



Università
Ca' Foscari
Venezia

**Department
of Management**

Working Paper Series

A. Collevocchio and M. LiCalzi

**The probability of nontrivial
common knowledge**

**Working Paper n. 6/2011
Revised: March 2012**

ISSN: 2239-2734



This Working Paper is published under the auspices of the Department of Management at Università Ca' Foscari Venezia. Opinions expressed herein are those of the authors and not those of the Department or the University. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional nature.

The probability of nontrivial common knowledge*

ANDREA COLLEVECCHIO
[collevec@unive.it]
Dept. of Management
Università Ca' Foscari Venezia

MARCO LICALZI
[licalzi@unive.it]
Dept. of Management
Università Ca' Foscari Venezia

(March 2012)

Abstract. We study the probability that two or more agents can attain common knowledge of nontrivial events when the size of the state space grows large. We adopt the standard epistemic model where the knowledge of an agent is represented by a partition of the state space. Each agent is endowed with a partition generated by a random scheme consistent with his cognitive capacity. Assuming that agents' partitions are independently distributed, we prove that the asymptotic probability of nontrivial common knowledge undergoes a phase transition. Regardless of the number of agents, when their cognitive capacity is sufficiently large, the probability goes to one; and when it is small, it goes to zero. Our proofs rely on a graph-theoretic characterization of common knowledge that has independent interest.

Keywords: common knowledge, epistemic game theory, random partitions.

JEL Classification Numbers: C72, D83.

Correspondence to:

Marco LiCalzi Department of Management
 Università Ca' Foscari Venezia
 S. Giobbe, Cannaregio 873
 30121 Venice, Italy
Phone: [+39] 041-234-6925
Fax: [+39] 041-234-7444
E-mail: licalzi@unive.it
Skype: mlicalzi

* We acknowledge useful comments from the audiences at UECE Lisbon 2011 and Amases Pisa 2011. We thank Lucia Milone for graciously sharing one of her amazingly fast computer codes.

1 Introduction

There is not much need to justify common knowledge as a theoretical construct of paramount interest. Since Aumann (1987) wrote that “the common knowledge assumption underlies all of game theory and much of economic theory”, an increasing appreciation of its importance and pervasiveness has been under way. Specialized fields of inquiry such as interactive epistemology or epistemic game theory have been spawned. Expository reviews from Brandenburger and Dekel (1989) and Geanakoplos (1992) have spread awareness among nonspecialists.

The question that falls within the purview of this paper inquires how restrictive is the assumption that two or more agents attain common knowledge of a nontrivial event. While a general answer is bound to depend on many fine details, it is possible to abstract them away and provide a clean baseline. To this aim, we consider a simple environment aligned to the standard model in Aumann (1976).

There is a *knowledge space* formed by a common finite state space S and by a partition profile over S that defines, for each agent, his knowledge. Each element in the partition profile is independently and uniformly drawn from a random distribution over the possible partitions of S and there are at least two agents. We study what is the probability that the realized partition profile yields at least one nontrivial event (that is, different from S itself) that is common knowledge among all agents. To put it more simply, we look at the probability of nontrivial common knowledge.

Our main result is the existence of a phase transition for a fairly general random scheme. As the size of the state space S grows large, the probability of a nontrivial event being common knowledge between two or more agents goes to zero (respectively, one) when the average size of the components of a knowledge partition is large (small). The logic of this result is described by two illustrative examples in Section 3.

The probabilistic method, of course, is unapt to capture all the specific circumstances that may favor or counter the attainment of common knowledge in real life. On the other hand, it provides a clear benchmark that elucidates the technical hurdles that a finite but large state space imposes in this respect. A well-known instance in the game-theoretic literature is the study of the probability that a finite game in strategic form has at least one Nash equilibrium in pure strategies, initiated by Goldman (1957) for zero-sum games and brought afore in a seminal paper by Goldberg et al. (1968); see Rinott and Scarsini (2000) for a review of the main results, as well as some advances that dispense with the assumption of stochastic dependence. A recent addition to this line of research by Daskalakis et al. (2011) takes into account the structure of the interaction graph.

Finally, we mention related literature that, while different in content, resonate well with the scope of our inquiry. Hellman and Samet (2012) investigate how restrictive is the assumption of common priors. Using topological rather than probabilistic techniques, they provide a condition on partition profiles that identifies another phase transition: the set of consistent type profiles that can be derived from a common prior is topologically large (respectively, small) when the partition profile is tight (not tight). Dimitri (1993) studies information processing skills and provides sufficient conditions for individual partitioned knowledge to emerge asymptotically from possibility correspondences.

The rest of the paper is organized as follows. Section 2 collects notation and mathe-

mathematical preliminaries; we tried our best to make sure that skipping the formalities of this short section should not impair accessibility for the rest of the paper. Section 3 sets out assumptions and provides two examples that illustrate the nature of our results. Section 4 states our main theorems and gives an overview of the demonstrative techniques. These are based on a graph-theoretic characterization of common knowledge that (to the best of our knowledge) is new and of independent interest; in particular, when agents' informational partitions are represented as graphs, the events that are common knowledge are generated by a simple operation known as graph sum. All proofs and associated lemmas are collected in the appendix.

2 Preliminaries and notation

Knowledge. The standard model used to describe the knowledge of an agent assumes a set S of states of the world endowed with a partition; see Geanakoplos (1994). The mutually disjoint and exhaustive classes constituting the partition are called *blocks*. If two states are in the same block, then the agent cannot distinguish them. The *possibility correspondence* $\pi : S \rightarrow 2^S \setminus \emptyset$ describes the *informational partition* of an agent: for each $s \in S$, $\pi(s)$ is the set of states that an agent thinks are possible when the true state of the world is s . Moreover, the possibility correspondence is *nondelusional*: for all s in S , $s \in \pi(s)$; that is, the agent never fails to believe that the true state is possible.

Any subset E of S is called an *event*. When the true state of the world is s and $s \in E$, we say that an event E occurs or is true. If $\pi(s) \subseteq E$, then every state that an agent thinks possible (given the true state s) entails E and we say that the agent *knows* E . In general, an agent may know E in some state s and may not know it in another state s' . If $\pi(s) \subseteq E$ for any s in E , we say that E is *self-evident* to the agent. Such event cannot be true without the agent knowing it.

The formal definition of common knowledge was introduced in Aumann (1976). Given S , consider a finite set A formed by $a \geq 2$ agents. Each agent i in A is associated with a partition π_i that represents his knowledge. The finest common coarsening of the *partition profile* $(\pi_i, i \in A)$ is another partition M called their *meet*. An event $E \subseteq S$ is *common knowledge* among the agents in A at s if and only if for any n and sequence (i_1, i_2, \dots, i_n) , it holds that $\pi_{i_n}(\pi_{i_{n-1}} \dots (\pi_{i_1}(s))) \subseteq E$. When S is finite, there is an equivalent definition that leads to a simple characterization. An event that is simultaneously self-evident to all agents in A is called a *public event*. An event E is common knowledge at s if and only if there is a public event occurring at s that entails E ; or, equivalently, if the (fictitious) agent whose informational partition is the meet M of the partition profile $(\pi_i, i \in A)$ knows E at s .

Partitions. Let S be a finite set with n elements; when useful, we write it as S_n to make the number of elements evident. Following custom, we list the blocks of a partition of S in increasing order of their least elements and the elements of each block in increasing order. For instance, the blocks of the partition $\{3, 4, 5\}, \{6, 1\}, \{2\}$ of a set with six elements are listed as 16|2|345.

The set of all partitions of S , ordered by refinement, is a lattice. We write $\pi \preceq \pi'$ to

denote that π is coarser than π' . The notation $\pi_1 \wedge \pi_2$ stands for the finest coarsening of π_1 and π_2 ; analogously, $\pi_1 \vee \pi_2$ denotes their coarsest refinement. The meet of the partition profile $(\pi_a, a \in A)$ is $M = \bigwedge_{a \in A} \pi_a$.

The number of partitions for a (finite) set S_n of n elements is given by the *Bell number* B_n . The first few Bell numbers are $B_0 = 1$, $B_1 = 1$, $B_2 = 2$, $B_3 = 5$, $B_4 = 15$, $B_5 = 52$, and $B_6 = 203$. The Bell numbers satisfy the recursive formula

$$B_{n+1} = \sum_{k=0}^n \binom{n}{k} B_k$$

as well as Dobinsky's formula

$$B_n = \sum_{k=0}^{+\infty} k^n \cdot \left(\frac{e^{-1}}{k!} \right)$$

according to which B_n is the n -th moment of a Poisson distribution with expected value 1.

Random partitions. A random partition of a finite set S_n is a random variable Π taking values in the set $\mathcal{P}(S_n)$ of partitions of S_n . The distribution of Π refers to the collection of probabilities $P(\Pi = \pi)$ as π ranges over $\mathcal{P}(S_n)$.

The simplest example is the *uniform random partition*, where Π has the uniform distribution $P(\Pi = \pi) = 1/B_n$ for each partition π in $\mathcal{P}(S_n)$. Let ξ_n be the number of blocks in a random partition. It is possible to show that the uniform random partition has $E(\xi_n) = B_{n+1}/B_n - 1$ and $\text{Var}(\xi_n) = B_{n+2}/B_n - (B_{n+1}/B_n)^2 - 1$. Moreover, using standard asymptotic techniques,

$$E(\xi_n) = \frac{n}{\ln n} (1 + o(1))$$

as $n \uparrow +\infty$; see Sachkov (1997).

A more general case is the following *random allocation scheme*; see Pitman (1997). Throw n numbered balls into m numbered urns and assume that all m^n possible allocations of balls into urns are equally likely. More formally, let X_i be the number of the urn containing the i -th ball for $1 \leq i \leq n$. Then the X_i are independent and uniformly distributed over the m urns and Π_{mn} is the partition of the n balls induced by the random equivalence relation $i \sim j$ if and only if $X_i = X_j$. The distribution of Π_{mn} induced by the uniform distribution over the m urns is

$$P(\Pi_{mn} = \pi) = \frac{m(m-1) \dots (m-k+1)}{m^n} \quad \text{if } \pi \text{ has } k \text{ blocks;}$$

see Pitman (1997). For $n \geq 3$ this random partition does not have a uniform distribution. However, it is possible to generate a uniformly distributed random partition by a suitable randomization of m ; see Stam (1983).

3 Illustrative examples

The question we investigate in this paper is the probability of nontrivial common knowledge among $a \geq 2$ agents when the finite state space grows large. This section provides two examples that help elucidate the logic of our approach and the import of our results.

We make the following general assumptions. There is a finite state space S_n with n elements and there are $a \geq 2$ agents. Each agent $i = 1, \dots, a$ is endowed with a partition π_i of S_n representing his knowledge. We assume that the partition profile for the set $A = \{1, \dots, a\}$ of agents is obtained by making a independent (but not necessarily identically distributed) random draws over the set of possible partitions.

3.1 The uniform model

Let $P_a(n)$ denote the probability that a group of $a \geq 2$ agents attains common knowledge of a nontrivial event over the state space S_n . We begin by explicitly computing $P_a(n)$ for the special case where the informational partition of each agent is i.i.d. according to the *uniform model*; in particular, for each agent any possible partition is equally likely to occur. For instance, assume that there are $a = 2$ agents and $n = 3$ elements in S . For simplicity, label the elements of S as integers so that $S = \{1, 2, 3\}$. The $B_3 = 5$ possible partitions of S are: 123; 1|23; 2|31; 3|12; 1|2|3. Under the uniform model, each of these five partitions has an identical probability $1/5$ of occurring for each agent.

Suppose that the knowledge of Agent 1 is represented by the partition 1|23 and the knowledge of Agent 2 by the partition 2|13. Then the meet of their partitions is 123, and only the trivial event S is common knowledge. On the contrary, suppose that Agent 1 has the partition 1|2|3 and Agent 2 has the partition 2|13; now, the nontrivial events $\{2\}$ and $\{1, 3\}$ are common knowledge. Clearly, the existence of a nontrivial event that is common knowledge depends on the partition profile for the two agents.

We list in Table 1 the $5 \cdot 5 = 25$ partition profiles that may occur and mark by \times those profiles for which some nontrivial event is common knowledge. Since 10 out of the 25 equally

	123	1 23	2 13	3 12	1 2 3
123					
1 23		\times			\times
2 13			\times		\times
3 12				\times	\times
1 2 3		\times	\times	\times	\times

Table 1: When the meet of two partitions is not the trivial partition.

likely cells are marked, we obtain a probability $P_2(3) = 10/25 = 0.4$ of attaining nontrivial common knowledge under the uniform model when $a = 2$ and $n = 3$.

Our question of interest is what happens to the probability $P_a(n)$ when the size n of the state space S_n grows large. That is, if two (or more) agents face a large state space and their partition profile is drawn according to the i.i.d. uniform model, what is the asymptotic prob-

ability that they can attain common knowledge of a nontrivial event? Perhaps surprisingly, the answer turns out to be zero for this special case.

Proposition 1 *Under the i.i.d. uniform model, $\lim_n P_a(n) = 0$ for any $a \geq 2$.*

The proof of this result is a straightforward corollary of Theorem 5 in Pittel (2000), where it is shown that

$$P_a(n) = O\left(\frac{\log^{a+1} n}{n^{a-1}}\right).$$

This result is obtained in a setup very different from ours; in particular, we warn the casual reader that the lattice of partitions in Pittel (2000) is the dual of what is commonly used in the game-theoretic literature; thus, his “join” is the equivalent of the “meet” defined in this paper. Moreover, his result is based on combinatorial techniques that do not extend to more general random schemes, and therefore are very different from the probabilistic methods employed in this paper.

3.2 The random allocation scheme

The i.i.d. uniform model is a convenient example, but it is not sufficiently flexible to provide a plausible benchmark. Two limitations are most prominent. First, there is no clear relationship between the assumption that all partitions are equally likely and the structure of agents’ knowledge. Second, since agents’ partitions are identically distributed, there is an implicit presumption that the randomness surrounding the generation of an informational partition is the same across agents. We introduce a richer *random allocation scheme* that avoids these two limitations. An extension left for further research is to make agents’ partitions stochastically dependent.

Concerning the first limitation, we assume that an agent cannot articulate his knowledge over more than m distinct partitional blocks; that is, any partition of his has at most m blocks. Let n be the number of states in S_n . Then the ratio $\kappa(m, n) = m/n$ can be viewed as a measure of the *cognitive capacity* of an agent relative to the size of the state space. An agent with $\kappa < 1$ has no access to sufficient cognitive capacity for accommodating perfect knowledge of each state, because the maximum number of blocks in his informational partition is strictly lower than the cardinality of S_n .

To generate a random partition consistent with this assumption, we let each of the n states be independently assigned to one of the m available blocks with equal probability, so that all the m^n possible allocations of states into blocks are equally likely. In particular, if we let N_i denote the number of states in block i , we find the multinomial distribution

$$\Pr\left[\bigcap_{i=1}^m (N_i = n_i)\right] = \frac{n!}{n_1!n_2!\dots n_m!} \cdot \left(\frac{1}{m}\right)^n$$

over any configuration (n_1, n_2, \dots, n_m) such that $\sum_{i=1}^m n_i = n$.

Let us consider an example. For instance, suppose $m = 2$ and $n = 3$ so that an agent has cognitive capacity $\kappa = 2/3$. There are $2^3 = 8$ equally likely allocations of three states over

two blocks, generating four possible partitions: 123; 1|23; 2|31; 3|12. Each of these partitions can be generated into exactly two different ways, and thus they all have the same probability $1/4$ of occurring. Instead, if we let $m = 3$ and increase the agent's cognitive capacity to $\kappa = 1$, then there are five possible partitions: 123; 1|23; 2|31; 3|12; 1|2|3. The first partition occurs with probability $3/27$ and the other four with probability $6/27$ each.

Concerning the second limitation, we allow heterogeneity in agents' informational partitions: each agent has a (possibly, different) bound m_i with a corresponding cognitive capacity $\kappa_i = \kappa(m_i, n)$. Similarly to the i.i.d. uniform model, we maintain the assumption that agents' random partitions are stochastically independent for tractability. Therefore, given a agents and a state space S_n , the random allocation scheme is defined by a profile $(\kappa_1, \dots, \kappa_a)$ of cognitive capacities. A special case of importance occurs when each agent has the same cognitive capacity $\kappa_i = \kappa$.

Continuing our example, suppose that there are $a = 2$ agents with the same cognitive capacity $\kappa = 2/3$. Under the random allocation scheme, their partitions are i.i.d. draws from the uniform distribution over the four partitions 123; 1|23; 2|31; 3|12. A reasoning similar to the example in Section 3.1 yields a probability $P = 3/16$ of attaining nontrivial common knowledge.

Suppose instead that the agents have different cognitive capacities, with $\kappa_1 = 2/3$ and $\kappa_2 = 1$. Dropping the first row from Table 1 and adding marginal probabilities, we obtain Table 2 where we mark by \times those profiles for which some nontrivial event is common knowledge. Adding up the products of the relevant marginal probabilities for all the cells

	123	1 23	2 13	3 12	1 2 3	
123						1/4
1 23		\times			\times	1/4
2 13			\times		\times	1/4
3 12				\times	\times	1/4
	$3/27$	$6/27$	$6/27$	$6/27$	$6/27$	

marked by \times , we find that the probability of nontrivial common knowledge is now $P = 1/3$. Note how increasing the cognitive capacity of the second agent raises the probability of nontrivial common knowledge. This property of monotonicity holds generally. We state first the formal result, and then the sharp and important intuition behind it.

Given the number n of states and the number a of agents, let $P_a(n; \kappa_1, \dots, \kappa_a)$ denote the probability that a group of a agents, each with his own cognitive capacity κ_i , attains common knowledge of a nontrivial event over the state space S_n . The following monotonicity property holds; see the Appendix for a proof.

Theorem 2 *The probability $P_a(n; \kappa_1, \kappa_2, \dots, \kappa_a)$ is (weakly) increasing in each κ_i .*

From a technical viewpoint, this result greatly simplifies the comparative statics concerning the probability of nontrivial common knowledge under the random allocation scheme. Given a profile $(\kappa_1, \dots, \kappa_a)$ of cognitive capacities, let $\kappa_* = \min_i \kappa_i$ and $\kappa^* = \max_i \kappa_i$. Theorem 2 implies

$$P_a(n; \kappa_*, \kappa_*, \dots, \kappa_*) \leq P_a(n; \kappa_1, \kappa_2, \dots, \kappa_a) \leq P_a(n; \kappa^*, \kappa^*, \dots, \kappa^*) \quad (1)$$

for any profile $(\kappa_1, \dots, \kappa_a)$. Hence, an adequate analysis of the behavior of $P_a(n; \kappa_1, \dots, \kappa_a)$ when $\kappa_i = \kappa$ for all i provides natural upper and lower bounds for its value when the profile of cognitive capacities is heterogeneous. This makes the assumption of identical cognitive capacities a convenient simplifying device. Unless otherwise stated, in the following we assume $\kappa_i = \kappa$ for all i and then write simply $P_a(n; \kappa)$ for $P_a(n; \kappa, \kappa, \dots, \kappa)$.

As n grows large, the ability of a group of agents to attain nontrivial common knowledge is heavily influenced by the minimum and maximum levels κ_* and κ^* of their cognitive capacities. When κ_* is sufficiently large (albeit possibly lower than 1), agents have an ample cognitive capacity that tends to generate knowledge partitions formed by many small blocks. The meet of partitions with many small blocks is more likely to contain a nontrivial block; hence, it is easier to find some nontrivial event that is common knowledge among the agents. Vice versa, when κ^* is sufficiently small, the reduced cognitive capacity favors knowledge partitions with few large blocks, whose meet is likely to be the trivial partition. This makes it unlikely for nontrivial events to be common knowledge.

Table 2 showcases this intuition by listing the probabilities $P_a(n; \kappa)$ of nontrivial common knowledge between two agents under the random allocation scheme, with identical cognitive capacities and two different choices of the parameters for $n = 4, 8, 12, 16, 20$. (These probabilities are empirical estimates generated over one million simulations.) The first row reports the probability values for $m = n$, with a constant cognitive capacity $\kappa = 1$. It is apparent

	$n = 4$	$n = 8$	$n = 12$	$n = 16$	$n = 20$
$m = n$	0.666	0.928	0.983	0.996	0.999
$m = \lfloor \sqrt{n} \rfloor$	0.109	0.008	0.008	0.008	0.001

Table 2: Probability of nontrivial common knowledge between two agents.

that, as n increases, the probabilities are increasing towards one: the cognitive capacity is sufficiently high to support nontrivial common knowledge when the state space grows large. The second row, instead, reports the probability values for $m = \lfloor \sqrt{n} \rfloor$, when the (common) cognitive capacity κ is approximately $1/\sqrt{n}$. Now, as n grows large, the cognitive capacity shrinks sufficiently fast that the probability of nontrivial common knowledge tends to 0.

We extend this intuition in the next section by proving that the asymptotic probability of attaining nontrivial common knowledge in the random allocation scheme undergoes a phase transition that depends (almost exclusively) on the limit behavior of the cognitive capacities. Loosely speaking, when the the minimum cognitive capacity κ_* does not decrease too fast, the asymptotic probability of attaining common knowledge between $a \geq 2$ agents goes to 1; on the other hand, if the maximum cognitive capacity κ^* declines sufficiently rapidly, then this probability goes to 0. Thus, depending on the strength of the epistemic assumptions supporting the model, the existence of nontrivial common knowledge may be viewed in the limit as an event that occurs with probability zero or one.

4 Results

This section states our two main results. Immediately after, we describe the general strategy used to develop the proofs. These latter ones are collected in the appendix.

Recall the assumptions made at the beginning of Section 3. There is a finite state space S_n and there are $a \geq 2$ agents. Each agent $i = 1, \dots, a$ is endowed with a partition π_i of S_n representing his knowledge. The partition profile for the set $A = \{1, \dots, a\}$ of agents is generated by a stochastically independent random draws over the set of partitions following the random allocation scheme.

When the size of the state space n grows large, we may view the parameter $m = m(n)$ as a function of n . Throughout the paper, we write m to simplify notation and switch to $m(n)$ only when it is necessary to emphasize the dependence on n . Accordingly, we slightly abuse notation and write κ for the cognitive capacity $\kappa(n) = m(n)/n$.

4.1 Statements

Let $P_a(n; \kappa)$ denote the probability that a group of $a \geq 2$ agents attains common knowledge of a nontrivial event over the state space S_n when they all have the same cognitive capacity κ . The first result is a sufficient condition under which the asymptotic probability $P_a(n; \kappa)$ of nontrivial common knowledge between $a \geq 2$ agents goes to 1.

Theorem 3 *If the sequence $\kappa(n)$ satisfies*

$$\liminf_{n \rightarrow \infty} \kappa(n) \cdot \ln n > a, \quad (2)$$

then

$$\lim_{n \rightarrow \infty} P_a(n, \kappa) = 1.$$

As anticipated, when the cognitive capacity is large and (2) is satisfied, nontrivial common knowledge asymptotically occurs with probability one. For instance, go back to the first example in Section 3.2 where we assume $m(n) = n$ and thus $\kappa(n) = 1$. Replacing this into (2), we see immediately that the condition holds for any $a \geq 2$ and thus the asymptotic probability of nontrivial common knowledge is one for any group of $a \geq 2$ agents. The special case of Theorem 3 for $m(n) = n$ was first claimed in Pittel (2000), with an explicit proof given in LiCalzi and Surucu (2012).

An immediate consequence of Theorem 2 is the following extension of Theorem 3 to the case of heterogenous cognitive capacities. Given the profile $(\kappa_1(n), \kappa_2(n), \dots, \kappa_a(n))$ of agents' cognitive capacities, let $k_*(n) = \min_i \kappa_i(n)$.

Corollary 4 *If the sequence $\kappa_*(n)$ satisfies*

$$\liminf_{n \rightarrow \infty} \kappa_*(n) \cdot \ln n > a, \quad (3)$$

then

$$\lim_{n \rightarrow \infty} P_a(n, \kappa_1, \kappa_2, \dots, \kappa_a) = 1.$$

The second result is a sufficient condition under which the asymptotic probability $P_a(n; \kappa)$ of nontrivial common knowledge between a agents with identical cognitive capacities is 0.

Theorem 5 *If the sequence $\kappa(n)$ satisfies*

$$\limsup_{n \rightarrow \infty} \kappa(n) \cdot \ln n < \frac{1}{8}, \quad (4)$$

then

$$\lim_{n \rightarrow \infty} P_a(n, \kappa) = 0.$$

Dual to the above, when the cognitive capacity is small and (4) is satisfied, nontrivial common knowledge almost surely fails to occur asymptotically. For instance, the second example in Section 3.2 assumes $m(n) = \lfloor \sqrt{n} \rfloor$ and thus $\kappa(n) \approx 1/\sqrt{n}$. Then (4) holds for any $a \geq 2$ and thus the asymptotic probability of nontrivial common knowledge is zero for any group of $a \geq 2$ agents. As a remark, we mention that it is possible to improve the constant $1/8$ in (4), but this would make the proof more complicated and fail to produce a sharp phase transition. Nonetheless, the results are sharp up to a multiplicative constant.

Similarly to before, Theorem 2 implies an analog of Corollary 4 for heterogenous cognitive capacities. Given the profile $(\kappa_1(n), \kappa_2(n), \dots, \kappa_a(n))$ of agents' cognitive capacities, let $k^*(n) = \max_i \kappa_i(n)$.

Corollary 6 *If the sequence $\kappa^*(n)$ satisfies*

$$\limsup_{n \rightarrow \infty} \kappa^*(n) \cdot \ln n < \frac{1}{8}, \quad (5)$$

then

$$\lim_{n \rightarrow \infty} P_a(n, \kappa_1, \kappa_2, \dots, \kappa_a) = 0.$$

Taken together, Theorems 3 and 5 show that the asymptotic probability of attaining common knowledge of some nontrivial event under homogeneous cognitive capacities undergoes a phase transition that is driven by the common cognitive capacity κ , with criticality reached at $\kappa(n) \approx 1/\ln n$. Corollaries 4 and 6 extend these results to the case where agents have heterogenous cognitive capacities.

4.2 The demonstrative approach

The key observation in our demonstrative approach is that any partition π of the set S_n is equivalent to a collection of connected components for an undirected graph over the vertex set $V_n = \{1, 2, \dots, n\}$. To see why, it suffices to place an edge between two (distinct) vertices i and j from V_n if and only if the corresponding elements from S_n belong to the same block. Then any block of π corresponds to a connected component over V_n , and the partition π corresponds to the union of all the connected components of the graph \mathcal{G} .

For instance, consider Figure 1. The central column uses the traditional set-theoretic representation to describe the informational partitions of two agents over a set S with six elements, and their corresponding meet. The right column replicates the same information

using a graph-theoretic visualization. Remarkably, using this latter representation, the graph of the meet is generated by joining the edge lists of all the agents' graphs. Formally speaking, this operation is known as *graph sum*.¹ When such graph sum is connected, as in the bottom-right corner of Figure 1, then the agents share only trivial common knowledge.

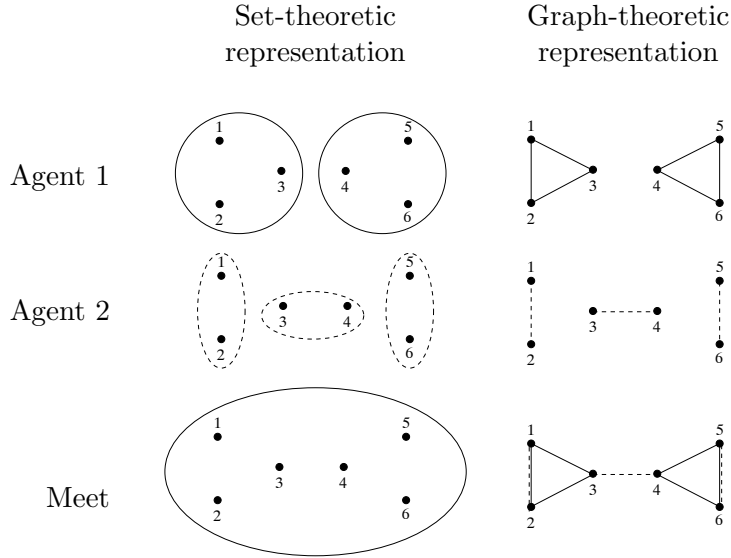


Figure 1: The partitions of two agents and their meet.

Thus, given a partition profile $(\pi_i, i \in A)$ for $a \geq 2$ agents over the state space S_n , we can equivalently represent it as a profile $(\mathcal{G}_i, i \in A)$ of corresponding informational graphs. The meet generated by the partition profile $(\pi_i, i \in A)$ is equivalent to the graph-sum $\mathcal{G}_a(n, m)$ over the graphs $(\mathcal{G}_i, i \in A)$. In particular, the graph $\mathcal{G}_a(n, m)$ is connected if and only if the meet of $(\pi_i, i \in A)$ is the trivial partition. When we add the assumption that the partition profile is generated by stochastically independent draws according to the random allocation scheme described in Section 2 and 3.2, the probability of attaining nontrivial common knowledge is the same as the probability that $\mathcal{G}_a(n, m)$ is *not* a connected graph.

This turns our goal into a study of the connectedness of the random graph $\mathcal{G}_a(n, m)$ when the number of blocks m for the random allocation scheme is a function of n and n grows large. In particular, we prove Theorem 3 by showing in the appendix that, when (2) holds, then

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{G}_a(n, m) \text{ is not connected}) = 1.$$

Analogously, Theorem 5 is proven by demonstrating that (4) implies

$$\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{G}_a(n, m) \text{ is connected}) = 1.$$

The proof of this second result is more challenging and necessitates a few intermediate propositions.

¹ We are not aware of any literature on graph-theoretic representations of common knowledge. Fagin et al. (1995) represent Kripke structures as labeled graphs.

Remarkably, our results state that there is a phase transition when $m(n) \sim n/\ln n$, similarly to a well-known result in the Erdős-Renyi theory of random graphs, where $m(n)$ can be interpreted as the inverse of the probability that an edge is open. However, the model studied in this paper has a richer structure than in the Erdős-Renyi's construction, due to the presence of stochastic dependence among the edges.

A Proofs

A.1 Proof of Theorem 2

As discussed in Section 4.2, the probability $P_a(n; \kappa_1, \dots, \kappa_a)$ that a group of a agents with heterogeneous cognitive capacities attains nontrivial common knowledge is equal to the probability that the graph sum $\mathcal{G}_a(n, \kappa_1, \dots, \kappa_a)$ over the profile $(\mathcal{G}_i, i \in A)$ of their informational graphs is not connected. Therefore, it suffices to prove that $\mathbb{P}(\mathcal{G}_a(n, \kappa_1, \dots, \kappa_a) \text{ is connected})$ is (weakly) decreasing in any κ_i . By symmetry, this is accomplished if we show that

$$\mathbb{P}(\mathcal{G}_a(n, \kappa'_1, \kappa_2, \dots, \kappa_a) \text{ is connected}) \geq \mathbb{P}(\mathcal{G}_a(n, \kappa_1, \kappa_2, \dots, \kappa_a) \text{ is connected})$$

when $\kappa'_1 < \kappa_1$. Recall that $m_1 = \kappa_1 n$. For simplicity, we drop all other arguments and write $\mathcal{G}_a(m_1)$ instead of $\mathcal{G}_a(n, \kappa_1, \kappa_2, \dots, \kappa_a)$ so that our goal becomes to prove that

$$\mathbb{P}(\mathcal{G}_a(m'_1) \text{ is connected}) \geq \mathbb{P}(\mathcal{G}_a(m_1) \text{ is connected}) \quad \text{for } m'_1 < m_1.$$

But, by induction, this follows if we prove that the above inequality holds for adjacent values of m_1 :

$$\mathbb{P}(\mathcal{G}_a(m_1 - 1) \text{ is connected}) \geq \mathbb{P}(\mathcal{G}_a(m_1) \text{ is connected}) \quad \text{for } m_1 \geq 2. \quad (6)$$

Let \mathcal{G}_{-1} denote the graph sum over the profile $(\mathcal{G}_i, i \in A \setminus \{1\})$; that is, \mathcal{G}_{-1} is the random graph generated by all agents except the first one. Since \mathcal{G}_{-1} is a subgraph of $\mathcal{G}_a(m_1)$, whenever \mathcal{G}_{-1} is connected it follows that $\mathcal{G}_a(m_1)$ is connected for any value of m_1 . On the other hand, suppose that \mathcal{G}_{-1} is not connected and consider $\mathcal{G}_a(m_1)$ as the graph sum of \mathcal{G}_{-1} and the random graph $\mathcal{G}_1(m_1)$ for Agent 1.

Clearly, $\mathcal{G}_a(m_1)$ is connected if $\mathcal{G}_1(m_1)$ carries enough links to fill the gaps among the connected components of \mathcal{G}_{-1} . Since $\mathcal{G}_1(m_1)$ is stochastically independent of \mathcal{G}_{-1} , the links are exchangeable and thus only their number matters to achieve connectedness. The greater the number of links added by \mathcal{G}_{-1} , the more likely $\mathcal{G}_a(m_1)$ is to be connected. Let $\mathcal{N}(m_1)$ denote the (random) number of links associated with $\mathcal{G}_1(m_1)$. Then (6) follows if we show that $\mathcal{N}(m - 1)$ stochastically dominates $\mathcal{N}(m_1)$ for any $m_1 \geq 2$.

Recall from Section 3.2 that, given m_1 , the random allocation scheme distributes n states over (at most) m_1 blocks and the profile $\mathbf{X} = (N_1, N_2, \dots, N_{m_1})$ of the number of states in each block has a multinomial distribution. Correspondingly, the random graph $\mathcal{G}_1(m_1)$ has (at most) m_1 connected components, and a component i with N_i nodes carries $(N_i^2 - N_i)/2$ links. Therefore,

$$\mathcal{N}(m_1) = \sum_{i=1}^{m_1} \frac{N_i^2 - N_i}{2} = \sum_{i=1}^{m_1} \left(\frac{N_i^2}{2} \right) - \frac{n}{2}$$

where we use the equality $\sum_i N_i = n$.

To prove that $\mathcal{N}(m-1)$ stochastically dominates $\mathcal{N}(m)$, we apply the theory of stochastic majorization. We say that a random vector \mathbf{X}' *stochastically majorizes* another random vector \mathbf{X} if $f(\mathbf{X}')$ stochastically dominates $f(\mathbf{X})$ for any Borel measurable Schur-convex function $f : \mathbb{R}^n \rightarrow \mathbb{R}$. A family \mathbf{X}_θ of random vectors indexed by a parameter vector θ in \mathbb{R}^m is a Schur family if whenever θ' majorizes θ then $\mathbf{X}_{\theta'}$ stochastically majorizes \mathbf{X}_θ . Application 4.2.a in Nevius et al. (1977) shows that the family of multinomial distributions over m_1 categories indexed by the vector $\theta = (\theta_1, \dots, \theta_m)$ of probabilities for each category is a Schur family, when $\theta_1 > 0$ and $\sum_i \theta_i = 1$. By continuity, their result extends to the case $\theta_{m_1} = 0$ which generates the the multinomial distribution over $m_1 - 1$ categories as a special case of the multinomial distribution over m_1 categories (under the obvious normalization $0^0 = 1$). Now, note that $\mathcal{N}(m-1)$ and $\mathcal{N}(m)$ are generated by the multinomial distributions respectively associated with the parameter vectors

$$\left(\frac{1}{m_1-1}, \frac{1}{m_1-1}, \dots, \frac{1}{m_1-1}, 0 \right) \quad \text{and} \quad \left(\frac{1}{m_1}, \frac{1}{m_1}, \dots, \frac{1}{m_1}, \frac{1}{m_1} \right),$$

where the first vector majorizes the second. Since the function $f(n_1, \dots, n_{m_1}) = \sum_i (n_i/2) - n$ is Schur-convex, we obtain that $\mathcal{N}(m_1-1)$ stochastically dominates $\mathcal{N}(m_1)$. \square

A.2 Proof of Theorem 3

As discussed in Section 4.2, we show that $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{G}_a(n, m) \text{ is not connected}) = 1$. We speak only of vertices (or, equivalently, nodes), but recall that each state in S_n corresponds to a vertex in $V_n = \{1, 2, \dots, n\}$. We say that a vertex is *single* for an agent k when it is not connected to other nodes in his informational graph \mathcal{G}_k ; and it is *everywhere single* when it is single for each agent. Given n nodes, let

$$H_i \stackrel{\text{def}}{=} \{\text{the } i\text{-th vertex is everywhere single}\},$$

be the event that the i -th vertex (out of n) is single for each of the agents. Let

$$X_n = \sum_{i=1}^n \mathbb{1}_{H_i}$$

be the random variable counting how many nodes are everywhere single. The strategy of the proof is the following. Since the graph $\mathcal{G}_a(n, m)$ cannot be connected when a vertex is everywhere single,

$$\{X_n \geq 1\} \subset \{\mathcal{G}_a(n, m) \text{ is not connected}\}$$

and thus

$$\mathbb{P}(X_n \geq 1) \leq \mathbb{P}(\mathcal{G}_a(n, m) \text{ is not connected}).$$

We state a lower bound for $\mathbb{P}(X_n \geq 1)$ and show that it converges to 1 when the sequence $m(n)$ is such that $\kappa(n)$ satisfies (2).

The lower bound is provided by the inequality

$$\mathbb{P}(X \geq 1) \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}, \quad (7)$$

that holds for any non-degenerate random variable X on the non-negative integers. This can be deduced from the Cauchy-Schwartz inequality

$$\mathbb{E}[X] = \mathbb{E}[X \mathbb{1}_{\{X \geq 1\}}] \leq \mathbb{E}[X^2]^{1/2} \mathbb{P}(X \geq 1)^{1/2},$$

after squaring both sides and rearranging. We need to show that the right-hand side of (7) converges to 1.

We begin with a few observations. The probability that a vertex i is everywhere single is

$$\mathbb{P}(H_i) = \left(\frac{m(m-1)^{n-1}}{m^n} \right)^a = \left(1 - \frac{1}{m} \right)^{a(n-1)}.$$

Hence,

$$\mathbb{E}[X_n] = n \left(1 - \frac{1}{m} \right)^{a(n-1)}.$$

Moreover, for $i \neq j$, we claim that

$$\mathbb{P}(H_i \cap H_j) = \left(1 - \frac{1}{m} \right)^{a(n-1)} \left(1 - \frac{1}{m-1} \right)^{a(n-2)}.$$

To see why, consider the following. Conditional on the first node i being everywhere single, the other $n-1$ vertices are uniformly distributed over the remaining $m-1$ components for each agent. Hence, the probability that the vertex j is everywhere single is $(1 - 1/(m-1))^{a(n-2)}$. Combined with the exchangeability of the H_i 's, this yields

$$\begin{aligned} \mathbb{E}[X_n^2] &= n\mathbb{P}(H_1) + n(n-1)\mathbb{P}(H_1 \cap H_2) \\ &= n \left(1 - \frac{1}{m} \right)^{a(n-1)} + n(n-1) \left(1 - \frac{1}{m} \right)^{a(n-1)} \left(1 - \frac{1}{m-1} \right)^{a(n-2)}. \end{aligned}$$

Finally, note that (2) implies that we can choose $\varepsilon > 0$ such that

$$\liminf_{n \rightarrow \infty} \kappa(n) \cdot \ln n \geq a + \varepsilon. \quad (8)$$

Consider now the right-hand side of (7). We have

$$\begin{aligned}
\frac{\mathbb{E}[X_n]^2}{\mathbb{E}[X_n^2]} &= \frac{n^2 \left(1 - \frac{1}{m}\right)^{2a(n-1)}}{n \left(1 - \frac{1}{m}\right)^{a(n-1)} + n(n-1) \left(1 - \frac{1}{m}\right)^{a(n-1)} \left(1 - \frac{1}{m-1}\right)^{a(n-2)}} \\
&= \frac{n \left(1 - \frac{1}{m}\right)^{a(n-1)}}{1 + (n-1) \left(1 - \frac{1}{m-1}\right)^{a(n-2)}} \geq \frac{n \left(1 - \frac{1}{m}\right)^{a(n-1)}}{1 + (n-1) \left(1 - \frac{1}{m-1}\right)^{-a} \left(1 - \frac{1}{m}\right)^{a(n-1)}} \quad (9) \\
&= \frac{n}{\left(1 - \frac{1}{m}\right)^{-a(n-1)} + (n-1) \left(1 - \frac{1}{m-1}\right)^{-a}} \geq \frac{n}{\left[\left(1 - \frac{1}{m}\right)^{-am}\right]^{\frac{n}{m}} + n \left(1 - \frac{1}{m-1}\right)^{-a}} \\
&= \frac{n}{O\left(n^{\frac{a}{a+\varepsilon}}\right) + n \left(1 - \frac{1}{m-1}\right)^{-a}} = \frac{1}{O\left(n^{\frac{-\varepsilon}{a+\varepsilon}}\right) + \left(1 - \frac{1}{m-1}\right)^{-a}},
\end{aligned}$$

where the first inequality follows from $1/m \leq 1/(m-1)$, the second inequality is justified by using $n-1 \leq n$ twice, and the next to last equality obtains because (8) implies

$$\left[\left(1 - \frac{1}{m}\right)^{-am}\right]^{\frac{n}{m}} = O\left([e^a]^{\frac{n}{m}}\right) = O\left(n^{\frac{a}{a+\varepsilon}}\right).$$

Since $O\left(n^{\frac{-\varepsilon}{a+\varepsilon}}\right) \rightarrow 0$ and $\left(1 - \frac{1}{m-1}\right)^{-a} \rightarrow 1$ as $n \uparrow \infty$, the last expression in (9) goes to 1. Hence, the right-hand side of (7) converges to 1, and this proves the theorem. \square

A.3 Proof of Theorem 5

As discussed in Section 4.2, we show that $\lim_{n \rightarrow \infty} \mathbb{P}(\mathcal{G}_a(n, m) \text{ is connected}) = 1$. The strategy of the proof is the following. We prove that with high probability the random graph $\mathcal{G}_a(n, m)$ contains a connected subgraph whose size is larger than $n/2$. Then, we show that this subgraph is very likely to connect all the vertices of $\mathcal{G}_a(n, m)$. Thus, we conclude that $\mathcal{G}_a(n, m)$ is very likely to be connected.

We begin with a key estimate. We say that two (nonempty) disjoint sets of states *overlap* when there is a block that contains states from both collections. Our first intermediate result is an exponential bound for the probability that two sets of states overlap. More precisely, consider the following experiment where each random draw is assumed to be stochastically independent. There are $k \geq 1$ white states and $n - k \geq 1$ black states. (Colors are used only for presentational purposes, and are otherwise irrelevant.) We randomly place each of the n colored states into the m blocks with uniform probability. Then we repeat the procedure afresh $r - 1$ more times. Let $W_r(k, n - k)$ be the event that the two sets of white and black states overlap; that is, in some of the r repetitions, at least one block contains states of different colors.

Proposition 7 *Suppose $n \geq 4m$. Then*

$$\mathbb{P}(W_r(k, n - k)) \geq 1 - (4.35)^r \exp\left\{-\frac{rn}{2m}\right\}.$$

Proof. By definition, it is clear from symmetry that $\mathbb{P}(W_r(k, n-k)) = \mathbb{P}(W_r(n-k, k))$. Hence, without loss of generality, we assume $1 \leq k \leq n/2$. We prove the lower bound by establishing an upper bound for the probability of the complementary event, denoted $W_r^c(k, n-k)$.

Label each block using a different integer from 1 to m . For any nonempty proper subset $M \subset \{1, \dots, m\}$ of the blocks, let $E_M^{(i)}$ denote the event that, in the i -th repetition, each block with a label in M is not empty and the union of the blocks with labels in M contains *all and only* the white states. For a fixed repetition i in $\{1, \dots, r\}$, the events $E_M^{(i)}$, with M running over all nonempty proper subsets of $\{1, \dots, m\}$, are disjoint. Thus, we have

$$W_r^c(k, n-k) = \bigcap_{i=1}^r \bigcup_M E_M^{(i)}.$$

Denote by $J(j, \ell)$, with $1 \leq j \leq \binom{m}{\ell}$, a given list of all the subsets of blocks whose cardinality is $1 \leq \ell \leq m-1$. For a fixed ℓ , all events $E_M^{(i)}$ with cardinality $|M| = \ell$ share the same probability. Hence, using the independence between repetitions,

$$\begin{aligned} \mathbb{P}(W_r^c(k, n-k)) &= \left(\sum_{\ell=1}^{(m-1) \wedge k} \sum_{j=1}^{\binom{m}{\ell}} \mathbb{P}(E_{J(1, \ell)}^{(1)}) \right)^r = \left(\sum_{\ell=1}^{(m-1) \wedge k} \binom{m}{\ell} \mathbb{P}(E_{J(1, \ell)}^{(1)}) \right)^r \\ &\leq \left(\sum_{\ell=1}^{(m-1) \wedge k} \binom{m}{\ell} \left(\frac{\ell}{m}\right)^k \left(1 - \frac{\ell}{m}\right)^{n-k} \right)^r. \end{aligned}$$

We now apply the inequality $\binom{m}{\ell} < (em/\ell)^\ell$, which holds for any $m \geq \ell \geq 1$. This can be proved using the Stirling approximation $n! \geq (n/e)^n \sqrt{2\pi n} e^{1/(12n+1)}$ as follows

$$\binom{m}{\ell} < \frac{m^\ell}{\ell!} \leq \frac{m^\ell}{(\ell/e)^\ell \sqrt{2\pi\ell} e^{1/(12\ell+1)}} = \frac{(em/\ell)^\ell}{\sqrt{2\pi\ell} e^{1/(12\ell+1)}} < (em/\ell)^\ell.$$

We thus have

$$\begin{aligned} \left[\mathbb{P}(W_r^c(k, n-k)) \right]^{1/r} &< \sum_{\ell=1}^{(m-1) \wedge k} \left(\frac{em}{\ell}\right)^\ell \left(\frac{\ell}{m}\right)^k \left(1 - \frac{\ell}{m}\right)^{n-k} \\ &\leq \sum_{\ell=1}^{(m-1) \wedge k} e^\ell \left(\frac{\ell}{m}\right)^{k-\ell} \exp\left\{-\frac{(n-k)\ell}{m}\right\} && \text{(as } 1 - \ell/m \leq e^{-\ell/m}\text{)} \\ &\leq \sum_{\ell=1}^{(m-1) \wedge k} \exp\left\{-\frac{(n-k)\ell}{m} + \ell\right\} && \text{(as } \ell/m \leq 1\text{)} \\ &\leq \sum_{\ell=1}^{(m-1) \wedge k} \exp\left\{-\frac{n\ell}{2m} + \ell\right\} && \text{(as } k \leq n/2\text{)} \\ &= e^{-n/(2m)+1} \sum_{\ell=0}^{(m-1) \wedge k-1} \exp\left\{-\frac{n\ell}{2m} + \ell\right\} \\ &\leq 1.6e^{-n/(2m)+1} \leq 4.35e^{-n/(2m)}, \end{aligned}$$

where the next to last inequality follows from

$$\begin{aligned} \sum_{\ell=0}^{(m-1)\wedge k-1} \exp\left\{-\frac{n\ell}{2m} + \ell\right\} &\leq \sum_{\ell=0}^{\infty} \left(\exp\left\{-\frac{n}{2m} + 1\right\}\right)^\ell = \frac{1}{1 - \exp\left\{-\frac{n}{2m} + 1\right\}} \\ &\text{(and, as } n \geq 4m\text{)} \\ &\leq \frac{1}{1 - \exp\left\{-\frac{4m}{2m} + 1\right\}} \leq \frac{1}{1 - e^{-1}} \leq 1.6. \end{aligned}$$

□

Recall that our ultimate goal is to show that the probability of the event $\{\mathcal{G}_a(n, m) \text{ is connected}\}$ goes to one. To this purpose, we produce a stochastic algorithm that generates an event Ψ_n whose occurrence implies that $\mathcal{G}_a(n, m)$ is connected. Once this event is defined, it suffices to show that $\lim_{n \rightarrow \infty} \mathbb{P}(\Psi_n) = 1$.

The construction of such event is accomplished as follows. Recall that there are a agents and m blocks for each of them. (We toss n labelled states into m blocks to generate the random partition of an agent.) Arrange these am blocks over a $(a \times m)$ grid \mathbf{A} , where the rows represent the agents and the columns represent the blocks. Let $A_{ij} = A_{ij}^n$ be the set of states found in the j -th block ($j = 1, \dots, m$) of the i -th agent ($i = 1, \dots, a$). Since we distribute a set $N = \{1, \dots, n\}$ of distinct states over the blocks in each row, it is clear that $\bigcup_{j=1}^m A_{ij} = N$ for any agent i ; moreover, $A_{ij} \cap A_{k\ell} \neq \emptyset$ only if $i \neq k$.²

Given the grid \mathbf{A} , we construct a second grid \mathbf{A}' . For any set of states A_{ij} in \mathbf{A} , we copy a subset of states from A_{ij} and put them into A'_{ij} : we count the number of states τ_{ij} in A_{ij} and choose $\lfloor \frac{\tau}{2} \rfloor$ of them, randomly and uniformly over the $\binom{\tau}{\lfloor \frac{\tau}{2} \rfloor}$ possibilities. In particular, if $\tau_{ij} = 0$ or 1, then we leave A'_{ij} empty.

We construct a special subset of states from the m blocks of the first agent such that the vertices associated with these states generate a connected component, and moreover they are very likely to be connected with the other vertices of $\mathcal{G}_a(n, m)$.

If necessary, relabel the blocks in $M = \{1, \dots, m\}$ to make sure that the first block is not empty. For simplicity, we identify each block with its label. Now, set $\ell(1) = 1$ and pick the first block. Let $M^t = M \setminus \{\ell(1), \dots, \ell(t-1)\}$ be the set of (yet) unpicked blocks, and recursively define $\ell(t)$ as follows. The subset

$$U^t \stackrel{\text{def}}{=} \left\{ k \in M^t : \exists i \in \{2, \dots, a\}, j \in \{1, \dots, m\} \text{ s.t. } A_{ij} \cap \left(\bigcup_{s=1}^{t-1} A'_{1\ell(s)} \right) \neq \emptyset \text{ and } A_{ij} \cap A'_{1k} \neq \emptyset \right\}$$

is not empty when there exists a block for some agent $i \neq 1$ that contains states in common with some block A'_{1s} with $s \in \{\ell(1), \dots, \ell(t-1)\}$, and at least one state from a block A'_k with $k \in M^t$. That is, U^t is not empty when there are unpicked blocks from $M \setminus M^t$ whose content is connected with the blocks in M^t by means of another agent's block. When U^t is not empty, we let $\ell(t)$ point to one of the unpicked blocks in $M \setminus M^t$, randomly chosen

² We abuse notation in the interest of simplicity. Formally speaking, one should speak of a sets of N states with identical labels and claim that only blocks in distinct rows may contain states with the same labels.

with uniform probability among all the blocks in U^t . When U^t is empty, we set $\ell(t) = \infty$ and terminate the procedure. The key point of this construction is that, for $\ell(t) < \infty$, the vertices associated with the states in $\bigcup_{s=1}^t A_{1\ell(s)}$ form a connected component in $\mathcal{G}(n, m)$.

Denote by

$$T \stackrel{\text{def}}{=} \inf\{t \geq 1 : \left| \bigcup_{s=1}^t A_{1\ell(s)} \right| \geq n/2\}.$$

the smallest time step in the algorithm when more than half of the states of Agent 1 are linked. (Set $T = \infty$ if the defining event never occurs.) Given a subset B of states from $N = \{1, \dots, n\}$, denote by $\mathcal{U}_i(B)$, with $2 \leq i \leq a$, the collection of blocks of the i -th agent that contain states from B . Formally speaking, for $2 \leq i \leq a$, we set

$$\mathcal{U}_i(B) \stackrel{\text{def}}{=} \{j \in \{1, \dots, m\} : A_{ij} \cap B \neq \emptyset\}.$$

We are now ready to define the event

$$\Psi_n \stackrel{\text{def}}{=} \left\{ T < \infty \text{ and } \max_{2 \leq i \leq a} \left| \mathcal{U}_i \left(\bigcup_{t=1}^T A_{1\ell(t)} \right) \right| = m \right\} \quad (10)$$

and prove the claim that its occurrence implies that $\mathcal{G}_a(n, m)$ is connected.

Proposition 8

$$\Psi_n \subseteq \{\mathcal{G}_a(n, m) \text{ is connected}\}.$$

Proof. Suppose that Ψ_n holds. For $T < \infty$, the vertices associated with the states in $\bigcup_{t=1}^T A_{1\ell(t)}$ form a connected component in $\mathcal{G}(n, m)$. When Ψ_n holds, there is at least an agent $i \neq 1$ for whom $|\mathcal{U}_i(\bigcup_{t=1}^T A_{1\ell(t)})| = m$; or, equivalently, we can find states from $\bigcup_{t=1}^T A_{1\ell(t)}$ in each of i 's blocks. Therefore, there exists a path in $\mathcal{G}_a(n, m)$ connecting any pair of vertices, and we can conclude that $\mathcal{G}_a(n, m)$ is connected. \square

We remark that the only role played by the $A'_{1\ell(s)}$ in the definition of U_t is to concatenate the $A_{1\ell(t)}$, with $t \in \{1, 2, \dots, T\}$. This concatenation makes the $A'_{1\ell(t)}$ dependent on each other. On the other hand, conditional on its cardinality, the set $A_{1\ell(t)} \setminus A'_{1\ell(t)}$ is independent of T because this latter random variable depends only on the collection A'_{1i} , with $1 \leq i \leq m$.

We now state a few intermediate results that lead up to estimating $\mathbb{P}(\Psi_n)$.

Proposition 9 *Let B a given set of $\lfloor n/4 \rfloor$ states. For any $i \in \{2, \dots, a\}$,*

$$\mathbb{P}\left(\left| \mathcal{U}_i \left(\bigcup_{t=1}^T A_{1\ell(t)} \right) \right| = m \mid T < \infty\right) \geq \mathbb{P}\left(\left| \mathcal{U}_i(B) \right| = m\right).$$

Proof. Suppose $T < \infty$. By the definition of T , we have $Q \stackrel{\text{def}}{=} \left| \bigcup_{t=1}^T A_{1\ell(t)} \right| \geq n/2$. Recall that by construction $|A'_{1j}| = \lfloor |A_{1j}|/2 \rfloor \leq |A_{1j}|/2$ for any $j = 1, \dots, m$. This yields

$$\left| \bigcup_{t=1}^T A'_{1\ell(t)} \right| = \sum_{t=1}^T |A'_{1\ell(t)}| = \sum_{t=1}^T \left\lfloor \frac{|A_{1\ell(t)}|}{2} \right\rfloor \leq Q/2,$$

and thus we obtain

$$\left| \bigcup_{t=1}^T (A_{1\ell(t)} \setminus A'_{1\ell(t)}) \right| = \left| \bigcup_{t=1}^T A_{1\ell(t)} \right| - \left| \bigcup_{t=1}^T A'_{1\ell(t)} \right| \geq Q - Q/2 = Q/2 \geq \frac{n}{4} \geq |B|. \quad (11)$$

Given the random variable Q , the random variables T and $\bigcup_{t=1}^T (A_{1\ell(t)} \setminus A'_{1\ell(t)})$ are independent. Together with (11), the monotonicity of $\mathcal{U}_i(\cdot)$ with respect to set inclusion yields that

$$\left| \mathcal{U}_i \left(\bigcup_{t=1}^T A_{1\ell(t)} \setminus A'_{1\ell(t)} \right) \right| \text{ is stochastically larger than } \left| \mathcal{U}_i(B) \right|.$$

Since these two random variables cannot exceed m , this proves the Proposition. \square

Proposition 10 *Define*

$$\gamma_n \stackrel{\text{def}}{=} \frac{\lfloor n/4 \rfloor}{m} - \ln m \quad (12)$$

and let B be a given set of $\lfloor n/4 \rfloor$ states. For any $i \in \{2, \dots, a\}$,

$$\mathbb{P} \left(\left| \mathcal{U}_i(B) \right| = m \right) \geq 1 - e^{-\gamma_n}.$$

Proof. Define the event

$$G_j \stackrel{\text{def}}{=} \{\text{the block } A_{ij} \text{ contains no states from } B\}.$$

Then

$$\left\{ \left| \mathcal{U}_i(B) \right| = m \right\} = \left(\bigcup_{j=1}^m G_j \right)^c$$

and the conclusion follows from the upper bound

$$\mathbb{P} \left(\bigcup_{j=1}^m G_j \right) \leq m \mathbb{P}(G_1) = m \left(1 - \frac{1}{m} \right)^{\lfloor n/4 \rfloor} \leq \exp \left\{ -(\lfloor n/4 \rfloor / m) + \ln m \right\} = e^{-\gamma_n}.$$

\square

The next result states a lower bound for the probability that $\mathcal{G}_a(n, m)$ is connected. Once this is in place, the main theorem will follow easily.

Theorem 11 *Suppose $n \geq 4m$. Then*

$$\mathbb{P} \left(\mathcal{G}_a(n, m) \text{ is connected} \right) \geq \left(1 - \exp \{ -(a-1)\gamma_n \} \right) \cdot \left(1 - (4.35)^{a-1} \exp \left\{ -\frac{(a-1) \lfloor n/4 \rfloor}{2m} \right\} \right)^{m-1}, \quad (13)$$

where γ_n is defined in (12).

Proof. Recall from Proposition 8 that $\Psi_n \subseteq \{\mathcal{G}_a(n, m) \text{ is connected}\}$. We can estimate $\mathbb{P}(\Psi_n)$ and provide a lower bound for the probability that $\mathcal{G}_a(n, m)$ is connected. By the definition of Ψ_n in (10), we have

$$\mathbb{P}(\Psi_n) = \mathbb{P}\left(\max_{2 \leq i \leq a} \left| \mathcal{U}_i\left(\bigcup_{t=1}^T A_{1\ell(t)}\right) \right| = m \mid T < \infty\right) \cdot \mathbb{P}(T < \infty). \quad (14)$$

We estimate the two terms on the right-hand side one at the time, beginning with the first one. Recall that $\max_{2 \leq i \leq a} \left| \mathcal{U}_i\left(\bigcup_{t=1}^T A_{1\ell(t)}\right) \right| \leq m$. We have

$$\begin{aligned} \mathbb{P}\left(\max_{2 \leq i \leq a} \left| \mathcal{U}_i\left(\bigcup_{t=1}^T A_{1\ell(t)}\right) \right| < m \mid T < \infty\right) &\leq \left[\mathbb{P}(|\mathcal{U}_i(B)| < m)\right]^{a-1} \\ &= \left[1 - \mathbb{P}\left(|\mathcal{U}_i(B)| = m\right)\right]^{a-1} \\ &\leq 1 - \exp\{-(a-1)\gamma_n\}, \end{aligned} \quad (15)$$

where the first inequality follows from Proposition 9 and the independence between agents' partitions, and the second inequality from Proposition 10.

Next, we estimate $\mathbb{P}(T < \infty)$. The event $\{T < \infty\}$ occurs if and only if, for each $t = 1, \dots, m$, it is true that either $\{\ell(t) < \infty \text{ and } T \geq t\}$ or $\{T < t\}$ take place. Hence, if we define the event

$$E_t \stackrel{\text{def}}{=} \bigcap_{k=1}^t \left\{ \{\ell(k) < \infty, T \geq k\} \cup \{T < k\} \right\}$$

for $t = 1, \dots, m$, then

$$\{T < \infty\} = E_m. \quad (16)$$

Note the initial value $\mathbb{P}(E_1) = 1$, because $\ell(1) = 1$. We get the recursive relation

$$\mathbb{P}(E_t) = \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid E_{t-1}\right) \cdot \mathbb{P}(E_{t-1}). \quad (17)$$

We begin by deriving a lower bound for the first probability in the right-hand side of the above equation. Recall that, for any triplet of events A, B, C , we have

$$\mathbb{P}(A \mid B) = \mathbb{P}(A \mid B \cap C) \cdot \mathbb{P}(B \mid C) + \mathbb{P}(A \mid B^c \cap C) \cdot \mathbb{P}(B^c \mid C).$$

Hence,

$$\begin{aligned}
& \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid E_{t-1}\right) \\
&= \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid T > t-1, E_{t-1}\right) \cdot \mathbb{P}(T > t-1 \mid E_{t-1}) \\
&\quad + \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid T \leq t-1, E_{t-1}\right) \cdot \mathbb{P}(T \leq t-1 \mid E_{t-1}) \\
&= \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid T > t-1, E_{t-1}\right) \cdot \mathbb{P}(T > t-1 \mid E_{t-1}) \\
&\quad + \mathbb{P}(T \leq t-1 \mid E_{t-1}) \\
&\geq \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid T > t-1, E_{t-1}\right) \cdot \\
&\quad \left[\mathbb{P}(T > t-1 \mid E_{t-1}) + \mathbb{P}(T \leq t-1 \mid E_{t-1})\right] \\
&= \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid T > t-1, E_{t-1}\right) \\
&= \mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid T > t-1, \ell(t-1) < \infty\right) \\
&= \mathbb{P}\left(\ell(t) < \infty \mid T > t-1, \ell(t-1) < \infty\right)
\end{aligned}$$

where the next to last equality uses $\{T > t-1, E_{t-1}\} = \{T > t-1, \ell(t-1) < \infty\}$, and the last equality ignores the redundant events. This shows that

$$\mathbb{P}\left(\{\ell(t) < \infty, T \geq t\} \cup \{T < t\} \mid E_{t-1}\right) \geq \mathbb{P}\left(\ell(t) < \infty \mid T > t-1, \ell(t-1) < \infty\right).$$

Combining this inequality with (17), we get

$$\begin{aligned}
\mathbb{P}(E_t) &\geq \mathbb{P}\left(\ell(t) < \infty \mid T > t-1, \ell(t-1) < \infty\right) \cdot \mathbb{P}(E_{t-1}) \\
&= \prod_{k=2}^t \mathbb{P}\left(\ell(k) < \infty \mid T > k-1, \ell(k-1) < \infty\right),
\end{aligned}$$

where the last step comes from recursion. Applying (16) yields

$$\mathbb{P}(T < \infty) = \mathbb{P}(E_m) \geq \prod_{t=2}^m \mathbb{P}\left(\ell(t) < \infty \mid T > t-1, \ell(t-1) < \infty\right). \quad (18)$$

It remains to estimate the probabilities in the last product above. Fix any $t \geq 2$ and suppose $\{T > t-1, \ell(t-1) < \infty\}$ holds. Pick one state at random in $\bigcup_{k=1}^{t-1} A'_{1\ell(k)}$ and color it white; paint in black all the states in the complementary set

$$\left(\bigcup_{k=1}^{t-1} A'_{1\ell(k)}\right)^c,$$

and ignore all other states from $\bigcup_{k=1}^{t-1} A'_{1\ell(k)}$. The total number of black and white states exceeds $\lfloor n/4 \rfloor$ because $T > t-1$. When the white state overlaps with the black ones, then

$\ell(t) < \infty$; that is, $W_{a-1}(1, \lfloor n/4 \rfloor) \subset \{\ell(t) < \infty\}$. Moreover, $W_{a-1}(1, \lfloor n/4 \rfloor)$ is independent of the event $\{T > t-1, \ell(t-1) < \infty\}$ because we chose the only white state at random. Hence, by virtue of Proposition 7 with $r = a-1$ repetitions, we get

$$\begin{aligned} \mathbb{P}(\ell(t) < \infty \mid T > t-1, \ell(t-1) < \infty) &\geq \mathbb{P}(W_{a-1}(1, \lfloor n/4 \rfloor)) \\ &\geq 1 - (4.35)^{a-1} \exp\left\{-\frac{(a-1)\lfloor n/4 \rfloor}{2m}\right\}. \end{aligned}$$

Replacing in (18), we obtain

$$\mathbb{P}(T < \infty) \geq \left(1 - (4.35)^{a-1} \exp\left\{-\frac{(a-1)\lfloor n/4 \rfloor}{2m}\right\}\right)^{m-1}. \quad (19)$$

Let us put everything together. By Proposition 8, we know $\Psi_n \subseteq \{\mathcal{G}_a(n, m) \text{ is connected}\}$. Replacing the two estimates (15) and (19) into (14), we get

$$\begin{aligned} &\mathbb{P}(\mathcal{G}_a(n, m) \text{ is connected}) \\ &\geq \mathbb{P}(\Psi_n) = \mathbb{P}\left(\max_{2 \leq i \leq a} \left| \mathcal{U}_i\left(\bigcup_{t=1}^T A_{1\ell(t)}\right) \right| = m \mid T < \infty\right) \cdot \mathbb{P}(T < \infty) \\ &\geq \left(1 - \exp\{-(a-1)\gamma_n\}\right) \cdot \left(1 - (4.35)^{a-1} \exp\left\{-\frac{(a-1)\lfloor n/4 \rfloor}{2m}\right\}\right)^{m-1} \end{aligned} \quad (20)$$

and the theorem is proved. \square

Proof of Theorem 5. It suffices to show that the last expression on the right-hand side of (20) goes to 1 as $n \uparrow \infty$. We check below that the two bracketed terms in this expression converge to 1. Note that, if $m(n)$ satisfies (4), then for all sufficiently large n we have $n > 8m \ln n$, and in particular $n \geq 4m$.

Consider the first bracketed term. By (4), for all sufficiently large n , we get

$$\gamma_n = \frac{\lfloor n/4 \rfloor}{m} - \ln m > \lfloor 2 \ln n \rfloor - \ln m > \ln n + \ln\left(\frac{n}{m}\right) - 1 > \ln n + \ln(8 \ln n) - 1.$$

As $n \rightarrow \infty$, the last expression on the right diverges and thus $\gamma_n \rightarrow \infty$. Hence, the first bracketed term on the right-hand side of (20) goes to 1.

Recall $a \geq 2$. Substituting $n > 8m \ln n$, the second bracketed term yields the asymptotic estimate

$$\exp\left\{-\frac{(a-1)\lfloor n/4 \rfloor}{2m}\right\} \leq \exp\left\{-\frac{(a-1)\lfloor 2m \ln n \rfloor}{2m}\right\} = O\left(\frac{1}{n^{a-1}}\right) = O\left(\frac{1}{n}\right),$$

and thus

$$\begin{aligned} \left(1 - (4.35)^{a-1} \exp\left\{-\frac{(a-1)\lfloor n/4 \rfloor}{2m}\right\}\right)^{m-1} &= \left(1 - (4.35)^{a-1} O\left(\frac{1}{n}\right)\right)^{m-1} \\ &= \left(1 - O\left(\frac{1}{n}\right)\right)^{m-1}. \end{aligned} \quad (21)$$

For sufficiently large n , the expression $1 - O(\frac{1}{n})$ in parenthesis is larger than $1/2$. Observe that, for some appropriately large constant C , the inequality $(1 - x) \geq e^{-Cx}$ holds for any $x < 1/2$. Thus, for sufficiently large n and C , the last expression in (21) is larger than

$$\exp\{-C(m-1)O(1/n)\},$$

and this converges to 1 because $\lim_{n \rightarrow \infty} m(n)/n = \lim_{n \rightarrow \infty} \kappa(n) = 0$ by (4). □

References

- [1] R. Aumann (1976), “Agreeing to disagree”, *The Annals of Statistics* **4**, 1236–1239.
- [2] R. Aumann (1987), “Game theory”, in J. Eatwell, M. Milgate, and P. Newman (eds.), *The New Palgrave: A Dictionary of Economics*, 460–482, Macmillan.
- [3] A. Brandenburger and E. Dekel (1989), “The role of common knowledge assumptions in game theory”, in F. Hahn (ed.), *The Economics of Missing Markets, Information, and Games*, 46–61, Clarendon Press.
- [4] C. Daskalakis, A.G. Dimakis, and E. Mossel (2011), “Connectivity and equilibrium in random games”, *The Annals of Applied Probability* **21**, 987–1016.
- [5] N. Dimitri (1993), “Learning partitions”, *Economics Letters* **42**, 195–199.
- [6] R. Fagin, J.Y. Halpern, Y. Moses, M.Y. Vardi (1995), *Reasoning about Knowledge*, The MIT Press.
- [7] J. Geanakoplos (1992), “Common knowledge”, *Journal of Economic Perspectives* **6**, 53–82.
- [8] J. Geanakoplos (1994), “Common knowledge”, in: R. J. Aumann and S. Hart (eds.), *Handbook of Game Theory with Economic Applications* **2**, chap. 40, 1437–1496, North-Holland.
- [9] K. Goldberg, A.J. Goldman, and M. Newman (1968), “The probability of an equilibrium point”, *Journal of Research. National Bureau of Standards. Section B.* **72**, 93–101.
- [10] A.J. Goldman (1957), “The probability of a saddlepoint”, *American Mathematical Monthly* **64**, 729–730.
- [11] Z. Hellman and D. Samet (2012), “How common are common priors?”, *Games and Economic Behavior* **74**, 517–525.
- [12] M. LiCalzi and O. Surucu (2012), “The power of diversity over large solution spaces”, *Management Science*, forthcoming.
- [13] S.E. Nevius, F. Proschan, and J. Sethuraman (1977), “Schur functions in statistics II. Stochastic majorization”, *Annals of Statistics* **5**, 263–273.

- [14] B. Pittel (2000), “Where the typical set partitions meet and join”, *Electronic Journal of Combinatorics* **7**, R5.
- [15] J. Pitman (1997), “Some probabilistic aspects of set partitions”, *American Mathematical Monthly* **104**, 201–209.
- [16] Y. Rinott and M. Scarsini (2000), “On the number of pure strategy Nash equilibria in random games”, *Games and Economic Behavior* **33**, 274–293.
- [17] V.N. Sachkov (1997), *Probabilistic Methods in Combinatorial Analysis*, Cambridge University Press.
- [18] A.J. Stam (1983), “Generation of a random partition of a finite set by an urn model”, *Journal of Combinatorial Theory, Series A*, **35**, 231–240.