Submitted to Econometrica

Value-Based Distance Between Information Structures

Fabien Gensbittel, Marcin Pęski, Jérôme Renault

April 11, 2020

VALUE-BASED DISTANCE BETWEEN INFORMATION STRUCTURES

Fabien Gensbittel, Marcin Pęski, Jérôme Renault*

AbstractWe define the distance between two information structures as the largest possible difference in the value across all zero-sum games. We provide a tractable characterization of the distance. We use it to discuss the relation between the value of information in games versus single-agent problems, the value of additional information, informational substitutes, complements, or joint information. The convergence to a countable information structure under the value-based distance is equivalent to the weak convergence of belief hierarchies, implying, among others, that for zero-sum games, the approximate knowledge is equivalent to the common knowledge. At the same time, the space of information structures under the value-based distance is large: there exists a sequence of information structures, where players acquire more and more information, and $\varepsilon > 0$ such that any two elements of the sequence have distance at least ε. This result answers by the negative the second (and last unsolved) of the three problems posed by J.F. Mertens in his paper "Repeated Games", ICM 1986.

1. INTRODUCTION

The role of information is of fundamental importance for the economic theory. It is well known that even small differences in information may

*April 11, 2020. F. Gensbittel and J. Renault : Toulouse School of Economics, University Toulouse Capitole. M. Pęski : Department of Economics, University of Toronto. We are grateful to Satoru Takahashi and Siyang Xiong for comments. We are also grateful for comments to Bart Lipman as the editor and the three anonymous referees. Our work has benefited from the AI Interdisciplinary Institute ANITI. ANITI is funded by the French "Investing for the Future - PIA3" program under the Grant agreement nÂ°ANR-19-PI3A-0004. J. Renault and F. Gensbittel gratefully acknowledge funding from the French National Research Agency (ANR) under the Investments for the Future (Investissements d'Avenir) program, grant ANR-17-EURE-0010. J. Renault gratefully acknowledges funding from ANR MaSDOL. M. Pęski gratefully acknowledges the financial support from the Insight Grant of the Social Sciences and Humanities Research Council of Canada and the hospitality of HEC Paris, where parts of this research were completed.

lead to significant differences in the behavior (Rubinstein (1989)). A recent literature on the strategic (dis)-continuities has studied these differences intensively and in full generality. A typical approach is to consider all possible information structures, modeled as elements of an appropriately defined universal space of information structures, and study the differences in the strategic behavior across all games.

A similar methodology has not been applied to study the relationship between the information, and the agent's bottom line, their payoffs. There are perhaps few reasons for this. First, following Dekel *et al.* (2006), Weinstein and Yildiz (2007) and others, the literature has focused on the interim rationalizability as the solution concept. Compared with the equilibrium, this choice has several advantages: it is easier to analyze, it is more robust from the decision-theoretic perspective, it can be factorized through the Mertens-Zamir hierarchies of beliefs (Dekel *et al.* (2006), Ely and Peski (2006)), and, it does not suffer from the existence problems (unlike the equilibrium - see Simon (2003)). However, the value of information is typically measured in the ex ante sense, where solution concepts like the Bayesian Nash equilibrium are more appropriate. Also, the multiplicity of solutions necessitates that the literature takes the set-based approach. This, of course, makes the quantitative comparison of the value of information difficult. Last but not least, the freedom in choosing games without any restriction makes the equilibrium payoff comparison between information structures trivial, where almost all (see Section 7 for a detailed discussion of this point).

Despite the challenges, we find the questions concerning the strategic value of a information as important and fascinating. How to measure the value of information on the universal type space? How much a player can gain (or lose) from an additional information? Which information structures are similar, in the sense that they always lead to the same payoffs? In order to address these questions, and given the last point in the previous paragraph, we must restrict the analysis to a class of games. In this paper, we propose to focus on zero-sum games. We do so for both substantive and pragmatic reasons. On one hand, the question of the value of information is of special importance when the players' interests are opposing. With zero-sum games, the information has natural comparative statics: a player is better off when her information improves and/or the opponent's information worsens (Peski (2008)). Such comparative statics are intuitive, they hold in the single-agent decision problems (Blackwell (1953)), but they do not hold for general games, where better information may worsen a player's strategic position, and players may have incentives to engage in a pre-game communication to manipulate the available information. Second, many of the constructions in the strategic discontinuities literature rely on special classes of games, like coordination games, or betting games (Rubinstein (1989), Morris (2002), Ely and Peski (2011), Chen and Xiong (2013) among others). This begs the question whether some of the surprising phenomena, like the difference between approximate knowledge and common knowledge, apply in other classes of games. Our restriction allows to clarify this issue for zero-sum games.

On the other hand, the restriction avoids all the problems mentioned above. Finite zero-sum games have always an equilibrium on common prior information structures (Mertens *et al.* (2015)) that depends only on the distribution over hierarchies of beliefs. The equilibrium has decent decision-theoretic foundations (Brandt (2019)), and, even if it is not unique, the ex ante payoff always is and it is equal to the value of the zero-sum game. Finally, as we demonstrate through numerous results and examples in the paper and in the Online Appendix, the restriction uncovers a rich internal structure of the universal type space.

We define the distance between two common prior information structures as the largest possible difference in the value across all zero-sum payoff functions that are bounded by a constant. This has a straightforward interpretation as a tight upper bound on the gain or loss from moving from one information structure to another. Our first result provides a characterization of the distance in terms of total variation distance between sets of information structures. This distance can be computed as a solution to a convex optimization problem.

The characterization is tractable in applications. In particular,we use it to describe the conditions under which the distance between information structures is maximized in single-agent problems (which are a subclass of zero-sum games). We provide bounds to measure the impact of the marginal distribution over the state. We also use it in a series of results on the comparison of the value of information. A tight upper bound on the value of an additional piece of information is defined as the distance between two type spaces, in one of which one or two players have access to new information. We give conditions when the value of new information is maximized in the single-agent problems. We describe the situations when the value of one piece of information decreases when the other piece of information becomes available, or, in other words, when the two pieces of information are substitutes. Similarly, we show that, under some conditions, the value of one piece of information increases when the other player receives an additional information, or in other words, that the pieces of information for opposing players are complements.[1] Finally, we show that the new information matters only if it is valuable to at least one of the players individually. The joint information contained in the correlation between players' signals is in itself not valuable in the zero-sum games.

The second main result shows that the space of information structures is large under the value-based distance: there exists an infinite sequence

[1] Hellwig and Veldkamp (2009) study when the information acquisition decisions are complements or substitutes in a beauty contest game.

of information structures u^n and $\varepsilon > 0$ such that the value-based distance between each pair of structures is at least ε. In particular, it is not possible to approximate the set of information structures with finitely many well-chosen information structures. In the proof, we construct a Markov chain with the first element of the chain correlated with the state of the world. We construct an information structure u^n so that player 1 observes the first n odd elements of the sequence and the other player observes the first n even elements. Our construction implies that in information structure u^{n+1}, each player gets an extra signal. Thus, having more and more information may lead... nowhere. This is unlike the single-player case, where more and more signals corresponds to a martingale and the values converge uniformly over bounded decision problems.

The Markov construction implies that all the information structures $n' \geq n$ have the same n-th order belief hierarchies (Mertens and Zamir (1985)). As a consequence, our distance is not robust with respect to the product convergence of belief hierarchies. This observation may sound familiar to a reader of the strategic (dis)continuities literature. However, we emphasize that the proof of our result is entirely novel. All earlier constructions heavily rely on either coordination games, or games with betting elements (Rubinstein (1989), Morris (2002), Ely and Peski (2011), Chen and Xiong (2013) among others). Such constructions do not work with zero-sum games.

More importantly, there are significant differences between strategic topologies and the topology induced by the value-based distance. For instance, the type spaces from the famous email game example of Rubinstein (1989), or any approximate knowledge spaces, converge to the common knowledge of the state for the value-based distance. More generally, we show that any sequence of countable information structures converges to a countable structure under value-based distance if and only if the associated hierarchies of beliefs converge in the product topology. The impact of the higher-order beliefs becomes significant only for uncountable information structures.

An important contribution is that our result leads to an answer to the last open problem posed in Mertens (1986)[2]. Specifically, his Problem 2 asks about the equicontinuity of the family of value functions over information structures across all (uniformly bounded) zero-sum games. The positive answer would have implied the equicontinuity of the discounted and the average value in repeated games, and it would have consequences for the convergence in the limits theorems[3]. Unfortunately, our results show that the answer to the problem is negative.

Our paper adds to the literature on the topologies of information structures. Dekel *et al.* (2006) (see also Morris (2002)) introduce *uniform-strategic topologies*, where two types are close if, for any (not necessarily zero-sum) game, the sets of (almost) rationalizable outcomes are (almost) equal.[4] There are two key differences between that and our approach. First, the uniform-strategic topology applies to all (including non-zero-sum) games. Our restriction allows us to show that some of the surprising phenomena studied in this literature, like the difference between approximate knowledge and common knowledge, are not relevant for zero-sum games. Second,

[2]Problem 1 asked about the convergence of the value, and it was proved false in Ziliotto (2016). Problem 3 asked about the equivalence between the existence of the uniform value and the uniform convergence of the value functions, and it was proved to be false by Monderer and Sorin (1993) and Lehrer and Monderer (1994).

[3] Equicontinuity of value functions is used to obtain limit theorems in several works such as Mertens and Zamir (1971), Forges (1982), Rosenberg and Sorin (2001), Rosenberg (2000), Rosenberg and Vieille (2000), Rosenberg *et al.* (2004), Renault (2006), Gensbittel and Renault (2015), Venel (2014), Renault and Venel (2017).

[4]Dekel *et al.* (2006) focus mostly on a weaker notion of *strategic topology* that differs from the uniform strategic in the same way that the pointwise convergence differs from uniform convergence. Chen *et al.* (2010) and Chen *et al.* (2016) provide a characterization of the strategic and the uniform-strategic topologies in terms of convergences of belief hierarchies.

we work with *ex ante* information structures and the equilibrium solution concept, whereas the uniform-strategic topology is designed to work on the *interim* level, with rationalizability. The ex ante equilibrium approach is more appropriate for value comparison and other related questions. For instance, in the information design context, the quality of the information structure is typically evaluated *before* players receive any information.

Finally, this paper contributes to a recent but rapidly growing field of information design (Kamenica and Gentzkow (2011), Ely (2017), Bergemann and Morris (2015), to name a few). In that literature, an agent designs or acquires an information that later will be used in either a single-agent decision problem or a strategic situation. In principle, the design of information may be divorced from the game itself. For example, a bank may acquire software to process and analyze large amounts of financial information before knowing what stock it is going to trade on, or, a spy master allocates resources to different tasks or regions before she understands the nature of future conflicts. The value-based distance is a tight upper bound on the willingness to pay for a change in information structure. Our results provide insight into a structure of the space of choices of the information designer, including its diameter and internal complexity.

2. MODEL

A (countable) *information structure* is an element $u \in \Delta(K \times \mathbb{N} \times \mathbb{N})$ of the space of probabilities over tuples (k, c, d), where K is a fixed finite set with $|K| \geq 2$, and $\mathbb{N}$ is the set of nonnegative integers[5]. The interpretation is that k is a state of nature, and c and d are the signals of, respectively, player 1 (maximizer) and player 2 (minimizer). In other words, an information

[5] All the results from Sections 3 and 4 extend to uncountable information structures. As this extension requires heavier notation and technical details that would distract the reader from the main messages of the present work, it is relegated to the Online Appendix.

structure is a 2-player common prior Harsanyi type space over K with at most countably many types. The set of information structures is denoted by $\mathcal{U} = \mathcal{U}(\infty)$, and for $L = 1, 2, ..., \mathcal{U}(L)$ denotes the subset of information structures where each player receives a signal smaller than or equal to $L-1$ with probability 1. If C and D are nonempty countable sets, we always interpret elements $u \in \Delta(K \times C \times D)$ as information structures, using fixed enumerations of C and D. In particular, if C and D are finite with cardinality at most L, we view $u \in \Delta(K \times C \times D)$ as an information structure in $\mathcal{U}(L)$. For each $u, v \in \mathcal{U}$ define the total variation norm as $\|u - v\| = \sum_{k,c,d} |u(k,c,d) - v(k,c,d)|$.

A *payoff function* is a map $g : K \times I \times J \to [-1, 1]$, where I, J are finite nonempty sets of actions. The set of payoff functions with action sets of cardinality $\leq L$ is denoted by $\mathcal{G}(L)$, and let $\mathcal{G} = \bigcup_{L \geq 1} \mathcal{G}(L)$ be the set of all payoff functions.

An information structure u and a payoff function g together define a zero-sum Bayesian game $\Gamma(u, g)$ played as follows: first, (k, c, d) is selected according to u, player 1 learns c, and player 2 learns d. Next, simultaneously, player 1 chooses $i \in I$ and player 2 chooses $j \in J$, and finally the payoff of player 1 is $g(k, i, j)$. The zero-sum game $\Gamma(u, g)$ has a value (the unique equilibrium, or the minmax, payoff of player 1), which we denote by $\mathrm{val}(u, g)$.

We define *the value-based distance* between two information structures as the largest possible difference in the value across all payoff functions:

$$(1) \qquad \mathbb{d}(u, v) = \sup_{g \in \mathcal{G}} |\,\mathrm{val}(u, g) - \mathrm{val}(v, g)|.$$

This has a straightforward interpretation as a tight upper bound on the gain or loss from moving from one information structure to another. Since all payoffs are in $[-1, 1]$, it is easy to see that $\mathbb{d}(u, v) \leq \|u - v\| \leq 2$.[6]

[6] The inequality is a property of zero-sum games. For every game $g \in \mathcal{G}$, let σ be an optimal strategy of player 1 in $\Gamma(u, g)$ and τ be an optimal strategy of player 2 in $\Gamma(v, g)$.

The distance (1) satisfies two axioms of a metric: the symmetry and the triangular inequality. However, it is possible that $\mathbb{d}(u,v) = 0$ for $u \neq v$. For instance, if we start from an information structure u and relabel the signals of the players, we obtain an information structure u' that is formally different from u but "equivalent" to u. Say that u and v are equivalent, and write $u \sim v$, if for all game structures g in $\mathcal{G}$, $\text{val}(u,g) = \text{val}(v,g)$. We let $\mathcal{U}^* = \mathcal{U} / \sim$ be the set of equivalence classes. Thus, $\mathbb{d}$ is a pseudo-metric on $\mathcal{U}$ and a metric on $\mathcal{U}^*$.

For each information structure $u \in \Delta(K \times C \times D)$, there is a unique belief-preserving mapping that maps signals c and d into induced Mertens-Zamir hierarchies of beliefs $\tilde{c} \in \Theta_1$ and $\tilde{d} \in \Theta_2$, where Θ_i is the universal space of player i's belief hierarchies over K (see Mertens *et al.* (2015)). The mapping induces a consistent probability distribution $\tilde{u} \in \Delta(K \times \Theta_1 \times \Theta_2)$ over the state and hierarchies of beliefs. Let $\Pi_0 = \{\tilde{u} : u \in \mathcal{U}\}$ be the space of all such distributions. The closure of Π_0 (in the weak topology, that is, the topology induced by the product convergence of belief hierarchies) is denoted as Π. Π is the space of consistent probability distributions induced by generalized (measurable, possibly uncountable) information structures. The space Π is compact under weak topology; Π_0 is dense in Π (see corollary III.2.3 and theorem III.3.1 in Mertens *et al.* (2015)). Note that for a payoff function g and $u \in \Pi$, one can similarly define the value $\text{val}(u,g)$ of the associated Bayesian game (see Proposition III.4.2 in Mertens *et al.* (2015)).

3. CHARACTERIZATION OF THE DISTANCE

We start with the notion of garbling, used by Blackwell to compare statistical experiments Blackwell (1953). A garbling is a map $q : \mathbb{N} \to \Delta(\mathbb{N})$. The

Using the saddle-point property of the value, the difference $\text{val}(u,g) - \text{val}(v,g)$ is not larger than the differences of payoffs in $\Gamma(u,g)$ and $\Gamma(v,g)$ when the players play (σ,τ) in both games. This difference is clearly not larger than $\|u - v\|$.

set of all garblings is denoted by $\mathcal{Q} = \mathcal{Q}(\infty)$, and for each $L = 1, 2, ..., \mathcal{Q}(L)$ denotes the subset of garblings $q : \mathbb{N} \to \Delta(\{0, ..., L-1\})$. Given a garbling q and an information structure u, we define the information structures $q.u$ and $u.q$ so that for each k, c, d,

$$q.u(k,c,d) = \sum_{c'} u(k,c',d)q(c|c') \text{ and } u.q(k,c,d) = \sum_{d'} u(k,c,d')q(d|d').$$

We will interpret garblings in two different ways. First, a garbling is seen as an information loss: suppose that (k, c', d) is selected according to u, c is selected according to the probability $q(c')$, and player 1 learns c (and player 2 learns d). The new information structure is exactly equal to $q.u$, where the signal received by player 1 has been deteriorated through the garbling q. Similarly, $u.q$ corresponds to the dual situation where the signal of player 2 has been deteriorated. Further, the garbling q can also be seen as a behavior strategy of a player in a Bayesian game $\Gamma(u, g)$: if the signal received is c, play the mixed action $q(c)$ (the sets of actions of g being identified with subsets of $\mathbb{N}$). The relation between the two interpretations plays an important role in the proof of Theorem 1 below.

THEOREM 1 *For each $L = 1, 2, ..., \infty$, each $u, v \in \mathcal{U}(L)$,*

$$\sup_{g \in \mathcal{G}} (\mathrm{val}(v,g) - \mathrm{val}(u,g)) = \min_{q_1,q_2 \in \mathcal{Q}(L)} \|q_1.u - v.q_2\|. \tag{2}$$

Hence, $$\mathbb{d}(u,v) = \max\left\{ \min_{q_1,q_2 \in \mathcal{Q}(L)} \|q_1.u - v.q_2\|, \min_{q_1,q_2 \in \mathcal{Q}(L)} \|u.q_1 - q_2.v\| \right\}.$$

If $L < \infty$, the supremum in (2) is attained by some $g \in \mathcal{G}(L)$.

We describe the idea of the proof. The starting point is to identify each garbling with a mixed strategy in the Bayesian game $\Gamma(u, g)$ induced from an information structure u. Using this identification, the expected payoff in this game can be written as $\langle g, q_1.u.q_2 \rangle$ where $\langle g, u \rangle = \sum_{k,c,d} g(k,c,d)u(k,c,d)$. Among others, each player can use the Id strategy which plays the received signal. Using the saddle-point property, the difference in values $\mathrm{val}(v, g) -$

$\mathrm{val}(u,g)$ is no less than the difference between player 2's optimal payoff against the strategy Id in v (i.e. $\inf_{q_2}\langle g, v.q_2\rangle$) and player 1's optimal payoff against the Id strategy in u (i.e. $\sup_{q_1}\langle g, q_1.u\rangle$). Since this holds for any game g, it follows that the value-based distance is bounded below by $\sup_g \inf_{q_1,q_2}\langle g, v.q_2 - q_1.u\rangle$. Moreover, using the monotony of the value with respect to information, we have that $\mathrm{val}(v,g) - \mathrm{val}(u,g) \leq \mathrm{val}(v.q_2, g) - \mathrm{val}(q_1.u, g) \leq \|v.q_2 - q_1.u\|$. Observing that $\|v.q_2 - q_1.u\| = \sup_g\langle g, v.q_2 - q_1.u\rangle$, we deduce that the value-based distance is also bounded above by $\inf_{q_1,q_2}\sup_g\langle g, v.q_2 - q_1.u\rangle$. Theorem 1 then follows from the Sion's Minimax Theorem. We leave the complete proof to the appendix.

The Theorem provides a characterization of the value-based distance between two information structures u and v for each player as a total variation distance between two sets obtained as garblings of the original information structures $\{q.u : q \in \mathcal{Q}\}$ and $\{v.q : q \in \mathcal{Q}\}$.

The result simplifies the problem of computing the value-based distance. First, it reduces the dimensionality of the optimization domain from payoff functions and strategy profiles (to compute the value) to a pair of garblings. More importantly, the solution to the original problem (1) is typically a saddle point as it involves finding optimal strategies in a zero-sum game. On the other hand, the function $\|q_1.u - v.q_2\|$ is convex in garblings (q_1, q_2), and, if $L < \infty$, the domains of the optimization problem $\{q.u : q \in \mathcal{Q}(L)\}, \{u.q : q \in \mathcal{Q}(L)\}$ are convex and compact. Thus, for finite structures, the right-hand side of (2) is a convex, compact, and finitely dimensional optimization problem.

For any $u, v \in \mathcal{U}$, say that *player 1 prefers u to v in every game*, write $u \succeq v$, if for all $g \in \mathcal{G}$, $\mathrm{val}(u,g) - \mathrm{val}(v,g) \geq 0$. By the monotony of the value with respect to information in zero-sum games, we have $q.u \preceq u \preceq u.q$ for each garbling q. Theorem 1 implies the following extension of Blackwell's theorem and the characterization from (Peski (2008)) to countable information structures.

COROLLARY 1 *$u \succeq v \iff$ there exists q_1, q_2 in $\mathcal{Q}$ s.t. $q_1.u = v.q_2$.*

4. APPLICATIONS

The characterization from Theorem 1 is quite tractable. This section contains few straightforward applications. The Online Appendix contains numerous examples to illustrate the computations and the subsequent results.

4.1. *The impact of the marginal over K*

Among many ways that two information structures can differ, the most obvious one is that they may have different distributions over the states k. In order to capture the impact of such differences, the next result provides tight bounds on the distance between two type spaces with a given distribution overs the states:

PROPOSITION 1 *For each $p, q \in \Delta K$, each $u, v \in \mathcal{U}$ such that $\operatorname{marg}_K u = p, \operatorname{marg}_K v = q$, we have*

$$(3) \qquad \sum_k |p_k - q_k| \leq \mathbb{d}(u, v) \leq 2\left(1 - \max_{p', q' \in \Delta K} \sum_k \min\left(p_k q'_k, p'_k q_k\right)\right).$$

If $p = q$, the upper bound is equal to $2\left(1 - \max_k p_k\right)$.

The bounds are tight. The lower bound in (3) is reached when the two information structures do not provide any information to any of the players. The upper bound is reached with information structures where one player knows the state perfectly and the other player does not know anything.

When $p = q$, Proposition 1 describes the diameter of the space of information structures with the same distribution p of states. The result is useful for, among others, information design questions, where such space is exactly the choice set when Nature fixes the distribution of states, and the designer of information chooses how much information to acquire. In such a case, the diameter has an interpretation of the (tight) upper bound on the potential gain/loss from moving between information structures.

4.2. *Single-agent problems*

A natural question is what games maximize the value-based distance $\mathbb{d}$. The next result characterizes the situations, when the maximum in (1) is attained by a special class of zero-sum games: the single-agent problems.

Formally, a payoff function $g \in \mathcal{G}(L)$ is a *single-agent (player 1) problem* if the set of actions of player 2 is a singleton, $J = \{*\}$. Let $\mathcal{G}_1 \subset \mathcal{G}$ be the set of player 1 problems. Then, for each $g \in \mathcal{G}_1$, each information structure u, $\mathrm{val}\,(g, u)$ is the maximal expected payoff of player 1 in problem g. Let

$$\mathbb{d}_1\,(u, v) := \sup_{g \in \mathcal{G}_1} |\mathrm{val}\,(u, g) - \mathrm{val}\,(v, g)| \leq \mathbb{d}\,(u, v)\,. \tag{4}$$

For any structure $u \in \Delta\,(K \times C \times D)$, we say that the players' information is *conditionally independent*, if, under u, signals c and d are conditionally independent given k.

Proposition 2 *Suppose that $u, v \in \Delta\,(K \times C \times D)$ are two information structures with conditionally independent information such that $\mathrm{marg}_{K \times D}\, u = \mathrm{marg}_{K \times D}\, v$. Then, $\mathbb{d}\,(u, v) = \mathbb{d}_1\,(u, v)\,.$*

Proposition 2 says that if two information structures differ only by an information of one player, and the players information are conditionally independent in both cases, then the maximum in value-based distance (1) is attained by a single-agent decision problem. Such problems form a relatively small subclass of games and they are easier to identify. In the Online Appendix, we apply the Proposition to compute exact distance between information structures induced by multiple Blackwell experiments.

The proof of the Proposition relies on the characterization from Theorem 1 and shows that the minimum in the optimization problem is attained by the same pair of garblings as in the single-agent version of the problem.

4.3. *Value of additional information: games vs. single agent*

Consider two information structures $u \in \Delta(K \times (C \times C') \times D)$ and $v = \mathrm{marg}_{K\times C\times D}\, u$. When moving from v to u, player 1 gains an additional signal c'. Because u represents more information, u is (weakly) more valuable, and the value of the additional information is defined as $\mathbb{d}(u, v)$, which is equal to the tight upper bound on the gain from the additional signal. A corollary to Proposition 2 shows that if the signals of the two players are independent conditional on the state, the gain from the new information is the largest in the single-agent problems.

Corollary 2 *Suppose that information in u (and therefore in v) is conditionally independent. Then,* $\mathbb{d}(u, v) = \mathbb{d}_1(u, v)$.

4.4. *Informational substitutes*

Next, we ask two questions about the impact of a piece of information on the value of another piece of information. In both cases, we use some conditional independence assumptions that are weaker than in Proposition 2. Suppose that

$$u \in \Delta(K \times (C \times C_1 \times C_2) \times D) \text{ and } v = \mathrm{marg}_{K\times(C\times C_1)\times D}\, u,$$
$$u' = \mathrm{marg}_{K\times(C\times C_2)\times D}\, u, \text{ and } v' = \mathrm{marg}_{K\times C\times D}\, u.$$

When moving from v' to u' or v to u, player 1 gains an additional signal c_2. The difference is that in the latter case, player 1 has more information that comes from signal c_1. The next result shows the impact of an additional signal on the value of information.

Proposition 3 *Suppose that, under u, c_1 is conditionally independent from (c, c_2, d) given k. Then,* $\mathbb{d}(u', v') \geq \mathbb{d}(u, v)$.

Given the assumptions, the marginal value of signal c_2 decreases when signal c_1 is also present. In other words, the two pieces of information are substitutes.

4.5. *Informational complements*

Another question is about the impact of an information of the other player on the value of information. Suppose that

$$u \in \Delta(K \times (C \times C_1) \times (D \times D_1)) \text{ and } v = \operatorname{marg}_{K \times C \times (D \times D_1)} u,$$
$$u' = \operatorname{marg}_{K \times (C \times C_1) \times D} u \text{ and } v' = \operatorname{marg}_{K \times C \times D} u.$$

When moving from v' to u' or v to u, in both cases, player 1 gains an additional signal c_1. However, in the latter case, player 2 has an additional piece of information that comes from signal d_1. The next result shows the impact of the opponent's signal on the value of information.

Proposition 4 *Suppose that (c, c_1) and d are conditionally independent given k. Then,* $\mathbb{d}(u', v') \leq \mathbb{d}(u, v)$.

Given the assumptions, signal c_1 becomes more valuable when the opponent also has access to additional information. Hence, the two pieces of information are complements.

4.6. *Value of joint information*

Finally, we consider a situation where two players receive additional information simultaneously. Consider a distribution $\mu \in \Delta(X \times Y \times Z)$ over countable spaces. We say that random variables x and y are ε-conditionally independent given z if

$$\sum_z \mu(z) \sum_{x,y} |\mu(x,y|z) - \mu(x|z)\mu(y|z)| \leq \varepsilon.$$

Let $u \in \Delta(K \times (C \times C_1) \times (D \times D_1))$ and $v = \mathrm{marg}_{K \times C \times D} u$. When moving from v to u, both players receive a piece of additional information.

PROPOSITION 5 *Suppose that d_1 is ε-conditionally independent from (k,c) given d, and c_1 is ε-conditionally independent from (k,d) given c. Then, $\mathbb{d}(u,v) \leq \varepsilon$.*

The Proposition considers a situation where the additional signal of each player does not provide this player any significant information about the state of the world and the original information of the other player. Such signals would be useless in a single-decision problem. Such signals may be useful in a strategic setting, as valuable information may be contained in their joint distribution.[7] Nevertheless, Proposition 5 says that the information that is jointly shared by the two players is not valuable in zero-sum games.

Although very simple, Proposition 5 has powerful consequences. Below, we use it to show that information structures with approximate knowledge of the state have also approximate common knowledge of the state. More generally, we use it in the proof of Theorem 3 below.

5. LARGE SPACE OF INFORMATION STRUCTURES

5.1. $(\mathcal{U}^*, \mathbb{d})$ *is not totally bounded*

In this section, we assume without loss of generality that $K = \{0,1\}$.

[7]How useful it is, it depends on the solution concept. The joint information is important for Bayesian Nash Equilibrium and Independent Interim Rationalizability - see the leading example of Ely and Peski (2006). The joint information is not important *by assumption* for the Bayes Correlated Equilibrium of Bergemann and Morris (2015) or Interim Correlated Rationalizability of Dekel *et al.* (2007).

THEOREM 2 *There exists $\varepsilon > 0$ and a sequence (u^l) of information structures such that $\mathbb{d}(u^l, u^p) > \varepsilon$ if $l \neq p$.*

The Theorem says that the space of information structures is large: it cannot be partitioned into finitely many subsets such that all structures in a subset are arbitrarily close to each other.

The proof, with an exception of one step that we describe below, is constructive. For fixed large N, we construct a probability μ over infinite sequences $k, c_1, d_1, c_2, d_2, ...$ that starts with a state k followed by alternating signals for each player. The sequence $c_1, d_1, c_2, d_2, ...$ follows a Markov chain on $\{1, ..., N\}$, and the state k only depends on c_1. In structure u^l, player 1 observes signals $(c_1, c_2, ..., c_l)$, and player 2 observes $(d_1, d_2, ..., d_l)$. Thus, the sequence of structures u^l can be understood as fragments of a larger information structure, where progressively more and more information is revealed to each player. The Theorem shows that the larger structure is not the limit of its fragments in the value-based distance. In particular, there is no analog of the martingale convergence theorem for the value-based distance for such sequences.

This has to be contrasted with two other settings, where the limits of information structures are well defined. First, in the single-player case, any sequence of information structures in which the player is receiving more and more signals converges for the distance $\mathbb{d}_1$. Second, the Markov property means that (a) the state is independent from all players' information conditionally on c_1, and (b) each new piece of information is independent from the previous pieces of information conditional on the most recent information of the other player. This ensures that the l-th level hierarchy of beliefs of any type in structure u^l is preserved by all consistent types in structures u^p for $p \geq l$. Therefore, Theorem 2 exhibits a sequence of type spaces in which belief hierarchies converge in the product topology. In particular, it shows that the knowledge of the l-th level hierarchy of beliefs for any arbitrarily high l is not sufficient to play ε-optimally in all finite zero-sum games.

5.2. *Last open problem of Mertens*

Recall that for each information structure u, $\tilde{u}$ denotes the associated consistent probability distribution over belief hierarchies. Because each finite-level hierarchy of beliefs becomes constant as we move along the sequence u^l, it must be that the sequence $\tilde{u}^l$ converges weakly in Π to the limit $\tilde{u}^l \to \tilde{\mu}$. The limit is the consistent probability obtained from the prior distribution μ. Theorem 2 shows that

$$\limsup_{l} \sup_{g \in \mathcal{G}} \left| \operatorname{val}(\mu, g) - \operatorname{val}\left(u^l, g\right) \right| \geq \varepsilon.$$

In particular, the family of all functions $(u \mapsto \operatorname{val}(u, g))_{g \in \mathcal{G}}$ is not equicontinuous on Π equipped with the weak topology. This answers negatively the second of the three problems posed by Mertens (1986) in his Repeated Games survey from ICM: "This equicontinuity or Lispchitz property character is crucial in many papers..." (see also footnote 2).

The importance of the Mertens question comes from the role that it plays in the limit theorems in the repeated games. The existence of a limit value has attracted a lot of attention since the first results by Aumann and Maschler (1995) and Mertens and Zamir (1971) for repeated games and by Bewley and Kohlberg (1976) for stochastic games. Once the fact that an appropriate family of value functions is equicontinuous is established, the existence of the limit value is typically obtained by showing that there is at most one accumulation point of the family (v_δ), for example, by showing that any accumulation point satisfies a system of variational inequalities admitting at most one solution (see e.g. the survey Laraki and Sorin (2015) and footnote 3 for related works).

5.3. *Comments on the proof*

Fix $\alpha < \frac{1}{25}$. We show that we can find even N high enough and a set $S \subseteq \{1, ..., N\}^2$ with certain mixing properties:

$$\begin{aligned} |\{i : (i,j) \in S\}| &\simeq \frac{N}{2}, \text{ for each } j, \\ |\{i : (i,j), (i,j') \in S\}| &\simeq \frac{N}{4}, \text{ for each } j, j', \\ |\{i : (i,j), (i,j'), (l,i) \in S\}| &\simeq \frac{N}{8}, \text{ for each } j, j', l, \end{aligned}$$

etc. The "$\simeq$" means that the left-hand side is within α-related distance to the right-hand side. Altogether, there are 8 properties of this sort (see Appendix C.3) that essentially mean that various sections of S are "uncorrelated" with each other.

We are unable to directly construct S with the required properties. Instead, we show the existence of set S using the probabilistic method of P. Erdős (for a general overview of the method, see Alon and Spencer (2008)). Suppose that the sets $S(i)$ for $i = 1, ..., N$ are chosen independently and uniformly from all $\frac{N}{2}$-element subsets of $\{1, ..., N\}$. We show that if $N \geq 10^8$, then the set $S = \{(i,j) : j \in S(i)\}$ satisfies the required properties with positive probability, proving that a set satisfying these properties exists. Our proof is not particularly careful about the optimal N (hence about the largest ε allowing for the conclusions of Theorem 2).

Given S, we construct the probability distribution μ. First, state k is chosen with equal probability, and c_1 is chosen so that $\frac{c_1}{N+1}$ is the conditional probability of $k = 1$. Next, inductively, for each $l \geq 1$, we choose

- d_l uniformly from set $S(c_l) = \{j : (c_l, j) \in S\}$ and conditionally independently from $k, ..., d_{l-1}$ given c_l, and
- c_{l+1} uniformly from set $S(d_l)$ and conditionally independently from $k, ..., c_l$ given d_l.

As a result, c_1, d_1, c_2, d_2,... follows a Markov chain.

To provide a lower bound on the distance between different information structures, we construct a sequence of games. In game g^p, player 1 is supposed to reveal the first p pieces of her information; player 2 reveals the first $p-1$ pieces. The payoffs are such that it is a dominant strategy for player 1 to precisely reveal her first order belief about the state, which amounts to truthfully reporting c_1. Furthermore, we verify whether the sequence of reports $\left(\hat{c}_1, \hat{d}_1, ..., \hat{c}_{p-1}, \hat{d}_{p-1}, \hat{c}_p\right)$ belongs to the support of the distribution of the Markov chain. If it does, then player 1 receives payoff $\varepsilon \sim \frac{1}{10(N+1)^2}$. If it does not, we identify the first report in the sequence that deviates from the support. The responsible player is punished with payoff -5ε (and the opponent receives 5ε).

The payoffs and the mixing properties of matrix S ensure that players have incentives to report their information truthfully. We check it formally, and we show that if $l > p$, then $\mathbb{d}\left(u^l, u^p\right) \geq \text{val}\left(u^l, g^{p+1}\right) - \text{val}\left(u^p, g^{p+1}\right) \geq 2\varepsilon$.

Our argument implies that the conclusion of the Theorem is true for $\varepsilon = 2.10^{-17}$. However, our argument is not optimized for the largest possible value of ε and we strongly suspect that the threshold ε is much larger.

6. VALUE-BASED TOPOLOGY

6.1. *Relation to the weak topology*

The previous sections discussed the quantitative aspect of the value-based distance. Now, we analyze its qualitative aspect: the topological information.

THEOREM 3 *Let u be in $\mathcal{U}^*$. A sequence (u_n) in $\mathcal{U}^*$ converges to u for the value-based distance if and only if the sequence $(\tilde{u}_n)$ converges weakly to $\tilde{u}$ in Π_0.*

The result says that a convergence in value-based topology to a countable structure is equivalent to the convergence in distribution of finite-order hierarchies of beliefs. Informally, around countable structures, the higher-order beliefs have diminishing importance.

We describe the idea of the proof. If u is finite, we surround the hierarchies $\tilde{c}$ for $c \in C$ by sufficiently small and disjoint neighborhoods, so that all hierarchies in the neighborhood of $\tilde{c}$ have similar beliefs about the state and the opponent. We do the same for the other player. The weak convergence ensures that the converging structures assign large probability to the neighborhoods. We show that any information about a player's hierarchy beyond the neighborhood to which it belongs is almost conditionally independent (in the sense of Section 4.6) from the information about the state and the opponents' neighborhoods. By Proposition 5, only the information about neighborhoods matters, and the latter is similar to the information in the limit structure u. If u is countable, we also show that it can be appropriately approximated by finite structures.

There are two reasons why Theorem 3 is surprising: (a) it seems to have the opposite message to the literature on strategic (dis)continuities, and (b) it seems to contradict our discussion of Theorem 2. We deal with these two issues in order.

6.1.1. *Strategic discontinuities*

For an illustration of the first issue, consider email-game information structures u from Rubinstein (1989). Player 1 always knows the state. Player 2's first-order belief attaches the probability of at least $\frac{1}{1+\varepsilon\frac{p}{1-p}}$ to one of the states, where $p < 1$ is the initial probability of one of the states and ε is the probability of losing the message. It is well-known that, as $\varepsilon \to 0$, the Rubinstein's type spaces converge in the weak topology to the common knowledge of the state. Theorem 3 implies that the Rubinstein's type spaces also converge under the value-based distance.

We can make the last point somehow more general. An information structure $u \in \Delta(K \times C \times D)$ exhibits ε-knowledge of the state if there is a mapping $\kappa : C \cup D \to K$ such that

$$u\Big(\{u(\{k = \kappa(c)\}|c) \geq 1 - \varepsilon\}\Big) \geq 1-\varepsilon \text{ and } u\Big(\{u(\{k = \kappa(d)\}|d) \geq 1 - \varepsilon\}\Big) \geq 1-\varepsilon.$$

In other words, the probability that any of the player player assigns at least $1 - \varepsilon$ to some state is at least $1 - \varepsilon$.

PROPOSITION 6 *Suppose that u exhibits ε-knowledge of the state and that $v \in \Delta(K \times K_C \times K_D)$, where $K_C = K_D = K$ and $\operatorname{marg}_K v = \operatorname{marg}_K u$, and $v(k = k_C = k_D) = 1$. (In other words, v is a common knowledge structure with the only information about the state.) Then,*

$$\mathbb{d}(u, v) \leq 20\varepsilon.$$

Thus, approximate knowledge structures are close to common knowledge structures. The convergence of approximate knowledge type spaces to the approximate common knowledge is a consequence of Theorem 3. The metric bound stated in the Proposition requires a separate (simple) proof based on Proposition 5.

The above results seem to go against the main message of the strategic discontinuities literature (Rubinstein (1989), Dekel *et al.* (2006), Weinstein and Yildiz (2007), Ely and Peski (2011), etc.), where the convergence of finite-order hierarchies does not imply strategic convergence even around finite structures. There are three important ways in which our setting differs. First, we rely on the ex ante equilibrium concept, rather than interim rationalizability. We are also interested in the payoff comparison rather than the behavior. Second, we restrict attention to zero-sum games. Finally, we only work with common prior type spaces.

We believe that each of these differences is important. First, if we worked with rationalizability, an argument due to Weinstein and Yildiz (2007) applies, and, given sufficient richness assumption, it can be used to show that the resulting topology is strictly finer than the weak topology[8]. Further, the ex ante focus and payoff comparison but without restriction to zero-sum games lead to a topology that is significantly finer than the weak topology (in fact, so fine that it can be useless - see Section 7 for a detailed discussion). The role of common prior is less clear. On one hand, Lipman (2003) imply that, at least from the interim perspective, common prior does not generate significant restrictions on finite-order hierarchies. On the other hand, we rely on the ex ante perspective, and common prior is definitely important for Proposition 5, which plays an important role in the proof.

6.1.2. *Relation to Theorem 2*

For the second issue, recall that Theorem 2 exhibits a sequence of countable information structures such that the hierarchies of beliefs converge in the weak topology along the sequence, but the sequence does not converge in the value-based distance. The limiting structure, namely the distribution of the realizations of the infinite Markov chain, is *uncountable*. On the other hand, Theorem 3 says that the convergence in the weak topology to a *countable* information structure is equivalent to the convergence in the value-based distance. Together, the two results imply that although the weak and value-based topologies are equivalent around countable structures $\mathcal{U}^*$, they differ beyond $\mathcal{U}^*$. The impact of the higher-order beliefs becomes significant only for uncountable information structures.

Another way to illustrate the relation between two results is to observe that, although the two topologies coincide on $\mathcal{U}^* \simeq \Pi_0$, and the latter has a compact closure Π under the weak topology, the completion of $\mathcal{U}^*$ with respect to $\mathbb{d}$ is not compact. This should not be confusing, as the "completion" is metric specific and not a purely topological notion and different metrics that induce the same topology can have different completions.

[8]We are grateful to Satoru Takahashi for clarifying this point.

6.2. *Pointwise value-based topology and completions*

An alternative way to define a topology on the space of information structures would be through the convergence of values. Say that a sequence of information structures (u_n) converges to u pointwise if for all payoff functions $g \in \mathcal{G}$, $\lim_{n\to\infty} \mathrm{val}(u_n, g) = \mathrm{val}(u, g)$. Clearly, if (u_n) converges to u for the value-based distance, then it also converges to u pointwise.

The topology of pointwise convergence is the weakest topology that makes the value mappings continuous. And since $\mathrm{val}(\mu, g)$ is also well defined for μ in Π, pointwise convergence is also well-defined on Π. Moreover by Theorem 12 of Gossner and Mertens (2001), the topology of pointwise convergence coincides with the topology of weak convergence on Π. Using Theorem 3, we obtain the following corollary:

COROLLARY 3 *On the set $\mathcal{U}^*$, the topology induced by the value-based distance, the topology of weak convergence and the topology of pointwise convergence coincide. In particular, let u in $\mathcal{U}^*$ and (u_n) be in $\mathcal{U}^*$. Then (u_n) converges to u for the value-based distance if and only if for every g in $\mathcal{G}$,* $\mathrm{val}(u_n, g) \xrightarrow[n\to\infty]{} \mathrm{val}(u, g)$.

A standard way to define a metric compatible with the pointwise topology is the following. Consider any sequence $(g_n)_n$ that is dense in the set of payoff functions $\cup_{L\geq 1}[-1,1]^{K\times L^2}$, in the sense[9] that for each g in $[-1,1]^{K\times L^2}$ and $\varepsilon > 0$, there exists n such that $|g(k,i,j) - g_n(k,i,j)| \leq \varepsilon$ for all $(k,i,j) \in K \times L^2$. The particular choice of $(g_n)_n$ will play no role in the sequel. Define now the distance $\mathbb{d}_W$ on $\mathcal{U}^*$ by:

$$\mathbb{d}_W(u,v) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\,\mathrm{val}(u, g_n) - \mathrm{val}(v, g_n)|.$$

[9]To construct such a sequence, one can for instance proceed as follows. For each positive integer L, consider a finite grid approximating $[-1,1]^{K\times L^2}$ up to $1/L$, then define $(g_n)_n$ by collecting the elements of all grids.

By density of $(g_n)_n$, we have $\mathbb{d}_W(u_l, u) \xrightarrow[l\to\infty]{} 0$ if and only if for all g, $\mathrm{val}(u_l, g) \xrightarrow[l\to\infty]{} \mathrm{val}(u, g)$. $\mathcal{U}^*$ equipped with $\mathbb{d}_W$ is a metric space, and we denote by $\mathcal{V}$ its completion for $\mathbb{d}_W$. For this distance, $\mathcal{U}^*$ is isometric to a dense subset of $\mathcal{V}$, so that $\mathcal{V}$ can be seen as the closure of $\mathcal{U}^*$. Using Theorem 12 of Gossner and Mertens (2001), we have the following result.

THEOREM 4 *$\mathcal{V}$ is homeomorphic to the space Π, endowed with the weak topology.*

PROOF: Define similarly the distance $\mathbb{d}_W$ on Π by $\mathbb{d}_W(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\mathrm{val}(\mu, g_n) - \mathrm{val}(\nu, g_n)|$. By construction, the map $(u \mapsto \tilde{u})$ from $\mathcal{U}^*$ to Π_0 is an isometry for $\mathbb{d}_W$. So $\mathcal{V}$ is isometric to the completion of Π_0 for $\mathbb{d}_W$. But on Π, the topology induced by $\mathbb{d}_W$ is the weak topology, and for this topology, Π is the closure of Π_0. So the completion of Π_0 for $\mathbb{d}_W$ is Π. *Q.E.D.*

As a consequence, $\mathcal{V}$ is compact and does not depend on the choice of (g_n). It contains not only the information structures with countably many types, but also the information structures with continuum of signals, obtained as limits of sequences of information structures with countably many types.

The main interest of Theorem 4 is that we can now view Π as the set $\mathcal{V}$. We can recover exactly the space $(\Pi, weak)$ using values of zero-sum Bayesian games and the completion of a metric space[10]. This may be seen as a duality result between games and information: Π is defined with hierarchies of beliefs but no reference to games and payoffs, whereas $\mathcal{V}$ is defined by values of zero-sum games, with no explicit reference to belief hierarchies. In particular, restricting attention to the values of zero-sum games is still sufficient to obtain the full space Π with the weak topology. Now

[10]We could have worked from the beginning with possibly uncountable information structures, that is with Borel probabilities over $K \times [0,1] \times [0,1]$. Endowing this set with a distance $\mathbb{d}_W$ yields a metric space directly homeomorphic to Π, with no need to go the completion since the space would already be complete. See online material.

the construction of $\mathcal{V}$ yields a new, alternative, interpretation of Π, and one might possibly hope to deduce properties of (Π, weak) by transfering, via the homeomorphism, properties first proven on $\mathcal{V}$.

Finally, although $\mathbb{d}_W$ and our value-based distance $\mathbb{d}$ induce the same topology on $\mathcal{U}^*$, their completions differ. Theorem 2 implies that the completion $\mathcal{W}$ of $\mathcal{U}^*$ for $\mathbb{d}$ is not compact. The space $\mathcal{W}$ also contains information structures with continuum of signals and represents a new space of incomplete information structures, with strong foundations based on the suprema of differences between values of Bayesian games.

7. PAYOFF-BASED DISTANCE

In this section, we consider a version of the distance (1) where the supremum is taken over all, including non-zero-sum, games. We show that such a payoff-based distance between information structures is, mostly, trivial.

A *non-zero sum payoff function* is a map $g : K \times I \times J \to [-1, 1]^2$ where I, J are finite sets. Let $\mathrm{Eq}\,(u, g) \subseteq \mathbb{R}^2$ be the set of Bayesian Nash Equilibrium (BNE) payoffs in game g on information structure u. Assume that the space $\mathbb{R}^2$ is equipped with the maximum norm $\|x - y\|_{\max} = \max_{i=1,2} |x_i - y_i|$ and the space of compact subsets of $\mathbb{R}^2$ with the induced Hausdorff distance $\mathbb{d}^H_{\max}$. Let

$$\text{(5)}\qquad \mathbb{d}_{NZS}\,(u, v) = \sup_{g \text{ is a non-zero-sum payoff function}} \mathbb{d}^H_{\max}\left(\mathrm{Eq}\,(u, g)\,, \mathrm{Eq}\,(v, g)\right).$$

Then, clearly as in our original definition, $0 \leq \mathbb{d}_{NZS}\,(u, v) \leq 2$.[11]

Contrary to the value in the zero-sum game, the BNE payoffs on information structure u cannot be factorized through the distribution $\tilde{u} \in \Pi$ over the hierarchies of beliefs induced by u. For this reason, we only restrict our

[11] A related approach to closeness of information structures is taken in Kajii and Morris (1998). They say that an information structure is close to another one if, for all bounded games, any equilibrium of one is close to an almost equilibrium of the other.

analysis to information structures that are non-redundant, or equivalently information structures induced by a consistent probability with countable support in Π_0. We do so because the dependence of the BNE on the redundant information is not yet well-understood[12].

Let $u \in \Delta(K \times C \times D)$ be an information structure. A subset $A \subseteq K \times C \times D$ is a *proper common knowledge component* if $u(A) \in (0,1)$ and for each signal $s \in C \cup D$, $u(A|s) \in \{0,1\}$. An information structure is *simple* if it does not have a proper common knowledge component. Each non-redundant information structure u has a representation as a convex combination of (non-redundant) simple information structures $u = \sum_\alpha p_\alpha u_\alpha$, where $\sum p_\alpha = 1, p_\alpha \geq 0$, and $p_\alpha > 0$ for at most countably many α.

THEOREM 5 *Suppose that u, v are non-redundant information structures. If u and v are simple, then*

$$\mathbb{d}_{NZS}(u,v) = \begin{cases} 0, & \textit{if } \tilde{u} = \tilde{v}, \\ 2 & \textit{otherwise.} \end{cases}$$

More generally, suppose that $u = \sum p_\alpha u_\alpha$ and $v = \sum q_\alpha v_\alpha$ are the decompositions into simple information structures. We can always choose the decompositions so that $\tilde{u}_\alpha = \tilde{v}_\alpha$ for each α. Then,

$$\mathbb{d}_{NZS}(u,v) = \sum_\alpha |p_\alpha - q_\alpha|.$$

The distance between the two non-redundant simple information structures is binary, either 0 if the information structures are equivalent, or 2 if they are not. In particular, the distance between all simple information structures that do not have the same hierarchies of beliefs is trivially equal

[12] See Sadzik (2008). An alternative approach would be to take an equilibrium solution concept that can be factorized through the hierarchies of beliefs. An example is Bayes Correlated Equilibrium from Bergemann and Morris (2015).

to its maximum possible value 2. The distance $\mathbb{d}_{NZS}$ between two non-redundant, but not necessarily simple information structures depends on how similar is their decomposition into the simple components. Theorem 5 implies that (5) is too fine measure of distance between information structures to be useful.

The proof in the case of two non-redundant and simple structures u and v is straightforward. Let $\tilde{u} \neq \tilde{v}$. First, it is well-known that there exist a finite game $g : K \times I \times J \to [-1, 1]^2$ in which each type of player 1 in the support of $\tilde{u}$ and $\tilde{v}$ reports her hierarchy of beliefs as the unique rationalizable action. Second, Lemma III.2.7 in Mertens *et al.* (2015) (or corollary 4.7 in Mertens and Zamir (1985)) shows that the supports of distributions $\tilde{u}$ and $\tilde{v}$ must be disjoint (it is also a consequence of the result by Samet (1998)). Thus, we can construct a game, in which, additionally to the first game, player 2 chooses between two actions $\{u, v\}$ and it is optimal for her to match the information structure to which player 1's reported type belongs. Finally, we multiply the so obtained game by $\varepsilon > 0$ and construct a new game, in which, additionally, player 1 receives payoff $1 - \varepsilon$ if player 2 chooses u and a payoff of $-1 + \varepsilon$ if player 2 chooses v. Hence, the payoff distance between the two information structures is at least $2 - \varepsilon$, where ε is arbitrary small. So-constructed game, has a BNE in the unique rationalizable profile.

8. CONCLUSION

In this paper, we have introduced and analyzed the value-based distance on the space of information structures. The main advantage of the definition is that it has a simple and useful interpretation as the tight upper bound on the loss or gain from moving between two information structures. This allows us to directly apply it to numerous questions about the value of information, the relation between the games and single-agent problems, comparison of information structures, etc. Additionally, we show that the distance contains an interesting topological information. On one hand, the topology induced on the countable information structures is equivalent to the topology of weak convergence of consistent probabilities over coherent hierarchies of beliefs. On the other hand, the set of countable information structures is not totally bounded for the value-based distance, which solves negatively the last open question raised in Mertens (1986), with deep implications for stochastic games.

By restricting our attention to zero-sum games, we were able to reexamine the relevance of many phenomena observed and discussed in the strategic discontinuities literature. On one hand, the distinction between the approximate knowledge and the approximate common knowledge is not important in situations of conflict. On the other hand, the higher order beliefs matter on some, potentially uncountably large structures. More generally, we believe that the discussion of the strategic phenomena on particular classes of games can be fruitful line of future research. It is not the case that each problem must involve coordination games. Interesting classes of games to study could be common interest games, potential games, etc. [13]

APPENDIX A: PROOF OF THEOREM 1

The proof of Theorem 1 relies on two main aspects: the two interpretations of a garbling (deterioration of signals and strategy) and the minmax theorem.

Part 1. We start with general considerations and first identify payoff functions with particular infinite matrices. For $1 \leq L < \infty$, let $G(L)$ be the set of maps from $K \times \mathbb{N} \times \mathbb{N}$ to $[-1, 1]$ such that $g(k, i, j) = -1$ if $i \geq L, j < L$, $g(k, i, j) = 1$ if $i < L, j \geq L$, and $g(k, i, j) = 0$ if $i > L, j > L$. Elements in $G(L)$ correspond to payoff functions with action set $\mathbb{N}$ for each

[13] As an example of work in this direction, Kunimoto and Yamashita (2018) studies an order on hierarchies and types induced by payoffs in supermodular games.

player, with any strategy $\geq L$ being weakly dominated. We define $G = G(\infty) = \bigcup_{L\geq 1} G(L)$, for each u, v in $\mathcal{U}$ the values $\mathrm{val}(u,g)$ and $\mathrm{val}(v,g)$ are well defined, and $\mathbb{d}(u,v) = \sup_{g\in G} |\,\mathrm{val}(u,g) - \mathrm{val}(v,g)|$.

For $u \in \mathcal{U}$ and $g \in G$, we denote by $\gamma_{u,g}(q_1,q_2)$ the payoff of player 1 in the zero-sum game $\Gamma(u,g)$ when player 1 plays $q_1 \in \mathcal{Q}$ and player 2 plays $q_2 \in \mathcal{Q}$. Extending as usual g to mixed actions, we have $\gamma_{u,g}(q_1,q_2) = \sum_{k,c,d} u(k,c,d)g(k,q_1(c),q_2(d))$. Notice that the scalar product $\langle g,u\rangle = \sum_{k,c,d} g(k,c,d)u(k,c,d)$ is well defined and corresponds to the payoff $\gamma_{u,g}(Id,Id)$, where $Id \in \mathcal{Q}$ is the strategy that plays with probability one the signal received. A straightforward computation leads to $\gamma_{u,g}(q_1,q_2) = \langle g, q_1.u.q_2\rangle$. Consequently,

$$\mathrm{val}(u,g) = \max_{q_1\in\mathcal{Q}}\min_{q_2\in\mathcal{Q}}\langle g, q_1.u.q_2\rangle = \min_{q_2\in\mathcal{Q}}\max_{q_1\in\mathcal{Q}}\langle g, q_1.u.q_2\rangle.$$

And for $L = 1,2,...,+\infty$ and $g \in G(L)$, the max and min can be obtained by elements of $\mathcal{Q}(L)$. Since both players can play the Id strategy in $\Gamma(u,g)$, we have for all $u \in \mathcal{U}$ and $g \in G(L)$ that $\inf_{q_2\in\mathcal{Q}(L)}\langle g, u.q_2\rangle \leq \mathrm{val}(u,g) \leq \sup_{q_1\in\mathcal{Q}(L)}\langle g, q_1.u\rangle$. Notice also that for all u, v in $\mathcal{U}(L)$, $\|u-v\| = \sup_{g\in G(L)}\langle g, u-v\rangle$.

Part 2. We now prove Theorem 1. Fix u, v in $\mathcal{U}(L)$, with $L = 1,2,...,+\infty$. For $g \in G(L)$, we have $\inf_{q_1,q_2\in\mathcal{Q}(L)}\langle g, v.q_2 - q_1.u\rangle \leq \mathrm{val}(v,g) - \mathrm{val}(u,g)$, so

$$\sup_{g\in G(L)} (\mathrm{val}(v,g) - \mathrm{val}(u,g)) \geq \sup_{g\in G(L)}\inf_{q_1,q_2\in\mathcal{Q}(L)}\langle g, v.q_2 - q_1.u\rangle. \tag{6}$$

For $g \in G$, $q_1, q_2 \in \mathcal{Q}(L)$, by monotony of the value with respect to information, we have $\mathrm{val}(v.q_2, g) \geq \mathrm{val}(v,g)$ and $\mathrm{val}(u,g) \geq \mathrm{val}(q_1.u, g)$. So $\mathrm{val}(v,g) - \mathrm{val}(u,g) \leq \mathbb{d}\,(q_1.u, v.q_2) \leq \|q_1.u - v.q_2\|$. Hence

(7)

$$\sup_{g\in\mathcal{G}} (\mathrm{val}(v,g) - \mathrm{val}(u,g)) \leq \inf_{q_1,q_2\in\mathcal{Q}(L)} \|q_1.u - v.q_2\| = \inf_{q_1,q_2\in\mathcal{Q}(L)}\sup_{g\in G(L)}\langle g, v.q_2 - q_1.u\rangle.$$

We are now going to show that

$$\text{(8)}\qquad \sup_{g\in G(L)}\inf_{q_1,q_2\in\mathcal{Q}(L)}\langle g, v.q_2-q_1.u\rangle=\min_{q_1,q_2\in\mathcal{Q}(L)}\sup_{g\in G(L)}\langle g, v.q_2-q_1.u\rangle.$$

Together with inequalities 6 and 7, it will give $\sup_{g\in\mathcal{G}}(\mathrm{val}(v,g)-\mathrm{val}(u,g))=\sup_{g\in G(L)}(\mathrm{val}(v,g)-\mathrm{val}(u,g))=\min_{q_1,q_2\in\mathcal{Q}(L)}\|q_1.u-v.q_2\|$.

To prove 8, we will apply a variant of Sion's theorem (see e.g., Mertens *et al.* (2015) Proposition I.1.3) to the zero-sum game with strategy spaces $G(L)$ for the maximizer, $\mathcal{Q}(L)^2$ for the minimizer, and payoff $h(g,(q_1,q_2))=\langle g, v.q_2-q_1.u\rangle$. The strategy sets $G(L)$ and $\mathcal{Q}(L)^2$ are convex, and h is bilinear.

Case 1: $L<+\infty$. Then $\Delta(\{0,...,L-1\})$ is compact, and $\mathcal{Q}(L)^2$ is compact for the product topology. Moreover, h is continuous, so by Sion's theorem, 8 holds. And $\sup_{g\in G(L)}(\mathrm{val}(v,g)-\mathrm{val}(u,g))$ is achieved, since $G(L)$ is compact.

Case 2: $L=+\infty$. We are going to modify the topology on $\mathcal{Q}$ in order to have $\mathcal{Q}(L)^2$ compact and h l.s.c. in (q_1,q_2). The idea is to identify 0 and $+\infty$ in $\mathbb{N}$. Formally given $q\in\Delta(\mathbb{N})$ and a sequence $(q_n)_n$ of probabilities over $\mathbb{N}$, we define: $(q_n)_n$ converges to q if and only if: $\forall c\geq 1$, $\lim_{n\to\infty}q_n(c)=q(c)$. It implies $\limsup_n q_n(0)\leq q(0)$.

$\Delta(\mathbb{N})$ is now compact, and we endow $\mathcal{Q}$ with the product topology, so that $\mathcal{Q}(L)^2$ is itself compact. Fix $g\in G$. We finally show that $\langle g, q.u\rangle$ is u.s.c. in $q\in\mathcal{Q}$ and $\langle g, v.q\rangle$ is l.s.c. in $q\in\mathcal{Q}$. For this, we take advantage of the particular structure of G: there exists L' such that $g\in G(L')$.

For each q in $\Delta(\mathbb{N})$, we have for each k in K and d in $\mathbb{N}$

$$\begin{aligned} g(k,q,d) &= \sum_{c\in\mathbb{N}} g(c)g(k,c,d)\\ &= g(k,0,d)+\sum_{c=1}^{L'-1}(g(k,c,d)-g(k,0,d))q(c)+\sum_{c\geq L'}(g(k,c,d)-g(k,0,d))q(c). \end{aligned}$$

And for each $c\geq L'$, we have $g(k,c,d)-g(k,0,d)\leq 0$. If $(q_n)_n$ converges to q for our new topology, $\lim_n\sum_{c=1}^{L'-1}(g(k,c,d)-g(k,0,d))q_n(c)=$

$\sum_{c=1}^{L'-1}(g(k,c,d)-g(k,0,d))q(c)$ and by Fatou's lemma $\limsup_n \sum_{c\geq L'}(g(k,c,d)-g(k,0,d))q_n(c) \leq \sum_{c\geq L'}(g(k,c,d)-g(k,0,d))q(c)$. As a consequence $\limsup_n g(k,q_n,d) \leq g(k,q,d)$. This is true for each k and d, and we easily obtain that $\langle g, q.u\rangle = \sum_{k,c,d} u(k,c,d)g(k,q(c),d)$ is u.s.c. in $q \in \mathcal{Q}$.

Similarly, for each $q \in \Delta(\mathbb{N})$, $k \in K$, and $c \in \mathbb{N}$, we can write $g(k,c,q) = g(k,c,0)+\sum_{d=1}^{L'-1}(g(k,c,d)-g(k,c,0))q(c)+\sum_{d\geq L'}(g(k,c,d)-g(k,c,0))q(c)$, with $g(k,c,d)-g(k,c,0) \geq 0$ for $d \geq L'$, and show that $\langle g, v.q\rangle$ is l.s.c. in $q \in \mathcal{Q}$.

APPENDIX B: PROOFS OF SECTION 4

B.1. *Proof of Proposition 1*

We prove the lower bound of (3). Let $g(k) = \mathbb{1}_{p_k>q_k} - \mathbb{1}_{p_k\leq q_k}$. Then,

$$\mathbb{d}(u,v) \geq \mathrm{val}(u,g) - \mathrm{val}(v,g) = \sum_{k\in K}(p_k-q_k)\,g(k) = \sum_{k\in K}|p_k-q_k|\,.$$

Let us prove the upper bound of (3). Define $\bar{u}$ and $\underline{v}$ in $\Delta(K\times K_C\times K_D)$ with $K = K_C = K_D$ such that $\bar{u}(k,c,d) = p_k\,\mathbb{1}_{c=k}\,\mathbb{1}_{d=k_0}$ for some fixed $k_0 \in K$ (complete information for player 1, trivial information for player 2, and the same prior about k as u) and $\underline{v}(k,c,d) = q_k\,\mathbb{1}_{c=k_0}\,\mathbb{1}_{d=k}$ for all (k,c,d) (trivial information for player 1, complete information for player 2, and the same beliefs about k as v). Since the value of a zero-sum game is weakly increasing with player 1's information and weakly decreasing with player 2's information, we have

$$\sup_{g\in\mathcal{G}}(\mathrm{val}(u,g)-\mathrm{val}(v,g)) \leq \sup_{g\in\mathcal{G}}(\mathrm{val}(\bar{u},g)-\mathrm{val}(\underline{v},g)) = \min_{q_1\in\mathcal{Q},q_2\in\mathcal{Q}}\|\bar{u}.q_2-q_1.\underline{v}\|,$$

where, according to Theorem 1, the minimum in the last expression is attained for garblings with values in ΔK. Since player 2 has a unique signal in $\bar{u}$, only $q_2(.|k_0) \in \Delta K$ matters. We denote it by $q' = q_2\,(.|k_0)$. Similarly, we define $p' = q_1(.|k_0) \in \Delta(K)$. Then,

$$\begin{aligned}\|\bar{u}.q_2 - q_1.\underline{v}\| &= \sum_{(k,c,d)\in K^3} |p_k\, \mathbb{1}_{c=k}\, q'_d - q_k\, \mathbb{1}_{d=k}\, p'_c| \\ &= \sum_{k\in K} |p_k q'_k - q_k p'_k| + p_k(1-q'_k) + q_k(1-p'_k) \\ &= 2 + \sum_{k\in K} |p_k q'_k - q_k p'_k| - p_k q'_k - q_k p'_k = 2\left(1 - \sum_{k\in K} \min\left(p_k q'_k, q_k p'_k\right)\right).\end{aligned}$$

A similar inequality holds by inverting the roles of u and v, and the upper bound follows from the fact that one can choose arbitrary p', q'.

If $p = q$, then $\sum_{k\in K} \min\left(p_k q'_k, q_k p'_k\right) = \sum_{k\in K} p_k \min\left(q'_k, p'_k\right) \leq \sum_{k\in K} p_k p'_k \leq \max_{k\in K} p_k$, where the latter is attained by $p'_k = q'_k = \mathbb{1}_{\{k=k^*\}}$ for some $k^* \in K$ such that $p_{k^*} = \max_{k\in k} p_k$.

B.2. *Proof of Proposition 2*

Let us start with general properties of $\mathbb{d}_1$. Let us define the set of single-agent information structures as $\mathcal{U}_1 = \Delta(K \times \mathbb{N})$ using the same convention that countable sets are identified with subsets of $\mathbb{N}$. Note that given $u \in \Delta(K \times C \times D)$, $\mathrm{marg}_{K\times C}\, u \in \mathcal{U}_1$. Let $\mathcal{G}'_1 = \{g' : K \times I \to \mathbb{R} \mid I \text{ finite }\}$ be the set of single-agent decision problems, and define for $u', v' \in \mathcal{U}_1$, $\mathbb{d}'_1(u', v') = \sup_{g'\in\mathcal{G}'_1} |\mathrm{val}(v', g') - \mathrm{val}(u', g')|$. It is easily seen that for any $u, v \in \Delta(K \times C \times D)$,

$$\text{(9)} \qquad \mathbb{d}_1(u, v) = \mathbb{d}'_1(u', v') = \max\{\min_{q\in\mathcal{Q}} \|u' - q.v'\|, \min_{q\in\mathcal{Q}} \|q.u' - v'\|\}$$

where $u' = \mathrm{marg}_{K\times C}\, u$, $v' = \mathrm{marg}_{K\times C}\, v$, $q.u'(k, c) = \sum_{s\in C} u'(k, s) q(s)(c)$ and where the last equality can be obtained by mimicking (and simplifying) the arguments of the proof of Theorem 1.

We now prove Proposition 2. Using the assumptions, we have $u(k) = v(k)$, $u\,(c, d|k) = u\,(c|k)\, u\,(d|k)$, and $v\,(c', d|k) = v\,(d|k)\, v\,(c'|k) = u(d|k) v\,(c'|k)$.

For any pair of garblings q_1, q_2

$$\begin{aligned}\left\|u.q_2 - q_1.v\right\| &= \sum_{k,c,d}\left|\sum_{\beta} u\left(k,c,\beta\right) q_2\left(d|\beta\right) - \sum_{\alpha} v\left(k,\alpha,d\right) q_1\left(c|\alpha\right)\right| \\ &= \sum_{k,c} u\left(k\right) \sum_{d}\left|u\left(c|k\right)\sum_{\beta} u\left(\beta|k\right) q_2\left(d|\beta\right) - \left(\sum_{\alpha} v\left(\alpha|k\right) q_1\left(c|\alpha\right)\right) u\left(d|k\right)\right| \\ &= \sum_{k,c} u\left(k\right)\sum_{d}\left|u\left(d|k\right)\Gamma\left(k,c\right) + \Delta\left(k,d\right) u\left(c|k\right)\right|,\end{aligned}$$

where $\Delta\left(k,d\right) = u\left(d|k\right) - \sum_{\beta} u\left(\beta|k\right) q_2\left(d|\beta\right)$, and $\Gamma\left(k,c\right) = \sum_{\alpha} v\left(\alpha|k\right) q_1\left(c|\alpha\right) - u\left(c|k\right)$. Because $|x+y| \geq |x| + \operatorname{sgn}(x) y$ for each $x, y \in \mathbb{R}$, we have

$$\begin{aligned}&\sum_{d}\left|u\left(d|k\right)\Gamma\left(k,c\right) + \Delta\left(k,d\right) u\left(c|k\right)\right| \\ \geq &\sum_{d} u\left(d|k\right)\left|\Gamma\left(k,c\right)\right| + \operatorname{sgn}\left(\Gamma\left(k,c\right)\right) u\left(c|k\right)\sum_{d}\Delta\left(k,d\right) = \sum_{d} u\left(d|k\right)\left|\Gamma\left(k,c\right)\right|.\end{aligned}$$

where the last equality comes from the fact that $\sum_d \Delta\left(k,d\right) = 0$. Thus, we obtain

$$\begin{aligned}\left\|u.q_2 - q_1.v\right\| &\geq \sum_{k,c,d} u\left(k\right)\left|u\left(d|k\right)\Gamma\left(k,c\right)\right| \\ &= \sum_{k,c,d} u\left(k\right)\left|u\left(d|k\right) u\left(c|k\right) - \sum_{\alpha} u\left(d|k\right) v\left(\alpha|k\right) q_1\left(c|\alpha\right)\right| = \left\|u - q_1.v\right\|.\end{aligned}$$

We deduce that $\min_{q_1,q_2}\left\|u.q_2 - q_1.v\right\| = \min_{q_1}\left\|u - q_1.v\right\|$. Inverting the roles of the players, we also have $\min_{q_1,q_2}\left\|v.q_2 - q_1.y\right\| = \min_{q_1}\left\|v - q_1.u\right\|$. We conclude that

$$\begin{aligned}\mathbb{d}(u,v) &= \max\{\min_{q_1,q_2}\left\|u.q_2 - q_1.v\right\|; \min_{q_1,q_2}\left\|v.q_2 - q_1.y\right\|\} \\ &= \max\{\min_{q_1}\left\|u - q_1.v\right\|; \min_{q_1}\left\|v - q_1.u\right\|\} = \mathbb{d}_1(u,v),\end{aligned}$$

where the last equality follows from (9) together with the fact that $\operatorname{marg}_{K\times D} u = \operatorname{marg}_{K\times D} v$.

B.3. *Proof of Proposition 3*

Because $u \succeq v$,

$$\mathbb{d}\left(u,v\right)=\min_{q_2\in\mathcal{Q}}\min_{q_1\in\mathcal{Q}}\left\|u.q_2-q_1.v\right\|\leq\min_{q_2\in\mathcal{Q}}\min_{q_1:C\to\Delta(C\times C_2)}\left\|u.q_2-\hat{q}_1.v\right\|,$$

where in the right-hand side of the inequality, we use a restricted set of player 1's garblings. Precisely, for every garbling $q_1 : C \to \Delta\left(C \times C_2\right)$, we associate the garbling $\hat{q}_1$ defined by $\hat{q}_1(c', c_1', c_2'|c, c_1) = \mathbb{1}_{\{c_1\}}(c_1')q_1(c', c_2'|c)$. Further, for any such q_1 and an arbitrary garbling q_2, we have

$$\left\|u.q_2-\hat{q}_1.v\right\|=\sum_{k,c,c_1,c_2,d}\left|\sum_{\beta}u\left(k,c,c_1,c_2,\beta\right)q_2\left(d|\beta\right)-\sum_{\alpha}u\left(k,\alpha,c_1,d\right)q_1\left(c,c_2|\alpha\right)\right|$$

$$=\sum_{k,c,c_1,c_2,d}u\left(k,c_1\right))\left|\sum_{\beta}u\left(c,c_2,\beta|k,c_1\right)q_2\left(d|\beta\right)-\sum_{\alpha}u\left(\alpha,d|k,c_1\right)q_1\left(c,c_2|\alpha\right)\right|.$$

Because of the conditional independence assumption, the above is equal to

$$=\sum_{k,c,c_2,d}\left(\sum_{c_1}u\left(k,c_1\right)\right)\left|\sum_{\beta}u\left(c,c_2,\beta|k\right)q_2\left(d|\beta\right)-\sum_{\alpha}u\left(\alpha,d|k\right)q_1\left(c,c_2|\alpha\right)\right|$$

$$=\sum_{k,c,c_2,d}\left|\sum_{\beta}u\left(k,c,c_2,\beta\right)q_2\left(d|\beta\right)-\sum_{\alpha}u\left(k,\alpha,d\right)q_1\left(c,c_2|\alpha\right)\right|=\left\|u'.q_2-q_1.v'\right\|.$$

Hence $\mathbb{d}\left(u,v\right)\leq\min_{q_2}\min_{q_1:C\to\Delta(C\times C_2)}\left\|u'.q_2-q_1.v'\right\|=\mathbb{d}\left(u',v'\right)$.

B.4. *Proof of Proposition 4*

We have $\mathbb{d}\left(u',v'\right)=\mathbb{d}_1\left(u',v'\right)=\mathbb{d}_1\left(u,v\right)\leq\mathbb{d}\left(u,v\right)$. The first equality comes from Proposition 2, the second from the fact that u and u' (resp. v and v') induce the same distribution on player 1 first order beliefs, and the inequality from the definition of the two distances.

B.5. *Proof of Proposition 5*

It is enough to show that if c_1 is ε-conditionally independent from (k, d) given c, then $\sup_{g\in\mathcal{G}}\mathrm{val}\left(u,g\right)-\mathrm{val}\left(v,g\right)\leq\varepsilon$.

For this, let $q_2 : D \times D_1 \to D$ be defined as $q_2(d, d_1)(d') = \mathbb{1}_{d'=d}$. Let $q_1 : C \to C \times C_1$ be defined as $q_1(c, c_1|c) = u(c_1|c)$. Then,

$$\begin{aligned} \|u.q_2 - q_1.v\| &= \sum_{k,c,c_1,d} |u(k, c, c_1, d) - u(k, c, d) u(c_1|c)| \\ &= \sum_{c} u(c) \sum_{k,c_1,d} |u(k, c_1, d|c) - u(k, d|c) u(c_1|c)| \le \varepsilon. \end{aligned}$$

The claim follows from Theorem 1.

APPENDIX C: PROOF OF THEOREM 2

N is a very large even integer to be fixed later, and we write $A = C = D = \{1, ..., N\}$, with the idea of using C while speaking of the actions or signals of player 1 and using D while speaking of the actions and signals of player 2. We fix ε and α, to be used later, such that $0 < \varepsilon < \frac{1}{10(N+1)^2}$ and $\alpha = \frac{1}{25}$. We will consider a Markov chain with law ν on A, satisfying the following:

- the law of the first state of the Markov chain is uniform on A,
- given the current state, the law of the next state is uniform on a subset of size $N/2$,
- and few more conditions, to be defined later.

A sequence $(a_1, ..., a_l)$ of length $l \ge 1$ is said to be *nice* if it is in the support of the Markov chain: $\nu(a_1, ..., a_l) > 0$. For instance, any sequence of length 1 is nice, and $N^2/2$ sequences of length 2 are nice.

The rest of the proof is split in 3 parts: we first define the information structures $(u^l)_{l\ge1}$ and some payoff structures $(g^p)_{p\ge1}$. Then we define two conditions $UI1$ and $UI2$ on the information structures and show that they imply the conclusions of Theorem 2. Finally, we show, via the probabilistic method, the existence of a Markov chain ν satisfying all our conditions.

C.1. *Information and payoff structures* $(u^l)_{l\geq 1}$ *and* $(g^l)_{l\geq 1}$

For $l \geq 1$, define the information structure $u^l \in \Delta(K \times C^l \times D^l)$ so that for each state k in K, signal $c = (c_1, ..., c_l)$ in C^l of player 1 and signal $d = (d_1, ..., d_l)$ in D^l for player 2,

$$u^l(k, c, d) = \nu(c_1, d_1, c_2, d_2, ..., c_l, d_l) \left(\frac{c_1}{N+1} \mathbf{1}_{k=1} + \left(1 - \frac{c_1}{N+1}\right) \mathbf{1}_{k=0} \right).$$

The following interpretation of u^l holds: first select $(a_1, a_2, ..., a_{2l}) = (c_1, d_1, ..., c_l, d_l)$ in A^{2l} according to the Markov chain ν (i.e., uniformly among the nice sequences of length $2l$), then tell $(c_1, c_2, ..., c_l)$ (the elements of the sequence with odd indices) to player 1 and $(d_1, d_2, ..., d_l)$ (the elements of the sequence with even indices) to player 2. Finally, choose the state $k = 1$ with probability $c_1/(N+1)$ and state $k = 0$ with the complement probability $1 - c_1/(N+1)$.

Notice that the definition is not symmetric among players: the first signal c_1 of player 1 is uniformly distributed and plays a particular role. The marginal of u^l on K is uniform, and the marginal of u^{l+1} over $(K \times C^l \times V^l)$ is equal to u^l.

Consider a sequence $(a_1, ..., a_l)$ of elements of A that is not nice (i.e., such that $\nu(a_1, ..., a_l) = 0$). We say that the sequence is *not nice because of player 1* if $\min\{t \in \{1, ..., l\}, \nu(a_1, ..., a_t) = 0\}$ is odd and *not nice because of player 2* if $\min\{t \in \{1, ..., l\}, \nu(a_1, ..., a_t) = 0\}$ is even. A sequence $(a_1, ..., a_l)$ is now nice, or not nice because of player 1, or not nice because of player 2. A sequence of length 2 is either nice, or not nice because of player 2.

For $p \geq 1$, define the payoff structure $g^p : K \times C^p \times D^{p-1} \to [-1, 1]$ such that for all k in K, $c' = (c'_1, ..., c'_p)$ in C^p, $d' = (d'_1, ..., d'_{p-1})$ in D^{p-1} :

$$g^p(k, c', d') = g_0(k, c'_1) + h^p(c', d'), \text{ where } g_0(k, c'_1) = -\left(k - \frac{c'_1}{N+1}\right)^2 + \frac{N+2}{6(N+1)},$$

$$h^p(c', d') = \begin{cases} \varepsilon & \text{if } (c'_1, d'_1, ..., c'_p) \text{ is nice,} \\ 5\varepsilon & \text{if } (c'_1, d'_1, ..., c'_p) \text{ is not nice because of player 2,} \\ -5\varepsilon & \text{if } (c'_1, d'_1, ..., c'_p) \text{ is not nice because of player 1.} \end{cases}$$

One can check that $|g^p| \leq 5/6 + 5\varepsilon \leq 8/9$. Regarding the g_0 part of the payoff, consider a decision problem for player 1 where c_1 is selected uniformly in A and the state is selected to be $k = 1$ with probability $c_1/(N+1)$ and $k = 0$ with probability $1 - c_1/(N+1)$. Player 1 observes c_1 but not k, and she chooses c'_1 in A and receives payoff $g_0(k, c'_1)$. We have $\frac{c_1}{N+1} g_0(1, c'_1) + (1 - \frac{c_1}{N+1}) g_0(0, c'_1) = \frac{1}{(N+1)^2}(c'_1(2c_1 - c'_1) + (N+1)((N+2)/6 - c_1))$. To maximize this expected payoff, it is well known that player 1 should play her belief on k, i.e. $c'_1 = c_1$. Moreover, if player 1 chooses $c'_1 \neq c_1$, her expected loss from not having chosen c_1 is at least $\frac{1}{(N+1)^2} \geq 10\varepsilon$. And the constant $\frac{N+2}{6(N+1)}$ has been chosen such that the value of this decision problem is 0.

Consider now $l \geq 1$ and $p \geq 1$. By definition, the Bayesian game $\Gamma(u^l, g^p)$ is played as follows: first, $(c_1, d_1, ..., c_l, d_l)$ is selected according to the law ν of the Markov chain, player 1 learns $(c_1, ..., c_l)$, player 2 learns $(d_1, ..., d_l)$, and the state is $k = 1$ with probability $c_1/(N+1)$ and $k = 0$ otherwise. Then, *simultaneously* player 1 chooses c' in C^p and player 2 chooses d' in D^{p-1}, and finally, the payoff to player 1 is $g^p(k, c', d')$. Notice that by the previous paragraph about g_0, it is always strictly dominant for player 1 to report correctly her first signal, i.e. to choose $c'_1 = c_1$. We will show in the next section that if $l \geq p$ and player 1 simply plays the sequence of signals she received, player 2 cannot do better than also truthfully reporting his own signals, leading to a value not lower than the payoff for nice sequences, which is ε. On the contrary, in the game $\Gamma(u^l, g^{l+1})$, player 1 has to report not only the l signals she has received but also an extra-signal c'_{l+1} that she has to guess. In this game, we will prove that if player 2 truthfully reports his own signals, player 1 will incur the payoff -5ε with a probability of at least (approximately) $1/2$, and this will result in a low value. These intuitions will prove correct in the next section, under some conditions $UI1$ and $UI2$.

C.2. *Conditions UI and values*

To prove that the intuitions of the previous paragraph are correct, we need to ensure that players have incentives to report their true signals, so we need additional assumptions on the Markov chain.

Notations and definition: Let $l \geq 1$, $m \geq 0$, $c = (c_1, ..., c_l)$ in C^l and $d = (d_1, ..., d_m)$ in D^m. We write

$$\begin{array}{lcll} a^{2q}(c,d) & = & (c_1, d_1, ..., c_q, d_q) \in A^{2q} & \text{for each } \; q \leq \min\{l, m\}, \\ a^{2q+1}(c,d) & = & (c_1, d_1, ..., c_q, d_q, c_{q+1}) \in A^{2q+1} & \text{for each } q \leq \min\{l-1, m\}. \end{array}$$

For $r \leq \min\{2l, 2m+1\}$, we say that c and d are *nice at level* r, and we write $c \smile_r d$, if $a^r(c,d)$ is nice.

In the next definition, we consider an information structure $u^l \in \Delta(K \times C^l \times D^l)$ and denote by $\tilde{c}$ and $\tilde{d}$ the respective random variables of the signals of player 1 and 2.

DEFINITION 1 *We say that the* conditions $UI1$ are satisfied *if for all* $l \geq 1$, *all* $c = (c_1, ..., c_l)$ *in* C^l *and* $c' = (c'_1, ..., c'_{l+1})$ *in* C^{l+1} *such that* $c_1 = c'_1$, *we have*

$$u^l\left(c' \smile_{2l+1} \tilde{d} \;\middle|\; \tilde{c} = c, c' \smile_{2l} \tilde{d}\right) \in [1/2 - \alpha, 1/2 + \alpha] \tag{10}$$

and for all $m \in \{1, ..., l\}$ *such that* $c_m \neq c'_m$, *for* $r = 2m-2, 2m-1$,

$$u^l\left(c' \smile_{r+1} \tilde{d} \;\middle|\; \tilde{c} = c, c' \smile_{r} \tilde{d}\right) \in [1/2 - \alpha, 1/2 + \alpha]. \tag{11}$$

We say that the *conditions* $UI2$ *are satisfied* if for all $1 \leq p \leq l$, for all $d \in D^l$, for all $d' \in D^{p-1}$, for all $m \in \{1, ..., p-1\}$ such that $d_m \neq d'_m$, for

$r = 2m-1, 2m$

$$u^l \left(\tilde{c} \smile_{r+1} d' | \tilde{d} = d, \tilde{c} \smile_r d' \right) \in [1/2-\alpha, 1/2+\alpha]. \tag{12}$$

To understand the conditions $UI1$, consider the Bayesian game $\Gamma(u^l, g^{l+1})$, and assume that player 2 truthfully reports his sequence of signals and that player 1 has received the signals $(c_1, ..., c_l)$ in C^l. (10) states that if the sequence of reported signals $(c'_1, \tilde{d}_1, ..., c'_l, \tilde{d}_l)$ is nice at level $2l$, then whatever the last reported signal c'_{l+1} is, the conditional probability that $(c'_1, \tilde{d}_1, ..., c'_l, \tilde{d}_l, c'_{l+1})$ is still nice is in $[1/2-\alpha, 1/2+\alpha]$, (i.e., close to $1/2$). Regarding (11), first notice that if $c' = c$, then by construction $(c'_1, \tilde{d}_1, ..., c'_l, \tilde{d}_l)$ is nice and $u^l \left(c' \smile_{r+1} \tilde{d} \mid \tilde{c} = c, c' \smile_r \tilde{d} \right) = u^l \left(c \smile_{r+1} \tilde{d} \mid \tilde{c} = c \right) = 1$ for each $r = 1, ..., 2l-1$. Assume now that for some $m = 1, ..., l$, player 1 misreports her m^{th}-signal (i.e., reports $c'_m \neq c_m$). (11) requires that given that the reported signals were nice so far (at level $2m-2$), the conditional probability that the reported signals are not nice at level $2m-1$ (integrating c'_m) is close to $1/2$, and moreover, if the reported signals are nice at this level $2m-1$, adding the next signal $\tilde{d}_m$ of player 2 has a probability close to $1/2$ of keeping the reported sequence nice. Conditions $UI2$ have a similar interpretation, considering the Bayesian games $\Gamma(u^l, g^p)$ for $p \leq l$, assuming that player 1 truthfully reports her signals and that player 2 plays d' after having received the signals d.

Proposition 7 *Conditions $UI1$ and $UI2$ imply*

$$\forall l \geq 1, \forall p \in \{1, ..., l\}, \qquad \mathrm{val}(u^l, g^p) \geq \varepsilon. \tag{13}$$

$$\forall l \geq 1, \quad \mathrm{val}(u^l, g^{l+1}) \leq -\varepsilon. \tag{14}$$

As a consequence of this proposition, under the existence of a Markov chain satisfying conditions $UI1$ and $UI2$, we obtain Theorem 2:

$$\text{If } l > p, \text{ then } d(u^l, u^p) \geq \mathrm{val}(u^l, g^{p+1}) - \mathrm{val}(u^p, g^{p+1}) \geq 2\varepsilon.$$

Proof of proposition 7. We assume that $UI1$ and $UI2$ hold. We fix $l \geq 1$, work on the probability space $K \times C^l \times D^l$ equipped with the probability u^l, and denote by $\tilde{c}$ and $\tilde{d}$ the random variables of the signals received by the players.

1) We first prove (13). Consider the game $\Gamma(u^l, g^p)$ with $p \in \{1, ..., l\}$. We assume that player 1 chooses the truthful strategy. Fix $d = (d_1, ..., d_l)$ in D^l and $d' = (d'_1, ..., d'_{p-1})$ in D^{p-1}, and assume that player 2 has received the signal d and chooses to report d'. Define the non-increasing sequence of events: $A_n = \{\tilde{c} \smile_n d'\}$. We will prove by backward induction that

$$(15) \qquad \forall n = 1, ..., p, \;\; \mathbb{E}[h^p(\tilde{c}, d')|\tilde{d} = d, A_{2n-1}] \geq \varepsilon.$$

If $n = p$, $h^p(\tilde{c}, d') = \varepsilon$ on the event A_{2p-1}, implying the result. Assume now that for some n such that $1 \leq n < p$, we have $\mathbb{E}[h^p(\tilde{c}, d')|\tilde{d} = d, A_{2n+1}] \geq \varepsilon$. Since we have a non-increasing sequence of events, $\mathbb{1}_{A_{2n-1}} = \mathbb{1}_{A_{2n+1}} + \mathbb{1}_{A_{2n-1}} \mathbb{1}_{A^c_{2n}} + \mathbb{1}_{A_{2n}} \mathbb{1}_{A^c_{2n+1}}$, so by definition of the payoffs, $h^p(\tilde{c}, d') \mathbb{1}_{A_{2n-1}} = h^p(\tilde{c}, d') \mathbb{1}_{A_{2n+1}} + 5\varepsilon \mathbb{1}_{A_{2n-1}} \mathbb{1}_{A^c_{2n}} - 5\varepsilon \mathbb{1}_{A_{2n}} \mathbb{1}_{A^c_{2n+1}}$.

First assume that $d'_n = d_n$. By construction of the Markov chain, $u^l(A_{2n+1}|A_{2n-1}, \tilde{d} = d) = 1$, implying that $u^l(A^c_{2n+1}|A_{2n-1}, \tilde{d} = d) = u^l(A^c_{2n}|A_{2n-1}, \tilde{d} = d) = 0$. As a consequence,

$$\begin{aligned}
\mathbb{E}[h^p(\tilde{c}, d')|\tilde{d} = d, A_{2n-1}] &= \mathbb{E}[h^p(\tilde{c}, d') \mathbb{1}_{A_{2n+1}} |\tilde{d} = d, A_{2n-1}] \\
&= \mathbb{E}[\mathbb{E}[h^p(\tilde{c}, d')|\tilde{d} = d, A_{2n+1}] \mathbb{1}_{A_{2n+1}} |\tilde{d} = d, A_{2n-1}] \geq \varepsilon.
\end{aligned}$$

Assume now that $d'_n \neq d_n$. Assumption UI2 implies that

$$\begin{aligned}
u^l(A^c_{2n}|A_{2n-1}, \tilde{d} = d) &\geq 1/2 - \alpha, \\
u^l(A_{2n} \cap A^c_{2n+1}|A_{2n-1}, \tilde{d} = d) &\leq (1/2 + \alpha)^2, \\
u^l(A_{2n+1}|A_{2n-1}, \tilde{d} = d) &\geq (1/2 - \alpha)^2.
\end{aligned}$$

It follows that

$$
\begin{aligned}
&\mathbb{E}[h^p(\tilde{c}, d'|\tilde{d}) = d, A_{2n-1}]\\
=&\mathbb{E}[\mathbb{E}[h^p(\tilde{c}, d')|\tilde{d} = d, A_{2n+1}]\, \mathbb{1}_{A_{2n+1}} | \tilde{d} = d, A_{2n-1}]\\
&+ 5\varepsilon u^l(A^c_{2n}|A_{2n-1}, \tilde{d} = d) - 5\varepsilon u^l(A_{2n} \cap A^c_{2n+1}|A_{2n-1}, \tilde{d} = d)\\
\geq&\varepsilon(\frac{1}{4} - \alpha + \alpha^2) + 5\varepsilon(\frac{1}{2} - \alpha) - 5\varepsilon(\frac{1}{4} + \alpha + \alpha^2) = \varepsilon(\frac{3}{2} - 11\alpha - 4\alpha^2) \geq \varepsilon,
\end{aligned}
$$

and (15) follows by backward induction.

Since A_1 is an event that holds almost surely, we deduce that $\mathbb{E}[h^p(\tilde{c}, d')|\tilde{d} = d] \geq \varepsilon$. Hence the truthful strategy of player 1 guarantees the payoff ε in $\Gamma(u^l, g^p)$.

2) We now prove (14). Consider the game $\Gamma(u^l, g^{l+1})$. We assume that player 2 chooses the truthful strategy. Fix $c = (c_1, ..., c_l)$ in C^l and $c' = (c'_1, ..., c'_{l-1})$ in C^{l-1}, and assume that player 1 has received the signal c and chooses to report c'. We will show that the expected payoff of player 1 is not larger than $-\varepsilon$, and assume w.l.o.g. that $c'_1 = c_1$. Consider the non-increasing sequence of events $B_n = \{c' \smile_n \tilde{d}\}$. We will prove by backward induction that $\forall n = 1, ..., l,\ \ \mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2n}] \leq -\varepsilon$.

If $n = l$, we have $\mathbb{1}_{B_{2l}} = \mathbb{1}_{B_{2l+1}} + \mathbb{1}_{B_{2l}}\, \mathbb{1}_{B^c_{2l+1}}$, and $h^{l+1}(c', \tilde{d})\, \mathbb{1}_{B_{2l}} = \varepsilon\, \mathbb{1}_{B_{2l+1}} - 5\varepsilon\, \mathbb{1}_{B_{2l}}\, \mathbb{1}_{B^c_{2l+1}}$. UI1 implies that $|u^l(B_{2l+1}|\tilde{c} = c, B_{2l}) - \frac{1}{2}| \leq \alpha$, and it follows that

$$
\begin{aligned}
\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2l}] &= \varepsilon\, u^l(B_{2l+1}|\tilde{c} = c, B_{2l}) - 5\varepsilon\, u^l(B^c_{2l+1}|u = \hat{u}, B_{2l})\\
&\leq \varepsilon\,(\frac{1}{2} + \alpha) - 5\varepsilon\,(\frac{1}{2} - \alpha) \leq -\varepsilon.
\end{aligned}
$$

Assume now that for some $n = 1, ..., l-1$, we have $\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2n+2}] \leq -\varepsilon$. We have $\mathbb{1}_{B_{2n}} = \mathbb{1}_{B_{2n+2}} + \mathbb{1}_{B_{2n}}\, \mathbb{1}_{B^c_{2n+1}} + \mathbb{1}_{B_{2n+1}}\, \mathbb{1}_{B^c_{2n+2}}$, and by definition of h^{l+1},

$$
h^{l+1}(c', \tilde{d})\, \mathbb{1}_{B_{2n}} = h^{l+1}(c', \tilde{d})\, \mathbb{1}_{B_{2n+2}} - 5\varepsilon\, \mathbb{1}_{B_{2n}}\, \mathbb{1}_{B^c_{2n+1}} + 5\varepsilon\, \mathbb{1}_{B_{2n+1}}\, \mathbb{1}_{B^c_{2n+2}}.
$$

First, assume that $c'_{n+1} = c_{n+1}$, then $u^l(B_{2n+2}|B_{2n}, \tilde{c} = c) = 1$. Then

$$
\begin{aligned}
\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2n}] &= \mathbb{E}[h^{l+1}(c', \tilde{d})\, \mathbb{1}_{B_{2n+2}} | \tilde{c} = c, B_{2n}],\\
&= \mathbb{E}[\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2n+2}]\, \mathbb{1}_{B_{2n+2}} | \tilde{c} = c, B_{2n}] \leq -\varepsilon.
\end{aligned}
$$

Assume on the contrary that $c'_{n+1} \neq c_{n+1}$. Assumption UI1 implies that

$$\begin{aligned} u^l(B^c_{2n+1}|B_{2n}, \tilde{c} = c) &\geq 1/2 - \alpha, \\ u^l(B_{2n+1} \cap B^c_{2n+2}|B_{2n}, \tilde{c} = c) &\leq (1/2 + \alpha)^2, \\ u^l(B_{2n+2}|B_{2n}, \tilde{c} = c) &\geq (1/2 - \alpha)^2. \end{aligned}$$

It follows that

$$\begin{aligned} \mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2n}] = \mathbb{E}[\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_{2n+2}] \, \mathbb{1}_{B_{2n+2}} \, |\tilde{c} = c, B_{2n}] \\ - 5\,\varepsilon\, u^l(B^c_{2n+1}|B_{2n}, \tilde{c} = c) + 5\,\varepsilon\, u^l(B_{2n+1} \cap B^c_{2n+2}|B_{2n}, \tilde{c} = c) \\ \leq -\varepsilon\,(\frac{1}{4} - \alpha + \alpha^2) - 5\,\varepsilon\,(\frac{1}{2} - \alpha) + 5\,\varepsilon\,(\frac{1}{4} + \alpha + \alpha^2) \leq -\varepsilon. \end{aligned}$$

By induction, we obtain $\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c, B_2] \leq -\varepsilon$. Since B_2 holds almost surely here, we get $\mathbb{E}[h^{l+1}(c', \tilde{d})|\tilde{c} = c] \leq -\varepsilon$, showing that the truthful strategy of player 2 guarantees that the payoff of the maximizer is less or equal to $-\varepsilon$, which concludes the proof.

C.3. *Existence of an appropriate Markov chain*

Here we conclude the proof of Theorem 2 by showing the existence of an even integer N and a Markov chain with law ν on $A = \{1, ..., N\}$ satisfying our conditions

1) the law of the first state of the Markov chain is uniform on A,

2) for each a in A, there are exactly $N/2$ elements b in A such that $\nu(b|a) = 2/N$ and

3) $UI1$ and $UI2$.

Denoting by $P = (P_{a,b})_{(a,b)\in A^2}$ the transition matrix of the Markov chain, we have to prove the existence of P satisfying 2) and 3). The proof is nonconstructive and uses the following probabilistic method, where we select independently for each a in A, the set $\{b \in A, P_{a,b} > 0\}$ uniformly among the subsets of A with cardinal $N/2$. We will show that when N goes to infinity, the probability of selecting an appropriate transition matrix become strictly positive and, in fact, converges to 1.

Formally, denote by $\mathcal{S}_A$ the collection of all subsets $S \subseteq A$ with cardinality $|S| = \frac{1}{2}N$. We consider a collection $(S_a)_{a\in A}$ of i.i.d. random variables uniform distributed over $\mathcal{S}_A$ defined on a probability space $(\Omega_N, \mathcal{F}_N, \mathbb{P}_N)$. For all a, b in A, let $X_{a,b} = \mathbb{1}_{\{b\in S_a\}}$ and $P_{a,b} = \frac{2}{N}X_{a,b}$. By construction, P is a transition matrix satisfying 2). Theorem 2 will now follow from the following proposition.

PROPOSITION 8

$$\mathbb{P}_N\left(\ P \textit{ induces a Markov chain satisfying UI1 and UI2}\ \right) \xrightarrow[N\to\infty]{} 1.$$

In particular, this probability is strictly positive for all sufficiently large N.

The rest of this section is devoted to the proof of proposition 8. We start with probability bounds based on Hoeffding's inequality.

LEMMA 1 *For any $a \neq b$, each $\gamma > 0$*

$$\mathbb{P}_N\left(\left||S_a\cap S_b| - \frac{1}{4}N\right| \geq \gamma N\right) \leq \frac{1}{2}e^4 N e^{-2\gamma^2 N}.$$

PROOF: Consider a family of i.i.d. Bernoulli variables $(\widetilde{X}_{i,j})_{i=a,b,\,j\in A}$ of parameter $\frac{1}{2}$ defined on a space $(\Omega, \mathcal{F}, \mathbb{P})$. For $i = a, b$, define the events $\widetilde{L}_i = \{\sum_{j\in A}\widetilde{X}_{i,j} = \frac{N}{2}\}$ and the set-valued variables $\widetilde{S}_i = \{j \in A \,|\, \widetilde{X}_{i,j} = 1\}$. It is straightforward to check that the conditional law of $(\widetilde{S}_a, \widetilde{S}_b)$ given $\widetilde{L}_a \cap \widetilde{L}_b$ under $\mathbb{P}$ is the same as the law of (S_a, S_b) under $\mathbb{P}_N$. It follows that

$$\begin{aligned}\mathbb{P}_N\left(\left||S_a\cap S_b| - \frac{1}{4}N\right| \geq \gamma N\right) &= \mathbb{P}\left(\left||\widetilde{S}_a\cap \widetilde{S}_b| - \frac{1}{4}N\right| \geq \gamma N \,\middle|\, \widetilde{L}_a\cap\widetilde{L}_b\right)\\ &\leq \frac{\mathbb{P}\left(\left||\widetilde{S}_a\cap \widetilde{S}_b| - \frac{1}{4}N\right| \geq \gamma N\right)}{\mathbb{P}\left(\widetilde{L}_a\cap\widetilde{L}_b\right)}.\end{aligned}$$

Using Hoeffding inequality, we have

$$\mathbb{P}\left(\left||\widetilde{S}_a\cap \widetilde{S}_b| - \frac{1}{4}N\right| \geq \gamma N\right) = \mathbb{P}\left(\left|\sum_{j\in A}\widetilde{X}_{a,j}\widetilde{X}_{b,j} - \frac{1}{4}N\right| \geq \gamma N\right) \leq 2e^{-2\gamma^2 N}.$$

On the other hand, using Stirling approximation[14], we have

$$\mathbb{P}\left(\widetilde{L}_a \cap \widetilde{L}_b\right) = \left(\frac{1}{2^N}\frac{N!}{\left(\frac{N}{2}!\right)^2}\right)^2 \geq \left(\frac{2^{N+1}N^{-\frac{1}{2}}}{2^N e^2}\right)^2 = \frac{4}{Ne^4}.$$

We deduce that $\mathbb{P}_N\left(\left||S_a \cap S_b| - \frac{1}{4}N\right| \geq \gamma N\right) \leq \frac{1}{2}e^4 N e^{-2\gamma^2 N}$. *Q.E.D.*

LEMMA 2 *For each $a \neq b$, for any subset $S \subseteq A$ and any $\gamma \geq \frac{1}{2N-2}$,*

$$\mathbb{P}_N\left(\left|\sum_{i\in S} X_{i,a} - \frac{1}{2}|S|\right| \geq \gamma N\right) \leq 2e^{-2N\gamma^2}, \text{ and } \mathbb{P}_N\left(\left|\sum_{i\in S} X_{i,a}X_{i,b} - \frac{1}{4}|S|\right| \geq \gamma N\right) \leq 2e^{-\frac{1}{8}N\gamma^2}.$$

PROOF: For the first inequality, notice that $X_{i,a}$ are i.i.d. Bernoulli random variables with parameter $\frac{1}{2}$. The Hoeffding inequality implies that

$$\mathbb{P}_N\left(\left|\sum_{i\in S} X_{i,a} - \frac{1}{2}|S|\right| \geq \gamma N\right) \leq 2e^{-2\gamma^2\frac{N^2}{|S|}} \leq 2e^{-2N\gamma^2}.$$

Q.E.D.

For the second inequality, let $Z_i = X_{i,a}X_{i,b}$. Notice that all variables Z_i are i.i.d. Bernoulli random variables with parameter $p = \frac{1}{2}\left(\frac{\frac{N}{2}-1}{N-1}\right) = \frac{1}{4} - \frac{1}{4N-4}$. The Hoeffding inequality implies that

$$\mathbb{P}_N\left(\left|\sum_{i\in S} Z_i - \frac{1}{4}|S|\right| \geq \gamma N\right) \leq \mathbb{P}_N\left(\left|\sum_{i\in S} Z_i - p|S|\right| \geq \frac{1}{2}\gamma N\right) \leq 2e^{-2\gamma^2\frac{N^2}{|S|}} \leq 2e^{-\frac{1}{2}N\gamma^2},$$

where we used that $|S||p - \frac{1}{4}| \leq \frac{N}{4N-4} \leq \frac{\gamma N}{2}$ for the first inequality.

For each $a \neq b$ and $c \neq d$, each $\gamma > 0$, define

$$\begin{array}{lll} Y_a = 2\sum_{i\in A} X_{i,a}, & Y^c = 2\sum_{i\in A} X_{c,i} = N, & \\ Y_{a,b} = 4\sum_{i\in A} X_{i,a}X_{i,b}, & Y_a^c = 4\sum_{i\in A} X_{i,a}X_{c,i}, & Y^{c,d} = 4\sum_{i\in A} X_{c,i}X_{d,i}, \\ Y_{a,b}^c = 8\sum_{i\in A} X_{i,a}X_{i,b}X_{c,i}, & Y_a^{c,d} = 8\sum_{i\in A} X_{i,a}X_{c,i}X_{d,i}, & Y_{a,b}^{c,d} = 16\sum_{i\in A} X_{i,a}X_{i,b}X_{c,i}X_{d,i}. \end{array}$$

LEMMA 3 *For each $a \neq b$ and $c \neq d$, each $\gamma \geq 64/N$, each of the variables $Z \in \{Y_a, Y^c, Y_{a,b}, Y^{c,d}, Y_a^c, Y_{a,b}^c, Y_a^{c,d}, Y_{a,b}^{c,d}\}$,*

$$\mathbb{P}_N\left(|Z - N| \geq \gamma N\right) \leq e^4 N e^{-\frac{N}{32}\left(\frac{\gamma}{10}\right)^2}.$$

[14] We have $n^{n+\frac{1}{2}}e^{-n} \leq n! \leq en^{n+\frac{1}{2}}e^{-n}$ for each n.

Proof: In case $Z = Y_a$ or $Y_{a,b}$, the bound follows from Lemma 2 (for $S = A$). If case $Z = Y^c$, the bound is trivially satisfied. If $Z = Y^{c,d}$, the bound follows from Lemma 1.

In case $Z = Y_{a,b}^{c,d}$, notice that $Y_{a,b}^{c,d} = 16 \sum_{i \in S_c \cap S_d} Z_i$ where $Z_i = X_{i,a} X_{i,b}$. All variables Z_i are i.i.d. Bernouilli random variables with parameter $p = \frac{1}{4} - \frac{1}{4N-4}$. Moreover, $\{Z_i\}_{i \neq c,d}$ are independent of $S_c \cap S_d$. Enlarging the probability space, we can construct a new collection of i.i.d. Bernoulli random variables Z_i' such that $Z_i' = Z_i$ for all $i \neq c, d$ and such that $\{(Z_i')_{i \in A}, S_c \cap S_d\}$ are all independent. Then, $\left| Y_{a,b}^{c,d} - 16 \sum_{i \in S_c \cap S_d} Z_i' \right| \leq 32$, and, because $\frac{1}{2}\gamma N \geq 32$, we have

$$\mathbb{P}_N\left(\left| Y_{a,b}^{c,d} - N \right| \geq \gamma N \right) \leq \mathbb{P}_N\left(\left| \sum_{i \in S_c \cap S_d} Z_i' - \frac{1}{16} N \right| \geq \frac{1}{32} \gamma N \right).$$

Define the events

$$A = \left\{ \left| \frac{1}{4} \left| S_c \cap S_d \right| - \frac{N}{16} \right| \geq \frac{1}{160} \gamma N \right\}, \quad B = \left\{ \left| \sum_{i \in S_c \cap S_d} Z_i' - \frac{1}{4} \left| S_c \cap S_d \right| \right| \geq \frac{1}{40} \gamma N \right\}.$$

Then, the probability can be further bounded by

$$\leq \mathbb{P}_N(A) + \mathbb{P}_N(B) \leq \frac{1}{2} e^4 N e^{-2N\left(\frac{1}{40}\gamma\right)^2} + 2e^{-\frac{1}{2}N\left(\frac{1}{40}\gamma\right)^2} \leq e^4 N e^{-\frac{N\gamma^2}{3200}}$$

where the first bound comes from Lemma 1 and the second from the second bound in Lemma 2.

The remaining bounds have proofs similar to (and simpler than) the case $Z = Y_{a,b}^{c,d}$. We omit the details in the interest of space. *Q.E.D.*

Finally, we describe an event E that collects these bounds. Recall that $\alpha = 1/25$, and define for each $a \neq b$ and $c \neq d$,

$$\begin{aligned} E_{a,b,c,d} \;=\; & \left\{ \left| \frac{Y_{a,b}}{Y_a} - 1 \right| \leq 2\alpha \right\} \cap \left\{ \left| \frac{Y_{a,b}^c}{Y_a^c} - 1 \right| \leq 2\alpha \right\} \cap \left\{ \left| \frac{Y_a^{c,d}}{Y_a^c} - 1 \right| \leq 2\alpha \right\} \cap \left\{ \left| \frac{Y_{a,b}^{c,d}}{Y_a^{c,d}} - 1 \right| \leq 2\alpha \right\} \\ & \cap \left\{ \left| \frac{Y^{c,d}}{Y^c} - 1 \right| \leq 2\alpha \right\} \cap \left\{ \left| \frac{Y_a^c}{Y^c} - 1 \right| \leq 2\alpha \right\} \cap \left\{ \left| \frac{Y_a^{c,d}}{Y^{c,d}} - 1 \right| \leq 2\alpha \right\}. \end{aligned}$$

Finally, let $E = \bigcap_{a,b,c,d: a \neq b \text{ and } c \neq d} E_{a,b,c,d}$.

Lemma 4 *We have*

$$\mathbb{P}_N(E) > 1 - 7e^4 N^5 e^{-\frac{N}{2163200}} \xrightarrow[n\to\infty]{} 1.$$

Proof: Take $\gamma = \frac{\alpha}{1+\alpha} = \frac{1}{26}$ and let

$$F_{a,b,c,d} = \bigcap_{Z \in \{Y_a, Y_{a,b}, Y^{c,d}, Y^{c,d}, Y^c_a, Y^c_{a,b}, Y^{c,d}_a, Y^{c,d}_{a,b}\}} \{|Z - N| \leq \gamma N\}.$$

It is easy to see that $F_{a,b,c,d} \subseteq E_{a,b,c,d}$. The probability that $F_{a,b,c,d}$ holds can be bounded from Lemma 3 (as soon as $N \geq \frac{64}{\gamma} = 1664$), as

$$\mathbb{P}_N(F_{a,b,c,d}) \geq 1 - 7e^4 N e^{-\frac{N}{32.(260)^2}}.$$

The result follows since there are fewer than N^4 ways of choosing (a, b, c, d). *Q.E.D.*

Computations using the bound of lemma 4 show that $N = 52.10^6$ is enough to have the existence of an appropriate Markov chain. So one can take $\varepsilon = 3.10^{-17}$ in the statement of Theorem 2. We conclude the proof of proposition 8 by showing that event E implies conditions $UI1$ and $UI2$.

Lemma 5 *If event E holds, then the conditions $UI1, UI2$ are satisfied.*

Proof: We fix the law ν of the Markov chain on A and assume that it has been induced, as explained at the beginning of section C.3, by a transition matrix P satisfying E. For $l \geq 1$, we forget about the state in K and still denote by u^l the marginal of u^l over $C^l \times D^l$. If $c = (c_1, ..., c_l) \in C^l$ and $d = (d_1, ..., d_l) \in D^l$, we have $u^l(c, d) = \nu(c_1, d_1, ..., c_l, d_l)$.

Let us begin with condition UI2, which we recall here: for all $1 \leq p \leq l$, for all $d \in D^l$, for all $d' \in D^{p-1}$, for all $m \in \{1, ..., p-1\}$ such that $d_m \neq d'_m$, for $r = 2m - 1, 2m$,

$$u^l\left(\tilde{c} \smile_{r+1} d' | \tilde{d} = d, \tilde{c} \smile_r d'\right) \in [1/2 - \alpha, 1/2 + \alpha], (12)$$

where $(\tilde{c}, \tilde{d})$ is a random variable selected according to u^l. The quantity $u^l\left(\tilde{c} \smile_{r+1} d' | \tilde{d} = d, \tilde{c} \smile_r d'\right)$ is thus the conditional probability of the event ($\tilde{c}$ and d' are nice at level $r+1$) given that they are nice at level r and that the signal received by player 2 is d. We divide the problem into different cases.

<u>Case $m > 1$ and $r = 2m-1$.</u> The events $\{\tilde{c} \smile_{2m} d'\}$ and $\{\tilde{c} \smile_{2m-1} d'\}$ can be decomposed as follows:

$$\begin{aligned} \{\tilde{c} \smile_{2m-1} d'\} &= \{\tilde{c} \smile_{2m-2} d'\} \cap \{X_{d'_{m-1}, \tilde{c}_m} = 1\}, \\ \{\tilde{c} \smile_{2m} d'\} &= \{\tilde{c} \smile_{2m-2} d'\} \cap \{X_{d'_{m-1}, \tilde{c}_m} = 1\} \cap \{X_{\tilde{c}_m, d'_m} = 1\}. \end{aligned}$$

So $u^l\left(\tilde{c} \smile_{2m} d' | \tilde{d} = d, \tilde{c} \smile_{2m-1} d'\right) = u^l\left(X_{\tilde{c}_m, d'_m} = 1 | \tilde{d} = d, \tilde{c} \smile_{2m-1} d'\right)$, and the Markov property gives

$$\begin{aligned} u^l\left(\tilde{c} \smile_{2m} d' | \tilde{d} = d, \tilde{c} \smile_{2m-1} d'\right) &= u^l\left(X_{\tilde{c}_m, d'_m} = 1 | X_{d'_{m-1}, \tilde{c}_m} = 1, X_{d_{m-1}, \tilde{c}_m} = 1, X_{\tilde{c}_m, d_m} = 1\right) \\ &= \frac{\sum_{i \in U} X_{i, d'_m} X_{d'_{m-1}, i} X_{d_{m-1}, i} X_{i, d_m}}{\sum_{i \in U} X_{d'_{m-1}, i} X_{d_{m-1}, i} X_{i, d_m}}. \end{aligned}$$

This is equal to $\frac{1}{2}\frac{Y^{d_{m-1}, d'_{m-1}}_{d_m, d'_m}}{Y^{d_{m-1}, d'_{m-1}}_{d_m}}$ if $d'_{m-1} \neq d_{m-1}$, and to $\frac{1}{2}\frac{Y^{d_{m-1}}_{d_m, d'_m}}{Y^{d_{m-1}}_{d_m}}$ if $d'_{m-1} = d_{m-1}$. In both cases, E implies (12).

<u>Case $r = 2m$.</u>

We have $u^l\left(\tilde{c} \smile_{2m+1} d' | \tilde{d} = d, \tilde{c} \smile_{2m} d'\right) = u^l\left(X_{d'_m, \tilde{c}_{m+1}} = 1 | \tilde{d} = d, \tilde{c} \smile_{2m} d'\right)$, and by the Markov property

$$\begin{aligned} & u^l\left(\tilde{c} \smile_{2m+1} d' | \tilde{d} = d, \tilde{c} \smile_{2m} d'\right) \\ = \;& u^l\left(X_{d'_m, \tilde{c}_{m+1}} = 1 | X_{d_m, \tilde{c}_{m+1}} = 1, X_{\tilde{c}_{m+1}, d_{m+1}} = 1\right) \\ = \;& \frac{\sum_{i \in U} X_{d'_m, i} X_{d_m, i} X_{i, d_{m+1}}}{\sum_{i \in U} X_{d_m, i} X_{i, d_{m+1}}} = \frac{1}{2}\frac{Y^{d'_m, d_m}_{d_{m+1}}}{Y^{d_m}_{d_{m+1}}} \in [1/2 - \alpha, 1/2 + \alpha]. \end{aligned}$$

Case $m=1$, $r=1$.

$$\begin{aligned} & u^l\left(\tilde{c}\smile_2 d'|\tilde{d}=d,\tilde{c}\smile_1 d'\right) \\ = \; & u^l\left(\tilde{c}\smile_2 d'|\tilde{d}=d\right) = u^l\left(X_{\tilde{c}_1,d_1'}=1|X_{\tilde{c}_1,d_1}=1\right), \\ = \; & \frac{\sum_{i\in U} X_{i,d_1'}X_{i,d_1}}{\sum_{i\in U} X_{i,d_1}} = \frac{1}{2}\frac{Y_{d_1,d_1'}}{Y_{d_1}} \in [1/2-\alpha, 1/2+\alpha]. \end{aligned}$$

The proof of condition $UI1$ being similar, it is omitted here. *Q.E.D.*

APPENDIX D: PROOFS OF THEOREM 3

D.1. *Theorem 3: the weak topology is contained in the value-based topology*

Assume that $u_n \in \Delta(K\times C_n\times D_n)$ and $u\in\Delta(K\times C\times D)$ are information structures such that $\mathbb{d}(u_n,u)\to 0$. Then, for all games g in $\mathcal{G}$, $|\operatorname{val}(\tilde{u_n},g)-\operatorname{val}(\tilde{u},g)| = |\operatorname{val}(u_n,g)-\operatorname{val}(u,g)| \to 0$. By Theorem 12 in Gossner and Mertens (2001), the functions $(\operatorname{val}(.,g))_g$ span the topology on Π. So $(\tilde{u}_n)_n$ converges weakly to $\tilde{u}$.

D.2. *Theorem 3: the value-based topology is contained in the weak topology*

Assume that $u_n \in \Delta(K\times C_n\times D_n)$ and $u\in\Delta(K\times C\times D)$ are information structures such that $\tilde{u}_n$ converges to $\tilde{u}$ in the weak topology. We will prove that

$$\limsup_{n\to\infty}\sup_{g\in\mathcal{G}}\left(\operatorname{val}(u_n,g)-\operatorname{val}(u,g)\right)\le 0. \tag{16}$$

Because we can switch the roles of players, this will suffice to establish that $\mathbb{d}(u_n,u)\to 0$.

Partitions of unity. We can without loss of generality assume that u is non-redundant and all signals c and d have positive probability. We can associate signals $c\in C\subseteq\mathbb{N}$ and $d\in D\subseteq\mathbb{N}$ with the corresponding hierarchies of beliefs in Θ_1 and Θ_2. In other words, we identify $C\subseteq\Theta_1$ as the (countable) support of $\tilde{u}$ and $D\subseteq\Theta_2$ as the smallest countable set such that for each $c \in C$, $\phi_1 (K \times D|c) = 1$ (i.e., D is the union of countable supports of all beliefs of hierarchies in C). For each $c \in C$ and $d \in D$, we denote the corresponding hierarchies under u as $\tilde{c}$ and $\tilde{d}$. Also, let $C^m = C \cap \{0, ..., m\}$ and $D^m = D \cap \{0, ..., m\}$.

Because Θ_2 is Polish, for each $m \in \mathbb{N}$ and each $d \in D^m$, we can find continuous functions $\kappa_d^m : \Theta_2 \to [0, 1]$ for $m \in \mathbb{N}, d \in \{0, ..., m\}$ such that $\kappa_d^m \left(\tilde{d}\right) = 1$ for each $d \in D^m$, $\kappa_d^m \equiv 0$ if $d \notin D$, and $\sum_{d=0}^m \kappa_d^m (\theta_2) = 1$ for each $\theta_2 \in \Theta_2$. In other words, for each m, $\{\kappa_d^m\}_{0 \le d \le m}$ is a continuous partition of unity on space Θ_2 with the property that for each $d \in D^m$, κ_d^m peaks at hierarchy $\tilde{d}$. Notice that for each $c \in C$, each $d \in D^p$, we have $\mathbb{E}_{\phi_1(\tilde{c})}[\mathbb{1}_{\{k\}}(.)\kappa_d^p(.)] \ge u (k, d|c)$, and

$$\sum_{k \in K} \sum_{d=0}^{p} \left| \mathbb{E}_{\phi_1(\tilde{c})}[\mathbb{1}_{\{k\}}(.)\kappa_d^p(.)] - u (k, d|c) \right| = u(D \setminus D^p|c).$$

Because all hierarchies $\tilde{c}, c \in C$ are distinct, for each m, there exists $p^m < \infty$ and $\varepsilon^m \in \left(0, \frac{1}{m}\right)$ such that for any $c, c' \in C^m$ such that $c \ne c'$,

$$\sum_{k \in K} \sum_{d=0}^{p^m} \left| \mathbb{E}_{\phi_1(\tilde{c})}[\mathbb{1}_{\{k\}}\, \kappa_d^{p^m}] - \mathbb{E}_{\phi_1(\tilde{c}')}[\mathbb{1}_{\{k\}}\, \kappa_d^{p^m}] \right| \ge 2\varepsilon^m.$$

$$\text{Let } \; h_c^m (\theta_1) = \sum_{k} \sum_{d=0}^{p^m} \left| \mathbb{E}_{\phi_1(\theta_1)}[\mathbb{1}_{\{k\}}\, \kappa_d^{p^m}] - \mathbb{E}_{\phi_1(\tilde{c})}[\mathbb{1}_{\{k\}}\, \kappa_d^{p^m}] \right|.$$

Then, h_c^m is a continuous function such that $h_c^m (\tilde{c}) = 0$ and such that if $h_c^m (\theta_1) \le \varepsilon^m$ for some $c \in C^m$, then $h_{c'}^m (\theta_1) \ge \varepsilon^m$ for any $c' \in C^m$ such that $c' \ne c$. For $0 \le c \le m + 1$, define continuous functions

$$\kappa_c^m (\theta_1) = \max \left(1 - \frac{1}{\varepsilon^m} h_c^m (\theta_1), 0\right) \text{ for } c \in C_m,$$

$$\kappa_c^m \equiv 0 \text{ if } c \notin C, \text{ and } \kappa_{m+1}^m (\theta_1) = 1 - \sum_{c=0}^{m} \kappa_c^m (\theta_1).$$

Then, for each m, $\sum_{c=0}^{m+1} \kappa_c^m \equiv 1$, and $\kappa_c^m (\theta_1) \in [0, 1]$ for each $c = 0, ..., m+1$, which implies that $\{\kappa_c^m\}_{0 \le c \le m+1}$ is a continuous partition of unity on space Θ_1 such that for each $c \in C^m$, $\kappa_c^m(\tilde{c}) = 1$.

Conditional independence. For each information structure $v \in \Delta(K \times C' \times D')$, define an information structure $K^m v \in \Delta(K \times C' \times \{0, ..., m+1\} \times D' \times \{0, ..., p^m\})$ so that $K^m v\left(k, c', \hat{c}, d', \hat{d}\right) = v(k, c', d')\, \kappa^m_{\hat{c}}(\tilde{c}')\, \kappa^{p^m}_{\hat{d}}\left(\tilde{d}'\right)$. Let $\delta^m v = 2\varepsilon^m + K^m v(\hat{c} = m+1)$. We are going to show that, under $K^m v$, signal c' is $\delta^m v$-conditionally independent from $\left(k, \hat{d}\right)$ given $\hat{c}$. Notice first that, if $K^m v\left(k, d', \hat{d}, c', \hat{c}\right) > 0$ for some $\hat{c} \in C^m$, then $h^m_{\hat{c}}(\tilde{c}') \leq \varepsilon^m$. It follows that

$$
\begin{aligned}
&\sum_k \sum_{\hat{d}=0}^{p^m} \left| K^m v\left(k, \hat{d} | \hat{c}, c'\right) - \mathbb{E}_{\phi_1\left(\tilde{\hat{c}}\right)}[\mathbb{1}_{\{k\}}\, \kappa^{p^*}_{\hat{d}}] \right| \\
&= \sum_k \sum_{\hat{d}=0}^{p^m} \left| K^m v\left(k, \hat{d} | c'\right) - \mathbb{E}_{\phi_1\left(\tilde{\hat{c}}\right)}[\mathbb{1}_{\{k\}}[\kappa^{p^m}_{\hat{d}}] \right| \\
&= \sum_k \sum_{\hat{d}=0}^{p^m} \left| \mathbb{E}_{\phi_1(\tilde{c}')}[\mathbb{1}_{\{k\}}\, \kappa^{p^m}_{\hat{d}}] - \mathbb{E}_{\phi_1\left(\tilde{\hat{c}}\right)}[\mathbb{1}_{\{k\}}\, \kappa^{p^m}_{\hat{d}}] \right| = h^m_{\hat{c}}(\tilde{c}') \leq \varepsilon^m.
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
&\sum_k \sum_{\hat{d}=0}^{p^m} \left| K^m v\left(k, \hat{d} | \hat{c}\right) - \mathbb{E}_{\phi_1\left(\tilde{\hat{c}}\right)}[\mathbb{1}_{\{k\}}\, \kappa^{p^*}_{\hat{d}}] \right| \\
&= \sum_k \sum_{\hat{d}=0}^{p^m} \left| \frac{1}{K^m v(\hat{c})} \sum_{c' \in C'} K^m v(c', \hat{c}) K^m v\left(k, \hat{d} | \hat{c}, c'\right) - \mathbb{E}_{\phi_1\left(\tilde{\hat{c}}\right)}[\mathbb{1}_{\{k\}}[\kappa^{p^m}_{\hat{d}}] \right| \\
&\leq \sum_{c' \in C'} \frac{K^m v(c', \hat{c})}{K^m v(\hat{c})} \sum_k \sum_{\hat{d}=0}^{p^m} \left| K^m v\left(k, \hat{d} | \hat{c}, c'\right) - \mathbb{E}_{\phi_1\left(\tilde{\hat{c}}\right)}[\mathbb{1}_{\{k\}}[\kappa^{p^m}_{\hat{d}}] \right| = h^m_{\hat{c}}(\tilde{c}') \leq \varepsilon^m.
\end{aligned}
$$

$$
\begin{aligned}
\text{Hence,} \quad &\sum_{\hat{c}=1}^{m+1} \sum_{c'} K^m v(\hat{c}, c') \sum_{k, \hat{d}} \left| K^m v\left(k, \hat{d} | \hat{c}, c'\right) - K^m v\left(k, \hat{d} | \hat{c}\right) \right| \\
&\leq 2\varepsilon^m \sum_{\hat{c}=1}^{m} K^m v(\hat{c}) + K^m v(\hat{c} = m+1) \leq \delta^m v.
\end{aligned}
$$

Define the information structure $L^m v = \operatorname{marg}_{K \times \{0, ..., p^m\} \times \{0, ..., m+1\}} K^m v$. Then, because $d(K^m v, v) = 0$, the proof of Proposition 5 implies that

$\sup_{g\in\mathcal{G}}\left(\mathrm{val}\left(v,g\right)-\mathrm{val}\left(L^{m}v,g\right)\right)\leq\delta^{m}v$.

Proof of claim (16). Observe that for each $k,\hat{c},\hat{d}$,

$$\left(L^{m}u_{n}\right)\left(k,\hat{c},\hat{d}\right)=\mathbb{E}_{\tilde{u}_{n}}\left(\kappa_{\hat{c}}^{m}\left(\theta_{1}\right)\mathbb{E}_{\phi_{1}\left(\theta_{1}\right)}[\mathbb{1}_{\{k\}}\,\kappa_{\hat{d}}^{p^{m}}]\right).$$

Because all the functions in the brackets above are continuous, the weak convergence $\tilde{u}_{n}\to\tilde{u}$ implies that $\left(L^{m}u_{n}\right)\left(k,\hat{c},\hat{d}\right)\to\left(L^{m}u\right)\left(k,\hat{c},\hat{d}\right)$ for each $k,\hat{c},\hat{d}$. Because the information structures $L^{m}u_{n}$ and $L^{m}u$ are described on the same and finite spaces of signals, the pointwise convergence implies $\mathbb{d}\left(L^{m}u_{n},L^{m}u\right)\leq\left\Vert L^{m}u_{n}-L^{m}u\right\Vert\to0$ as $n\to\infty$. Moreover, if $\hat{c}\in C^{m}$ and $\hat{d}\in D^{p^{m}}$, the definitions imply that $\left(L^{m}u\right)\left(k,\hat{c},\hat{d}\right)\geq u\left(k,\hat{c},\hat{d}\right)$. Thus,

$$\mathbb{d}\left(L^{m}u,u\right)\leq\left\Vert L^{m}u-u\right\Vert\leq2\left(u\left(C\backslash C^{m}\right)+u\left(D\backslash D^{p^{m}}\right)\right)\underset{n\to\infty}{\longrightarrow}0.$$

It follows that $\delta^{m}u_{n}=\left(K^{m}u_{n}\right)\left(\hat{c}=m+1\right)\underset{n\to\infty}{\longrightarrow}\left(L^{m}u\right)\left(\hat{c}=m+1\right)$, and

$$\left(L^{m}u\right)\left(\hat{c}=m+1\right)=1-(L^{m}u)(C^{m}\times D^{p^{m}})\leq1-u(C^{m}\times D^{p^{m}})\leq u\left(C\backslash C^{m}\right)+u\left(D\backslash D^{p^{m}}\right).$$

Together, we obtain for each m,n

$$\begin{aligned}\sup_{g\in\mathcal{G}}\left(\mathrm{val}\left(u_{n},g\right)-\mathrm{val}\left(u,g\right)\right)\leq&\sup_{g\in\mathcal{G}}\left(\mathrm{val}\left(u_{n},g\right)-\mathrm{val}\left(L^{m}u_{n},g\right)\right)\\&+\sup_{g\in\mathcal{G}}\left(\mathrm{val}\left(L^{m}u_{n},g\right)-\mathrm{val}\left(L^{m}u\right)\right)+\sup_{g\in\mathcal{G}}\left(\mathrm{val}\left(L^{m}u\right)-\mathrm{val}\left(u,g\right)\right)\\\leq&\delta^{m}u_{n}+\left\Vert L^{m}u_{n}-L^{m}u\right\Vert+\left(u\left(C\backslash C^{m}\right)+u\left(D\backslash D^{p^{m}}\right)\right).\end{aligned}$$

Hence, $\limsup_{n\to\infty}\sup_{g\in\mathcal{G}}\left(\mathrm{val}\left(v,g\right)-\mathrm{val}\left(L^{m}v,g\right)\right)\leq3\left(u\left(C\backslash C^{m}\right)+u\left(D\backslash D^{p^{m}}\right)\right)$. When $m\to\infty$, the right hand side converges to 0 as well.

APPENDIX E: PROOF OF PROPOSITION 6

Let $u'\in\Delta\left(K\times\left(K_{C}\times C\right)\times\left(K_{D}\times D\right)\right)$ be defined so that $u=\mathrm{marg}_{K\times c\times D}\,u'$ and $u'(\{k_{C}=\kappa\left(c\right),k_{D}=\kappa\left(d\right)\})=1$. Because u' does not have any new information, we verify (for instance using Proposition 5) that $\mathbb{d}\left(u,u'\right)=0$.

We are going to show that C is 16ε-conditionally independent from $K \times K_D$ given K_C. Notice that because u exhibits ε-knowledge,

$$\begin{aligned} u'\left\{k_C \neq k \text{ or } k_D \neq k\right\} &\leq u'\left\{k_C \neq k\right\} + u'\left\{k_D \neq k\right\} \\ &\leq 2\varepsilon + 2\varepsilon = 4\varepsilon. \end{aligned}$$

Thus,

$$\begin{aligned} &\sum_{k,k_C,k_D} u'\left(k_C\right) \sum_c \left|u'\left(k,k_D,c|k_C\right) - u'\left(k,k_D|k_C\right) u'\left(c|k_C\right)\right| \\ =&\sum_{k,k_C,k_D} u'\left(k,k_C,k_D\right) \sum_c \left|u'\left(c|k,k_C,k_D\right) - \sum_{k',k_D{}'} u'\left(c|k',k_C,k'_D\right) u'\left(k',k'_D|k_C\right)\right| \\ \leq& \sum_k u'\left(k,k,k\right) \sum_c \left|u'\left(c|k,k,k\right) - \sum_{k',k_D{}'} u'\left(c|k',k_C=k,k'_D\right) u'\left(k',k'_D|k_C=k\right)\right| \\ &+ 2u'\left\{k_C \neq k \text{ or } k_D \neq k\right\} \\ \leq& \sum_k u'\left(k,k,k\right) \sum_c \left|u'\left(c|k,k,k\right) - u'\left(c|k,k,k\right) \frac{u'\left(k,k,k\right)}{u'\left(k_C=k\right)}\right| \\ &+ \sum_k u'\left(k,k,k\right) \sum_c \sum_{k'\neq k, \text{ or } k'_D\neq k} \left|u'\left(c|k,k_C=k,k_D\right) u'\left(k',k'_D|k_C=k\right)\right| \\ &+ 2u'\left\{k_C \neq k \text{ or } k_D \neq k\right\} \\ \leq& \sum_k u'\left(k,k,k\right) \left|1 - \frac{u'\left(k,k,k\right)}{u'\left(k_C=k\right)}\right| + 3u'\left\{k_C \neq k \text{ or } k_D \neq k\right\} \\ \leq& \sum_k \left|u'\left(k_C=k\right) - u'\left(k,k,k\right)\right| + 3u'\left\{k_C \neq k \text{ or } k_D \neq k\right\} \\ \leq& 4u'\left\{k_C \neq k \text{ or } k_D \neq k\right\} \leq 16\varepsilon. \end{aligned}$$

Because an analogous result applies to the information of the other player, Proposition 5 shows that

$$\mathbb{d}\left(u',v'\right) \leq 16\varepsilon,$$

where $v' = \mathrm{marg}_{K\times K_C\times K_D}$. Because

$$\begin{aligned}\mathbb{d}(v,v') \leq &\sum_{k,k_C,k_D} |v(k,k_C,k_D) - v'(k,k_C,k_D)| \\ &\leq 2v'\{k_C \neq k \text{ or } k_D \neq k\} = 2u'\{k_C \neq k \text{ or } k_D \neq k\} \leq 4\varepsilon,\end{aligned}$$

the triangle inequality implies that

$$\mathbb{d}(u,v) \leq \mathbb{d}(u,u') + \mathbb{d}(u',v') + \mathbb{d}(v,v') \leq 20\varepsilon.$$

APPENDIX F: PROOF OF THEOREM 5

Suppose that u and v are two simple, and non-redundant information structures. Let $\tilde{u}$ and $\tilde{v}$ be the associated probability distributions over belief hierarchies of player 1. It is easy to show that if two non-redundant information structures induce the same distributions over hierarchies of beliefs $\tilde{u} = \tilde{v}$, then they are equivalent from any strategic point of view, and, in particular, they induce the same set of ex ante BNE payoffs. Hence, we assume that $\tilde{u} \neq \tilde{v}$.

Let $H_u = \mathrm{supp}\tilde{u}$ and $H_v = \mathrm{supp}\tilde{v}$. Lemma III.2.7 in Mertens *et al.* (2015) implies that the sets H_u and H_v are disjoint.

It is well known that there exists a non-zero sum payoff function $g^{(0)}$: $K \times (I \times I_0) \times J \to [-1,1]^2$ such that $I_0 = H_u \cup H_v$ and such that the set of rationalizable actions for player 1 of type $c \in C$ with hierarchy $h(c)$ is contained in the set $I \times \{h(c)\}$. In particular, in a Bayesian Nash equilibrium, each type of player 1 will report its hierarchy. Construct game $g^{(1)} : K \times (I \times I_0) \times (J \times \{u,v\}) \to [-1,1]^2$ with payoffs

$$g_1^{(1)}(k,i,i_0,j,j_0) = g_1^{(0)}(k,i,i_0,j),$$

$$g_2^{(1)}(k,i,i_0,j,j_0) = \frac{1}{2}g_2^{(0)}(k,i,i_0,j) + \begin{cases}\frac{1}{2}, & \text{if } j_0 = u \text{ and } i_0 \in H_u \\ -\frac{1}{2}, & \text{if } j_0 = u \text{ and } i_0 \notin H_u, \\ 0, & \text{if } j_0 = v.\end{cases}$$

Then, the rationalizable actions of player 2 in game $g^{(1)}$ are contained in $J \times \{u\}$ for any type in type space u and in $J \times \{v\}$ for any type in type space v.

Finally, for any $\varepsilon \in (0,1)$,construct a game $g^{\varepsilon} : K \times (I \times I_0) \times (J \times \{u, v\}) \to [-1,1]^2$ with payoffs

$$g_1^{\varepsilon}(k, i, i_0, j, j_0) = \varepsilon g_1^{(0)}(k, i, i_0, j, j_0) + (1-\varepsilon) \begin{cases} 1, & \text{if } j_0 = u, \\ -1, & \text{if } j_0 = v, \end{cases},$$

$$g_2^{\varepsilon} \equiv g_2^{(1)}.$$

Then, the Bayesian Nash equilibrium payoff of player belongs to $[1-\varepsilon, 1]$ on the structure u and $[-1, -1+\varepsilon]$ on the structure v. It follows that the payoff distance between the two type spaces is at least $2 - 2\varepsilon$, for arbitrary $\varepsilon > 0$.

Next, suppose that u and v are two non-redundant information structures with the decomposition $u = \sum_{\alpha} p_{\alpha} u_{\alpha}$ and $v = \sum_{\alpha} q_{\alpha} v_{\alpha}$ and such that $\tilde{u}_{\alpha} = \tilde{v}_{\alpha}$ for each α. Let g be a non-zero sum payoff function. Let σ_{α} be an equilibrium on u_{α} with payoffs $g_{\alpha} = g(\sigma_a) \in \mathbb{R}^2$. Let s_{α} be the associated equilibrium on v_{α} (that can be obtained by mapping the hierarchies of beliefs through an appropriate bijection) with the same payoffs g_{α}. The distance between payoffs is bounded my

$$\begin{aligned} \left\| \sum p_{\alpha} g(\sigma_{\alpha}) - q_{\alpha} g(s_{\alpha}) \right\|_{\max} &= \left\| \sum (p_{\alpha} - q_{\alpha}) g_a \right\|_{\max} \\ &\leq \sum |p_{\alpha} - q_{\alpha}| \, \|g_{\alpha}\|_{\max} \leq \sum |p_{\alpha} - q_{\alpha}|, \end{aligned}$$

where the last inequality comes from the fact that payoffs are bounded.

On the other hand, let $A = \{\alpha : p_{\alpha} > q_{\alpha}\}$. Using a similar construction as above, we can construct a game $g^{(1)}$ such that player 2's actions have a form $J \times \{u_A, u_B\}$, and his rationalizable actions are contained in set $J \times \{u_A\}$ for any type in type space $u_{\alpha}, \alpha \in A$ and in $J \times \{u_B\}$ otherwise. Further, we construct a game $g^{(\varepsilon)}$ as above. Then, any player 1's equilibrium $g^{(\varepsilon)}_{1,\alpha}$ payoff is at least $1-\varepsilon$ for any type in type space $u_\alpha, \alpha \in A$, and $-1+\varepsilon$ for any type in type space u_α for $\alpha \notin A$. Denoting the equilibrium payoff of player 2 as $g^{\varepsilon}_{2,\varepsilon}$, the payoff distance in game g^{ε} is at least

$$\max\left(\left|\sum_\alpha (p_\alpha - q_\alpha)\, g_{1,\alpha}\right|, \left|\sum_\alpha (p_\alpha - q_\alpha)\, g_{2,\alpha}\right|\right) \geq \left|\sum_\alpha (p_\alpha - q_\alpha)\, g_{1,\alpha}\right|$$
$$\geq \left[\sum_{\alpha\in A} (p_\alpha - q_\alpha) - \sum_{\alpha\notin A} (p_\alpha - q_\alpha)\right](1-\varepsilon) \geq (1-\varepsilon)\sum |p_\alpha - q_\alpha|\,.$$

Because the $\varepsilon > 0$ is arbitrary, the two above inequalities show that the payoff distance is equal to $\sum |p_\alpha - q_\alpha|$.

REFERENCES

Alon, N. and Spencer, J. H. (2008). *The Probabilistic Method*. Hoboken, N.J: Wiley-Interscience, 3rd edn. 5.3

Aumann, R. J. and Maschler, M. (1995). *Repeated Games with Incomplete Information*. Cambridge, Mass: The MIT Press. 5.2

Bergemann, D. and Morris, S. (2015). Bayes Correlated Equilibrium and the Comparison of Information Structures in Games. *Theoretical Economics*, **11** (2), 487–522. 1, 7, 12

Bewley, T. and Kohlberg, E. (1976). The asymptotic theory of stochastic games. *Mathematics of Operations Research*, **1** (3), 197–208. 5.2

Blackwell, D. (1953). Equivalent comparisons of experiments. *The annals of mathematical statistics*, pp. 265–272. 1, 3

Brandt, F. B. F. (2019). Justifying optimal play via consistency. *Theoretical Economics*. 1

Chen, Y.-C., di Tillio, A., Faingold, E. and Xiong, S. (2010). Uniform topologies on types. *Theoretical Economics*, **5** (3), 445–478. 4

—, —, — and — (2016). Characterizing the strategic impact of misspecified beliefs. *The Review of Economic Studies*. 4

— and Xiong, S. (2013). The e-mail game phenomenon. *Games and Economic Behavior*, **80**. 1

Dekel, E., Fudenberg, D. and Morris, S. (2006). Topologies on types. *Theoretical Economics*, **1** (3), 275–309. 1, 4, 6.1.1

—, — and — (2007). Interim correlated rationalizability. *Theoretical Economics*. 7

Ely, J. C. (2017). Beeps. *American Economic Review*, **107** (1), 31–53. 1

— and Peski, M. (2006). Hierarchies of belief and interim rationalizability. *Theoretical Economics*, **1** (1), 19–65. 1, 7

— and — (2011). Critical types. *Review of Economic Studies*, **78** (3). 1, 6.1.1

Forges, F. (1982). Infinitely repeated games of incomplete information: Symmetric case with random signals. *International Journal of Game Theory*, **11** (3-4), 203–213. 3

Gensbittel, F. and Renault, J. (2015). The value of markov chain games with incomplete information on both sides. *Mathematics of Operations Research*, **40** (4), 820–841. 3

Gossner, O. and Mertens, J.-F. (2001). The value of information in zero-sum games. 6.2, 6.2, D.1

Hellwig, C. and Veldkamp, L. (2009). Knowing what others know: Coordination motives in information acquisition. *The Review of Economic Studies*, **76** (1), 223–251. 1

Kajii, A. and Morris, S. (1998). Payoff continuity in incomplete information games. *Journal of Economic Theory*. 11

Kamenica, E. and Gentzkow, M. (2011). Bayesian Persuasion. *American Economic Review*, **101** (6), 2590–2615. 1

Kunimoto, T. and Yamashita, T. (2018). *Order on Types based on Monotone Comparative Statics*. TSE Working Papers 18-942, Toulouse School of Economics (TSE). 13

Laraki, R. and Sorin, S. (2015). Advances in zero-sum dynamic games. In *Handbook of game theory with economic applications*, vol. 4, Elsevier, pp. 27–93. 5.2

Lehrer, E. and Monderer, D. (1994). Discounting versus Averaging in Dynamic Programming. *Games and Economic Behavior*, **6** (1), 97–113. 2

Lipman, B. (2003). Finite order implications of common priors. *Econometrica*, **71** (4), 1255–1267. 6.1.1

Mertens, J.-F. (1986). *Repeated games*. Proceedings of the International Congress of Mathematicians 788, Berkeley, USA. 1, 5.2, 8

—, Sorin, S. and Zamir, S. (2015). *Repeated games*, vol. 55. Cambridge University Press. 1, 2, 7, A, F

— and Zamir, S. (1971). The value of two-person zero-sum repeated games with lack of information on both sides. *International Journal of Game Theory*, **1** (1), 39–64. 3, 5.2

— and — (1985). Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, **14**, 1–29. 1, 7

MONDERER, D. and SORIN, S. (1993). Asymptotic properties in dynamic programming. *International Journal of Game Theory*, **22** (1), 1–11. 2

MORRIS, S. (2002). Typical Types. 1

PESKI, M. (2008). Comparison of information structures in zero-sum games. *Games and Economic Behavior*, **62** (2), 732–735. 1, 3

RENAULT, J. (2006). The Value of Markov Chain Games with Lack of Information on One Side. *Mathematics of Operations Research*, **31** (3), 490–512. 3

— and VENEL, X. (2017). Long-term values in markov decision processes and repeated games, and a new distance for probability spaces. *Mathematics of Operations Research*, **42** (2), 349–376. 3

ROSENBERG, D. (2000). Zero sum absorbing games with incomplete information on one side: asymptotic analysis. *SIAM Journal on Control and Optimization*, **39** (1), 208–225. 3

—, SOLAN, E. and VIEILLE, N. (2004). Stochastic games with a single controller and incomplete information. *SIAM journal on control and optimization*, **43** (1), 86–110. 3

— and SORIN, S. (2001). An operator approach to zero-sum repeated games. *Israel Journal of Mathematics*, **121** (1), 221–246. 3

— and VIEILLE, N. (2000). The maxmin of recursive games with incomplete information on one side. *Mathematics of Operations Research*, **25** (1), 23–35. 3

RUBINSTEIN, A. (1989). The Electronic Mail Game: Strategic Behavior under "Almost Common Knowledge". *American Economic Review*, **79** (3), 385–391. 1, 6.1.1, 6.1.1

SADZIK, T. (2008). Beliefs revealed in bayesian equilibrium. *Ekonomia. Rynek, Gospodarka, Społeczeństwo*, (20), 136–138. 12

SAMET, D. (1998). Iterated expectations and common priors. *Games and economic Behavior*, **24** (1-2), 131–141. 7

SIMON, R. (2003). Games of incomplete information, ergodic theory, and the measurability of equilibria. *Israel Journal of Mathematics*, **138** (1), 73–92. 1

VENEL, X. (2014). Commutative stochastic games. *Mathematics of Operations Research*, **40** (2), 403–428. 3

WEINSTEIN, J. and YILDIZ, M. (2007). A structure theorem for rationalizability with application to robust predictions of refinements. *Econometrica*, **75** (2), 365–400. 1, 6.1.1

ZILIOTTO, B. (2016). Zero-sum repeated games: Counterexamples to the existence of the asymptotic value and the conjecture maxmin=lim v_n. *The Annals of Probability*, **44** (2), 1107–1133. 2