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 of nature $\theta$ 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 $\theta$. The signals are independent and identically distributed across time but may be correlated among agents. We are interested in how the degree of commonality of agents' beliefs—that is, how close they are to achieving common knowledge of $\theta$—is affected by the degree of commonality (correlation) of their information.

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

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

In what follows, "commonality of information" is formalized using a multivariate version of "more correlated," defined in the next section. "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.

**Binary and Conclusive Signals** We begin by considering a case where agents' signals are (i) *binary* (either "good news" or "bad news," as in the example above); and (ii) *conclusive,* in the sense that even one piece of good news reveals that the state is $G$ (again, as in the example). This special case is useful because first-order uncertainty—that is, concerning the state of nature $\theta$—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 whether or not $G$ can be common $p$-believed is completely determined by a single parameter $q$. This parameter is the belief about whether 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 $G$. To summarize, the event that $G$ is common $p$-believed exhibits a "bang-bang" property: if $p \leq q$, this event is as large as possible and if $p > q$, it is empty (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, second-most pessimistic type is the 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

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

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.

**General Signals** In Section 4 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 and the "bang-bang" property does not hold in general. In particular, when $p > q$, it may be that the event that $G$ is common $p$-believed is not empty. As a consequence, our main result in the general model (Theorem 2) reaches a weaker conclusion than Theorem 1 does in the binary and conclusive model. In two special cases, however, the "bang-bang" property re-emerges. These are (a) when the signals are *binary* but now non-conclusive; and (b) when the signals are *almost-public* in that the probability that different agents have different information is small.[2]

**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 $Q$ 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 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.[3] 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 (2023) study how common learning is affected by the underlying signal process. Consider joint distributions over states

[2]In both cases, some additional conditions are needed as well (see Section 4.3).

[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.

of nature and signals, $P$ and $Q$, such that $P$ is more informative about the state $\theta$ 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 $\theta$ but $Q$ is more correlated than $P$. We show that when $T$ is large enough, greater 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). 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.

Somewhat more distant is the 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 $\theta$. The increase in "commonality" of the sort mentioned above affects agents' beliefs about each other only via the change in their beliefs about $\theta$. In our paper, the increase in commonality increases the correlation among agents' signals while keeping their beliefs about the fundamental state $\theta$ fixed.

# 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 remaining periods. At each time $t \geq 1$, each agent $i$ receives a *private* signal that is informative about the state of nature $\theta$. The set of possible signals is a finite, ordered set $\mathcal{X} = \{0, 1, 2, ..., K\}$ . 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 $G$ is $\rho$. 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 $\theta$. 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 $\theta$.

2. Conditional on $\theta$, 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 $(\theta', \boldsymbol{x}')$,

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

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 $\theta$, 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 $\theta$, 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 $\theta$, 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 $\rho$. 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). It is useful to list some desirable properties a partial order "more correlated than" should satisfy.

First, since we are interested in isolating the effects of increased correlation, we will compare only distributions $P$ and $Q$ with *identical univariate marginals* conditional on $\theta$, that is, for all $k \in \mathcal{X}$ and $\theta \in \Theta$,

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

In other words, the conditional distributions $P^\theta(\cdot)$ and $Q^\theta(\cdot)$ have the same univariate marginals. This implies that agents' beliefs about the state of nature $\theta$ are the same with signals $\boldsymbol{X}$ as with $\boldsymbol{Y}$. Let $\mu^\theta \in \Delta(\mathcal{X})$ denote the common univariate marginal conditional on $\theta$.

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

Second, since signals have no inherent cardinal meaning—they only serve to update beliefs—the notion of more correlated should be preserved by monotone transformations of the variables. In other words, if the variables $(Y_1, Y_2, ..., Y_I)$ are more correlated 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 $\phi_i$ is an increasing function.[7]

Third, the notion should be 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 $\boldsymbol{X}_J = (X_i)_{i \in J}$.

In what follows, we will use the following notion[8]:

**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* $\boldsymbol{z} \in \mathcal{X}^I$,

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

*and*

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

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).

If $\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$, then their distributions have identical univariate marginals, that is, condition (2) is automatically satisfied. The PQD order also satisfies the two desired properties listed above—it is preserved by monotone transformations of the variables and is also preserved for subsets of variables. It is also the case that 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 or the weak associative order. Since our results will be of the form: "$\boldsymbol{Y} \succcurlyeq_{PQD} \boldsymbol{X}$, then ...", the last feature means that the result will remain true for all standard notions of greater positive interdependence.

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 (i)

[7] 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 $\text{Cov}(\phi_1(Y_1), \phi_2(Y_2)) < \text{Cov}(\phi_1(X_1), \phi_2(X_2))$.

[8] 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".

the inequality (3) is strict for any $\boldsymbol{z}$ such that for at least two indices $i$, $z_i < K$; and (ii) the inequality (4) is strict for any $\boldsymbol{z}$ such that for at least two indices $i$, $z_i > 0$.[9]

**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 $\theta$ 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 (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 $p < 1$,

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

[9] If (i) is not satisfied, then the equality of univariate marginals implies that (3) is an equality. Similarly, if (ii) is not satisfied, then (4) is an equality.

# 3 Binary and Conclusive Signals

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

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

Note that binary and conclusive 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, with binary signals, a type $\mathbf{x}_i$ can be represented simply as $n_i = \sum_t x_i^t$ and so types 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 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, $\Omega^+ \subseteq \Omega^G$. Define

$$q = \Pr\left[\Omega^+ \mid N_i = 1\right] \tag{5}$$

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 $\Omega^+$ and $q$ depend on $T$ although we have suppressed this dependence to reduce the notational burden.

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

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

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 Main 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 (5) and let $q_{\boldsymbol{Y}} = \Pr_{\boldsymbol{Y}}[\Omega^+ \mid \sum_t Y_i^t = 1]$ be the analogous belief derived from signals $\boldsymbol{Y}$.[10]

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

Define

$$\rho_0 = \Pr\left[\Omega^G \mid N_i = 0\right] \tag{7}$$

to be the belief about $G$ of an agent who receives only 0-signals in each of the $T$ periods. As $T$ increases, $\rho_0$ goes to zero. Note also that $\Omega$, $q_{\boldsymbol{X}}$, $q_{\boldsymbol{Y}}$, as well as $\rho_0$ all depend on $T$ although we have suppressed this dependence, again to avoid notational clutter.

The main result of this section is[11]:

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

*(1) 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* $\boldsymbol{Y}$ *whereas* $G$ *is common* $p$*-believed with* $\boldsymbol{X}$ *whenever everyone knows* $G$*.*

*(2) If* $\boldsymbol{Y}^G \succ_{PQD} \boldsymbol{X}^G$*, 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 $\boldsymbol{X}$. In this sense, greater commonality of information *reduces* the commonality of beliefs.

A few remarks on the theorem are in order.

First, it is easy to verify that if $q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$, then for all $p$, $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subseteq C^p_{\boldsymbol{X}}\left(\Omega^G\right)$. This does not mean that the probability of the former is less than that of the latter—they are measured using differernt distributions $Q$ and $P$, respectively. But it does imply that for any $p$, $C^p_{\boldsymbol{X}}\left(\Omega^G\right) = \varnothing$ and $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \neq \varnothing$ is impossible.

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 $\varepsilon$, 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 $\boldsymbol{Y}$.

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.

[11] $C^p_{\boldsymbol{X}}\left(\Omega^G\right)$ is the set of states of the world in which $\Omega^G$ 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.

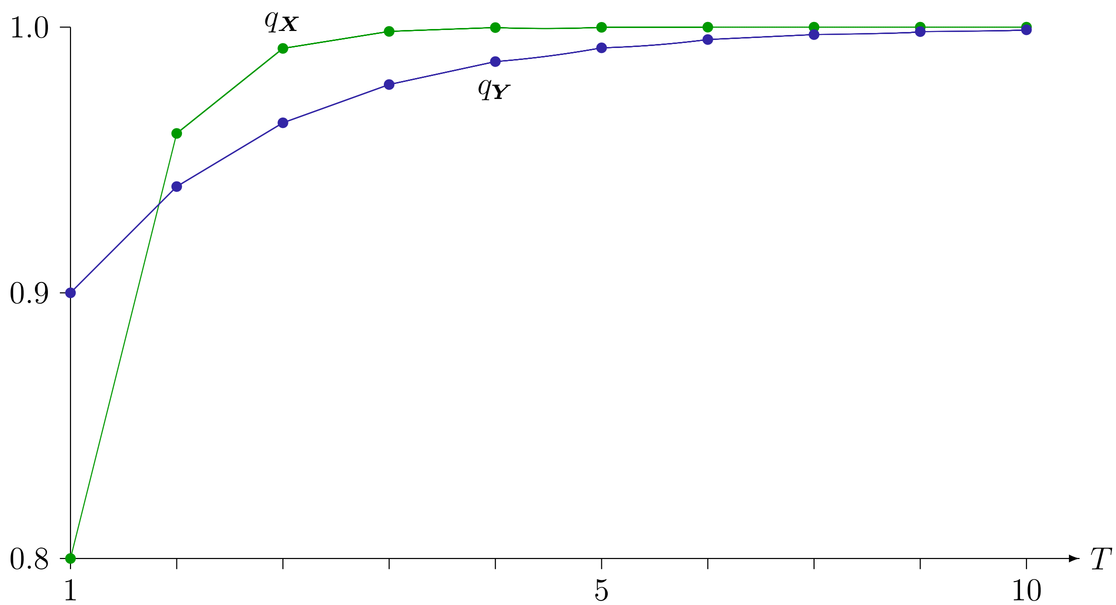


Figure 1: Threshold Beliefs for the Two Signals in Example 1

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 binary and 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 $p$, $C^p_{\boldsymbol{X}}\left(\Omega^G\right) \subseteq C^p_{\boldsymbol{Y}}\left(\Omega^G\right)$.

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 $L > R$, $T_0 = \lceil \ln L - \ln R \rceil$ 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 binary, 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 $C^p(\Omega^G)$ is as large as possible—any state of the world in which everyone knows that $\theta = G$ is included. But if $p > q$, $C^p(\Omega^G)$ is empty. Thus, $C^p(\Omega^G)$ 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 $C^p(\Omega^G)$ is derived in the following proposition.

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

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

$$C^p(\Omega^G) = \Omega^+$$

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

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

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 $C^p(\Omega^G)$. 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 the first effect is always stronger and so the probability of the event that all $n_j \geq n$ assigned by $n_i = n$ 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. What is crucial for this argument is that because signals are binary, types can be linearly ordered by the number of positive signals.

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

$$B^q\left(\Omega^G\right) = \{\omega : \forall j, n_j \ge 1\} = \Omega^+ \tag{8}$$

Moreover, (6) implies that all types with $n_i \ge 1$ assign at least probability $q$ to the event $\Omega^+$ that everyone got at least one positive signal. Formally, $\{\omega : n_i \ge 1\} \subseteq B_i^q\left(\Omega^+\right)$ and since $\Omega^+ = \{\omega : \forall j, n_j \ge 1\} \subset \{\omega : n_i \ge 1\}$, we have

$$\Omega^+ \subseteq B^q\left(\Omega^+\right) \tag{9}$$

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

Now (8) 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 (9), $\Omega^+ \subseteq B^{q,\ell}\left(\Omega^G\right)$.

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

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

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

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

as the set of states of the world $\omega$ 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 $n \le T+1$,

$$C^p\left(\Omega^G\right) \subseteq \Gamma^{(n)} \tag{10}$$

First, since $\Gamma^{(0)} = \{\omega : \forall j, n_j \ge 0\} = \Omega$, (10) 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 $\omega'$, there is an $i$ with $n_i = n$, that is, $i$ gets exactly $n$ positive signals and since $C^p\left(\Omega^G\right) \subseteq \Gamma^{(n)}$,

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

Lemma B.1 now implies that

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

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 (10).

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 binary and conclusive signals, the maximum commonality of beliefs—that is, the highest $p$ for which $\Omega^G$ can be common $p$-believed—is exactly $q$, the belief of the second-most pessimistic. In this section, we compare two signal distributions such that $\boldsymbol{Y}^G \succ_{PQD} \boldsymbol{X}^G$.[12] We show that a change from $\boldsymbol{X}^G$ to $\boldsymbol{Y}^G$ increases the pessimism of type $n_i = 1$.

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

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

**Proof.** Follows from Lemma A.3 and Lemma C.1 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, 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 $\boldsymbol{X}$,

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

at the same time

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

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

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

# 4 General Model

The sharp result in Theorem 1 was derived for the case of binary and 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 $\theta$—also plays a role. Moreover, if signals are not binary, the types cannot be linearly ordered. For instance, with three signals and $T = 2$, it is unclear whether type $(1, 1)$ is more or less optimistic than type $(2, 0)$.

We now 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 distribution $P$ has *full support*. Recall that since $P$ is affiliated, higher signals are more indicative that $\theta = G$; that is, $\Pr[G \mid X_i = k]$ is increasing in $k$.

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 and

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

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 at least one non-zero signal. Note that because of affiliation, type $\mathbf{e}^1$ is the *second-most* pessimistic type about both $\Omega^G$ and $\Omega^+$. Only type $\mathbf{0}$ is more pessimistic.

Once again we will compare signals $\boldsymbol{X}$ coming from $P$ with $\boldsymbol{Y}$ coming from $Q$ such that $\boldsymbol{Y}^\theta \succ_{PQD} \boldsymbol{X}^\theta$ for $\theta = G, B$. Let $q_{\boldsymbol{Y}}$ be defined in a manner analogous to $q_{\boldsymbol{X}}$.

As in (7), 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 $G$. Both $\rho_0$ and $\rho_1$ are the same for $\boldsymbol{X}$ and $\boldsymbol{Y}$ since $\boldsymbol{Y}^\theta \succ_{PQD} \boldsymbol{X}^\theta$ and so the univariate marginal distributions $\mu^\theta$ of $\boldsymbol{X}^\theta$ and $\boldsymbol{Y}^\theta$ are the same (see (2)). Moreover, the prior probability $\rho$ of $G$ is the same. Note that if a 1-signal is conclusive, as in last section, then $\rho_1 = 1$.

## 4.1 Main result

In the general model, we have the following result:

**Theorem 2**

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

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

*that is, if $G$ is common $p$-believed with signals $\boldsymbol{Y}$, it is also common $p$-believed with $\boldsymbol{X}$. Moreover, there are states of the world in which $G$ is common $p$-believed with $\boldsymbol{X}$ but not with $\boldsymbol{Y}$.*

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

Like Theorem 1, Theorem 2 says that, under the identified circumstances, greater commonality of information *reduces* the commonality of beliefs. Before proving the theorem, it is useful to compare its conclusions to those of Theorem 1.

First, in part (1) of Theorem 2, while $C_{\boldsymbol{Y}}^{p}\left(\Omega^{G}\right)$ is a strict subset of $C_{\boldsymbol{X}}^{p}\left(\Omega^{G}\right)$, the former may be non-empty and there is no guarantee that $\Pr_{\boldsymbol{Y}}\left[C_{\boldsymbol{Y}}^{p}\left(\Omega^{G}\right)\right] < \Pr_{\boldsymbol{X}}\left[C_{\boldsymbol{X}}^{p}\left(\Omega^{G}\right)\right]$. This is because the two ex ante probabilities are determined from different distributions $Q$ and $P$, respectively. In Section 4.3, however, we provide two environments in which under some weak conditions, the stronger conclusion that $C_{\boldsymbol{Y}}^{p}\left(\Omega^{G}\right) = \varnothing$ while $C_{\boldsymbol{X}}^{p}\left(\Omega^{G}\right) = \Omega^{+}$ obtains. Now, of course, the probability of the former is zero. These are

1. *binary* and non-conclusive signals (Section 4.3.1)

2. *almost-public* signals—that is, when the probability of events in which the signals received by the agents differ is small (Section 4.3.2).

Second, in the general model 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 binary and conclusive model of Section 3 where $\rho_1 = 1$. What if $\rho_1 \leq q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$? It can then be argued that for $p \in (\rho_0, q_{\boldsymbol{Y}})$,

$$C_{\boldsymbol{Y}}^{p}\left(\Omega^{G}\right) = C_{\boldsymbol{X}}^{p}\left(\Omega^{G}\right) = \Omega^{+}$$

and in this case correlation has no effect on $C^{p}\left(\Omega^{G}\right)$.

Finally, recall that with binary and conclusive signals, when $T = 1$, it was the case that $q_{\boldsymbol{X}} < q_{\boldsymbol{Y}}$. Where there are more than two signals, it may be that even for $T = 1$, $q_{\boldsymbol{Y}} < q_{\boldsymbol{X}}$ so that the conclusion of Theorem 2 may hold even for $T = 1$.[13]

[13] An example is available from the authors.

## 4.2 Proof of Theorem 2

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

We first prove, for general signals, an analog of Proposition 3.1. With general signals, however, the conclusion reached is weaker. Of course, $C^p\left(\Omega^G\right)$ gets smaller as $p$ increases and the proposition shows that it is *strictly* smaller as $p$ crosses the threshold $q$. But unlike in the case of binary and conclusive signals, $C^p\left(\Omega^G\right)$ may be non-empty even when $p$ exceeds $q$. In other words, the bang-bang property does not hold in general.

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

**Proposition 4.1** *For any $T$,*

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

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

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

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

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

$$C^p\left(\Omega^G\right) \subseteq \Omega^+ \tag{11}$$

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

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

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 (11), 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 $\mathbf{x}_i \neq \mathbf{0}$, $\Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right] \leq \Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{x}_i\right]$ 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 $\Omega^+$ is *evident p-belief* (or is *p-evident,* for short). Proposition 3 in Monderer and Samet (1989) now implies that $\Omega^+$ 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)** Next we argue that if $p > q$, the inclusion in (11) 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 $\omega' \notin C^p\left(\Omega^G\right)$. ■

### 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. Here we show that, modulo some minor qualifications, the same is true in general—that is, even when signals are not conclusive.

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

**Proposition 4.2** *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, $\rho_0$ goes to zero. Now for large enough $T$, $\rho_0 < q_{\mathbf{Y}}$ and part (i) of Proposition 4.1 applies to $C_{\mathbf{X}}^p\left(\Omega^G\right)$ and part (ii) to $C_{\mathbf{Y}}^p\left(\Omega^G\right)$.

**Example 2** We now consider an example in which signals are neither binary nor conclusive. In the example, all the assumptions in part (1) of Theorem 2 hold when $T = 2$. Specifically, $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}} < \rho_1$. The example illustrates the following. First, part (1) cannot be strengthened to read that for $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$, $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) = \varnothing$. Second, while Theorem 2 only says that $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subsetneq C^p_{\boldsymbol{X}}\left(\Omega^G\right)$, it is the case that the ex ante probability of the event $C^p_{\boldsymbol{Y}}\left(\Omega^G\right)$ according to $Q$ is smaller than the ex ante probability of the event $C^p_{\boldsymbol{X}}\left(\Omega^G\right)$ according to $P$.

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

Consider signals $\boldsymbol{Y}$ with the following joint distributions conditional on $\theta$, where $\varepsilon > 0$ is a small number.

$Q^G =$

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

$Q^B =$

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

When $\varepsilon$ is small, the distribution $Q$ is affiliated. In fact, all the (non-trivial) affiliation inequalities are strict.

Now consider signals $\boldsymbol{X}$ where for each $\theta$, 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 $\theta$, the signals $\boldsymbol{X}$ are independently distributed.

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

$$\begin{aligned}
\rho_0 &= \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{0}\right] \approx 0.031 \\
q_{\boldsymbol{Y}} &= \Pr\left[\Omega^+ \mid \mathbf{Y}_i = \mathbf{e}^1\right] \approx 0.936 \\
q_{\boldsymbol{X}} &= \Pr\left[\Omega^+ \mid \mathbf{X}_i = \mathbf{e}^1\right] \approx 0.954 \\
\rho_1 &= \Pr\left[\Omega^G \mid \mathbf{X}_i = \mathbf{e}^1\right] \approx 0.986
\end{aligned}$$

and since $\rho_0 < q_{\boldsymbol{Y}} < q_{\boldsymbol{X}} < \rho_1$, the conditions of Theorem 2 part (1) are met.

If $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$, then from Theorem 2,

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

and it may be verified that

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

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

Moreover, it can be verified that for $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$, when $\varepsilon \approx 0$,

$$\Pr_{\boldsymbol{Y}}\left[C^p_{\boldsymbol{Y}}\left(\Omega^G\right)\right] \approx 0.039 \text{ and } \Pr_{\boldsymbol{X}}\left[C^p_{\boldsymbol{X}}\left(\Omega^G\right)\right] \approx 0.468$$

Thus, in the example, for $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$ we not only have $\varnothing \neq C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subsetneq C^p_{\boldsymbol{X}}\left(\Omega^G\right)$ (part (1) of Theorem 2) 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. Thus, for $p \in (q_{\boldsymbol{Y}}, q_{\boldsymbol{X}})$ it is more likely that $\Omega^G$ will be common $p$-believed with signals $\boldsymbol{X}$ than with $\boldsymbol{Y}$. This last feature is not a consequence of Theorem 2.

## 4.3 Bang-bang property: two sufficient conditions

While Theorem 2 part (1) gives sufficient conditions for $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) \subsetneq C^p_{\boldsymbol{X}}\left(\Omega^G\right)$, this of course does not imply that $\Pr_{\boldsymbol{Y}}\left[C^p_{\boldsymbol{Y}}\left(\Omega^G\right)\right] < \Pr_{\boldsymbol{X}}\left[C^p_{\boldsymbol{X}}\left(\Omega^G\right)\right]$. Here we provide two sets of sufficient conditions such that in Theorem 2, the *stronger* conclusion that $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) = \varnothing$ while $\Omega^+ \subseteq C^p_{\boldsymbol{X}}\left(\Omega^G\right)$ obtains (as in Theorem 1).

One sufficient condition concerns the case where signals are *binary* (but perhaps not conclusive). The other sufficient condition concerns *almost public* signals—that is, the probability that different agents get different signals is small.

### 4.3.1 Binary signals

The following result generalizes the "bang-bang" result of Proposition 3.1 in that it allows for *non-conclusive* binary signals. Now while types can still be linearly ordered, there is first-order uncertainty—about the state of nature $\theta$. We show how the conclusion of Proposition 3.1 can be reached in this more permissive setting as well.

**Proposition 4.3** *Suppose signals are binary. For any $T$,*

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

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

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

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

Part (i) of Proposition 4.3 needs no additional proof since it follows from part (i) of 4.1. The proof of part (ii) of Proposition 4.3 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.

Proposition 4.3 implies that for binary signals, under the assumptions of Theorem 2, its conclusion in part (1) can be sharpened to read that $C^{p}_{\boldsymbol{Y}}\left(\Omega^{G}\right)=\varnothing$.

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

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

Consider signals $\boldsymbol{Y}$ with the following joint distributions conditional on $\theta$,

$Q^G =$

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

$Q^B =$

| | $Y_2=0$ | $Y_2=1$ |
|---|---|---|
| $Y_1=0$ | 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 $\theta$, $P^{\theta}(x_1,x_2)=\mu^{\theta}(x_1)\,\mu^{\theta}(x_2)$, that is, $P^{\theta}$ 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 $C^{p}_{\boldsymbol{X}}\left(\Omega^{G}\right)=\Omega^{+}$.

### 4.3.2 Almost-public signals

Here we consider signals $\boldsymbol{Y}$ that are so highly correlated that they are almost public. In other words, the probability that agents' signals disagree in any period is very small. We will show that under a weak condition, the conclusion of Theorem 2 can again be strengthened.

For $\varepsilon>0$, an $\varepsilon$-public signal distribution is constructed as follows. For any vector of signals $\boldsymbol{y}\in\mathcal{X}^{I}$, define

$$d(\boldsymbol{y})=\min_{k}\textstyle\sum_i |y_i-k|$$

as the distance $\boldsymbol{y}$ from to the nearest *diagonal* vector $(k,k,..,k)$. Note that $d(\boldsymbol{y})>0$ if and only if $\boldsymbol{y}$ is not diagonal.

**Definition 2** *$Q$ is* $\varepsilon$-public *if for all non-diagonal* $\boldsymbol{y}$,

$$Q^{\theta}(\boldsymbol{y})=\varepsilon^{d(\boldsymbol{y})}$$

The definition says that the probability of any vector that is not diagonal is small and decreases, by an order of magnitude, the further a signal vector is from being diagonal.

An $\varepsilon$-public distribution $Q$ can be interpreted as follows. A public signal $y=(k,k,...,k)$ is generated. Each agent, however, "misunderstands" the signal $k$ with error. Specifically, an agent thinks the signal is $k+1$ or $k-1$ each with probability $\varepsilon$; thinks that the signal is $k+2$ or $k-2$, each with probability $\varepsilon^2$; etc. Thus larger misunderstandings occur with smaller probabilities. Misunderstandings occur independently across agents.[14]

Let $\mu^{\theta} \in \Delta(\mathcal{X})$ denote the univariate marginal distribution of agents' signals in state of nature $\theta$ derived from $Q$. We will assume the following

**Condition 1** *For all $k > 1$, and $\theta, \theta' \in \{G, B\}$,*

$$\mu^{\theta}(1) > \mu^{\theta'}(k)$$

The condition is rather weak and requires only that the marginal probability of signal $k = 1$ in either state of nature $G$ or $B$ is not too small relative to those of signals $k > 1$ in *both* states, $G$ and $B$.

Fix a full-support univariate marginal $\mu \in \Delta(\mathcal{X})$ and let $\mathcal{Q}(\mu) \subset \Delta(\mathcal{X}^I)$ be the set of (symmetric) full-support joint distributions with univariate marginals equal to $\mu$.

For $\varepsilon$-public signals satisfying Condition 1, for all $\varepsilon$ small enough, the "bang-bang" property holds:

**Proposition 4.4** *Let $Q \in \mathcal{Q}(\mu)$ be an $\varepsilon$-public distribution and suppose that the univariate marginals $\mu^{\theta}$ satisfy Condition 1. For any $T$,*
*(i)if $\rho_0 < p \leq q \leq \rho_1$ then*

$$C^p(\Omega^G) = \Omega^+$$

*(ii)if $\rho_0 < q < p$, then for all $\varepsilon$ small enough,*

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

**Proof. Part (i)** follows from part (i) of Proposition 4.1 since this is just a special case.

**Part (ii)** Let $n$ be an integer satisfying $0 \leq n \leq KT + 1$ and define

$$S^{(n)} = \left\{ \mathbf{y}_i : \sum_{t=1}^{T} y_i^t = n \right\}$$

as the set of types $\mathbf{y}_i$ such that the sum of signals is $n$. Finally, define

$$\Gamma^{(n)} = \cup_{m \geq n} S^{(m)}$$

as the set of types $\mathbf{y}_i$ such that the signal-sum is at least $n$. Note that by definition, $\Gamma^{(1)} = \Omega^+$.

[14] This exact specification is not needed in what follows. All that is needed is that all non-diagonal $\boldsymbol{y}$ with $d(\boldsymbol{y}) = 1$ are of the same magnitude and that other non-diagonal $\boldsymbol{y}$'s are of lower magnitude. This last feature is in fact a necessary consequence of affiliation.

Lemma D.1 in Appendix D shows that for small enough $\varepsilon$, for any $\mathbf{y}_i \in S^{(n)}$,

$$\Pr\left[\Gamma^{(n)} \mid \mathbf{Y}_i = \mathbf{y}_i\right] \leq q$$

that is, a type with signal-sum $n$ is more pessimistic about $\Gamma^{(n)}$ than type with signal-sum 1 is about $\Gamma^{(1)}$.[15]

The remainder of the proof is identical to that of part (ii) of Proposition 3.1. ■

Proposition 4.4 implies that for $\varepsilon$-public signals, under the assumptions of Theorem 2, its conclusion in part (1) can again be sharpened to read that $C^p_{\mathbf{Y}}\left(\Omega^G\right) = \varnothing$ while $C^p_{\mathbf{X}}\left(\Omega^G\right) = \Omega^+$.

**Example 4** This example illustrates the workings of Theorem 2 when the signal distribution $Q$ is $\varepsilon$-public and satisfies Condition 1 and $P$ is (conditionally) independent. Signals are neither binary nor conclusive but the stronger conclusion that $C^p_{\mathbf{Y}}\left(\Omega^G\right) = \varnothing$ still obtains.

Suppose that the set of signals $\mathcal{X} = \{0, 1, 2\}$. There are two agents and the prior probability $\rho = 0.85$. Suppose also that $T = 2$.

Let the signals $\mathbf{Y}$ be distributed according to an $\varepsilon$-public distribution $Q$ with conditional distributions

$Q^G =$

| | $Y_2 = 0$ | $Y_2 = 1$ | $Y_2 = 2$ |
|---|---|---|---|
| $Y_1 = 0$ | $0.4 - \varepsilon - \varepsilon^2$ | $\varepsilon$ | $\varepsilon^2$ |
| $Y_1 = 1$ | $\varepsilon$ | $0.315 - 2\varepsilon$ | $\varepsilon$ |
| $Y_1 = 2$ | $\varepsilon^2$ | $\varepsilon$ | $0.285 - \varepsilon - \varepsilon^2$ |

$Q^B =$

| | $Y_2 = 0$ | $Y_2 = 1$ | $Y_2 = 2$ |
|---|---|---|---|
| $Y_1 = 0$ | $0.5 - \varepsilon - \varepsilon^2$ | $\varepsilon$ | $\varepsilon^2$ |
| $Y_1 = 1$ | $\varepsilon$ | $0.3 - 2\varepsilon$ | $\varepsilon$ |
| $Y_1 = 2$ | $\varepsilon^2$ | $\varepsilon$ | $0.2 - \varepsilon - \varepsilon^2$ |

so that the two marginal distributions are $\mu^G = (0.4, 0.315, 0.285)$ and $\mu^B = (0.5, 0.3, 0.2)$.

Let $\boldsymbol{X}$ be such that the conditional distributions are independent with the given marginals, that is, $P^\theta(x_1, x_2) = \mu^\theta(x_1) \times \mu^\theta(x_2)$.

Routine calculations show that for this example, when $\varepsilon = 0.07$,

$$\begin{aligned} \rho_0 &\simeq 0.784 \\ q_{\mathbf{X}} &\simeq 0.807 \\ q_{\mathbf{Y}} &\simeq 0.824 \\ \rho_1 &\simeq 0.826 \end{aligned}$$

[15] Note that unlike in the case of binary signals, there is no claim that different types with the same sum of signals have the same beliefs. For instance, if $T = 2$, type $\mathbf{x}_i = (1, 1)$ may have different beliefs than type $\mathbf{x}'_i = (2, 0)$ even though both have the same sum.

and since $\rho_0 < q_{\boldsymbol{X}} < q_{\boldsymbol{Y}} < \rho_1$, the conditions of Theorem 2 part (1) are met .

It may be verified that in this example, we have that for all $p \in (q_{\boldsymbol{X}}, q_{\boldsymbol{Y}})$, $C^p_{\boldsymbol{Y}}\left(\Omega^G\right) = \varnothing$ 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 binary nor need they 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 $P^\theta$ be the joint distribution of signals conditional on $\theta$. For fixed $\theta$, 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 $X_1$.[16]

Now consider another distribution $Q$ of states of nature and signals and again let $Q^\theta$ be the joint distribution of signals conditional on $\theta$. As above, for fixed $\theta$, $Q^\theta\left(Y_1 \mid Y_2\right)$ is also a Blackwell experiment.

We will say that

**Definition 3** *Suppose $I = 2$. The signals $\boldsymbol{Y}$ are* mutually more informative *than $\boldsymbol{X}$ if for all $\theta$, $Q^\theta\left(Y_j \mid Y_i\right)$ is Blackwell more informative than $P^\theta\left(X_j \mid X_i\right)$.*

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 $\theta$, $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same univariate marginal distributions.

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

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

**Proof.** Fix $\theta$. 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 the posteriors from $\boldsymbol{X}$.

[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 $(X_2, X_3)$ 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 $\boldsymbol{X}$.

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

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

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 $k$,

$$\boldsymbol{p}^k \in \text{co}\{ \boldsymbol{q}^m : m \in \mathcal{X} \}$$

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 $\boldsymbol{p}^k \in \Delta(\mathcal{X})$ (first-order) *stochastically dominates* the distribution $\boldsymbol{p}^0 \in \Delta(\mathcal{X})$. Similarly, for any $k > 0$, the distribution $\boldsymbol{q}^k \in \Delta(\mathcal{X})$ *stochastically dominates* the distribution $\boldsymbol{q}^0 \in \Delta(\mathcal{X})$.

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

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

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 $L \in \mathcal{X}$,

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

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

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

$$p_0^0 \leq q_0^0$$

which is equivalent to

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

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 that "$\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$" can be replaced with the condition "$\boldsymbol{Y}$ is mutually more informative than $\boldsymbol{X}$," provided that the inequality in (12) is strict. This is because Lemmas C.1 and C.2 only require (the strict version) of the inequality.

# 6 Common $p$-beliefs about $B$

Theorems 1 and 2 concern the effects of increased correlation on the common learning of the state of nature $G$. In many applications (such as the investment game in Example 1, games of regime change, currency attacks, etc.) one is interested in discovering the prospects of coordinated actions in one of the states and not the other. Nevertheless, one may well ask what effect increased correlation has on the common learning of $B$ and we now turn to this question.

In the general model of Section 4, there is nothing special about the state of nature $G$ versus $B$. In particular, all signals have full support in both states of nature. Thus all the results of Section 4 that concern the beliefs about $G$ have counterparts that concern the beliefs about $B$. The counterpart of $\Omega^+ = \{\omega : \forall j, \mathbf{X}_j \neq \mathbf{0}\}$ is $\Omega^- = \{\omega : \forall j, \mathbf{X}_j \neq (K, K, ..., K)\}$ since type $(K, K, ..., K)$ is most optimistic about $G$ and hence most pessimistic about $B$. The counterpart of $q = \Pr[\Omega^+ \mid \mathbf{X}_i = (1, 0, ..., 0)]$ is $r = \Pr[\Omega^- \mid \mathbf{X}_i = (K-1, K, ..., K)]$. Just as type $(1, 0, ..., 0)$ is the second-most pessimistic about $G$, type $(K-1, K, ..., K)$ is the second-most pessimistic about $B$. Finally, analogous to $\rho_0$ and $\rho_1$, let $\sigma_K = \Pr\left[\Omega^B \mid \mathbf{X}_i = (K, K, ..., K)\right]$ and $\sigma_{K-1} = \Pr\left[\Omega^B \mid \mathbf{X}_i = (K-1, K, ..., K)\right]$. These corresponding objects can then be used to derive results about $\theta = B$ that are qualitatively the same as those derived in Section 4 about $\theta = G$. Mimicking the arguments in Theorem 2, we would deduce (1) that if $\sigma_K < r_{\boldsymbol{Y}} < r_{\boldsymbol{X}} < \sigma_{K-1}$, then $C^p_{\boldsymbol{Y}}\left(\Omega^B\right) \subsetneq C^p_{\boldsymbol{X}}\left(\Omega^B\right)$; and (2) for $T$ large enough, $\sigma_K < r_{\boldsymbol{Y}} < r_{\boldsymbol{X}}$.

With *binary and conclusive* signals, however, there is an important asymmetry between the two states of nature $G$ and $B$. The signals do not have full support in state $B$ and even one 1-signal reveals that the state is $G$. So we now turn to consider this special case.

Analogous to the definition of $\Omega^G$, let

$$\Omega^B = \{\omega : \theta = B\}$$

as the set of states of the world in which the state of nature is $B$ and define

$$\Omega^0 = \{\omega : \forall j, \mathbf{x}_j = \mathbf{0}\}$$

as the set of states of the world in which *every* agent gets only 0-signals in every period.

We then have the following "bang-bang" result for agents' beliefs about $\Omega^B$.[17]

**Proposition 6.1** *Suppose signals are binary and conclusive.*

*(i) If $p \leq 1 - \rho_0$, then*

$$C^p\left(\Omega^B\right) = \Omega^0$$

*(ii) If $1 - \rho_0 < p$, then*

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

**Proof.** (i) Since type $\mathbf{0}$ assigns probability $1 - \rho_0$ to $B$, $\{\omega : \mathbf{x}_i = \mathbf{0}\} \subseteq B_i^p\left(\Omega^B\right)$ and hence $\Omega^0 \subseteq B^p\left(\Omega^B\right)$.

Also, since signals are conclusive, $\Omega^B \subset \Omega^0$, and so $1 - \rho_0 = \Pr\left[\Omega^B \mid \mathbf{X}_i = \mathbf{0}\right] < \Pr\left[\Omega^0 \mid \mathbf{X}_i = \mathbf{0}\right]$. Thus, $p < \Pr\left[\Omega^0 \mid \mathbf{X}_i = \mathbf{0}\right]$ and so $\{\omega : \mathbf{x}_i = \mathbf{0}\} \subseteq B_i^p\left(\Omega^0\right)$ and hence $\Omega^0 \subseteq B^p\left(\Omega^0\right)$. Thus, $\Omega^0 \subseteq B^p\left(B^p\left(\Omega^B\right)\right)$. Proceeding in this way, we obtain that $\Omega^0 \subseteq C^p\left(\Omega^B\right)$. Moreover, any type $\mathbf{x}_i \neq \mathbf{0}$ assigns probability 0 to the event $\Omega^B$ (since even a single 1-signal indicates that $\theta = G$). Thus, $\{\omega : \mathbf{x}_i \neq \mathbf{0}\} \cap B_i^p\left(\Omega^B\right) = \varnothing$ and this implies that $\{\omega : \mathbf{x}_i \neq \mathbf{0}\} \cap C^p\left(\Omega^B\right) = \varnothing$ as well. Thus, $C^p\left(\Omega^B\right) = \Omega^0$.

(ii) Since type $\mathbf{0}$ assigns a probability less than $p$ to $\Omega^B$, all types $\mathbf{x}_i$ do so as well. Thus, $B^p\left(\Omega^B\right) = \varnothing$ and so $C^p\left(\Omega^B\right) = \varnothing$. ■

Now consider two binary, signals such that $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$. Since they have the same univariate marginals, it is the case that $\Pr\left[\Omega^B \mid \mathbf{X}_i = \mathbf{0}\right] = \Pr\left[\Omega^B \mid \mathbf{Y}_i = \mathbf{0}\right] = 1-\rho_0$.

Proposition 6.1 now implies that with binary and conclusive signals, if $p \leq 1 - \rho_0$, then $C_{\boldsymbol{X}}^p\left(\Omega^B\right) = C_{\boldsymbol{Y}}^p\left(\Omega^B\right) = \Omega^0$ and if $p > 1 - \rho_0$, then $C_{\boldsymbol{X}}^p\left(\Omega^B\right) = C_{\boldsymbol{Y}}^p\left(\Omega^B\right) = \varnothing$. In other words, increased correlation does not affect the set $C^p\left(\Omega^B\right)$.

Finally, $\boldsymbol{Y} \succ_{PQD} \boldsymbol{X}$ implies that $\Pr_{\boldsymbol{X}}\left[\Omega^0\right] < \Pr_{\boldsymbol{Y}}\left[\Omega^0\right]$, that is, increased correlation *does not decrease* the likelihood that $\Omega^B$ is common $p$-believed.

# A Appendix: Affiliation and the PQD Order

Recall that a joint 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.*

[17] This result was suggested by a referee.

**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* $\boldsymbol{X}$ *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 $\Omega^+$ 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* $\mathbf{Y}^\theta \succ_{PQD} \boldsymbol{X}^\theta$. *Then*

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

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

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

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

# B Appendix: Binary Signals

**Lemma B.1** *Suppose signals are binary. 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.** Since signals are serially independent, without loss of generality, suppose that the conditioning events are such that $\sum_{t=1}^{T-1} X_i^t = n$ 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 $T$.

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 $\boldsymbol{m}_{-i}$,

$$\begin{aligned} & \Pr\left[\forall N_j \geq n+1, \boldsymbol{M}_{-i} = \boldsymbol{m}_{-i} \mid M_i = n, X_i^T = 1\right] \\ \leq \; & \Pr\left[\forall N_j \geq n, \boldsymbol{M}_{-i} = \boldsymbol{m}_{-i} \mid M_i = n, X_i^T = 0\right] \end{aligned} \tag{13}$$

This is because if the left-hand side of (13) 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 (13) 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 the probability distribution of $\boldsymbol{M}_{-i} = \sum_{t=1}^{T-1} \boldsymbol{X}_{-i}^t$ 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

We are interested in how correlation affects the probability $\Pr\left[\Omega^+ \mid \mathbf{X}_1 = \mathbf{e}^1\right]$ that type $\mathbf{e}^1 = (1, 0, ..., 0) \in \mathcal{X}^T$ assigns to the event that all others get at least one positive signal.

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

$$\begin{aligned} P\left[\cup_{j\neq 1} A_j\right] &= \tbinom{I-1}{1} P\left(A_2\right) - \tbinom{I-1}{2} P\left(A_2 \cap A_3\right) + \tbinom{I-1}{3} P\left(A_2 \cap A_3 \cap A_4\right) - ... \\ &= \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} P\left(A_2 \cap A_3 \cap ... \cap A_l\right) \end{aligned} \quad (14)$$

Now, since conditional on $\theta$, the signals are independent over time

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

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

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

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

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

and

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

and so we can rewrite (14) more compactly as

$$P\left[\cup_{j\neq 1} A_j\right] = \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} \left(\rho \alpha_l^G \left(\beta_l^G\right)^{T-1} + (1-\rho)\, \alpha_l^B \left(\beta_l^B\right)^{T-1}\right) \quad (15)$$

Note that for $\theta = G, B$, both $\alpha_l^\theta$ and $\beta_l^\theta$ 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 $\theta$, signals have full support, then $\alpha_l^\theta$ and $\beta_l^\theta$ are *strictly* decreasing.

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

$$Q\left[\cup_{j\neq 1} A_j\right] = \sum_{l=2}^{I} (-1)^l \tbinom{I-1}{l-1} \left(\rho\overline{\alpha}_l^G(\overline{\beta}_l^G)^{T-1} + (1-\rho)\,\overline{\alpha}_l^B(\overline{\beta}_l^B)^{T-1}\right) \tag{16}$$

where $\overline{\alpha}_l^\theta$ and $\overline{\beta}_l^\theta$ are defined in the same manner as $\alpha_l^\theta$ and $\beta_l^\theta$ but for the probability distribution $Q$ of $\mathbf{Y}$. 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 binary and conclusive. If*

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

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

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

**Proof.** First, since the signals $\mathbf{X}$ and $\mathbf{Y}$ 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 (15) and (16) 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, (17) 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,

$$\frac{\Pr\left[\mathbf{X}_1=\mathbf{e}^1,\exists j,\mathbf{X}_j=\mathbf{0}\right]}{\Pr\left[\mathbf{Y}_1=\mathbf{e}^1,\exists j,\mathbf{Y}_j=\mathbf{0}\right]}=\frac{P\left(\cup_{j\neq 1}A_j\right)}{Q\left(\cup_{j\neq 1}A_j\right)}<1 \tag{18}$$

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 (18)

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

■

**Lemma C.2** *Suppose $P$ and $Q$ are full-support distributions such that for $\theta=G,B$, and $i\neq j$,*

$$\Pr\left[Y_i=0,Y_j=0\mid\theta\right]>\Pr\left[X_i=0,X_j=0\mid\theta\right] \tag{19}$$

*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 (15) and (16) 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

$$\frac{P\left(\cup_{j\neq 1}A_j\right)}{Q\left(\cup_{j\neq 1}A_j\right)}=\frac{\sum_{l=2}^{I}(-1)^l\binom{I-1}{l-1}\left(\rho\alpha_l^G\left(\frac{\beta_l^G}{\overline{\beta}_2^B}\right)^{T-1}+(1-\rho)\,\alpha_l^B\left(\frac{\beta_l^B}{\overline{\beta}_2^B}\right)^{T-1}\right)}{\sum_{l=2}^{I}(-1)^l\binom{I-1}{l-1}\left(\rho\overline{\alpha}_l^G(\frac{\overline{\beta}_l^G}{\overline{\beta}_2^B})^{T-1}+(1-\rho)\,\overline{\alpha}_l^B(\frac{\overline{\beta}_l^B}{\overline{\beta}_2^B})^{T-1}\right)} \tag{20}$$

and observe that since both $(\theta,\boldsymbol{X})$ and $(\theta,\boldsymbol{Y})$ 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, (19) 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 $\beta_l^B$ 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 (20), 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 (20) 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,

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

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 (18)

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

■

# D Appendix: Almost-public signals

Recall that $S^{(n)} = \{\mathbf{y}_i : \sum_t y_i^t = n\}$ is the set of types whose signal-sum is $n$ and $\Gamma^{(n)} = \{\mathbf{y} : \forall j, \sum_t y_j^t \geq n\}$ is the set of types of all agents such that the signal-sum of *each* agent is at least $n$. Also, recall that $d(\boldsymbol{y}) = \min_k \sum_i |y_i - k|$ is the distance of $\boldsymbol{y} \in \mathcal{X}^I$ from to the nearest *diagonal* vector $(k, k, .., k) \in \mathcal{X}^I$.

In what follows, we will use the following (Landau) asymptotic notation to denote that a function $f(\varepsilon)$ is of *lower order* than $\varepsilon$.

**Definition 4** *A function $f(\varepsilon) = o(\varepsilon)$ if $\lim_{\varepsilon\to 0}(f(\varepsilon)/\varepsilon) = 0$.*

**Lemma D.1** *Let $Q \in \mathcal{Q}(\mu)$ be an $\varepsilon$-public distribution and suppose that the univariate marginals $\mu^\theta$ satisfy Condition 1. For $\varepsilon > 0$ small enough, for all $n > 1$ and for all $\mathbf{y}_i \in S^{(n)}$,*

$$\Pr\left[\Gamma^{(n)} \mid \mathbf{Y}_i = \mathbf{y}_i\right] < q \equiv \Pr\left[\Gamma^{(1)} \mid \mathbf{Y}_i = \mathbf{e}^1\right] \tag{21}$$

**Proof.** We will develop estimates for both sides of (21) that are such that the error in both estimates is $o(\varepsilon)$.

Define

$$R_{-i}^{t}(\mathbf{y}_i) = \left\{\mathbf{y}_{-i} : \forall s \neq t,\ d(\boldsymbol{y}^s) = 0 \text{ and } \forall j, y_j^t \geq y_i^t,\ d\left(\boldsymbol{y}^t\right) = 1\right\}$$

where $\mathbf{y}_{-i}$ is a vector of types of agents other than $i$. Note that if $\mathbf{y}_{-i} \in R_{-i}^{t}(\mathbf{y}_i)$, there is a *single* mismatch of signals only in period $t$. Note also that for all $t$, $R_{-i}^{t}(\mathbf{y}_i) \subset \Gamma^{(n)}$ and for $t \neq t'$ the sets $R_{-i}^{t}(\mathbf{y}_i)$ and $R_{-i}^{t'}(\mathbf{y}_i)$ are disjoint. Thus, for any $\mathbf{y}_i \in S^{(n)}$,

$$\Pr\left[\Gamma^{(n)}, \mathbf{Y}_i = \mathbf{y}_i\right] = \Pr\left[\forall j, \mathbf{Y}_j = \mathbf{y}_i\right] + \sum_{t=1}^{T} \Pr\left[R_{-i}^{t}(\mathbf{y}_i), \mathbf{Y}_i = \mathbf{y}_i\right] + o(\varepsilon) \qquad (22)$$

The first term is the probability that all agents $j \neq i$ are of the same type as $i$. The second term is on the probability of events in which there is *single* mismatch—not all signals are identical. The probabilities of all other events in $\Gamma^{(n)}$—involving either multiple mismatches in the same period or across periods—are of lower order and $o(\varepsilon)$.

The first term in (22)

$$\Pr\left[\forall j, \mathbf{Y}_j = \mathbf{y}_i\right] = \prod_{t=1}^{T} Q^{\theta}\left(y_i^t, ..., y_i^t\right)$$

We now proceed to calculate the second term in (22). We claim that for all $t$,

$$\Pr\left[R_{-i}^{t}(\mathbf{y}_i), \mathbf{Y}_i = \mathbf{y}_i \mid \theta\right] = \begin{cases} \prod_{s\neq t} Q^{\theta}(y_i^s, ..., y_i^s)\, I\varepsilon & \text{if } 0 \leq y_i^t < K \\ 0 & \text{if } y_i^t = K \end{cases} \qquad (23)$$

This is because, if $y_i^t < K$, then there are exactly $I$ elements of $R_{-i}^{t}(\mathbf{y}_i)$ since either (i) there is only one $h \neq i$ with $y_h^t = y_i^t + 1$; or (ii) $\forall j \neq i$, $y_j^t = y_i^t + 1$. There are $I - 1$ events of type (i) and only one event of type (ii).[18] Each has a probability of $\prod_{s\neq t} Q^{\theta}(y_i^s, ..., y_i^s)\, \varepsilon$ in state $\theta$. If $y_i^t = K$, then $R_{-i}^{t}(\mathbf{y}_i) = \varnothing$.

Let $\mathcal{T}_0$ be the set of periods in which $y_i^t = 0$ and $\mathcal{T}_K$ be the set of periods in which $y_i^t = K$. Let $\mathcal{T}_R$ be the set of remaining periods—those in which $0 < y_i^t < K$ and thus the second term in (22) is: for $I > 2$,

$$\sum_{t=1}^{T} \Pr\left[R_{-i}^{t}(\mathbf{y}_i), \mathbf{Y}_i = \mathbf{y}_i\right] = \sum_{t\notin \mathcal{T}_K} \prod_{s\neq t} Q^{\theta}(y_i^s, ..., y_i^s)\, I\varepsilon$$

and when $I = 2$, $I\varepsilon$ is replaced with $\varepsilon$.

Thus, we can rewrite (22), for $I > 2$,

$$\begin{aligned} \Pr\left[\Gamma^{(n)}, \mathbf{Y}_i = \mathbf{y}_i \mid \theta\right] &= \prod_{t=1}^{T} Q^{\theta}\left(y_i^t, ..., y_i^t\right) \\ &\quad + \sum_{t\notin \mathcal{T}_K} \prod_{s\neq t} Q^{\theta}(y_i^s, ..., y_i^s)\, I\varepsilon + o(\varepsilon) \end{aligned} \qquad (24)$$

[18] If $I = 2$, then the event in (i) is the same as the event in (ii) and so the total number of events of this kind is only $I - 1 = 1$.

and when $I = 2$, the $I\varepsilon$ is replaced with $\varepsilon$.

We now proceed to further simplify (24).

Observe that the marginal probability of $y_i^t$,

$$\mu^\theta(y_i^t) = Q^\theta\left(y_i^t, ..., y_i^t\right) + \begin{cases} I\varepsilon + o(\varepsilon) & \text{if } y_i^t = 0 \text{ or } K \\ 2I\varepsilon + o(\varepsilon) & \text{if } 0 < y_i^t < K \end{cases}$$

and it is useful to write this compactly as

$$\mu^\theta(y_i^t) = Q^\theta\left(y_i^t, ..., y_i^t\right) + \left(2 - \mathcal{I}_{\mathcal{T}_0 \cup \mathcal{T}_K}(t)\right) I\varepsilon + o(\varepsilon)$$

where $\mathcal{I}_{\mathcal{T}_0 \cup \mathcal{T}_K}$ is the indicator function of $\mathcal{T}_0 \cup \mathcal{T}_K$. This is because now if $y_i^t \neq 0$ or $K$, the $2I$ events with only one mismatch are (i) there is only one $h \neq i$ with $y_h^t = y_i^t \pm 1$; or (ii) $\forall j \neq i$, $y_j^t = y_i^t + 1$ or $\forall j \neq i$, $y_j^t = y_i^t - 1$. If $y_i^t = 0$ or $K$, then there are only $I$ such events.

Thus, the first term on the right-hand side of (24)

$$\begin{aligned} & \textstyle\prod_{t=1}^T Q^\theta\left(y_i^t, ..., y_i^t\right) \\ = & \textstyle\prod_{t \in \mathcal{T}_0 \cup \mathcal{T}_K}\left(\mu^\theta(y_i^t) - I\varepsilon\right)\prod_{t \in \mathcal{T}_R}\left(\mu^\theta(y_i^t) - 2I\varepsilon\right) + o(\varepsilon) \\ = & \textstyle\prod_{t=1}^T \mu^\theta(y_i^t) - 2I\varepsilon \displaystyle\sum_{t \in \mathcal{T}_R} \textstyle\prod_{s \neq t} \mu^\theta(y_i^s) - I\varepsilon \displaystyle\sum_{t \in \mathcal{T}_0 \cup \mathcal{T}_K} \textstyle\prod_{s \neq t} \mu^\theta(y_i^s) + o(\varepsilon) \end{aligned} \quad (25)$$

and the second term on the right-hand side of (24)

$$\begin{aligned} \textstyle\prod_{s \neq t} Q^\theta\left(y_i^s, ..., y_i^s\right) I\varepsilon &= \textstyle\prod_{s \neq t}\left(\mu^\theta(y_i^s) - \left(2 - \mathcal{I}_{(\mathcal{T}_0 \cup \mathcal{T}_K)}(s)\right) I\varepsilon\right) I\varepsilon + o(\varepsilon) \\ &= \textstyle\prod_{s \neq t} \mu^\theta(y_i^s) I\varepsilon + o(\varepsilon) \end{aligned} \quad (26)$$

Substituting from (25) and (26) in (24), we obtain for $I > 2$,

$$\Pr\left[\Gamma^{(n)}, \mathbf{Y}_i = \mathbf{y}_i \mid \theta\right] = \textstyle\prod_{t=1}^T \mu^\theta(y_i^t) - I\varepsilon \displaystyle\sum_{t \notin \mathcal{T}_0} \textstyle\prod_{s \neq t} \mu^\theta(y_i^s) + o(\varepsilon)$$

and if $I = 2$, then $I\varepsilon$ is replaced with just $\varepsilon$.

Writing $\rho^G = \rho$ and $\rho^B = 1 - \rho$, it now follows that

$$\Pr\left[\Gamma^{(n)}, \mathbf{Y}_i = \mathbf{y}_i\right] = \sum_\theta \rho^\theta \textstyle\prod_{t=1}^T \mu^\theta(y_i^t) - I\varepsilon \displaystyle\sum_{t \notin \mathcal{T}_0} \textstyle\prod_{s \neq t} \mu^\theta(y_i^s) + o(\varepsilon)$$

Also, since $\Pr\left[\mathbf{Y}_i = \mathbf{y}_i \mid \theta\right] = \prod_{t=1}^T \mu^\theta(y_i^t)$, for $I > 2$, the conditional probability

$$\Pr\left[\Gamma^{(n)} \mid \mathbf{Y}_i = \mathbf{y}_i\right] = 1 - \frac{\sum_\theta \rho^\theta \sum_{t \notin \mathcal{T}_0} \prod_{s \neq t} \mu^\theta(y_i^s)}{\sum_\theta \rho^\theta \prod_{t=1}^T \mu^\theta(y_i^t)} I\varepsilon + o(\varepsilon) \quad (27)$$

and if $I = 2$, then $I\varepsilon$ is replaced with $\varepsilon$.

Now if $\mathbf{y}_i = \mathbf{e}^1$ so that $n = 1$, we obtain from (27) that for $I > 2$

$$q \equiv \Pr\left[\Gamma^{(1)} \mid \mathbf{Y}_i = \mathbf{e}^1\right] = 1 - \frac{\sum_\theta \rho^\theta \mu^\theta (0)^{T-1}}{\sum_\theta \rho^\theta \mu^\theta (1) \mu^\theta (0)^{T-1}} I\varepsilon + o(\varepsilon) \tag{28}$$

and if $I = 2$, then $I\varepsilon$ is replaced with $\varepsilon$.

Because of (27) and (28), we have the following: when $\varepsilon$ is small, $\Pr\left[\Gamma^{(n)} \mid \mathbf{Y}_i = \mathbf{y}_i\right] < \Pr\left[\Gamma^{(1)} \mid \mathbf{Y}_i = \mathbf{e}^1\right]$ is implied by the inequality,

$$\frac{\sum_\theta \rho^\theta \sum_{t \notin \mathcal{T}_0} \prod_{s \neq t} \mu^\theta(y_i^s)}{\sum_\theta \rho^\theta \prod_{t=1}^T \mu^\theta(y_i^t)} > \frac{\sum_\theta \rho^\theta \mu^\theta(0)^{T-1}}{\sum_\theta \rho^\theta \mu^\theta(1) \mu^\theta(0)^{T-1}}$$

This, in turn, is implied by for all $\theta, \theta' \in \{G, B\}$

$$\begin{aligned} \frac{\sum_{t \notin \mathcal{T}_0} \prod_{s \neq t} \mu^\theta(y_i^s)}{\prod_{t=1}^T \mu^\theta(y_i^t)} &> \frac{\mu^{\theta'}(0)^{T-1}}{\mu^{\theta'}(1)\mu^{\theta'}(0)^{T-1}} \\ \sum_{t \notin \mathcal{T}_0} \frac{1}{\mu^\theta(y_i^t)} &> \frac{1}{\mu^{\theta'}(1)} \end{aligned}$$

and this is guaranteed by Condition 1.

Since there are only a finite number of types $\mathbf{y}_i$'s, for small enough $\varepsilon$, the conclusion is true for all $\mathbf{y}_i$.

This completes the proof. ■

# 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 (2023): "Learning Efficiency of Multi-agent Information Structures," forthcoming, *Journal of Political Economy.*

[5] Iachan, F., and P. Nenov (2015): "Information Quality and Crises in Regime-Change Games," *Journal of Economic Theory,* 158, 739–768.

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

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

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

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

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

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

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

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

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