Value-Based Distance Between Information Structures *

Fabien Gensbittel †, Marcin Pęski‡, Jérôme Renault§

June 24, 2021

Abstract

We define the distance between two information structures as the largest possible difference in value across all zero-sum games. We provide a tractable characterization of distance and 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 value-based distance is equivalent to the weak convergence of belief hierarchies, implying, among other things, that for zero-sum games, approximate knowledge is equivalent to common knowledge. At the same time, the space of information structures

*We are grateful to Satoru Takahashi and Siyang Xiong for comments. 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 n°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 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.

†Fabien Gensbittel: Toulouse School of Economics, University Toulouse Capitole, fabien.gensbittel@tse-fr.eu

‡M. Pęski: Department of Economics, University of Toronto, marcin.peski@utoronto.ca

§Jérôme Renault: Toulouse School of Economics, University Toulouse Capitole, jerome.renault@tse-fr.eu

under the value-based distance is large: there exists a sequence of information structures where players acquire increasingly more information, and $\varepsilon > 0$ such that any two elements of the sequence have distance of 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 lead to significant differences in the behavior (Rubinstein, 1989). A recent literature on 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 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 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 a 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 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 Bayesian Nash equilibrium are more appropriate. Also, the multiplicity of solutions necessitates that the literature take a set-based approach. This, of course, makes a quantitative comparison of the value of information difficult. Last but not least, the freedom to choose games without any restrictions makes the equilibrium payoff comparison between information structures trivial (see Section 7 for a detailed discussion of this point).

Despite the challenges, we find the questions concerning the strategic value of information to be important and fascinating. How can we measure the value of information on the universal type space? How much can a player gain (or lose) from 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. We propose to focus on zero-sum games. We do so for both substantive and pragmatic reasons. First, the question of the value of information is of special importance when 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, and although they hold in single-agent decision problems (Blackwell (1953)), they do not hold for general non-zero-sum games, where better information may worsen a player's strategic position, and where players may have incentive to engage in pre-game communication to manipulate information available to others. 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 for the clarification of this issue for zero-sum games.

On the other hand, the restriction avoids all the problems mentioned above. Finite zero-sum games always have 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 unique and equal to the value of the zero-sum game. Finally, as we demonstrate through numerous results and examples, 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 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 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 single-agent problems. We describe the situations in which the value of one piece of information decreases when another piece of information becomes available, i.e. the opposing players' 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 piece of information, i.e. the opposing players' pieces of information 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 not valuable in the zero-sum games.

The second main result shows that the space of information structures is large under value-based distance: there exists an infinite sequence 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 wherein the first element of the chain is correlated with the state of the world. We construct an information structure u^n so that one player 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 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 conse-

[1] Hellwig and Veldkamp (2009) studies the complementarity and substitutability of information on acquisition decisions in a beauty contest game.

quence, 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. Among other reasons, many earlier constructions heavily rely on coordination games (Rubinstein (1989), Morris (2002), Ely and Peski (2011), Chen and Xiong (2013) among others). Such constructions cannot be done with zero-sum games.

More importantly, there are significant differences between strategic topologies and the topology induced by 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 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 higher-order beliefs becomes significant only for uncountable information structures.

An important contribution is that our result leads to the solution of 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 average values in repeated games, which would have consequences for convergence in the limit theorems[3]. Our results, however, indicate 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

[2]Problem 1 asks about the convergence of the value, and it has been proved false in Ziliotto (2016). Problem 3 asks about the equivalence between the existence of the uniform value and uniform convergence of the value functions, and it has been 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), and Renault and Venel (2017).

(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, we work with *ex-ante* information structures and the equilibrium solution concept, whereas 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 is later 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 will trade, or, a spy master allocates resources to different tasks or regions before she understands the nature of future conflicts. Value-based distance is a tight upper bound on the willingness to pay for a change in information structure. Our results provide insight into the structure of the information designer's choice space, 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

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

the set of nonnegative integers[5]. The interpretation is that k is a state of nature, and c and d are the signals of player 1 (maximizer) and player 2 (minimizer), respectively. In other words, an information 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 of at most L, we view $u \in \Delta(K \times C \times D)$ as an information structure in $\mathcal{U}(L)$. For each $u, v \in \mathcal{U}$, let us 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 mapping $g : K \times I \times J \to [-1, 1]$, where I, J are finite nonempty action sets. 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.

Information structure u and 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 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 value across all payoff functions:

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

This has a straightforward interpretation as the 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]

[5] All the results from Sections 3 and 4 can be extended to uncountable information structures, the proof is relegated to Appendix L).

[6] The inequality is a property of zero-sum games. For every game $g \in \mathcal{G}$, let σ be an optimal

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 player signals, 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, where 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 in Blackwell (1953) to compare statistical experiments. A garbling is a mapping $q : \mathbb{N} \to \Delta(\mathbb{N})$. The 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

strategy of player 1 in $\Gamma(u,g)$ and let τ be an optimal strategy of player 2 in $\Gamma(v,g)$. Using the saddle-point property of the value, the difference $\text{val}(u,g) - \text{val}(v,g)$ is no 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 no larger than $\|u - v\|$.

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 in the Blackwell's comparison of experiments sense: suppose that (k, c', d) is selected according to u, c is selected according to 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 is deteriorated from garbling q. Similarly, $u.q$ corresponds to the dual situation where player 2's signal is deteriorated. Further, garbling q can also be seen as a behavioral 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 $L = 1, 2, ..., \infty$, and each $u, v \in \mathcal{U}(L)$,*

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

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 $g \in \mathcal{G}(L)$.

The first part of Theorem 1 finds a tight upper bound on the difference of value between all zero-sum games played on information structures v and u. It is equal to a total variation distance between two sets of garblings of the original information structures: the set $\mathcal{Q}.u = \{q.u : q \in \mathcal{Q}\}$ of structures obtained from u by deteriorating information of player 1 and the set $v.\mathcal{Q} = \{v.q : q \in \mathcal{Q}\}$ obtained from v by deteriorating information of player 2. The two sets are illustrated on Figure 1. (The direction up is better for player 1, and worse for player 2.)

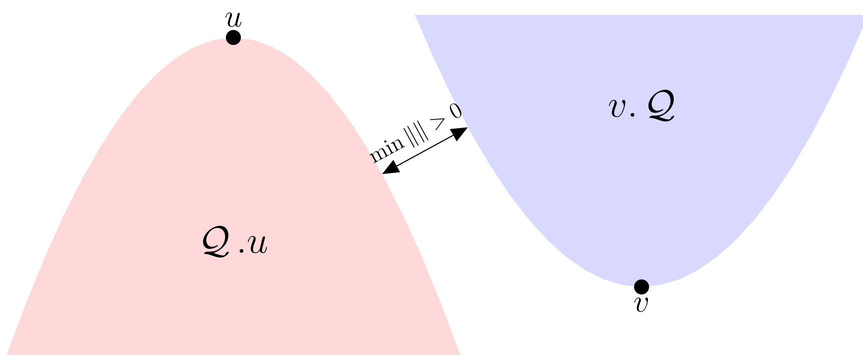


Figure 1:

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.

Theorem 1 is closely related to the comparison of information structures. Say *player 1 prefers* u *to* v in every game; i.e., write $u \succeq v$, if for all $g \in \mathcal{G}$, $\mathrm{val}(u, g) - \mathrm{val}(v, g) \geq 0$. The definition extends Blackwell's comparison of experiments to zero-sum games. If the two sets from Figure 1 have non-empty intersection, the distance between them is equal to 0, and player 1 prefers u to v. Conversely, if the two sets do not have an intersection, then there are games for which the difference in value on v and u is positive (and arbitrarily close to the total variation distance) and player 1 does not prefer u to v. Hence Theorem 1 implies the following

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

The above result extends (Peski, 2008) to countable information structures ((Peski, 2008) was stated only for finite structures).

The intuition of the proof of Theorem 1 is as follows. For simplicity, suppose that the information structure is finite. We first show that the maximum of the difference in values over all games g, i.e., the right-hand side of (2), is not smaller than the left-hand side of (2). The starting point is to identify each garbling with a mixed strategy in Bayesian game $\Gamma(u, g)$ induced from 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 $\langle g, u\rangle = \sum_{k,c,d} g(k, c, d)u(k, c, d)$ and q_i are garbling/strategies. Among others, each player can use strategy Id which plays the received signal. Because such a strategy is always available to both players, using the minmax property of equilibrium, the difference in values $\text{val}(v, g) - \text{val}(u, g)$ is no less than the difference between player 2's optimal payoff against the Id strategy of player 1 in v (i.e. $\inf_{q_2}\langle g, v.q_2\rangle$) and player 1's optimal payoff against the Id strategy of player 2 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 $\sup_g \inf_{q_1,q_2}\langle g, v.q_2 - q_1.u\rangle$. The latter is equal to the left-hand side of (2).

Next, we show that, for each game g, the difference in values is not higher than the left-hand side of (2). Using the monotony of the value with respect to information, we have that

$$\text{val}(v, g) - \text{val}(u, g) \leq \text{val}(v.q_2, g) - \text{val}(q_1.u, g) \leq \|v.q_2 - q_1.u\|$$

for arbitrary garblings q_1 and q_2. The last inequality was stated above (see footnote 6). 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.

4 Applications

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

4.1 The impact of the marginal over K

Among the many ways in which two information structures can differ, the most obvious one is that they may have different distributions over 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*

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

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

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 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: single-agent problems.

Formally, a payoff function $g \in \mathcal{G}(L)$ is a *single-agent (player 1) problem* if player 2's action set is a singleton, $J = \{*\}$. Let $\mathcal{G}_1 \subset \mathcal{G}$ be the set of player 1 problems.

Then, for each $g \in \mathcal{G}_1$ and 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 one player's additional piece of information, and the players' information are conditionally independent in both cases, then the maximum value-based distance (1) is attained in a single-agent decision problem. Such problems form a relatively small subclass of games and they are easier to identify. In Appendix H, we apply the proposition to compute the exact distance between information structures induced by multiple Blackwell experiments.

The proof of Proposition 2 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 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

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

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 additional piece of information for the other player on the value of information. Suppose that

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

When moving from v' to u' or v to u, in both cases, player 1 gains additional signal c_1. However, in the latter case, player 2 obtains an additional piece of information 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 simultaneously receive additional information. 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 $v = \text{marg}_{K \times C \times D} u$. 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 the potential scenario where the additional signal for each player does not provide the respective player with any significant information about the state of the world or the original information of the other player. While such signals would be useless in a single-agent decision problem, they 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

[7] How useful it is 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).

by the two players is not valuable in zero-sum games.

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

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 $K = \{0, 1\}$.

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 the 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 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 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 progressively more signals converges for 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 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

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

In particular, the family of all functions $(u \mapsto \text{val}(u, g))_{g \in \mathcal{G}}$ 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 of 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 equicontinuity of an appropriate family of value functions 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 a sufficiently high, even-valued N and a set $S \subseteq \{1, ..., N\}^2$ with certain mixing properties:

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

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 one another.

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 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 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 optimal N (or about the largest ε allowing for the conclusions of Theorem 2).

Given S, we construct 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 for 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 Theorem 2 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

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

Theorem 3. *Let u be in $\mathcal{U}^*$. A sequence (u_n) in $\mathcal{U}^*$ converges to u for 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.

The intuition of the proof is as follows: 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. 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 information about the state and the opponents' neighborhoods. By Proposition 5, only information about neighborhoods matters, and the latter is similar to the information in 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: It seems to (a) convey a message that is opposite to the literature on strategic (dis)continuities, and (b) 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)[8]. Player 1 always knows the state. Player 2's first-order belief attaches 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 type spaces converge in the weak topology to the common knowledge of the state. Theorem 3 implies that the Rubinstein type

[8]The email game information structure u_ε is as follows: there are two states 0 and 1, the latter with probability p. Player 1 knows the state. If the state is 1, a message is sent to the other player, who, upon receiving it, immediately sends it back. The message travels back and forth until it is lost, which happens with i.i.d. probability $\varepsilon > 0$ each time it travels. The signal of each player in u_ε is the number of messages she received. In Rubinstein (1989), a non-zero-sum coordination game is considered, and shown to have the property that the set of (Bayesian Nash) equilibrium payoffs with u_ε does not converge to the set of equilibrium payoffs with u_0, where u_0 is simply the structure corresponding to common knowledge of the state.

spaces also converge under 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 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 $K_C = K_D = K$, $\text{marg}_K v = \text{marg}_K u$, 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,*

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

Therefore, approximate knowledge structures are close to common knowledge structures. The convergence of approximate knowledge type spaces to common knowledge is a consequence of Theorem 3. The metric bound stated in Proposition 6 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 payoff comparison rather than 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, assuming sufficient richness, it can be used to show that the resulting topology is strictly finer than weak topology[9]. Further, the ex-ante focus and payoff comparison (but without restriction to zero-sum games) lead to a topology that is significantly finer than 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 the one hand, Lipman (2003) implies that, at least from the interim perspective, a 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.

Let us also mention that Proposition 6 is also related to Kajii and Morris (1997)(and additional results in Morris and Ui (2005) Morris-Ui and Oyama and Tercieux (2010)). In that paper, the authors fix a game with complete information and study the robustness of equilibria to perturbation of complete information in similar way to our notion of ε-knowledge. They show that if the game has a unique correlated equilibrium, then for sufficiently small ε, nearby games have an equilibrium with behavior close to the equilibrium of the complete information game. Proposition 6 says that, for zero-sum games that may have multiple equilibria, equilibrium payoffs are robust to small amounts of incomplete information.

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 convergence in weak topology to a *countable* information structure is equivalent to convergence in value-based distance. Together, the two results imply that, although weak and value-based topologies are equivalent around countable structures $\mathcal{U}^*$, they differ beyond $\mathcal{U}^*$. The impact of higher-order beliefs becomes significant only for uncountable information structures.

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

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 compact closure Π under 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.

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 pointwise to u if, for all payoff functions $g \in \mathcal{G}$, $\lim_{n\to\infty} \operatorname{val}(u_n, g) = \operatorname{val}(u, g)$. Clearly, if (u_n) converges to u for 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. Since $\operatorname{val}(\mu, g)$ 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 set $\mathcal{U}^*$, the topology induced by value-based distance, the topology of weak convergence, and the topology of pointwise convergence coincide. In particular, let u be in $\mathcal{U}^*$ and (u_n) be in $\mathcal{U}^*$. Then (u_n) converges to u for value-based distance if and only if for every g in $\mathcal{G}$, $\operatorname{val}(u_n, g) \xrightarrow[n\to\infty]{} \operatorname{val}(u, g)$.*

A standard way to define a metric compatible with 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[10] 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

[10] 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.

of $(g_n)_n$ will play no role in the sequel. Define now the distance $\mathbb{d}_W$ on $\mathcal{U}^*$ by:

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

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, $\operatorname{val}(u_l, g) \xrightarrow[l\to\infty]{} \operatorname{val}(u, g)$. $\mathcal{U}^*$ 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 weak topology.*

Proof. Define similarly distance $\mathbb{d}_W$ on Π as $\mathbb{d}_W(\mu, \nu) = \sum_{n=1}^{\infty} \frac{1}{2^n} |\operatorname{val}(\mu, g_n) - \operatorname{val}(\nu, g_n)|$. By construction, the mapping $(u \mapsto \tilde{u})$ 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 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. □

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 a continuum of signals, obtained as limits of sequences of information structures with countably many types.

The main point of interest in Theorem 4 is that we can now view Π as set $\mathcal{V}$. We can recover the exact space (Π, weak) using values of zero-sum Bayesian games and the completion of a metric space[11]. This may be seen as a duality result between games and information: Π is defined with hierarchies of beliefs but with 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 weak

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

topology. Now the construction of $\mathcal{V}$ yields a new, alternative, interpretation of Π, and one might possibly hope to deduce properties of (Π, weak) by transferring, 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 completion $\mathcal{W}$ of $\mathcal{U}^*$ for $\mathbb{d}$ is not compact. Space $\mathcal{W}$ also contains information structures with a 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 distance (1) where the supremum is taken over all (including non-zero-sum) games. Rubinstein's e-mail game ((Rubinstein, 1989)) shows the relevance of almost conditional independent information for non-zero-sum games (in the sense of Section 4.6). Therefore similar statements as Theorem 3 and Proposition 6 do not hold when considering non-zero-sum games. We show the stronger result that such a payoff-based distance between information structures is mostly trivial.

A *non-zero sum payoff function* is a mapping $g : K \times I \times J \to [-1,1]^2$ 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 that the space of compact subsets of $\mathbb{R}^2$ is equipped with the induced Hausdorff distance $\mathbb{d}^H_{\max}$. Let

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

Then, clearly as in our original definition, $0 \leq \mathbb{d}_{NZS}(u,v) \leq 2$.[12]

[12] A very similar approach to closeness of information structures is taken in MONDERERSAMET(96) and Kajii and Morris (1998). MONDERERSAMET(96) define a notion of distance d_{MS} on common prior information structures that relies on closeness of common belief events. They show that if two information structures are ε-close, then any equilibrium on one of them is ε-close to an ε-equilibrium on the other structure. Kajii and Morris (1998) use the last property as their

Contrary to the value in the zero-sum game, the BNE payoffs on information structure u cannot be factorized through distribution $\tilde{u} \in \Pi$ over the hierarchies of beliefs induced by u. For this reason, we only restrict our analysis to information structures that are non-redundant, which means that two different signals (occurring with positive probability) induce tow different hierarchies of beliefs. We do so because the dependence of the BNE on redundant information is not yet well-understood[13]. For convenience, we also restrict ourselves to information structures with finite support.

Let $u \in \Delta(K \times C \times D)$ be an information structure with finite support. 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. As it follows from Lemma III.2.7 in Mertens *et al.* (2015), each non-redundant information structure u with finite support has a representation as a finite convex combination of (non-redundant) simple information structures[14] $u = \sum_\alpha p_\alpha u_\alpha$, where $\sum p_\alpha = 1, p_\alpha \geq 0$.

Theorem 5. *Suppose that u, v are non-redundant information structures with finite*

definition. They say that an information structure is ε-close to another one if, for all bounded games, any equilibrium of one is ε-close to an ε-equilibrium of the other. Thus, the main difference between the approach on these two papers and our metric $\mathbb{d}_{NZS}$ is that we require ε-closeness to a (proper) equilibrium. Our metric $\mathbb{d}_{NZS}$ is closer in spirit to the value-based distance defined using zero-sum games, and, arguably, it is easier to interpret for values of ε that are far away from 0.

[13] 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).

[14] Let us sketch the argument for this result. For each signal (type) s of a player in the support of u, we can define $N_1(s)$ as the support of $u(.|s)$. Then, we repeat the construction for every signal in $N_1(s)$ and define $N_2(s)$ as the union of $N_1(s)$ and all the sets obtained this way, $N_2(s) = N_1(s) \cup (\cup_{\tilde{s} \in N_1(s)} N_1(\tilde{s}))$. Repeating this process, the sequence is eventually stationary, i.e. $N_{t+1}(s) = N_t(s)$ for some integer t. We obtain a finite set $N(s) = N_t(s)$ having the property that the conditional distribution of $u(.|N(s))$ is a (non-redundant) simple information structure with support $N(s)$. There are finitely many different sets $N(s)$ when s ranges through all signals in the support of u and they form a partition of the support of u. The representation as a convex combination follows directly from the construction.

support. If u and v are simple, then

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

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 to its maximum possible value 2. The distance between two non-redundant but not necessarily simple information structures, $\mathbb{d}_{NZS}$, depends on the similarity of their simple components after decomposition. Theorem 5 implies that (5) is too fine a 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 $\tilde{u} \neq \tilde{v}$. Earlier results have shown that there exists 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 (see lemma 4 in Dekel *et al.* (2006) and lemma 11 in Ely and Peski (2011)). 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 in Samet (1998)). Therefore, 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 resultant 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 arbitrarily small. The resultant game has a BNE in the unique rationalizable profile. [15]

8 Conclusion

In this paper, we have introduced and analyzed 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 apply it directly to numerous questions about the value of information, the relation between games and single-agent problems, a comparison of information structures, etc. Additionally, we show that the distance contains interesting topological information. On the one hand, the topology induced on 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 entirely bounded for value-based distance, which negatively solves 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 re-examine the relevance of many phenomena observed and discussed in the strategic discontinuities literature. While the distinction between approximate knowledge and approximate common knowledge is not important in situations of conflict, higher order beliefs may 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 a 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.[16]

[15] This construction relies creating new games by adding externality to payoffs of one player that depend only on the actions of the other player. Such techniques are available with non-zero-sum games, but not with zero-sum games. We are grateful to a referee for pointing it out.

[16] 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.

A Proof of Theorem 1

The proof of Theorem 1 relies on two main aspects: the two interpretations of 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 player, with any strategy $\geq L$ that is 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 $g \in G$, let $\gamma_{u,g}(q_1, q_2)$ denote 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 g to mixed actions, as usual, 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 payoff $\gamma_{u,g}(Id, Id)$, where $Id \in \mathcal{Q}$ is the strategy that plays the signal received with probability one. A straightforward computation leads to $\gamma_{u,g}(q_1, q_2) = \langle g, q_1.u.q_2 \rangle$. Consequently,

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

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 monotonicity 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 $\mathrm{val}(v, g) - \mathrm{val}(u, g) \leq \mathbb{d}\left(q_1.u, v.q_2\right) \leq \|q_1.u - v.q_2\|$. Therefore,

$$\sup_{g \in \mathcal{G}} \left(\mathrm{val}(v, g) - \mathrm{val}(u, g)\right) \leq \inf_{q_1, q_2 \in \mathcal{Q}(L)} \|q_1.u - v.q_2\| = \inf_{q_1, q_2 \in \mathcal{Q}(L)} \sup_{g \in G(L)} \langle g, v.q_2 - q_1.u \rangle. \tag{7}$$

We are now going to show that

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

Together with inequalities 6 and 7, it will give

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

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 $h(g, (q_1, q_2)) = \langle g, v.q_2 - q_1.u \rangle$. Strategy sets $G(L)$ and $\mathcal{Q}\left(L\right)^2$ are convex, and h is bilinear.

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

Case 2: $L = +\infty$. We are going to modify the topology on $\mathcal{Q}$ to have compact $\mathcal{Q}\left(L\right)^2$ and lower semi-continuous h on (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 $\mathcal{Q}\left(L\right)^2$ 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 $\mathbb{N}$:

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

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 $q \in \mathcal{Q}$.

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,

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

Now, 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 p', q' arbitrarily.

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 $u, v \in \Delta(K\times C\times D)$,

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

where $u' = \text{marg}_{K\times C}\, u$, $v' = \text{marg}_{K\times C}\, v$, $q.u'(k,c) = \sum_{s\in C} u'(k,s)q(s)(c)$ 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 $u(k) = v(k)$, $u\left(c,d|k\right) = u\left(c|k\right)u\left(d|k\right)$, and $v\left(c',d|k\right) = v\left(d|k\right)v\left(c'|k\right) = u(d|k)v\left(c'|k\right)$. For any pair of garblings q_1, q_2,

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

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 $|x+y| \geq |x| + \text{sgn}(x)y$ for each $x, y \in \mathbb{R}$, we have

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

where the last equality comes from the fact that $\sum_{d}\Delta\left(k,d\right) = 0$. Therefore, we obtain

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

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

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

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 $u \succeq v$,

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

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 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

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

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

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

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

B.4 Proof of Proposition 4

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

B.5 Proof of Proposition 5

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

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,

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

The claim follows from Theorem 1.

C Proof of Theorem 2

N is a very large even-valued 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 a few more conditions, to be defined later.

A sequence $(a_1, ..., a_l)$ of length $l \geq 1$ 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 remainder of the proof is split into 3 parts: we first define information structures $(u^l)_{l\geq 1}$ and payoff structures $(g^p)_{p\geq 1}$. 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 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 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: player 1's first signal c_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 of elements $(a_1, ..., a_l)$ 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. 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 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 $\frac{1}{(N+1)^2} \geq 10\varepsilon$. Furthermore, 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 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, player 1 chooses c' in C^p and player 2 chooses d' in D^{p-1} *simultaneously*, and finally, player 1's payoff is $g^p(k,c',d')$. Notice that, by the previous paragraph about g_0, it is always strictly dominant for player 1 to truthfully report her first signal, i.e. 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 has 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 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 payoff of -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 conditions $UI1$ and $UI2$.

C.2 Conditions UI and values

To prove that the intuition of the previous paragraph is 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

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

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 information structure $u^l \in \Delta(K \times C^l \times D^l)$ and let $\tilde{c}$ and $\tilde{d}$ denote the respective random variables of the signals of player 1 and player 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 $r = 2m-2, 2m-1$,

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

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 signals $(c_1, ..., c_l)$ in C^l. Equation (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 $c'_m \neq c_m$). Equation (11) requires that, given that the reported signals were nice thus 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 level $2m-1$, adding the next signal $\tilde{d}_m$ for 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 d signals.

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:

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

Proof of Proposition 7. We assume that $UI1$ and $UI2$ hold. We fix $l \geq 1$, work on probability space $K \times C^l \times D^l$ equipped with probability u^l, and let $\tilde{c}$ and $\tilde{d}$ denote 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 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

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

If $n = p$, $h^p(\tilde{c}, d') = \varepsilon$ on 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

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

and (15) follows by backward induction.

Since A_1 is an event that holds almost surely, we deduce that $\mathbb{E}[h^p(\tilde{c},d')|\tilde{d}=d]\geq\varepsilon$. Therefore, player 1's truthful strategy guarantees 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 signal c and chooses to report c'. We will show that player 1's expected payoff is no 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 player 2's truthful strategy guarantees that the payoff of the maximizer is less than 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-valued 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 $UI2$.

Letting $P = (P_{a,b})_{(a,b)\in A^2}$ denote the transition matrix of the Markov chain, we must 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 becomes strictly positive and, in fact, converges to 1.

Formally, let $\mathcal{S}_A$ denote 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 uniformly distributed over $\mathcal{S}_A$ defined on 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.

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

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

The remainder 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 space $(\Omega, \mathcal{F}, \mathbb{P})$. For $i = a, b$, define events $\widetilde{L}_i = \{\sum_{j\in A}\widetilde{X}_{i,j} = \frac{N}{2}\}$ and 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's 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's approximation[17], 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 $\mathbb{P}_N\left(\left||S_a \cap S_b| - \frac{1}{4}N\right| \geq \gamma N\right) \leq \frac{1}{2}e^4 N e^{-2\gamma^2 N}$. □

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

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

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

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

□

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 $p = \frac{1}{2}\left(\frac{\frac{N}{2}-1}{N-1}\right) = \frac{1}{4} - \frac{1}{4N-4}$. Hoeffding's 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 $|S||p - \frac{1}{4}| \leq \frac{N}{4N-4} \leq \frac{\gamma N}{2}$ for the first inequality.

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

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

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

Lemma 3. *For each $a \neq b$ and $c \neq d$, each $\gamma \geq 64/N$, each of the variables*

$$Z \in \{Y_a, Y^c, Y_{a,b}, Y^{c,d}, Y_a^c, Y_{a,b}^c, Y_a^{c,d}, Y_{a,b}^{c,d}\},$$

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

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 bound comes 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. ∎

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

$$E_{a,b,c,d} = \left\{\left|\frac{Y_{a,b}}{Y_a} - 1\right| \leq 2\alpha\right\} \cap \left\{\left|\frac{Y^c_{a,b}}{Y^c_a} - 1\right| \leq 2\alpha\right\} \cap \left\{\left|\frac{Y^{c,d}_a}{Y^c_a} - 1\right| \leq 2\alpha\right\} \cap \left\{\left|\frac{Y^{c,d}_{a,b}}{Y^{c,d}_a} - 1\right| \leq 2\alpha\right\}$$
$$\cap \left\{\left|\frac{Y^{c,d}}{Y^c} - 1\right| \leq 2\alpha\right\} \cap \left\{\left|\frac{Y^c_a}{Y^c} - 1\right| \leq 2\alpha\right\} \cap \left\{\left|\frac{Y^{c,d}_a}{Y^{c,d}} - 1\right| \leq 2\alpha\right\}.$$

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). □

Computations using the bound of Lemma 4 show that $N = 52.10^6$ is sufficient for the existence of an appropriate Markov chain. Therefore, 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 conditions $UI1, UI2$ are satisfied.*

Proof. We fix 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 let u^l denote 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.

Case $m > 1$ and $r = 2m-1$. 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).

Case $r = 2m$.

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. □

D Proofs of Theorem 3

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

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

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 assume without loss of generality that u is non-redundant and that 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, $\{\kappa_d^m\}_{0\leq d\leq m}$ 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 $c \in C$ and each $d \in D^p$, we have $\mathbb{E}_{\phi_1(\tilde{c})}[\mathbb{1}_{\{k\}}(.)\kappa_d^p(.)] \geq u(k, d|c)$, 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, there exists $p^m < \infty$ and $\varepsilon^m \in \left(0, \frac{1}{m}\right)$

for each m such that, for any $c, c' \in C^m$ such that $c \neq c'$,

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

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

Then, h_c^m is a continuous function such that $h_c^m(\tilde{c}) = 0$ and such that, if $h_c^m(\theta_1) \leq \varepsilon^m$ for some $c \in C^m$, then $h_{c'}^m(\theta_1) \geq \varepsilon^m$ for any $c' \in C^m$ such that $c' \neq c$. For $0 \leq c \leq m+1$, 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 $\{\kappa_c^m\}_{0 \leq c \leq m+1}$ 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 information structure $K^m v \in \Delta(K \times C' \times \{0, ..., m+1\} \times D' \times \{0, ..., p^m\})$ so that $K^m v\left(k, c', \hat{c}, d', \hat{d}\right) = v(k, c', d') \kappa_{\hat{c}}^m(\tilde{c}') \kappa_{\hat{d}}^{p^m}\left(\tilde{d}'\right)$. 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 $h_{\hat{c}}^{m}(\tilde{c}') \leq \varepsilon^{m}$. It follows that

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

On the other hand,

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

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

Define information structure $L^m v = \operatorname{marg}_{K\times\{0,\ldots,p^m\}\times\{0,\ldots,m+1\}} K^m v$. Then, because $d\left(K^m v, v\right) = 0$, the proof of Proposition 5 implies that $\sup_{g\in\mathcal{G}}\left(\operatorname{val}(v,g) - \operatorname{val}\left(L^m v, g\right)\right) \leq \delta^m v$.

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

$$\left(L^m u_n\right)\left(k,\hat{c},\hat{d}\right) = \mathbb{E}_{\tilde{u}_n}\left(\kappa_{\hat{c}}^{m}\left(\theta_1\right)\mathbb{E}_{\phi_1(\theta_1)}[\mathbb{1}_{\{k\}}\,\kappa_{\hat{d}}^{p^m}]\right).$$

Because all the functions in brackets above are continuous, weak convergence $\tilde{u}_n \to \tilde{u}$ implies that $(L^m u_n)\left(k, \hat{c}, \hat{d}\right) \to (L^m u)\left(k, \hat{c}, \hat{d}\right)$ 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 $\mathbb{d}(L^m u_n, L^m u) \leq \|L^m u_n - L^m u\| \to 0$ as $n \to \infty$. Moreover, if $\hat{c} \in C^m$ and $\hat{d} \in D^{p^m}$, the definitions imply that $(L^m u)\left(k, \hat{c}, \hat{d}\right) \geq u\left(k, \hat{c}, \hat{d}\right)$. Therefore,

$$\mathbb{d}(L^m u, u) \leq \|L^m u - u\| \leq 2\left(u(C \backslash C^m) + u\left(D \backslash D^{p^m}\right)\right) \underset{n\to\infty}{\longrightarrow} 0.$$

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

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

Together, we obtain for each m, n

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

Therefore, $\lim_{n\to\infty} \sup \sup_{g\in\mathcal{G}}(\text{val}(v, g) - \text{val}(L^m v, g)) \leq 3\left(u(C \backslash C^m) + u\left(D \backslash D^{p^m}\right)\right).$

When $m \to \infty$, the right-hand side converges to 0 as well.

E Proof of Proposition 6

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

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

Therefore,

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

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

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

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

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

the triangle inequality implies that

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

F Proof of Theorem 5

Suppose that u and v are two simple and non-redundant information structures with finite support. Let $\tilde{u}$ and $\tilde{v}$ be the associated probability distributions over player 1's belief hierarchies. 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 $H_u = \text{supp}\tilde{u}$ and $H_v = \text{supp}\tilde{v}$. Lemma III.2.7 in Mertens *et al.* (2015) implies that the sets H_u and H_v are disjoint.

By adapting the construction made in Lemma 4 of Dekel *et al.* (2006) (see also Lemma 11 in Ely and Peski (2011)), 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 player 1's rationalizable actions of type $c \in C$ with hierarchy $h(c)$ is contained in set $I \times \{h(c)\}$. In particular, in a BNE, 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 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 BNE payoff of player belongs to $[1 - \varepsilon, 1]$ on structure u and $[-1, -1 + \varepsilon]$ on 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 decomposition $u = \sum_\alpha p_\alpha u_\alpha$ and $v = \sum_\alpha q_\alpha v_\alpha$ 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 by

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

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 game $g^{(1)}$ such that player 2's actions have 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 game $g^{(\varepsilon)}$ as above. Then, any player 1's equilibrium payoff $g_{1,\alpha}^{(\varepsilon)}$ 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 player 2's equilibrium payoff as $g_{2,\varepsilon}^{\varepsilon}$, 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 $\varepsilon > 0$ is arbitrary, the two above inequalities show that the payoff distance is equal to $\sum |p_\alpha - q_\alpha|$.

G Examples and Counterexamples

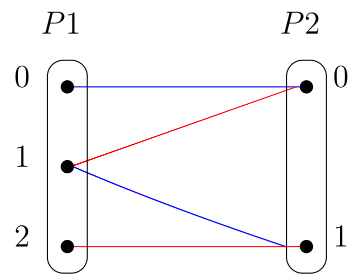
This section contains various examples and counterexamples.

G.1 Computing the distance

We begin with two examples of computations of the value-based distance based on the characterization from Theorem 1.

We assume that there are two states, $K = \{\mathit{Blue}, \mathit{Red}\}$. We consider information structures which are uniform over a finite subset of $K \times \mathbb{N} \times \mathbb{N}$.

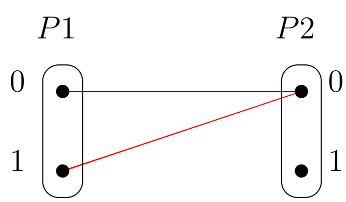
The next picture illustrates information structure u_1:



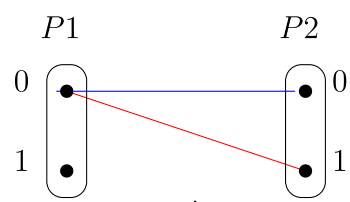
u_1

Each line corresponds to a pair of signals from the support of the information structure; each such a pair has equal probability. The color of the line represents the state. For example, conditional on both players receiving signal 0, the state is Blue. Note that signals 0 and 1 convey different information for player 2. After receiving signal 0, player 2 knows that if the state is Blue, then player 1 knows it, and if the state is Red, player 1's belief on the state is uniform. Whereas after receiving signal 1, player 2 knows that : if the state is Blue, then player 1's belief on the state is uniform, and if the state is Red, player 1 knows it.

The next figure presents information structures u_2 and u_2':



u_2



u_2'

Under information structure u_2, player 1 knows the state, but player 2 does not. Under u_2', player 2 knows the state, but player 1 does not. We have the following observation:

Proposition 9. $\mathbb{d}(u_1, u_2) = \frac{1}{2}$. $\mathbb{d}(u_1, u_2') = 1$.

It follows that u_1 is closer to u_2 than to u_2'.

Proof. We have $u_2 \succeq u_1$, hence

$$\mathbb{d}(u_2, u_1) = \min_{q_1 \in \mathcal{Q}, q_2 \in \mathcal{Q}} \|q_1.u_1 - u_2.q_2\|.$$

Define q_1 in $\mathcal{Q}$ such that $q_1(0) = \delta_0$, $q_1(1) = q_1(2) = \delta_1$, and q_2 in $\mathcal{Q}$ satisfying $q_2(0) = 1/2\,\delta_0 + 1/2\,\delta_1$. The information structures $q_1.u_1$ and $u_2.q_2$ can be represented as follows:

Notice that $u_2.q_2 \sim u_2$, whereas $q_1.u_1 \preceq u_1$. $\|q_1.u_1 - u_2.q_2\| = 1/2$, hence $d(u_2, u_1) \leq 1/2$.

Consider now the payoff structure g given by $\left\{ \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} \right\}$, where the color corresponds to the payoffs in a given state. In the game (u_2, g), it is optimal for player 1 to play Top if 0 and Bottom if 1, and $\mathrm{val}(u_1, g) = 1/2$. In the game (u_1, g) it is optimal for player 2 to play Left if 0 and Right if 1, and $\mathrm{val}(u_1, g) = 0$. Consequently, $\mathbb{d}(u_2, u_1) \geq 1/2$, and we obtain $\mathbb{d}(u_2, u_1) = 1/2$.

Notice that $u_2' \sim u_2''$, with u_2'' obtained from u_2' by exchanging the signals 0 and 1 for each player, and $\|u_1 - u_2''\| = 1$. Considering the payoff structure given by $\left\{ \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \right\}$ gives $\mathbb{d}(u_2', u_1) = 1$. □

G.2 Impact of the marginal over states

We illustrate an application of Proposition 1 in the binary case $K = \{0, 1\}$. Fix $p, q \in \Delta K$. In such a case, one easily checks that the maximum in the right hand side of inequalities (3) is attained by either $p' = q' = (1, 0)$, or $p' = q' = (0, 1)$, or

$p' = p, q' = q$. It follows that for any two information structures u, v,

$$\mathbb{d}\left(u,v\right) \leq 2\left(1-\max\left(\min\left(p_0,q_0\right),\min\left(p_1,q_1\right),p_0q_0+p_1q_1\right)\right).$$

Example 1. *The bound is attained when* $u\left(k,c,d\right) = p_k\, \mathbb{1}_{c=k}\, \mathbb{1}_{d=0}$ *and* $v\left(k,c,d\right) = q_k\, \mathbb{1}_{c=0}\, \mathbb{1}_{d=k}$ *for each* $k, c, d \in \{0,1\}$.

G.3 Value of additional information: games vs single-agent problems

The next example illustrates that the thesis of Proposition 2 does not hold without a conditional independence.

Example 2. *Suppose that* $C = \{*\}, K = D = C' = \{0,1\}$, *and let* $u \in \Delta(K \times (C \times C') \times D)$ *be defined by*

$$u\left(k,(*,c'),d\right) = \begin{cases} \frac{1}{4}\frac{k+c'}{2}, & \text{if } d = 1, \\ \frac{1}{4}\left(1-\frac{k+c'}{2}\right), & \text{if } d = 0. \end{cases}$$

Let v *denote the marginal of* u *over* $K \times C \times D$. *Going from* v *to* u, *the new information* c' *is independent from* k, *but* d *is both correlated with* k *and* c'. *Because* c' *is independent from the state, the value of new information in player 1 single-agent problem is 0:*

$$\mathbb{d}_1\left(u,v\right) = 0.$$

However, signal c' *provides non-trivial information about the signal of the other player, hence it is valuable in some games, which implies that* $\mathbb{d}(u,v) > 0$. *(Indeed, using Theorem 1, since* u *gives player* 1 *more information, we have*

$$\mathbb{d}(u,v) = \min_{q_1,q_2} \|u.q_2 - q_1.v\|,$$

where $q_2 : \{0,1\} \to \Delta\{0,1\}$ *and* $q_1 \in \Delta\{0,1\}$. *The existence of a pair* (q_1, q_2) *such* *that* $\|u.q_2 - q_1.v\| = 0$ *is equivalent to the system of equations*

$$\forall (k,d,c') \in \{0,1\}^3,\ u.q_2(k,d,c') = v.q_1(k,d,c'),$$

where the unknowns are q_1, q_2*, and one can check that this system does not admit any solution. In other words, the information that would be useless in a single-agent decision problem is valuable in a strategic setting.*

G.4 Informational substitutes

Here, we discuss the assumptions of Proposition 3. The conditional independence assumption is equivalent to two simpler assumptions (a) c_1 is conditionally independent from (c, c_2) given (k, d), and (b) c_1 and d are conditionally independent given k. Both (a) and (b) are important as it is illustrated in the two subsequent examples.

Example 3. *Violation of (a). Suppose that* $C = D = \{*\}$*,* $K = C_1 = C_2 = \{0,1\}$*,* c_1 *and* c_2 *are uniformly and independently distributed, and* $k = c_1 + c_2 \mod 2$*. Then, signal* c_2 *is itself independent from the state, hence useless without* c_1*. Knowing* c_1 *and* c_2 *means knowing the state, which is, of course, very valuable. Thus, the value of* c_2 *increases when* c_1 *is also present.*

Example 4. *Violation of (b). Suppose that* $C = \{*\}$*,* $K = C_1 = C_2 = D = \{0,1\}$*,* c_1 *and* d *are uniformly and independently distributed,* $c_2 = d$*, and* $k = c_1 + d \mod 2$*. Notice that part (a) of the assumption holds (given* (k, d)*, both signals* c_1 *and* c_2 *are constant, hence, independent), but part (b) is violated. Again, signal* c_2 *is useless alone, but together with* c_1 *it allows the determine the state and the information of the other player.*

G.5 Informational complements

The next example shows that the independence assumptions in Proposition 4 are important.

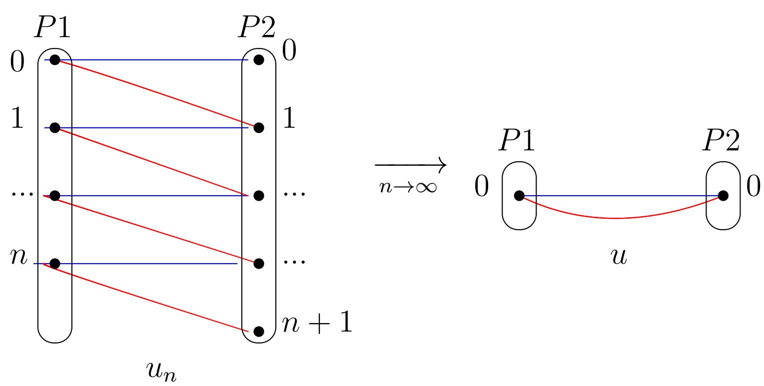
Example 5. *Suppose that $C = \{*\}$ and $K = D = D_1 = C_1 = \{0, 1\}$. The state is drawn uniformly. Signal d is equal to the state with probability $\frac{2}{3}$ and signal d_1 is equal to the state for sure. Finally, $c_1 = 1$ iff $d = k$.*

In other words, player 2's signal d is an imperfect information about the state. Signal c_1 carries information about the quality of the signal of player 1. Such signal is valuable in some games, hence $\mathbb{d}\,(u', v') > 0$. In the same time, if player 2 learns the state perfectly, signal c becomes useless, and $\mathbb{d}\,(u, v) = 0$. (The last claim can be formally proven using Proposition 5.)

G.6 Convergence to simple structures

Next, we provide an example of a convergent sequence of information structures.

Example 6. *Consider a sequence of information structures (which are all uniform over a finite subset of $K \times \mathbb{N} \times \mathbb{N}$):*



Signals 0 and $n+1$ perfectly reveal the state for player 2; no other signals provide any information about the state. When n is large, with a high probability, none of the players learns anything about the state (even if they learn about the beliefs of the other player). Notice that the signal of each player is $\frac{1}{n+1}$-independent from the state of the world. Proposition 5 implies that $\mathbb{d}\,(u_n, u) \leq \frac{2}{n+1} \to 0$. In particular, the information structures u_n converge to the structure u, where no player receive any information.

It is instructive to provide an elementary argument for the convergence. Consider garblings q_1, q_2, such that $q_1(0)$ is uniform on $\{0,...,n\}$, and $q_2(c) = \delta_0$ for each c. Then $q_1.u = u_n.q_2$. We obtain $u \succeq u_n$, and $\mathbb{d}(u, u_n) = \min_{q_1', q_2' \in \mathcal{Q}} \|q_1'.u_n - u.q_2'\|$. Consider now $q_1' = q_2$ and q_2' such that $q_2'(0)$ is uniform on $\{0,...,n+1\}$. We get $\|q_1'.u_n - u_n.q_2'\| \leq 1/(n+1) \xrightarrow[n\to\infty]{} 0$.

H Value of experiments

In this section, we compute the value of multiple (conditionally) independent Blackwell experiments.

H.1 Blackwell experiments

We work with a binary case $k = \{0, 1\}$ and, to simplify the analysis, we suppose that two states of the world are equally probable. A (binary Blackwell) experiment is defined as a probability $p \in \left(\frac{1}{2}, 1\right)$, with the interpretation that the agent observes signal $s = k$ with probability p and signal $s = 1 - k$ with probability $1 - p$.

For each $n, m \geq 0$ and each $p, r \in \left(\frac{1}{2}, 1\right)$, we define an information structure $u_{n,m}$ in which player 1 observes outcomes of n conditionally independent copies of experiment p and player 2 observes m conditionally independent copies of experiment q. Formally, for each $k = 0, 1, c \leq n, d \leq m$, let

$$u_{n,m}(k, c, d) = \begin{cases} \frac{1}{2}\left(p^{n-c}(1-p)^c\right)\left(r^{n-d}(1-r)^d\right), & k = 0 \\ \frac{1}{2}\left(p^c(1-p)^{n-c}\right)\left(r^d(1-r)^{n-d}\right) & k = 1 \end{cases},$$

Here, c is the number of 1s observed by player 1 and d is the number of 1s observed by player 2. The goal of this section is to compute

$$\mathbb{d}(u_{l,m}, u_{n,m})$$

for some $l < n$ and m. By Proposition 2, we have

$$\mathbb{d}(u_{l,m}, u_{n,m}) = \mathbb{d}_1(u_l, u_n), \tag{17}$$

where u_l and u_n are defined from $u_{l,m}$ and $u_{n,m}$ by taking marginals over state and information of player 1.

H.2 Result

We define S_n a random variable following a binomial $\mathcal{B}(n,p)$, and $D_n = 2S_n - n$. The interpretation is that S_n represents the successes of the Blackwell experiment p (i.e., the number of outcomes that are equal to the state) and $D_n = S_n - (n - S_n)$ is the difference between the number of successes and failures. It tuns out that we can represent (17) in terms of the probability distribution of D_n.

Proposition 10. *For each n, m, we have*

$$\mathbb{d}_1(u_n, u_l) = \max_{d \in \{0,\ldots,l\}} \gamma^d_{n,l}, \text{ where} \tag{18}$$
$$\gamma^d_{n,l} = 2(1 - q_d)(\mathbb{P}(D_n > d) - \mathbb{P}(D_l > d)) - 2q_d(\mathbb{P}(D_n < -d) - \mathbb{P}(D_l < -d)).$$

If either (a) n and l have the same parity or (b) n odd and l even, the maximum in (18) is achieved for $d^* = 0$:

$$\mathbb{d}_1(u_n, u_l) = \mathbb{P}(D_n > 0) - \mathbb{P}(D_l > 0) - \mathbb{P}(D_n < 0) + \mathbb{P}(D_l < 0).$$

Finally, if $n = 2$ and $l = 1$, then the maximum is obtained for $d^* = 1$, and

$$\mathbb{d}_1(u_1, u_2) = 2p(1 - p)(2p - 1).$$

We conjecture, although we are not able to show it, that if n is even and l is odd, then the maximum in (18) is achieved for $d^* = 1$:

$$\mathbb{d}_1(u_n, u_l) = 2(1 - p)(\mathbb{P}(D_n > 1) - \mathbb{P}(D_l > 1)) - 2p(\mathbb{P}(D_n < -1) - \mathbb{P}(D_l < -1)).$$

We have some more explicit formulas for small number of experiments. Quite surprisingly, although $u_1 \preceq u_2 \preceq u_3 \preceq u_4$, we have:

$$\mathbb{d}_1(u_2, u_1) = \mathbb{d}_1(u_3, u_1) = \mathbb{d}_1(u_3, u_2) = \mathbb{d}_1(u_4, u_2) = 2p(1-p)(2p-1).$$

And simple computations give:

$$\begin{aligned}
\mathbb{d}_1(u_4, u_1) &= 2p(1-p)(2p-1)(1+3p-3p^2), \\
\mathbb{d}_1(u_4, u_3) &= 6p^2(1-p)^2(2p-1), \\
\forall n, \mathbb{d}_1(u_n, u_0) &= \mathbb{P}(D_n > 0) - \mathbb{P}(D_n < 0), \\
\forall n \geq 1,\ \mathbb{d}_1(u_n, u_1) &= 2(1-p)\mathbb{P}(D_n > 1) - 2p\mathbb{P}(D_n < -1).
\end{aligned}$$

H.3 Proof of Proposition 10

We associate each element of $\Delta(K)$ with the probability $p \in [0,1]$ that the state is 1. Define for each integer d in $\mathbb{Z}$,

$$q_d = \frac{p^d}{p^d + (1-p)^d} = 1 - q_{-d} \in [0,1].$$

q_d is the belief of player 1 that $k = 1$ after observing s positive experiments and $s - d$ negative experiments, whatever is s. Under u_n, the law of the difference d is given by:

$$\mathbb{P}_n(d) = \frac{1}{2}\binom{n}{\frac{n+d}{2}} p^{(n-d)/2}(1-p)^{(n-d)/2}(p^d + (1-p)^d) = \mathbb{P}_n(-d) = \frac{1}{2}(\mathbb{P}(D_n = d) + \mathbb{P}(D_n = -d)),$$

for all integers d in $[-n, n]$ such that $n + d$ is even.

By (23), we know that:

$$\mathbb{d}_1(u_l, u_n) = \max_{f \in D} \left| \tilde{u}_n(f) - \tilde{u}_l(f) \right|, \tag{19}$$

where $\tilde{u}_n(f) = \sum_{d=-n}^{n} f(q_d)\mathbb{P}_n(d)$ is the expectation of f with respect to the probability induced by u_n on the a posteriori. Similarly, $\tilde{u}_l(f) = \sum_{d=-n}^{n} f(q_d)\mathbb{P}_l(d)$, so $\tilde{u}_n(f) - \tilde{u}_l(f) = \sum_{d=-n}^{n} f(q_d)(\mathbb{P}_n(d) - \mathbb{P}_l(d))$. Let us first consider the optimization problem :

$$e(u_l, u_n) = \max_{f \in E} \Big| \sum_{d=-n}^{n} f(q_d)(\mathbb{P}_n(d) - \mathbb{P}_l(d)) \Big|, \tag{20}$$

where E is the set of convex 1-Lipschitz functions from $(\Delta(K), \|.\|_1)$ to $\mathbb{R}$. Since $u_n \succeq u_l$, $e(u_l, u_n) = \max_{f \in E} \sum_{d=-n}^{n} f(q_d)(\mathbb{P}_n(d) - \mathbb{P}_l(d))$. W.l.o.g we can restrict attention to functions f in E such that $f(1/2) = 0$ and $f(q_d) = f(q_{-d})$ for each d (otherwise consider $h(p) = (f(p) + f(1-p))/2$ for each p). Define for each d in $\{0, ..., n-1\}$, the right slope of f at q_d:

$$t_d = \frac{f(q_{d+1}) - f(q_d)}{q_{d+1} - q_d}.$$

Since $n > l$, it is optimal to choose $t_d = 1$ for $d \geq l$. The quantity $\sum_{d=-n}^{n} f(q_d)(\mathbb{P}_n(d) - \mathbb{P}_l(d))$ only depends on f though the slopes $t_0, ..., t_{n-1}$, and is affine in $t_0, ..., t_{n-1}$. The constraints are $t_d \in [0, 1]$ and $t_{d+1} \geq t_d$ for each d, so for problem (20) there is an optimal f and $d^* \in \{0, ..., l\}$ such that $t_d = 0$ for $0 \leq d < d^*$ and $t_d = 1$ for $d \geq d^*$. Consequently, we can restrict attention to the family of functions (f_{d^*}) from $\Delta(K)$ to $[0, 1]$ such that:

$$f_{d^*}(q) = \begin{cases} |2q - 1| & \text{if} \quad |q - 1/2| \geq |q_{d^*} - 1/2| \\ 2q_{d^*} - 1 & \text{if} \quad |q - 1/2| \leq |q_{d^*} - 1/2| \end{cases}$$

Since every function in the family belongs to D, we obtain:

$$\mathbb{d}_1(u_l, u_n) = \max_{d^* = 0, \ldots, l} \sum_{d=-n}^{n} f_{d^*}(q_d)(\mathbb{P}_n(d) - \mathbb{P}_l(d)). \tag{21}$$

It turns out that $\sum_{d=-n}^{n} f_{d^*}(q_d)\mathbb{P}_n(d)$ is easy to compute:

Lemma 6.

$$\sum_{d=-n}^{n} f_{d^*}(q_d)\mathbb{P}_n(d) = (2q_{d^*} - 1) + 2\mathbb{P}(D_n > d^*)(1 - q_{d^*}) - 2q_{d^*}\mathbb{P}(D_n < -d^*).$$

Proof.

$$\begin{aligned}
\sum_{d=-n}^{n} f_{d^*}(q_d)\mathbb{P}_n(d) &= 2q_{d^*} - 1 + 4\sum_{d=d^*+1}^{n}(q_d - q_{d^*})\mathbb{P}_n(d) \\
&= 2q_{d^*} - 1 - 2q_{d^*}\mathbb{P}_n(|d| > d^*) + 2\mathbb{P}(D_n > d^*), \\
&= 2q_{d^*} - 1 - 2q_{d^*}(\mathbb{P}(D_n > d^*) + \mathbb{P}(D_n < -d^*)) + 2\mathbb{P}(D_n > d^*), \\
&= 2q_{d^*} - 1 + 2\mathbb{P}(D_n > d^*)(1 - q_{d^*}) - 2q_{d^*}\mathbb{P}(D_n < -d^*).
\end{aligned}$$

□

Remark: Consider the decision problem where: (k, c) is selected according to u_n, player 1 receives c and has to choose i in $\{-1, 0, 1\}$, with payoff $g(k, 0) = 2q_{d^*} - 1$, $g(1, 1) = g(0, 0) = 1$ and $g(1, 0) = g(0, 1) = -1$. The following strategy is optimal: given c, compute the belief q that $k = 1$. Then if $q \in [q_{-d^*}, q_{d^*}]$ play the safe action 0, if $q > q_{d^*}$ play $i = 1$ and if $q < q_{-d^*}$ play $i = -1$. The payoff of this strategy is precisely $\sum_{d=-n}^{n} f_{d^*}(q_d)\mathbb{P}_n(d)$.

The characterization (18) follows from equation (21) and lemma 6.
Assume now that n and l have the same parity, that is $n - l$ is even. For all d in $\{-l, ..., +l\}$, $\mathbb{P}_n(d) \leq \mathbb{P}_l(d)$, so the maximum in 21 is simply achieved for $d^* = 0$.

Finally, suppose that n is odd and l is even. We show that the maximum in equation (21) is achieved for $d^* = 0$. We only need to consider d even in $\{0, .., l\}$, and for each such d we define $h_d = f_d - f_0$. We have

$$\tilde{u}_n(h_d) = 2\sum_{d'=1}^{d-1}\mathbb{P}(|D_n| = d')(q_d - q_{d'}),\ \tilde{u}_l(h_d) = 2\sum_{d'=0}^{d-2}\mathbb{P}(|D_l| = d')(q_d - q_{d'}),$$

and it is enough to show that $\tilde{u}_n(h_d) \leq \tilde{u}_l(h_d)$. We are going to show that :

$$\forall d' = 0, 2, ..., d-2, \;\; \mathbb{P}(|D_l| = d')(q_d - q_{d'}) \geq \mathbb{P}(|D_n| = d'+1)(q_d - q_{d'+1}) \qquad (22)$$

We have

$$\begin{aligned}
\frac{\mathbb{P}(|D_l| = d')(q_d - q_{d'})}{\mathbb{P}(|D_n| = d'+1)(q_d - q_{d'+1})} &= \frac{q_d - q_{d'}}{q_d - q_{d'+1}} \frac{(p^{d'} + (1-p)^{d'})}{(p^{d'+1} + (1-p)^{d'+1})} (p(1-p))^{(l-n+1)/2} \frac{\binom{l}{(l+d')/2}}{\binom{n}{(n+d'+1)/2}}, \\
&\geq \frac{1 - q_{d'}}{1 - q_{d'+1}} \frac{(p^{d'} + (1-p)^{d'})}{(p^{d'+1} + (1-p)^{d'+1})} (p(1-p))^{(l-n+1)/2} \frac{\binom{l}{(l+d')/2}}{\binom{n}{(n+d'+1)/2}}, \\
&= (1-p)^{-1} (p(1-p))^{(l-n+1)/2} \frac{\binom{l}{(l+d')/2}}{\binom{n}{(n+d'+1)/2}}.
\end{aligned}$$

Since $p \in [1/2, 1]$, the minimum in p is achieved for $p = 1/2$, so that:

$$\frac{\mathbb{P}(|D_l| = d')(q_d - q_{d'})}{\mathbb{P}(|D_n| = d'+1)(q_d - q_{d'+1})} \geq 2^{n-l} \frac{\binom{l}{(l+d')/2}}{\binom{n}{(n+d'+1)/2}}.$$

We finally show that $2^{n-l}\binom{l}{(l+d')/2} \geq \binom{n}{(n+d'+1)/2}$ for all $d' = 0, ..., l-2$. This inequality is true for $n = l+1$, and $\frac{1}{2^n}\binom{n}{(n+d'+1)/2}$ decreases in n (for n even not smaller than $d'+1$).

Assume now that $n = 2$ and $l = 1$. The optimal f in equation 21 should put maximal weight for $q = q_0, q_2, q_{-2}$ and minimal weight for q_{-1}, q_1. One easily checks that the maximum is obtained for $d^* = 1$, and $d(u_1, u_2) = 2p(1-p)(2p-1)$.

I Special cases

We defined the distance $\mathbb{d}$ between *any* information structures as a maximum of value difference across *all* zero-sum games. In this Section, we consider versions of the definition, where we either restrict the space of games or information structures. Given the restrictions, we obtain a tighter characterization of the distance. Additionally, in all examples below, we show that the restricted spaces of information structures have compact completion, which has important consequences for the limit theorems in repeated games.

I.1 Single-agent decision problems

We introduced metric $\mathbb{d}_1$ to analyze the distance in single-agent decision problems. It is easy to see that the distance $\mathbb{d}_1$ depends only on the information of player 1: for any u, u', v, v' such that $\text{marg}_{K\times C}\, u = \text{marg}_{K\times C}\, u'$ with an analogous relation for v and v', we have $\mathbb{d}_1(u, v) = \mathbb{d}_1(u', v')$.

Let $\mathcal{U}_1 = \Delta(K \times \mathbb{N})$ be the set of probabilities over states and signals for player 1. From now on, we assume that u, v are elements of $\mathcal{U}_1$. Following the same method as in Theorem 1, we show that

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

and the Blackwell characterization : $u \succeq v \Leftrightarrow \exists q \in Q, q.u = v$.

Further, notice that the distance depends only on the induced distributions of conditional beliefs over K. Finally, we can show that if D is the set of suprema of affine functions from $\Delta(K)$ to $[-1, 1]$, then

$$\mathbb{d}_1(u, v) = \sup_{f\in D} \left| \int_{p\in\Delta(K)} f(p) d\tilde{u}(p) - \int_{p\in\Delta(K)} f(p) d\tilde{v}(p) \right|, \tag{23}$$

where $\tilde{u}$ and $\tilde{v}$ denote the distributions of first-order beliefs of Player 1 induced by the information structures u and v. Hence, $\mathcal{U}_1$ under $\mathbb{d}_1$ is totally bounded. Its completion $\overline{\mathcal{U}_1}$ is compact and

$$\overline{\mathcal{U}_1} \simeq \Delta(\Delta(K)).$$

I.2 One-sided full information

Let $\mathcal{U}_{OF}$ be the subset of $\mathcal{U}$ where with probability 1, the signal of player 1 reveals both the state and the signal of player 2. For $u \in \mathcal{U}_{OF}$, all what matters is the law $\tilde{u}$ of the belief of player 2 about the state. From Renault and Venel (2017), another characterization of $\mathbb{d}$ can be given for elements of $\mathcal{U}_{OF}$.

$$\mathbb{d}(u,v) = \sup_{f \in D_1} \left(\int_{p \in \Delta(K)} f(p) d\tilde{u}(p) - \int_{p \in \Delta(K)} f(p) d\tilde{v}(p) \right),$$

where

$$D_1 = \{f, \forall p, q \in \Delta(K), \forall a, b \geq 0,\ af(p) - bf(q) \leq \|ap - bq\|_1\}.$$

Hence, $\mathcal{U}_{OF}$ under $\mathbb{d}$ is totally bounded; its completion $\overline{\mathcal{U}_{OF}}$ is compact and

$$\overline{\mathcal{U}_{OF}} \simeq \Delta(\Delta(K)).$$

I.3 Public signals

Set $\mathcal{U}_P$ of information structures where both players receive the same signal: $\overline{\mathcal{U}_P}$ is compact, and homeomorphic to $\Delta(\Delta(K))$. Here given u in $\mathcal{U}_P$, what matters is the induced law $\tilde{u}$ on the common a posteriori of the players on K.

I.4 One player more informed

A generalization of the two previous cases is the set $\mathcal{U}_{OM}$ of information structures where player 1 knows the signal of player 2, i.e. when the signal of player 1 is enough to deduce the signal of player 2. $\overline{\mathcal{U}_{OM}}$ is compact, and homeomorphic to $\Delta(\Delta(\Delta(K)))$ (see Mertens (1986), Gensbittel *et al.* (2014)).

I.5 Conditionally independent signals

Set $\mathcal{U}_{CI}$ of independent information structures : $\mathcal{U}_{CI}$ is the set of u in $\mathcal{U}$ such that $u(c,d|k) = u(c|k)u(d|k)$ (the signals c and d are conditionally independent given k). Here $\overline{\mathcal{U}_{CI}}$ is homeomorphic to $\Delta(\Delta(K) \times \Delta(L))$.

J Value-based distance and bounds on equilibrium payoffs in non-zero-sum games

In this section, we show that the value-based distance between information structures contains information that is useful in some questions that are relevant for non-zero-sum games.

The section is divided into four parts. In the first part, we focus on the distance between the sets of feasible payoffs induced by an information structure and a game. More precisely, given an information structure u in $\mathcal{U}$ and a non zero-sum payoff function $g : K \times I \times J \to [-1,1]^2$ (with I, J finite), one naturally defines the non zero-sum Bayesian game $\Gamma(u,g)$ and one can ask whether $\mathbb{d}$ can be used to measure the distance between the feasible payoffs in these games. We provide a positive answer in three different cases (conditionally independent signals, public signals and one-sided full information).

Next, we present two applications. First, we show that for any game g with common interests (common payoffs for both players), if structures u and v belong to any of the three cases mentioned above, the best equilibrium payoff in $\Gamma(u,g)$ and $\Gamma(v,g)$ differ by at most $3\mathbb{d}(u,v)$.

The second application is concerned with the infinite repetition of the games $\Gamma(u,g)$ and $\Gamma(v,g)$. As is well-known, in such a repeated game the set of feasible and individually rational payoffs is the limit, when $\delta \to 1$, of the sets of δ-discounted Nash equilibrium payoffs (and under mild assumptions the limit of the sets of δ-discounted subgame-perfect Nash equilibrium payoffs). The second corollary establishes a strong connection, depending on $\mathbb{d}(u,v)$, between the sets of feasible and individually rational payoffs in $\Gamma(u, g)$ and $\Gamma(v, g)$.

Finally, we show by means of counterexample, that some assumptions on the information structures are necessary to obtain bounds on the non-zero-sum equilibrium payoffs in terms of value-based distance.

J.1 Bounds on distance between feasible payoffs

We let $F(u, g) \subset \mathbb{R}^2$ be the set of feasible payoffs of $\Gamma(u, g)$ (the convex hull of the feasible payoffs with pure strategies), and use the Hausdorff distance between non empty compact sets of $\mathbb{R}^2$, endowed with $\|.\|_{\max}$.

Proposition 11. *If the information structures u and v have conditionally independent signals, the distance between $F(u, g)$ and $F(v, g)$ is at most $3\mathrm{d}(u, v)$.*

Proof. Write $\varepsilon = \mathrm{d}(u, v)$. By Theorem 1, there exist garblings q_1, q_2, q_3, q_4 such that $\|q_1.u - v.q_2\| \leq \varepsilon$ and $\|q_3.v - u.q_4\| \leq \varepsilon$. We first show that u can be approximated by $q_3.v.q_2$.

Define $q_3.q_1$ for the garbling which given any signal c, selects c' according to $q_1(c)$ then finally c'' according to $q_3(c')$. Similarly, let $q_2.q_4(d) = \sum_{d'} q_4(d)(d')q_2(d')$ for all d. Notice that $(q_3.q_1).v = q_3.(q_1.v)$, $u.(q_2.q_4) = (u.q_4).q_2$ and $\|q.u' - q.v'\| \leq \|u' - v'\|$ for all q, u', v', we obtain:

$$
\begin{aligned}
\|q_3.q_1.u - u.(q_2.q_4)\| &\leq \|q_3.q_1.u - q_3.v.q_2\| + \|q_3.v.q_2 - (u.q_4).q_2\|, \\
&\leq \|q_1.u - v.q_2\| + \|q_3.v - u.q_4\|, \\
&\leq 2\varepsilon.
\end{aligned}
$$

In particular, projecting over states and signals for player 1 gives $\sum_{k,c} |q_3.q_1.u(k, c) -$

$u(k,c)| \leq 2\varepsilon$. Since u has conditionally independent signals, we have:

$$\begin{aligned}\|q_3.q_1.u(k,c) - u\| &= \sum_{k,c,d} |q_3.q_1.u(k,c,d) - u(k,c,d)|,\\ &= \sum_{k,c,d} |q_3.q_1.u(k,c)u(d|k) - u(k,c)u(d|k)|,\\ &= \sum_{k,c} |q_3.q_1.u(k,c) - u(k,c)| \leq 2\varepsilon.\end{aligned}$$

And we obtain $\|u - q_3.v.q_2\| \leq \|u - q_3.q_1.u\| + \|q_3.q_1.u - q_3.v.q_2\| \leq 3\varepsilon$.

Consider now a non zero-sum payoff function $g = (g_1, g_2)$ with all payoffs in $[-1, 1]$. We assume without loss of generality that for some L, the set of actions for each player in g is $I = J = \{0, ..., L-1\}$. Define the compact subsets $\mathcal{W}_L(u) = \{q'_1.u.q'_2,\ q'_1, q'_2 \in \mathcal{Q}(L)\}$ and $\mathcal{W}_L(v) = \{q'_1.v.q'_2, q'_1,\ q'_2 \in \mathcal{Q}(L)\}$ of $\Delta(K \times I \times J)$. The set of feasible payoffs can be written:

$$F(u,g) = \{(\langle g_1, w\rangle, \langle g_2, w\rangle), w \in \operatorname{conv} \mathcal{W}_L(u)\},$$

$$F(v,g) = \{(\langle g_1, w\rangle, \langle g_2, w\rangle), w \in \operatorname{conv} \mathcal{W}_L(v)\}.$$

To show that the distance between $F(u,g)$and $F(v,g)$ is at most 3 ε, it is enough to prove that the distance (for $\|.\| = \|.\|_1$) between $\operatorname{conv} \mathcal{W}_L(u)$ and $\operatorname{conv} \mathcal{W}_L(v)$, or simply between $\mathcal{W}_L(u)$ and $\mathcal{W}_L(v)$, is at most 3 ε.

To conclude, let u' be in $\mathcal{W}_L(u)$, u' can be written $u' = q'_1.u.q'_2$ for some q'_1, q'_2 in $\mathcal{Q}(L)$. Since $\|u - q_3.v.q_2\| \leq 3\varepsilon$, we have $\|u' - q'_1.q_3.v.(q'_2.q_2)\| \leq 3\varepsilon$ and the distance from u' to $\mathcal{W}_L(v)$ is[18] at most 3 ε. □

We have an analog of the proposition for information structures with public signals.

[18]The proof shows that the distance between the sets of feasible payoffs using mixed strategies in $\Gamma(u,h)$ and $\Gamma(v,h)$ also differ by at most $3\mathrm{d}(u,v)$.

Proposition 12. *If the information structures u and v have public signals, the distance between $F(u,g)$ and $F(v,g)$ is at most $\mathbb{d}(u,v)$.*

Proof. We assume that both players receive the same signals in each information structure u, v. From the proof of proposition 11, it is sufficient to show that for each L and $u' = q_1'.u.q_2'$ with q_1', q_2' in $\mathcal{Q}(L)$, the distance, with respect to $\|.\| = \|.\|_1$, from u' to the set $\operatorname{conv} \mathcal{W}_L(v)$ is at most $\mathbb{d}(u,v)$.

We have the existence of garblings q, q' such that $\|q.v - u.q'\| \leq \mathbb{d}(u,v)$, and projecting over states and signals for player 1 gives : $\sum_{k,c} |q.v(k,c) - u(k,c)| \leq \mathbb{d}(u,v)$. Define now the information structures w and u' by:

$$\forall k,c,d,\ w(k,c,d) = q.v(k,c)\,\mathbb{1}_{d=c} \quad \text{and} \quad v' = q_1'.w.q_2'.$$

We have $\|u-w\| = \sum_{k,c} |w(k,c)-u(k,c)| \leq \mathbb{d}(u,v)$, so w is a good approximation of u. Now, $\|u' - v'\| \leq \|u - w\| \leq \mathbb{d}(u,v)$, and it is enough to show that v' belongs to $\operatorname{conv} \mathcal{W}_L(v)$.

We first assume that the set of signals having positive probability under v is finite and written C. The garbling q selects, independently for each signal c in C, an element of $\mathbb{N}$ with probability $q(c)$. Define now for any map $s : C \to \mathbb{N}$,

$$\lambda_s = \prod_{c \in C} q(c)(s(c)).$$

λ_s is the probability that for each c in C, q chooses $s(c)$ if the signal is c. Since C is finite there are countably many such maps s, and $\sum_s \lambda_s = 1$. Define also $v_s = s.v.s$ in $\mathcal{W}_\infty(v)$ as: (k, c', d') is selected according to v then the signals received by the players are $c = s(c'), d = s(d')$. We have:

$$\begin{aligned}\sum_s \lambda_s v_s(k,c,d) &= \sum_s \lambda_s \sum_{c'} v(k,c')\, \mathbb{1}_{c=d=s(c')}, \\ &= \mathbb{1}_{c=d} \sum_{c'} v(k,c') \sum_s \lambda_s\, \mathbb{1}_{c=s(c')}, \\ &= \mathbb{1}_{c=d} \sum_{c'} v(k,c') q(c')(c), \\ &= w(k,c,d).\end{aligned}$$

Then $v' = q_1'.w.q_2' = \sum_s \lambda_s\, q_1'.v_s.q_2'$. For each s, $q_1'.v_s.q_2'$ belongs to $\mathcal{W}_L(v)$, hence $v' \in \operatorname{conv} \mathcal{W}_L(v)$.

Finally, if the set of signals in v is infinite, we consider a sequence (v_n) with finite support such that $\|v - v_n\| \to 0$, and the corresponding sequence (v_n') will have a limit point v' in $\operatorname{conv} \mathcal{W}_L(v)$ satisfying $\|u' - v'\| \leq \mathrm{d}(u,v)$. □

We also have an analog of the proposition for information structures with one-sided full information.

Proposition 13. *If the information structures u and v have one-sided full information, the distance between $F(u,g)$ and $F(v,g)$ is at most $\mathrm{d}(u,v)$.*

Proof. Consider u, v with one-sided full information for player 1. As in the previous proofs, it is sufficient to prove that for each L and q_1', q_2' in $\mathcal{Q}(L)$, the distance between $u' = q_1'.u.q_2'$ and $\operatorname{conv} \mathcal{W}_L(v)$ is at most $\mathrm{d}(u,v)$, and we assume w.l.o.g. that the set of signals for player 2 under v is the finite set D. There exist maps $f : \mathbb{N} \to K$, $h : \mathbb{N} \to D$ such that under v, if player 1's signal is c then the state is $k = f(c)$ and the signal of player 2 is $d = h(c)$. What is important here is the law induced over states and signals for player 2, and one might think as if player 1's signal was the pair (state, signal for player 2).

By Theorem 1 we have the existence of garblings q, q' such that $\|v.q - q'.u\| \leq \mathrm{d}(u,v)$, and projecting gives : $\sum_{k,d} |v.q(k,d) - u(k,d)| \leq \mathrm{d}(u,v)$. Define now the information structures w and v' by:

$$\forall k, c, d, w(k,c,d) = v.q(k,d)\ u(c|k,d) \text{ and } v' = q_1'.w.q_2'.$$

$\|u' - v'\| \leq \|u - w\| = \sum_{k,c,d} u(c|k,d)|u(k,d) - v.q(k,d)| \leq \mathbb{d}(u,v)$, and the goal is now to write v' as an element of conv $\mathcal{W}_L(v)$.

Given a map $s : D \to \mathbb{N}$, define:

- the information structure v_s in $\mathcal{W}_\infty(v)$ by: first select (k, c', d') according to v, then the state is k, the signal for player 2 is $s(d')$ and the signal of player 1 is chosen according to u conditionally on the state being $f(c')$ and the signal of player 2 being $s(h(c'))$, i.e. player 1 receives the signal c with probability $u(c|f(c'), s(h(c'))$.

- the probability $\lambda_s = \prod_{d \in D} q(d)(s(d))$ that for each d in D, q chooses $s(d)$ if the signal is d.

$$\begin{aligned}
\sum_s \lambda_s v_s(k,c,d) &= \sum_s \lambda_s \sum_{c',d'} v(k,c',d')\, \mathbb{1}_{d=s(d')}\, u(c|f(c'), s(h(c')), \\
&= \sum_{c',d'} v(k,c',d') \sum_s \lambda_s\, \mathbb{1}_{d=s(d')}\, u(c|k,d), \\
&= \sum_{d'} v(k,d')\ q(d')(d)\ u(c|k,d), \\
&= w(k,c,d).
\end{aligned}$$

For each s, $q_1'.v_s.q_2'$ belongs to $\mathcal{W}_L(v)$ and we obtain $v' = \sum_s \lambda_s q_1'.v_s.q_2' \in \text{conv}\, \mathcal{W}_L(v)$, completing the proof. □

J.2 Application to common interest games

Propositions 11, 12 and 13 directly imply the following result for games with common interests, where the best equilibrium payoff and the best feasible (in pure, mixed or correlated strategies) payoff coincide.

Corollary 4. *Consider a non zero-sum payoff function $g = (g_1, g_1)$ with common* *payoffs for the players, and information structures u and v satisfying the assumptions of at least one of the propositions 11, 12, 13. The best equilibrium payoff for player 1 in $\Gamma(u,g)$ is at most $3\mathbb{d}(u,v)$ from the best equilibrium payoff for player 1 in $\Gamma(v,g)$.*

J.3 Application to repeated non-zero-sum games

The propositions also imply that the sets of feasible and individually rationals payoffs of $\Gamma(u,g)$ and $\Gamma(v,g)$ are closely related if $\mathbb{d}(u,v)$ is small. In the following corollary, we denote by $m_1(u,g) = \text{val}(u,g_1)$ and $m_2(u,g) = -\text{val}(u,-g_2)$ the respective independent minmax of the players in the game $\Gamma(u,g)$.

Corollary 5. *Consider a non zero-sum payoff function g, and information structures u and v satisfying the assumptions of at least one of the propositions 11, 12, 13. Let $x = (x_1, x_2)$ be a feasible payoff in the game $\Gamma(u,g)$ satisfying $x_i \geq m_i(u,g)+4\mathbb{d}(u,v)$ for $i = 1,2$. Then x is $3\mathbb{d}(u,v)$-close to a payoff which is feasible and individually rational in $\Gamma(v,g)$.*

Proof. By definition of the value-based distance, $|m_i(u,g) - m_i(v,g)| \leq \mathbb{d}(u,v)$ for each player i. By one of the propositions, x is $3\mathbb{d}(u,v)$-close to a payoff y in $F(v,g)$. For each $i = 1,2$, $y_i \geq x_i - 3\mathbb{d}(u,v) \geq m_i(u,g)+\mathbb{d}(u,v) \geq m_i(v,g)$, so y is individually rational in the game $\Gamma(v,g)$. □

Remark: From the proofs of the propositions, one can see that the conclusion of corollary 5 holds more generally as soon as : u has conditionally independent signals and v is arbitrary, or v has one-sided full information and u is arbitrary, or both u and v have public signals. Same for corollary 4 with the conclusion: the best equilibrium payoff in $\Gamma(v,g)$ is at least the best equilibrium payoff in $\Gamma(u,g)$ minus $3\mathbb{d}(u,v)$.

J.4 Counterexample

We finally provide a counter-example to the three propositions and the two corollaries when under u player 1 is not more informed and the signals are conditionally dependent.

Example 7. $K = C = D = \{0, 1\}$. *Under* u, *the signals* c *and* d *are uniformly and independently distributed, and* $k = c + d \mod 2$. *Under* v, $c = d = 0$, *and* k *is uniformly selected. It is easy to see that* $\mathbb{d}(u, v) = 0$. *However, consider a non zero-sum payoff function* h *with* $g_1(k, i, j) = g_2(k, i, j) = 1$ *if* $k = i + j \mod 2$ *and* $g_1(k, i, j) = g_2(k, i, j) = -1$ *otherwise. The payoff* $x = (1, 1)$ *is feasible in the game* $\Gamma(u, g)$, *but no payoff with positive coordinates is feasible in* $\Gamma(v, g)$.

K Other properties of characterization

K.1 Comparison of information

The following Corollary is a direct corollary of Theorem 1:

Corollary 6. *For all information structures* u, v,

$$\sup_{g \in \mathcal{G}} \left(\text{val}(v, g) - \text{val}(u, g)\right) = \inf_{u' \preceq u, v' \succeq v} \|u' - v'\|. \tag{24}$$

This observation provides an additional interpretation to the characterization from Theorem 1: the maximum gain from replacing information structure u by v is equal to the minimum total variation distance between the set of information structures that are worse than u and those that are better than v.

K.2 Optimal strategies

Another useful property of Theorem 1 is that the garblings in equation 2 can be used to transform optimal strategies in one structure to approximately optimal strategies on another structure. Moreover, the transformation does not depend on the particular payoffs considered.

More precisely, we say that strategy σ of player 1 is ε-optimal in game g on structure u if for any strategy τ of player 2, the payoff of σ against τ is no smaller than $\text{val}\,(u, g) - \varepsilon$. We similarly define ε-optimal strategies for player 2.

For a strategy $\sigma \in \mathcal{Q}$ and a garbling $q_1 \in \mathcal{Q}$, define $\sigma.q_1$ in $\mathcal{Q}$ by $\sigma.q_1(c) = \sum_{c'} q_1(c'|c)\sigma(c')$ for each signal c: player 1 receives signal c, then selects χ according to $q_1(c)$ and plays $\sigma(\chi)$. We have

Proposition 14. *Fix u, v in $\mathcal{U}$ and let q_1 and q_2 in $\mathcal{Q}$ satisfy*

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

Then, if σ is an optimal strategy in g on v, then $\sigma.q_1$ is a $2\mathbb{d}\,(u, v)$-optimal strategy in g on u. Similarly, if τ is optimal for player 2 in g on u, then $\tau.q_2$ is $2\mathbb{d}\,(u, v)$-optimal for player 2 in g on v.

Proof. Consider σ an optimal strategy in g on v, and τ in $\mathcal{Q}$ arbitrary. The payoff induced by $(\sigma.q_1, \tau)$ in the game (u, g) is:

$$\begin{aligned}
\sum_{k,c,d} u(k, c, d) g(k, \sigma.q_1(c), \tau(d)) &= \sum_{k,c,d} q_1.u(k, c, d) g(k, \sigma(c), \tau(d)), \\
&\geq \sum_{k,c,d} v.q_2(k, c, d) g(k, \sigma(c), \tau(d)) - \|q_1.u - v.q_2\|, \\
&\geq \sum_{k,c,d} v(k, c, d) g(k, \sigma(c), \tau.q_2(d)) - d(u, v), \\
&\geq \mathrm{val}(v, g) - \mathbb{d}(u, v), \\
&\geq \mathrm{val}(u, g) - 2\mathbb{d}(u, v).
\end{aligned}$$

The dual property is proved similarly. □

L Uncountable information structures

L.1 Uncountable information structures

Let K be a compact metric space endowed with its Borel σ-algebra $\mathcal{B}(K)$. An (uncountable) information structure over K is defined as a pair of measurable spaces $(S_1, \mathcal{A}_1)$ and $(S_2, \mathcal{A}_2)$ and a probability measure u over the product measurable space

$K \times S_1 \times S_2$. Here, S_1 denotes the set of signals observed by Player 1, and S_2 the set of signals observed by Player 2.

Due to the construction of Mertens and Zamir (1985) of the universal belief space $\Omega = K \times \Theta_1 \times \Theta_2$, one can associate to u a unique consistent probability $P \in \Pi \subset \Delta(\Omega)$, such that the induced canonical information structure on the universal belief space (in which the signal of player i is her hierarchy of beliefs in Θ_i) is equivalent to u in the sense that

$$\forall g \in \mathcal{G}, \mathrm{val}(u, g) = \mathrm{val}(P, g).$$

(see e.g. Theorem III.2.4 and Propositions III.4.2 and III.4.4 in Mertens *et al.* (2015). In particular, the value is well-defined for any consistent probability and for any information structure.) Thus, in order to capture all the equivalence classes of arbitrary measurable information structures (possibly uncountable), it is sufficient to consider canonical ones, with signals being a hierarchy of beliefs. However, considering consistent probabilities is sometimes confusing in the sense that when modifying a canonical information structure with a garbling from Θ_i to Θ_i, the modified information structure is in $\Delta(\Omega)$ but not in general a consistent probability. Instead of working with information structures in $\Delta(\Omega)$ (as it is often done in Mertens *et al.* (2015) or Gossner and Mertens (2001)) and in order to avoid any confusions, we find convenient to use the set $\mathcal{U}_c$ of information structures with signals in $[0, 1]$. This set is sufficiently large in the sense that any measurable information structure is equivalent to an element of $\mathcal{U}_c$ (since it is equivalent to a canonical information structure in Π, which is itself equivalent to an element of $\mathcal{U}_c$ by relabeling the signals), and is stable by transformations based on garblings from $[0, 1]$ to $[0, 1]$.

Define $\mathcal{U}_c = \Delta(K \times [0, 1] \times [0, 1])$ as the set of information structures where signals of the players are in $[0, 1]$. Let $\mathcal{U}_c^*$ be the set of equivalence classes for the relation

$$u \simeq v \Leftrightarrow \forall g \in \mathcal{G}, \mathrm{val}(u, g) = \mathrm{val}(v, g).$$

$\mathcal{U}^*$ can be identified as the subset of $\mathcal{U}_c^*$ made by equivalence classes which contain an element with countable support.

Let Φ be the map from $\mathcal{U}_c$ to Π which associates to the information structure u the induced consistent probability over the (Mertens-Zamir) universal belief space Ω. This map is constant over equivalence classes since the value $\mathrm{val}(u, g)$ depends only on $\Phi(u)$ for every $g \in \mathcal{G}$, and thus Φ induces a map from $\mathcal{U}_c^*$ to Π, also denoted Φ. Moreover, Φ is one-to-one since the value functions of finite games separate points in Π (see Theorem 12 in Gossner and Mertens (2001)). We claim that this map is also onto. Indeed, for every $\mu \in \Pi$, one can associate a canonical information structure. Then, since the set of coherent belief hierarchies Θ_i is compact metric, there exists a Borel isomorphism ψ_i from Θ_i to $[0, 1]$[19], which allows to relabel the signals. The probability u over $(k, \psi_1(\theta_1), \psi_2(\theta_2))$ is such that $\Phi(u) = \mu$ by construction and this proves the claim.

L.2 Total variation norm and topologies on garblings

Let us recall a few facts about the total variation distance.

Let X denote a compact metric space, $\mathcal{F}_1(X)$ denote the set of (Borel) measurable functions from X to $[-1, 1]$ and $\mathcal{C}_1(X)$ denote the set of continuous functions from X to $[-1, 1]$. Then, the total variation norm satisfies

$$\begin{aligned}
\forall \mu, \nu \in \Delta(X),\ \|\mu - \nu\|_{TV} &= \sup_{A \in \mathcal{B}(X)} |\mu(A) - \nu(A)| \\
&= \frac{1}{2} \sup_{f \in \mathcal{F}_1(X)} \int_X f d(\mu - \nu) \\
&= \frac{1}{2} \sup_{f \in \mathcal{C}_1(X)} \int_X f d(\mu - \nu)
\end{aligned}$$

The first equality is the definition, the second is a classical exercise, and the third is obtained by using a standard approximation argument (e.g. by using that continuous functions are dense in $L^1(\mu + \nu)$ and a truncation argument).

[19] According to the Borel isomorphism theorem, all the uncountable standard Borel spaces are Borel isomorphic. Standard Borel spaces being Borel subsets of complete separable metric spaces, compact metric spaces are standard Borel spaces.

We also have a useful formula. Let π be another positive σ-finite measure such that μ and ν are absolutely continuous with respect to π (e.g., choose $\pi = \mu + \nu$). Then

$$2\|\mu - \nu\|_{TV} = \int_X \left| \frac{d\mu}{d\pi} - \frac{d\nu}{d\pi} \right| d\pi.$$

Indeed, the right-hand side is an upper bound since for any $f \in \mathcal{F}_1(X)$

$$\int_X f d(\mu - \nu) = \int_X f \left(\frac{d\mu}{d\pi} - \frac{d\nu}{d\pi} \right) d\pi,$$

and one may choose $f = sgn(\frac{d\mu}{d\pi} - \frac{d\nu}{d\pi})$ to prove that it is attained.

Remark: When X is countable, we have $\|\mu - \nu\|_{TV} = \frac{1}{2}\|\mu - \nu\|_1$.

Let now X and Y denote compact metric spaces and fix $\mu \in \Delta(X)$. Let $\mathcal{T}$ denote the set of equivalence classes of transitions probabilities from X to Y with respect to the relation of equality μ almost everywhere. Recall that a transition probability is simply a Borel measurable map from X to $\Delta(Y)$ when $\Delta(Y)$ is endowed with the weak topology.

$\mathcal{T}$ is a compact metrizable space when endowed with the topology τ_μ defined by

$$q_n \to q \Leftrightarrow \forall f \in \mathcal{L}, \int_X \left(\int_Y f(x,y) dq^n(x)(y) \right) d\mu(x) \to \int_X \left(\int_Y f(x,y) dq(x)(y) \right) d\mu(x),$$

where $\mathcal{L}$ denotes the set of bounded measurable maps on $X \times Y$ that are continuous with respect to the second variable (see e.g. Theorems 2.2 and 2.3 in Balder (1988)).

L.3 Extension of Theorem 1

Let $C = D = [0,1]$, and let us consider the class of game payoffs $\mathcal{G}_c$ of continuous maps from $K \times C \times D$ to $[-1,1]$. Recall that from Proposition III.4.2 in Mertens *et al.* (2015), for every $u \in \mathcal{U}_c$ and every $g \in \mathcal{G}_c$, the value $\text{val}(u,g)$ is well defined and both players have optimal strategies.

Let $\mathcal{Q}_c$ denote the set of measurable functions from $[0,1]$ to $\Delta([0,1])$.

Given $u \in \mathcal{U}_c$ and $q \in \mathcal{Q}_c$, define $q.u$ and $u.q$ as before, i.e. as the unique probability measures that satisfy for every bounded measurable function f:

$$\int_{K\times C\times D} f d(q_1.u) = \int_{K\times C\times D} \left(\int_C f(k,c',d)dq_1(c'|c)\right) du(k,c,d).$$

$$\int_{K\times C\times D} f d(u.q_2) = \int_{K\times C\times D} \left(\int_D f(k,c,d')dq_2(d'|d)\right) du(k,c,d).$$

Theorem 6. ***(Extension of Theorem 1)*** *For all $u, v \in \mathcal{U}_c$:*

$$\begin{aligned}\sup_{g\in\mathcal{G}}(\mathrm{val}(v,g) - \mathrm{val}(u,g)) &= \sup_{g\in\mathcal{G}_c}(\mathrm{val}(v,g) - \mathrm{val}(u,g)) \\ &= 2 \min_{q_1,q_2\in\mathcal{Q}_c} \|q_1.u - v.q_2\|_{TV}.\end{aligned}$$

As for Theorem 1, the result of Peski (2008) can be extended to arbitrary information structures. Precisely, for any $u, v \in \mathcal{U}_c$, 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 $q \in \mathcal{Q}_c$ and Theorem 6 implies the following result.

Corollary 7. *For all $u, v \in \mathcal{U}_c$, $u \succeq v \iff$ there exists q_1, q_2 in $\mathcal{Q}_c$ s.t. $q_1.u = v.q_2$.*

Remark 1. *It is important to recall that signals sets can be replaced by any other standard Borel space (by relabeling) for the following equality to hold*

$$\sup_{g\in\mathcal{G}}(\mathrm{val}(v,g) - \mathrm{val}(u,g)) = 2 \min_{(q_1,q_2)\in\mathcal{Q}_1\times\mathcal{Q}_2} \|q_1.u - v.q_2\|_{TV},$$

where $\mathcal{Q}_1$ and $\mathcal{Q}_2$ are sets of transition probabilities between appropriate spaces of signals. The same remark applies to Corollary 7.

L.3.1 Proof of Theorem 6

We first prove the second equality. Note that for any $g \in \mathcal{G}_c$, the duality product $\langle g, u\rangle = \int_{K\times C\times D} g du$ is well defined and corresponds to the payoff $\gamma_{u,g}(Id, Id)$, where

$Id \in \mathcal{Q}_c$ 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,

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

Since both players can play the Id strategy in $\Gamma(u, g)$, we have for all $u \in \mathcal{U}_c$ and $g \in \mathcal{G}_c$ that

$$\inf_{q_2 \in \mathcal{Q}_c} \langle g, u.q_2 \rangle \leq \text{val}(u, g) \leq \sup_{q_1 \in \mathcal{Q}_c} \langle g, q_1.u \rangle.$$

Notice also that for all u, v in $\mathcal{U}_c$,

$$2\|u - v\|_{TV} = \sup_{g \in \mathcal{G}_c} \langle g, u - v \rangle.$$

Fix u, v in $\mathcal{U}_c$. For $g \in \mathcal{G}_c$, we have

$$\inf_{q_1, q_2 \in \mathcal{Q}_c} \langle g, v.q_2 - q_1.u \rangle \leq \text{val}(v, g) - \text{val}(u, g),$$

so

$$\sup_{g \in \mathcal{G}_c} (\text{val}(v, g) - \text{val}(u, g)) \geq \sup_{g \in \mathcal{G}_c} \inf_{q_1, q_2 \in \mathcal{Q}_c} \langle g, v.q_2 - q_1.u \rangle. \tag{25}$$

For $g \in \mathcal{G}_c$, $q_1, q_2 \in \mathcal{Q}_c$, by monotony of the value with respect to information, we have

$$\text{val}(v.q_2, g) \geq \text{val}(v, g) \text{ and } \text{val}(u, g) \geq \text{val}(q_1.u, g),$$

so that

$$\text{val}(v, g) - \text{val}(u, g) \leq d\,(q_1.u, v.q_2) \leq 2\|q_1.u - v.q_2\|_{TV}.$$

Hence

$$\sup_{g\in\mathcal{G}_c}\left(\mathrm{val}(v,g)-\mathrm{val}(u,g)\right) \leq \inf_{q_1,q_2\in\mathcal{Q}_c} 2\|q_1.u - v.q_2\|_{TV} \tag{26}$$

$$= \inf_{q_1,q_2\in\mathcal{Q}_c}\sup_{g\in\mathcal{G}_c}\langle g, v.q_2 - q_1.u\rangle. \tag{27}$$

We are now going to show that

$$\sup_{g\in\mathcal{G}_c}\inf_{q_1,q_2\in\mathcal{Q}_c}\langle g, v.q_2 - q_1.u\rangle = \inf_{q_1,q_2\in\mathcal{Q}_c}\sup_{g\in\mathcal{G}_c}\langle g, v.q_2 - q_1.u\rangle. \tag{28}$$

Together with inequalities 25 and 27, it will imply the result.

To prove 28, we will apply a variant of Sion's theorem (see e.g., Proposition I.1.3 in Mertens *et al.* (2015)) to the zero-sum game with strategy spaces $\mathcal{G}_c$ for the maximizer, $\mathcal{Q}_c^2$ for the minimizer, and payoff $h(g,(q_1,q_2)) = \langle g, v.q_2 - q_1.u\rangle$.

At first, note that the strategy sets $\mathcal{G}_c$ and $\mathcal{Q}_c^2$ are convex, and that h is bilinear.

To avoid any confusion, let us denote $\mathcal{Q}_c^2 = Q_1 \times Q_2$. Let u_C denote the marginal distribution of u over the set C of signals of player 1 and v_D the marginal distribution of v over the set D of signals of player 2. Note that the payoff function of the above game do not change if we replace q_1 by another map which is equal u_C-almost everywhere to q_1, and a similar remark holds for q_2. Therefore, we may consider that Q_1 is the set of equivalence classes of transitions w.r.t to equality u_C almost everywhere, endowed with the topology τ_{u_C} defined in section L.2, and that Q_2 is the set of equivalence classes of transitions w.r.t to equality v_D almost everywhere, endowed with the topology τ_{v_D}. The set $Q_1 \times Q_2$ is thus metric compact for the associated product topology.

It remains to check that for every $g \in \mathcal{G}_c$, the map $(q_1,q_2) \to h(g,(q_1,q_2))$ is continuous. Note first that is is the sum of a function of q_1 and a function of q_2.

Let $\nu : C \to \Delta(K \times D)$ denote a version of the conditional law of (k,d) given c under the probability u. Then, we have

$$\begin{aligned}\langle g, q_1.u\rangle &= \int_{K\times C\times D}\left(\int_C g(k,c',d)dq_1(c'|c)\right)du(k,c,d)\\&= \int_C\int_{K\times D}\left(\int_C g(k,c',d)dq_1(c'|c)\right)d\nu(k,d|c)du_C(c)\\&= \int_C\int_C\left(\int_{K\times D} g(k,c',d)d\nu(k,d|c)\right)dq_1(c'|c)du_C(c)\end{aligned}$$

and the above expression is continuous with respect to q_1 using the definition of the topology τ_{u_C} since using bounded convergence, the function

$$f(c,c') := \int_{K\times D} g(k,c',d)d\nu(k,d|c),$$

is continuous with respect to the second variable.

A similar argument holds for q_2 and this concludes the proof.

It remains to prove the first equality. Let $g \in \mathcal{G}_c$ such that $\text{val}(u,g) > \text{val}(v,g) + 3\varepsilon$ for some $\varepsilon > 0$. According to Proposition III.4.2 in Mertens *et al.* (2015) (up to relabeling the signals), Player 1 has an ε-optimal strategy in the game $\Gamma(u,g)$ taking values almost surely in some finite set $A_1 \subset C$. Similarly, player 2 has an ε-optimal strategy in the game $\Gamma(v,g)$ taking values almost surely in some finite set $A_2 \subset D$. It results that the game with action spaces A_1, A_2 and payoff $\hat{g} = g|_{K\times A_1\times A_2}$ is such that

$$\text{val}(u,\hat{g}) \geq \text{val}(v,\hat{g}) + \varepsilon.$$

The conclusion follows.

L.4 Extensions of the results of section 4

We now explain how to extend the 5 Propositions of section 4 to uncountable information structures.

At first, note that Propositions 1 and 4 hold with the same proofs. Then, Proposition 2 can be generalized by using the same ideas and a few technical adaptations described below.

Let us start with general properties of the single-agent distance $\mathbb{d}_1$. Let $\mathcal{G}_{1,c} = \{g : K \times C \to [-1,1], | \, g \text{ continuous } \}$, which can be identified with a subset of $\mathcal{G}_c$ made of functions that do not depend on d. Then, as in the second part of the above proof of Theorem 1, we have

$$\forall u, v \in \mathcal{U}_c, \; \mathbb{d}_1(u,v) = \sup_{g \in \mathcal{G}_1} |\operatorname{val}(u,g) - \operatorname{val}(v,g)| = \sup_{g \in \mathcal{G}_{1,c}} |\operatorname{val}(u,g) - \operatorname{val}(v,g)|.$$

Recall that $C = D = [0,1]$ and define the set of single-agent information structures as $\mathcal{U}_{1,c} = \Delta(K \times C)$. Note that given $u \in \mathcal{U}_c$, $\operatorname{marg}_{K\times C} u \in \mathcal{U}_{1,c}$. Define for $u', v' \in \mathcal{U}_{1,c}$, $\mathbb{d}'_1(u',v') = \sup_{g \in \mathcal{G}_{1,c}} |\operatorname{val}(v',g) - \operatorname{val}(u',g)|$. For any $u, v \in \mathcal{U}_c$,

$$\mathbb{d}_1(u,v) = \mathbb{d}'_1(u',v') = \max\{\min_{q \in \mathcal{Q}_c} 2\|u' - q.v'\|_{TV}, \min_{q \in \mathcal{Q}_c} 2\|q.u' - v'\|_{TV}\} \tag{29}$$

where $u' = \operatorname{marg}_{K\times C} u$, $v' = \operatorname{marg}_{K\times C} v$, and $q.u'$ is the probability defined by

$$\int_{K\times C} f d(q.u') = \int_{K\times C} \int_C f(k,c') dq(c'|c) du'(k,c),$$

for all bounded measurable function f. This result can be obtained by mimicking (and simplifying) the arguments of the proof of Theorem 1.

L.4.1 Extension of Proposition 2

Proposition 15. *Suppose that $u, v \in \overline{\mathcal{U}}$ are two information structures with conditionally independent information such that* $\operatorname{marg}_{K\times D} u = \operatorname{marg}_{K\times D} v$. *Then,* $\mathbb{d}(u,v) = \mathbb{d}_1(u,v)$.

Let us first introduce some additional notation. Let $u_S = \operatorname{marg}_S(u)$ for $S = K, K\times C, K\times D$. Let $u_1 : K \to \Delta(C)$ denote a version of the conditional law of c given k induced by u, and $u_2 : K \to \Delta(D)$ denote a version of the conditional law of d given k induced by u. Let v_S, $v_1 : K \to \Delta(C)$ and $v_2 : K \to \Delta(D)$ be defined as above, and note that by assumptions $v_K = u_K$ and $u_2 = v_2$, and that the conditional law of (c,d) given k induced by u is given by $u_1(.|k) \otimes u_2(.|k)$ (and similarly for v).

Let us fix a pair of garblings q_1, q_2. Consider the measure $\pi_2 = \frac{1}{2}(u_{K\times D} + (u.q_2)_{K\times D})$ on $K \times D$ and the Radon-Nikodym densities

$$\phi_2 = \frac{du_{K\times D}}{d\pi_2},\ \psi_2 = \frac{d(u.q_2)_{K\times D}}{d\pi_2}.$$

Similarly, let $\pi_1 = \frac{1}{2}(u_{K\times C} + (q_1.v)_{K\times C})$ and

$$\phi_1 = \frac{du_{K\times C}}{d\pi_1},\ \psi_1 = \frac{d(q_1.v)_{K\times C}}{d\pi_1}.$$

Note that $\text{marg}_K(\pi_1) = \text{marg}_K(\pi_2) = u_K$.

For any and any bounded measurable function f we have using the conditional independence assumption:

$$\begin{aligned}
\int_{K\times C\times D} f d(u.q_2) &= \int_{K\times D}\int_C f(k,c,d)du_1(c|k)d(u.q_2)_{K\times D}(k,d) \\
&= \int_{K\times D}\int_C f(k,c,d)du_1(c|k)\psi_2(k,d)d\pi_2(k,d) \\
&= \int_{K\times C}\int_D f(k,c,d)\psi_2(k,d)d\pi_2(d|k)du_{K\times C}(k,c) \\
&= \int_{K\times C}\int_D f(k,c,d)\psi_2(k,d)d\pi_2(d|k)\phi_1(k,c)d\pi_1(k,c) \\
&= \int_K\int_D\int_C f(k,c,d)\psi_2(k,d)\phi_1(k,c)d\pi_1(c|k)d\pi_2(d|k)du_K(k).
\end{aligned}$$

Similarly, one has

$$\int_{K\times C\times D} f d(q_1.v) = \int_K\int_D\int_C f(k,c,d)\phi_2(k,d)\psi_1(k,c)d\pi_1(c|k)d\pi_2(d|k)du_K(k).$$

We obtain

$$\int_{K\times C\times D} f d(u.q_2 - q_1.v)$$
$$= \int_K \int_D \int_C f(k,c,d)\left[\psi_2(k,d)\phi_1(k,c) - \phi_2(k,d)\psi_1(k,c)\right] d\pi_2(d|k)d\pi_1(c|k)du_K(k),$$

so that the supremum over all measurable f bounded by 1 is equal to

$$2\|u.q_2 - q_1.v\|_{TV} = \int_K \int_D \int_C |\psi_2(k,d)\phi_1(k,c) - \phi_2(k,d)\psi_1(k,c)|\, d\pi_2(d|k)d\pi_1(c|k)du_K(k)$$
$$= \int_K \int_D \int_C |(\psi_2(k,d) - \phi_2(k,d))\phi_1(k,c) + \phi_2(k,d)(\phi_1(k,c) - \psi_1(k,c))|\, d\pi_2(d|k)d\pi_1(c|k)du_K(k)$$

Because $|x+y| \geq |x| + \text{sgn}(x)y$ for each $x, y \in \mathbb{R}$, we have

$$\begin{aligned}
&2\|u.q_2 - q_1.v\|_{TV} \\
&\geq \int_K \int_D \int_C \phi_2(k,d)\,|\phi_1(k,c) - \psi_1(k,c)|\, d\pi_2(d|k)d\pi_1(c|k)du_K(k) \\
&+ \int_K \int_D \int_C (\psi_2(k,d) - \phi_2(k,d))\phi_1(k,c)\text{sgn}(\phi_1(k,c) - \psi_1(k,c))d\pi_2(d|k)d\pi_1(c|k)du_K(k) \\
&= \int_K \int_D \int_C \phi_2(k,d)\,|\phi_1(k,c) - \psi_1(k,c)|\, d\pi_2(d|k)d\pi_1(c|k)du_K(k) \\
&= 2\|u - q_1.v\|_{TV}
\end{aligned}$$

where the last equality is obtained exactly as above with $2\|u.q_2 - q_1.v\|_{TV}$.

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

$$\begin{aligned}
\mathbb{d}(u,v) &= \max\{\min_{q_1,q_2} 2\,\|u.q_2 - q_1.v\|_{TV}\,; \min_{q_1,q_2} 2\,\|v.q_2 - q_1.y\|_{TV}\} \\
&= \max\{\min_{q_1} 2\,\|u - q_1.v\|_{TV}\,; \min_{q_1} 2\,\|v - q_1.u\|_{TV}\} = \mathbb{d}_1(u,v),
\end{aligned}$$

where the last equality follows from (29) together with the fact that $\text{marg}_{K\times D}\, u =$

$\text{marg}_{K\times D}\, v$.

L.4.2 Extension of Proposition 3

Proposition 3 holds using the same ideas and a few technical adaptations described below.

Let $C = C_1 = C_2 = D = [0,1]$.

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

Proposition 16. *Suppose that, under u, c_1 is conditionally independent from (c, c_2, d) given k. Then, $\mathbb{d}\left(u', v'\right) \geq \mathbb{d}\left(u, v\right)$.*

Recall that Theorem 1 holds for more general signal spaces. Because $u \succeq v$,

$$\mathbb{d}\left(u,v\right) = \min_{q_2}\min_{q_1} 2\left\|u.q_2 - q_1.v\right\|_{TV} \leq \min_{q_2}\min_{q_1:C\to\Delta(C\times C_2)} 2\left\|u.q_2 - \hat{q}_1.v\right\|_{TV},$$

where q_1 ranges through the set of garblings from $C \times C_1$ to $\Delta(C \times C_1 \times C_2)$, q_2 ranges through the set of garblings from D to $\Delta(D)$ and where in the right-hand side of the inequality, we use a restricted set of player 1's garblings. Precisely, given $q_1 : C \to \Delta\left(C \times C_2\right)$, we define the garbling $\hat{q}_1$ by the relation

$$\int_{C'\times C_1'\times C_2'} f d\hat{q}_1(c, c_1) = \int_{C'\times C_2'} f(c', c_1, c_2') dq_1(c', c_2'|c)$$

where f is an arbitrary bounded measurable function. Further, for any such q_1, an arbitrary garbling q_2 and an arbitrary measurable function f bounded by 1, we have

$$\begin{aligned}
&\int f d(u.q_2 - \hat{q}_1.v) \\
&= \int_{K\times C\times C_1\times C_2\times D}\left[\int_{D'} f(k,c,c_1,c_2,d')dq_2(d'|d) - \int_{C'\times C_2'} f(k,c',c_1,c_2',d)dq_1(c',c_2'|c)\right]du(k,c,c_1,c_2,d) \\
&= \int_K\int_{C\times C_2\times D}\int_{C_1}\left[\int_{D'} f(k,c,c_1,c_2,d')dq_2(d'|d) - \int_{C'\times C_2'} f(k,c',c_1,c_2',d)dq_1(c',c_2'|c)\right]du(c_1|k)du(c,c_2,d|k)du(k) \\
&= \int_K\int_{C\times C_2\times D}\left[\int_{D'} h(k,c,c_2,d')dq_2(d'|d) - \int_{C'\times C_2'} h(k,c',c_2',d)dq_1(c',c_2'|c)\right]du(c,c_2,d|k)du(k) \\
&= \int_{K\times C\times C_2\times D} hd(u'.q_2 - q_1.v')
\end{aligned}$$

where we used the conditional independence assumption, and where the function h is a measurable function bounded by 1 and defined by

$$h(k,c,c_2,d) = \int_{C_1} f(k,c,c_1,c_2,d)du(c_1|k).$$

We deduce that

$$\min_{q_2}\min_{q_1:C\to\Delta(C\times C_2)} 2\left\|u.q_2 - \hat{q}_1.v\right\|_{TV} \leq 2\left\|u'.q_2 - q_1.v'\right\|_{TV}.$$

Hence

$$\mathbb{d}(u,v) \leq \min_{q_2}\min_{q_1:C\to\Delta(C\times C_2)} 2\left\|u'.q_2 - q_1.v'\right\|_{TV} = \mathbb{d}(u',v'),$$

and this concludes the proof.

L.4.3 Extension of Proposition 5

Proposition 5 holds by adapting the definitions.

Precisely, we define ε conditional independence using the total variation norm. Consider a distribution $\mu \in \Delta(X\times Y\times Z)$ over compact metric spaces. We say that the random variables x and y are ε-conditionally independent given z under μ if

$$\int_Z \int_{X\times Y} 2\left\|\mu\left(x,y|z\right)-\mu\left(x|z\right)\otimes\mu\left(y|z\right)\right\|_{TV} d\mu(z)\leq\varepsilon.$$

Let $C=C_1=D=D_1=[0,1]$. Let $u\in\Delta\left(K\times\left(C\times C_1\right)\times\left(D\times D_1\right)\right)$ and $v=\text{marg}_{K\times C\times D}u$.

Proposition 17. *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, $\mathbb{d}\left(u,v\right)\leq\varepsilon$.*

It is enough to show that if c_1 is ε-conditionally independent from (k,d) given c, then

$$\sup_{g\in\mathcal{G}}\text{val}\left(u,g\right)-\text{val}\left(v,g\right)\leq\varepsilon.$$

For this, let $q_2:D\times D_1\to D$ be defined as $q_2\left(d,d_1\right)=\delta_d$ and let $q_1:C\to C\times C_1$ be defined as $q_1\left(c,c_1|c\right)=u\left(c_1|c\right)$. Then, for any measurable function f bounded by 1:

$$\begin{aligned}
&\int fd(u.q_2-q_1.v)\\
&\quad=\int_{K\times C\times C_1\times D}\left[f(k,c,c_1,d)-\int_{C_1'}f(k,c,c_1',d)du(c_1'|c)\right]du(k,c,c_1,d)\\
&\quad=\int_C\left[\int_{K\times C_1\times D}f(k,c,c_1,d)du(k,c_1,d|c)-\int_{K\times D}\int_{C_1'}f(k,c,c_1',d)du(c_1'|c)du(k,d|c)\right]du(c)\\
&\quad\leq\int_C 2\|u(k,c_1,d|c)-u(c_1'|c)\otimes u(k,d|c)\|_{TV}du(c)\\
&\quad\leq\varepsilon.
\end{aligned}$$

The claim follows from Theorem 1.

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

BALDER, E. J. (1988). Generalized equilibrium results for games with incomplete information. *Mathematics of Operations Research*, **13** (2), 265–276. L.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, 13

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, 7, F

—, — 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, 7, F

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., Oliu-Barton, M. and Venel, X. (2014). Existence of the uniform value in zero-sum repeated games with a more informed controller. *Journal of Dynamics & Games*, **1** (3), 411–445. I.4

— 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, L.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. (1997). The robustness of equilibria to incomplete information. *Econometrica*, **65** (6), 1283–1309. 6.1.1

— and — (1998). Payoff continuity in incomplete information games. *Journal of Economic Theory*. 12

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). 16

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, I.4

—, Sorin, S. and Zamir, S. (2015). *Repeated games*, vol. 55. Cambridge University Press. 1, 2, 7, 7, A, F, L.1, L.3, L.3.1

— 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, L.1

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

— and Ui, T. (2005). Generalized potentials and robust sets of equilibria. *Journal of Economic Theory*, **124** (1), 45–78. 6.1.1

Oyama, D. and Tercieux, O. (2010). Robust equilibria under non-common priors. *Journal of Economic Theory*, **145** (2), 752–784. 6.1.1

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

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, 8, 6.1.1, 7

Sadzik, T. (2008). Beliefs revealed in bayesian equilibrium. *Ekonomia. Rynek, Gospodarka, Społeczeństwo*, (20), 136–138. 13

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