Lexicographic Agreeing to Disagree and Perfect Equilibrium*

Christian W. Bach[†,‡] and Jérémie Cabessa[§]

Abstract. Aumann's seminal agreement theorem deals with the impossibility for agents to acknowledge their distinct posterior beliefs. We consider agreeing to disagree in an extended framework with lexicographic probability systems. A weak agreement theorem in the sense of identical posteriors only at the first lexicographic level obtains. Somewhat surprisingly, a possibility result does emerge for the deeper levels. Agents can agree to disagree on their posteriors beyond the first lexicographic level. By means of mutual absolute continuity as an additional assumption, a strong agreement theorem with equal posteriors at every lexicographic level ensues. Subsequently, we turn to games and provide epistemic conditions for the classical solution concept of perfect equilibrium. Our lexicographic agreement theorems turn out to be pivotal in this endeavour. The hypotheses of mutual primary belief in caution, mutual primary belief in rationality, and common knowledge of conjectures characterize perfect equilibrium epistemically in our lexicographic framework.

Keywords: agreeing to disagree; agreement theorems; common prior assumption; epistemic game theory; interactive epistemology; lexicographic Aumann structures; lexicographic beliefs; lexicographic conjectures; lexicographic probability systems; mutual absolute continuity; perfect equilibrium; solution concepts; static games; strong agreement theorem; weak agreement theorem

JEL Classifications: C72

*Preliminary versions of this work were presented at the 13th Conference on Logic and the Foundations of Game and Decision Theory (LOFT13), Milan, July 2018, as well as at the 6th World Congress of the Game Theory Society (GAMES2020), Budapest, July 2021. We are grateful to Adam Brandenburger, Andrés Carvajal, Robert Edwards, Amanda Friedenberg, Stephan Jagau, Andrés Perea, Burkhard Schipper, Elias Tsakas, and two anonymous referees for useful as well as constructive comments.

†Department of Economics, University of Liverpool Management School, University of Liverpool, Chatham Street, Liverpool, L69 7ZH, UNITED KINGDOM

‡EPICENTER, School of Business and Economics, Maastricht University, 6200 MD Maastricht, THE NETHERLANDS

§DAVID Laboratory, University of Versailles Saint-Quentin-en-Yvelines, 45 Avenue des États-Unis, 78035 Versailles, FRANCE

1 Introduction

The impossibility for two agents to agree to disagree is established by Aumann (1976)'s seminal agreement theorem. More precisely, if two Bayesian agents with a common prior receive private information and have common knowledge of their posterior beliefs, then these posteriors must be equal. In other words, distinct posterior beliefs cannot be common knowledge among Bayesian agents with the same prior beliefs. In this sense, agents cannot agree to disagree.[1] The impossibility of agreeing to disagree has important implications for any interactive situation where, loosely speaking, the mutual acknowledgement of distinct views or assessments is relevant, e.g. trade, speculation, political positions, or legal judgements.[2] The array of potential applications for the agreement theorem is vast.

Here, we explore agreeing to disagree in an extended framework with lexicographic beliefs. A lexicographic belief is a sequence of beliefs, where the different beliefs are given in descending order of importance.[3] The sequence's first component can be viewed as the agent's primary doxastic attitude, its second component as his secondary doxastic attitude, etc. Intuitively, a lexicographically-minded agent deems his first belief fundamentally more likely than his secondary belief, which in turn is fundamentally more likely than his tertiary belief, etc. Lexicographic beliefs resolve the problem of conditioning on events with probability zero. Revising beliefs based on hypotheses that are initially deemed impossible is relevant to hypothetical reasoning. An apt example are games. It can be important for a player to consider what would happen, if an opponent were to pick an unexpected choice, in order to act rationally himself.

In game theory, lexicographic beliefs do play a prominent role and have effectively been put into action to model caution and trembles.[4] In particular, they shed essential light on the foundations of weak dominance arguments and have served to unravel a

[1] An extensive literature on agreeing to disagree has emerged. Most contributions reconsider Aumann's impossibility theorem in more general frameworks. Notably, Bonanno and Nehring (1997) as well as Ménager (2012) provide comprehensive surveys on this literature. Some more recent contributions to the agreeing to disagree literature include Dégrement and Roy (2012), Hellman and Samet (2012), Bach and Perea (2013), Heifetz et al. (2013), Hellman (2013), Demey (2014), Lehrer and Samet (2014), Chen et al. (2015), Dominiak and Lefort (2015), Tarbush (2016), Bach and Cabessa (2017), Gizatulina and Hellman (2019), Pacuit (2018), Tsakas (2018), Liu (2019), as well as Contreras-Tejada et al. (2021).

[2] A prominent analysis of economic consequences of agreeing to disagree is Milgrom and Stokey's (1982) so-called no-trade theorem. Accordingly, if two traders agree on a prior efficient allocation of goods, then upon receiving private information it cannot be common knowledge that they both have an incentive to trade.

[3] Formally, lexicographic beliefs are modelled in their most general form by lexicographic probability systems due to Blume et al. (1991a).

[4] By now lexicographic beliefs have become a widespread tool in game theory and have been used, for instance, by Kreps and Wilson (1982), Kreps and Ramey (1987), Blume et al. (1991b), Brandenburger (1992a), Börgers (1994), Stahl (1995), Mailath et al. (1997), Asheim (2001, 2002), Govindan and Klumpp (2003), Asheim and Perea (2005), Brandeburger et al. (2008), Yang (2015), Dekel et al. (2016), Lee (2016), as well as Cantonini and De Vito (2018, 2020).

fundamental game-theoretic paradox: the so-called inclusion-exclusion problem.[5] The paradox arises whenever a player is required to *include* all, yet to *exclude* some, choices for an opponent. This startling tension is inherent in (iterated) weak dominance, also called (iterated) admissibility, which constitutes one of the most long-standing ideas in game theory going back at least to Gale (1953).

For an illustration of the inclusion-exclusion problem, consider the two player game depicted in Figure 1 with players *Alice* and *Bob*, where *Alice* chooses a "row" (a or b) and Bob picks a "column" (y or z). The unique strategy for *Alice* in line with weak dominance is a. Intuitively, against all choices of *Bob*, a never yields less than b, and against the particular strategy y of *Bob*, a induces a strictly higher payoff than b. For *Bob*, y is strictly worse than z against all of *Alice*'s choices. However, it seems impossible to support a with consistent beliefs, since on the one hand, *Alice* needs to assign positive probability to both y and z to render a uniquely optimal for her, while on the other hand, she should assign probability zero to the never optimal choice y for *Bob*. The remedy to the paradox lies in lexicographic beliefs. They are capable of not excluding any choice from consideration yet at the same time deeming some choices much more – indeed infinitely more – likely than others. With lexicographic beliefs, the inclusion-exclusion riddle evaporates. In the preceding example, a lexicographic belief for *Alice* that assigns probability one to z in its first level and probability one to y in its second level would already form a consistent doxastic attitude filtering out a as her unique optimal strategy.

	y	z
a	$1, 0$	$0, 1$
b	$0, 0$	$0, 1$

Fig. 1. A two player game

In terms of Aumann's impossibility theorem the question of whether agreeing to disagree is possible or not gains in depth if lexicographic beliefs are admitted and hypothetical reasoning can thereby be captured. For example, consider merchants forming beliefs about the arrival of a sea shipment. A primary contingency could revolve around the usual meteorological conditions that can affect the length of sea travel. Suppose that a secondary contingency would include fundamentally less likely factors affecting arrival like a pirate attack. If common knowledge of their posterior beliefs implies agents to agree on their beliefs given the primary contingency, then they could possibly still disagree with regards to the secondary contingency. Whether or not the agents do, could have different implications for the actions they take based on their (lexicographic) beliefs.

In general, given the importance of lexicographic beliefs in game theory on the one hand, and given Aumann's seminal impossibility result on agreeing to disagree on the

[5] The inclusion-exclusion problem has first been identified by Samuelson (1992), when showing that the solution concept of iterated weak dominance can be inconsistent with common knowledge assumptions.

other hand, it seems intriguing to ask how the agreement theorem is affected if standard probabilities are replaced by lexicographic probability systems. To address this question we define the notion of lexicographic Aumann structure, where the agents hold a sequence of priors on the basis of which they compute a sequence of posteriors in the style of Blume et al. (1991a). In our framework, a weak agreement theorem in the sense of merely identical first level posteriors obtains. However, we provide a disagreement result establishing that agents can actually agree to disagree on their posteriors beyond the first lexicographic level. Aumann's impossibility theorem does therefore not directly generalize to full-fledged lexicographic reasoning. Based on this observation, we introduce a condition which essentially states that every lexicographic level prior either neglects or considers the agents' private information synchronically. This condition can be viewed as a variant of standard mutual absolute continuity from probability theory. With the assistance of mutual absolute continuity, we provide a strong agreement theorem which establishes the impossibility of agreeing to lexicographically disagree.

Naturally, the question arises whether our lexicographic agreement theorems can be applied to game theory. It would be particularly illuminating to gain novel insights about classical solution concepts based on lexicographic agreeing to disagree. A prominent class of solution concepts in game theory is based on the idea of trembles. Intuitively, with a very small probability a player may make a mistake – "his hand might tremble" – in implementing his optimal strategy. So-called tremble equilibria formalize this intuition by postulating equilibrium behaviour as the limiting case when the trembles vanish. The most fundamental solution concept of this kind is Selten's (1975) perfect equilibrium.[6] A typical feature of tremble equilibria requires all trembles to satisfy some full support condition. In this sense, tremble equilibria also formalize cautious players, which suggests a link to lexicographic beliefs. Indeed, Blume et al. (1991b) investigate this link and provide a reformulation of perfect equilibrium as well as of proper equilibrium in terms of lexicographic conjectures, which are lexicographic beliefs about choices.

However, a characterization of tremble equilibria in terms of interactive thinking is still missing. Such an endeavour would imperatively involve higher-order beliefs, thereby moving beyond the basic doxastic layer of conjectures. Full interactive reasoning is modelled by imposing conditions on belief hierarchies which in turn assemble different layers of iterated beliefs. Conjectures, as beliefs about (opponents') choices, only constitute the first such layer. In order to fully describe the interactive thinking of players, it is crucial to also model their beliefs about their opponents' conjectures, their beliefs about their opponents' beliefs about their opponents' conjectures, etc. Due to their infinite nature belief hierarchies are cumbersome objects, but fortunately they can be represented in a compact way by means of epistemic models due to Harsanyi (1967-68). The epistemic program in game theory has employed such models to unveil the interactive reasoning assumptions implicitly endorsed by solution concepts in games.

[6] Other tremble equilibria have been proposed in the literature, for instance, Myerson's (1978) proper equilibrium, van Damme's (1984) quasi-perfect equilibrium, as well as Harsanyi and Selten's (1988) uniformly perfect equilibrium.

Our lexicographic agreement theorems are capable of shedding some light on the interactive reasoning underlying perfect equilibrium in games. Our framework of lexicographic Aumann structures with a common prior is capable of shedding some light on the interactive reasoning assumptions underlying tremble equilibria. Indeed, we provide epistemic conditions for perfect equilibrium. The epistemic hypotheses of mutual primary belief in caution, mutual primary belief in rationality, and common knowledge of conjectures characterize perfect equilibrium in terms of interactive reasoning. Our lexicographic agreement theorems play a prominent role in attaining our epistemic foundation. By means of the weak agreement theorem, all opponents of any given player can be ensured to hold the same marginal lexicographic conjecture about him. The strong agreement theorem is used to derive an independence property of the players' lexicographic conjectures.

We proceed as follows. The remainder of this section demarcates our model and results from the related literature. In Section 2, Blume et al.'s (1991a) lexicographic probability systems are incorporated into state-based interactive epistemology. Core notation is fixed and key concepts are defined. Section 3 contains a weak agreement theorem (**WAT**) with lexicographic probability systems, while Section 4 brings the deeper lexicographic levels into focus. Incongruity can obtain beyond the first level as our disagreement result (**DIS**) shows. In Section 5, under mutual absolute continuity, a lexicographically strong agreement theorem (**SAT**) is developed. We subsequently turn to games. In Section 6, Selten's (1975) seminal solution concept of perfect equilibrium is presented. A reformulation of this tremble equilibrium by means of lexicographic conjectures is furnished along the lines of Blume et al. (1991b) in Section 7. Epistemic conditions that characterize perfect equilibrium are put forth in Section 8. Finally, Section 9 offers some concluding remarks.

1.1 Related Literature

By establishing agreement theorems with lexicographic beliefs and providing epistemic conditions for perfect equilibrium, our contribution is twofold. On the one hand, we are connected to the literature on agreeing to disagree that has emerged since Aumann's seminal (1976) impossibility result. On the other hand, the application of our lexicographic agreement theorems to epistemically characterize perfect equilibrium adds to the foundations of game theory.

Our framework extends standard Aumann structures (Aumann, 1974 and 1976) by modelling the agents' beliefs with Blume et al.'s (1991a) lexicographic probability systems instead of mere probability distributions. Within this enriched set-up, we explore agreeing to disagree. Aumann's (1976) agreement theorem obtains as a special case of **WAT**, if the lexicographic common prior is truncated at the first level.

A lexicographic approach to agreeing to disagree is also taken by Bach and Perea (2013). Notably, their framework admits lexicographic beliefs as priors yet delivers a standard posterior for every agent. In contrast, by using lexicographic probability systems, we also model the posteriors as lexicographic beliefs. This does not only formally but also conceptually make an essential difference, as the agents' decision-relevant beliefs are the posteriors which are extended in our framework. A further restriction of

Bach and Perea (2013) is a non-overlapping support requirement on lexicographic priors, which we do not impose. The agreement theorem of Bach and Perea (2013) is implied as another special case of **WAT**, if the lexicographic posteriors are truncated at the first level.

Once lexicographic posteriors enter the picture novel insights emerge. Somewhat surprisingly, our possibility result **DIS** establishes that agents can actually agree to disagree with a lexicographic mindset. In fact, if a non-overlapping support requirement on lexicographic priors were to be desired, **DIS** would still remain valid. The additional assumption of mutual absolute continuity brings about our impossibility result **SAT**, which can be viewed as a *lexicographic* agreement theorem in sensu stricto.

In general, lexicographic probability systems deal with the problem of how to proceed if something is learned to which initially probability zero was assigned. An alternative tool for extending probabilities to handle conditioning on measure zero events are conditional probability systems due to Rényi (1955). They have prominently been used in game theory to define the reasoning concept of common strong belief in rationality for extensive forms by Battigalli and Sinaicalchi (2002). Lexicographic probability systems can be related to conditional probability systems and equivalences have been established under certain conditions (e.g. Hammond, 1994; Halpern, 2010; Tsakas, 2014). Lexicographic agreeing to disagree is thus indirectly also related to Tsakas (2018), who establishes two agreement theorems with conditional probability systems. However, his results cannot be directly compared to ours, since the models are too different. While we extend Aumann's partitional model by lexicographic probability systems, Tsakas (2018) uses type structures in the style of Battigalli and Siniscalchi (1991). In particular, the way in which the agents' posteriors enter the picture is inherently distinct. In Tsakas' (2018) framework, the agreement concerns a single posterior per agent, while our agreement theorems deal with lexicographic posteriors. Besides, already the computation of the first level posterior in our framework depends on which prior assigns positive probability to the conditioning event (i.e. the respective agent's information cell in lexicographic Aumann structures). In contrast, the determination of the conditioning event to derive the posterior in Tsakas' (2018) model is independent from the prior.

In the game-theoretic part of our paper, we explore the epistemic foundation of Selten's (1975) solution concept of perfect equilibrium. A reformulation of perfect equilibrium by means of lexicographic conjectures constitutes the first step. Although such a reformulation has already been established by Blume et al. (1991b), our Lemma 1 provides a similar construction for the sake of completeness and self-containedness. Being concerned with the players' interactive reasoning, epistemic foundations go beyond conjectures into the players' belief hierarchies. Our Theorems 3 and 4 provide an epistemic characterization of perfect equilibrium. They can be viewed as developping Blume et al.'s (1991a) analysis of perfect equilibrium in terms of lexicographic conjectures further into the full game-theoretic reasoning realm. In some sense, our relation to Blume et al. (1991b) with regard to perfect equilibrium is analogous to the relation of Aumann and Brandenburger (1995) to Harsanyi (1973) with regard to Nash equilibrium: while Harsanyi (1973) has proposed the interpretation of Nash equilibrium in terms of conjectures, Aumann and Brandenburger (1995) have taken this crucial insight into an epistemic framework, unveiling the underlying interactive reasoning assumptions of Nash equilibrium. Our game-theoretic results could be perceived of as generalizing Aumann and Brandenburger (1995) from Nash equilibrium to perfect equilibrium.[7]

For the special case of two players, perfect equilibrium has been characterized epistemically by Perea (2012). The supply of epistemic conditions for perfect equilibrium involving any finite number of players has still been an open question though, which our Theorems 3 and 4 address. An epistemic analysis of equilibrium notions faces two considerable challenges once more than two players are considered. Firstly, for a given player, all opponents have to share the same belief about the player's choice ("problem of projection"). Secondly, any player's belief about his opponents' choices needs to be independent ("problem of independence"). Our lexicographic agreement theorems turn out to be pivotal in resolving these intricacies. Besides his restriction to the two player case, Perea's (2012) type-based framework is distinct from our state-based lexicographic Aumann structures with a common prior. Epistemic conditions for the special setting of two players are provided by our Proposition 2, which can thus be juxtaposed with Perea (2012). Our hypotheses of mutual primary belief in caution and mutual primary belief in rationality are weaker variants of his common full belief in caution and common full belief in primary belief in rationality, respectively. Furthermore, mutual knowledge of lexicographic conjectures embodies a correct beliefs assumption among our epistemic conditions. In contrast, Perea's (2012) correct beliefs assumption essentially states that each player believes his opponent to only lexicographically deem possible the player's actual lexicographic belief hierarchy. While his epistemic operator is thus doxastic and the uncertainty is spanned by the full belief hierarchies, our correct beliefs assumption uses the stronger operator of knowledge but only concerns the players' conjectures in terms of uncertainty. Finally, Perea's (2012) notion of caution is more restrictive than ours. A player is cautious according to Perea (2012), whenever, if he lexicographically deems possible a type for any opponent, then he also lexicographically deems possible any strategy for that type. In contrast, a player already satisfies caution in our game-theoretic framework, whenever his lexicographic conjecture deems possible any strategy for all of his opponents.

2 Preliminaries

In state-based interactive epistemology, knowledge and beliefs are modelled within the framework of Aumann structures. Formally, an Aumann structure

$$\mathcal{A} := \big(\Omega, (\mathcal{I}_i)_{i \in I}, p\big)$$

consists of a finite set Ω of possible worlds (also called states of the world), a finite set I of agents, a possibility partition $\mathcal{I}_i$ of Ω for every agent $i \in I$, and a common

[7] There are some significant differences though. While Aumann and Brandenburger (1995) define knowledge as probability one belief in type-based structures, we use the standard notion of knowledge in state-based Aumann models to define common knowledge of conjectures. Also, our proofs critically build on (lexicographic) agreeing to disagree, whereas the proofs of Aumann and Brandenburger take a different route without using (standard) agreeing to disagree.

prior $p : \Omega \to [0,1]$ such that $\sum_{\omega\in\Omega} p(\omega) = 1$. The cell of $\mathcal{I}_i$ containing the world ω is denoted by $\mathcal{I}_i(\omega)$ and assembles those worlds deemed possible by agent i at world ω. It is standard to impose the so-called non-null information assumption which ensures that no information is excluded a priori, i.e. $p\big(\mathcal{I}_i(\omega)\big) > 0$ for all $i \in I$ and for all $\omega \in \Omega$.

Agents reason about events which are defined as sets of possible worlds. The common prior p naturally extends to a measure $p : 2^{\Omega} \to [0,1]$ on the event space by setting $p(E) = \sum_{\omega\in E} p(\omega)$ for all $E \in 2^{\Omega}$. Agents are Bayesians and consequently update the common prior with their private information as follows: the posterior belief of agent i in event E at world ω is given by

$$p\big(E \mid \mathcal{I}_i(\omega)\big) = \frac{p\big(E \cap \mathcal{I}_i(\omega)\big)}{p\big(\mathcal{I}_i(\omega)\big)}$$

and forms the decision-relevant belief of the agent.

Knowledge is formalized in terms of events. The event of agent i knowing event E, denoted by $K_i(E)$, is defined as

$$K_i(E) := \{\omega \in \Omega : \mathcal{I}_i(\omega) \subseteq E\}.$$

If $\omega \in K_i(E)$, then i is said to know E at ω. Mutual knowledge is given by

$$K(E) := \bigcap_{i\in I} K_i(E).$$

Setting $K^0(E) := E$, higher-order mutual knowledge is inductively defined by

$$K^m(E) := K\big(K^{m-1}(E)\big)$$

for all $m > 0$. Mutual knowledge can also be denoted as 1-order mutual knowledge. The conjunction of all higher-order mutual knowledge yields common knowledge, which is formally defined as

$$CK(E) := \bigcap_{m>0} K^m(E)$$

for all $E \in 2^{\Omega}$. This is often called the iterative definition of common knowledge. An equivalent formulation due to Aumann (1976) is based on the meet of the agents' possibility partitions and typically denoted as the meet definition of common knowledge.[8] Accordingly, common knowledge is constructed as

$$CK(E) := \Big\{\omega \in \Omega : \big(\bigwedge_{i\in I} \mathcal{I}_i\big)(\omega) \subseteq E\Big\}$$

[8] Given two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ of some set S, the partition $\mathcal{P}_1$ is called *finer* than the partition $\mathcal{P}_2$ (or $\mathcal{P}_2$ *coarser* than $\mathcal{P}_1$), if each cell of $\mathcal{P}_1$ is a subset of some cell of $\mathcal{P}_2$. Given n partitions $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n$ of S, the finest partition that is coarser than $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n$ is called the *meet* of $\mathcal{P}_1, \mathcal{P}_2, \ldots, \mathcal{P}_n$ and is denoted by $\bigwedge_{i=1}^n \mathcal{P}_i$. Moreover, given $x \in S$, the cell of the meet $\bigwedge_{i=1}^n \mathcal{P}_i$ containing x is denoted by $\big(\bigwedge_{i=1}^n \mathcal{P}_i\big)(x)$.

for all $E \in 2^{\Omega}$, where $(\bigwedge_{i \in I} \mathcal{I}_i)(\omega)$ is the cell of the meet that contains the world ω.[9]

Lexicographic beliefs are modelled in line with Blume et al. (1991a)'s notion of lexicographic probability systems. The following definition provides a direct adaptation of Blume et al. (1991a, Definition 3.1) to the interactive setting with multiple agents.

Definition 1. *Let Ω be a set of possible worlds, I be a set of agents, and $M_i > 0$ be some integer. A* lexicographic probability system for agent $i \in I$ (i-LPS) *is a tuple*

$$\rho_i = (p_i^1, \ldots, p_i^{M_i}),$$

where $p_i^m \in \Delta(\Omega)$ for all $m \in \{1, \ldots, M_i\}$.

Lexicographic beliefs are thus sequences of standard beliefs. The index numbers of a lexicographic probability system are also referred to as lexicographic levels.

Incorporating lexicographic probability systems into Aumann structures gives rise to the notion of lexicographic Aumann structures.

Definition 2. *A* lexicographic Aumann structure *is a tuple*

$$\mathcal{A}_L = \big(\Omega, I, (\mathcal{I}_i)_{i \in I}, (\rho_i)_{i \in I}\big),$$

where

- *Ω is a set of possible worlds,*
- *I is a set of agents,*
- *$\mathcal{I}_i \subseteq 2^{\Omega}$ is a possibility partition of Ω for every agent $i \in I$,*
- *$\rho_i = (p_i^1, \ldots, p_i^{M_i})$ is an i-LPS for every agent $i \in I$,*
- *for every agent $i \in I$ and for every world $\omega \in \Omega$, there exists a lexicographic level $m \in \{1, \ldots, M_i\}$ such that $p_i^m\big(\mathcal{I}_i(\omega)\big) > 0$.*

The fifth item of Definition 2 ensures that no information is excluded a priori, and formally reflects the idea of caution. Actually, this condition can be seen as the lexicographic analogue to Aumann (1976)'s requirement for all information cells to be non-null events in the standard framework of Aumann structures. Caution could also be modelled as follows: for all $i \in I$ and for all $\omega \in \Omega$ there exists $m \in \{1, \ldots, M_i\}$ such that $p_i^m(\omega) > 0$. Such a condition is stronger, as it requires that every world – as opposed to only the information received – is deemed possible at some lexicographic level. The fifth item of Definition 2 is thus preferable.

Agents use their information to reason lexicographically about events. Formally, we adjust Blume et al. (1991a, Definition 4.2) to the context of lexicographic Aumann structures.

[9] In fact, Brandenburger and Dekel (1987) propose a more general definition of common knowledge that can be used without the non-null information assumption holding (e.g. in situations where the set Ω of possible worlds is uncountable). They require posterior beliefs to be proper regular conditional probabilities and modify the agents' possibility partitions appropriately in the case of null cells. Their notion of common knowledge is iterative and based on knowledge as probability 1 posterior belief.

Definition 3. *Let $\mathcal{A}_L$ be a lexicographic Aumann structure, $\omega \in \Omega$ be some world, and $i \in I$ be some agent. The* conditional lexicographic probability system of agent i given his information at world ω (ω-conditional i-LPS) *is the tuple*

$$\rho_i^\omega = \Big(p_i^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega)\big), \dots, p_i^{m_L}\big(\cdot \mid \mathcal{I}_i(\omega)\big)\Big)$$

where

- *the finite sequence of indices $(m_l)_{l=0}^L$ is inductively defined by $m_0 := 0$ and $m_l := \min\big\{m \in \mathbb{N} : m_{l-1} < m \leqslant M_i$ and $p_i^m\big(\mathcal{I}_i(\omega)\big) > 0\big\}$ if $l > 0$;*
- *$p_i^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big) = \frac{p_i^{m_l}\big(E \cap \mathcal{I}_i(\omega)\big)}{p_i^{m_l}\big(\mathcal{I}_i(\omega)\big)}$ for all $E \in 2^\Omega$ and for all $l \in \{1, \dots, L\}$.*

An essential difference between lexicographic Aumann structures and the standard framework resides in the former equipping agents with multiple levels of – and not unique – posteriors beliefs. Technically, the sequence $(m_l)_{l=1}^L$ of indicies belonging to the ω-conditional i-LPS ρ_i^ω depends on both i and ω and should thus strictly speaking be written as $(m_{i,\omega,l})_{l=1}^{L_{i,\omega}}$. For the sake of simplicity, the shortcut notation $(m_l)_{l=1}^L$ is adopted, whenever the dependence on i and ω is clear from the context. Furthermore, attention is restricted to the first L lexicographic posterior levels, where $L := \min\{L_{i,\omega} > 0 : i \in I \text{ and } \omega \in \Omega\}$, in order to ensure that the conditional lexicographic probability systems of every agent at every world have the same length. This restriction is only imposed for technical reasons, so that the lexicographic level posteriors the agents interactively reason about exist for all agents. Otherwise events such as "equal posteriors at all lexicographic levels" could not be properly defined. Besides, note that the lexicographic character of lexicographic probability systems actually crystallizes in two ways: an agent's prior as well as posterior are furnished with a lexicographic structure.

The common prior assumption in Aumann structures can be directly generalized to the lexicographic setting.

Definition 4. *Let $\mathcal{A}_L$ be a lexicographic Aumann structure. The lexicographic Aumann structure $\mathcal{A}_L$ satisfies the* common prior assumption (CPA)*, if there exists $\rho = (p^1, \dots, p^M) \in \big(\Delta(\Omega)\big)^M$ such that $M = \min\{M_i \in \mathbb{N} : i \in I\}$ and $p_i^m = p^m$ for all $i \in I$ and for all $m \in \{1, \dots, M\}$. In this case, the tuple ρ is called* common prior *and $\mathcal{A}_{LCP} = \big(\Omega, I, (\mathcal{I}_i)_{i \in I}, \rho\big)$ is called* lexicographic Aumann structure with a common prior*.*

With the existence of a common prior, the ω-conditional i-LPS thus becomes:

$$\rho_i^\omega = \rho(\cdot \mid \mathcal{I}_i(\omega)) = \Big(p^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega)\big), \dots, p^{m_L}\big(\cdot \mid \mathcal{I}_i(\omega)\big)\Big)$$

Analogously to the case of subjective priors, the sequence $(m_l)_{l=1}^L$ of indices should strictly speaking be written as $(m_{i,\omega,l})_{l=1}^L$, which we refrain from doing whenever the dependence on i and ω is clear from the context.

To preempt any potential confusion about the lexicographic notation: the prior levels are denoted by $m \in \{1, \dots, M\}$, while the posterior levels are represented by

$l \in \{1, \dots, L\}$. The l-th posterior level corresponds to the prior level $m_l \in \{1, \dots, M\}$ for all $l \in \{1, \dots, L\}$.

According to so-called Harsanyi consistency, differences in agents' beliefs are to be attributed entirely to differences in the agents' information. This doctrine extends to our more general set-up with lexicographic beliefs. Indeed, Definition 3 ensures that posterior heterogeneity is already excluded in the case of the common prior assumption being satisfied, if the agents face symmetric information (i.e. receive precisely the same information). Consequently, distinct posteriors need to be due to information variety.

As an illustration of our formal framework as embodied by Definitions 1 to 4, consider again the sea shipment allusion from Section 1. A lexicographic Aumann structure (cf. Definition 2) would represent a situation, where different merchants hold contingent prior beliefs and are equipped with private information about the arrival of some sea shipment. Suppose that $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}$ comprises eight worlds. The eight worlds describe eight possible scenarios that are conceivable by all the merchants:

- the shipment arrives in fine weather with no pirate attack occurring ($\omega_1 \in \Omega$),
- the sea shipment does not arrive in fine weather with no pirate attack occurring ($\omega_2 \in \Omega$),
- the shipment arrives in adverse weather with no pirate attack occurring ($\omega_3 \in \Omega$),
- the shipment does not arrive in adverse weather with no pirate attack occurring ($\omega_4 \in \Omega$),
- the shipment arrives in fine weather with pirates attacking ($\omega_5 \in \Omega$),
- the shipment does not arrive in fine weather with pirates attacking ($\omega_6 \in \Omega$),
- the shipment arrives in adverse weather with pirates attacking ($\omega_7 \in \Omega$),
- the shipment does not arrive in adverse weather with pirates attacking ($\omega_8 \in \Omega$).

Suppose that some merchant $i \in I$ deems it substantially more likely that a pirate attack does not occor. In fact, he only considers the latter to be a hypothetical contingency but he nonetheless does not discard it entirely from his thinking. Suppose further that i enjoys access to a reliable meteorological source which is signalling fine weather conditions. Such a state of mind could be modelled in our framework as follows. Merchant i's information partition could be given by $\mathcal{I}_i = \big\{\{\omega_1, \omega_2, \omega_5, \omega_6\}, \{\omega_3, \omega_4, \omega_7, \omega_8\}\big\}$ and suppose that his subjective prior would be given by an i-LPS (cf. Definition 1) as follows: $\rho_i = (p_i^1, p_i^2)$ such that $p_i^1(\omega_1) = \frac{4}{9}$, $p_i^1(\omega_2) = p_i^1(\omega_3) = \frac{1}{9}$, and $p_i^1(\omega_4) = \frac{3}{9}$, as well as $p_i^2(\omega_5) = \frac{1}{4}$, $p_i^2(\omega_6) = \frac{1}{8}$, $p_i^2(\omega_7) = \frac{1}{8}$, and $p_i^2(\omega_8) = \frac{1}{2}$. Assume that the shipment does arrive under fine weather conditions while withstanding a pirates' attack. Formally speaking, ω_5 becomes the actual state of the world. The relevant posterior of merchant i is the ω_5-conditional i-LPS (cf. Definition 3) which then obtains as $\rho_i^{\omega_5} = \Big(p_i^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega_5)\big), p_i^{m_2}\big(\cdot \mid \mathcal{I}_i(\omega_5)\big)\Big)$ such that $p_i^{m_1}\big(\omega_1 \mid \mathcal{I}_i(\omega_5)\big) = \frac{4}{5}$ and $p_i^{m_1}\big(\omega_2 \mid \mathcal{I}_i(\omega_5)\big) = \frac{1}{5}$, as well as $p_i^{m_2}\big(\omega_5 \mid \mathcal{I}_i(\omega_5)\big) = \frac{2}{3}$ and $p_i^{m_2}\big(\omega_6 \mid \mathcal{I}_i(\omega_5)\big) = \frac{1}{3}$. Moreover, in the case of the merchants being like-minded – for instance due to similar relevant past experiences with sea shipments – a common prior (cf. Definition 4) could be imposed. The sequence of prior beliefs would then be the same for all merchants, i.e. there would exist $\rho = (p^1, \dots, p^M)$ such that $\rho_j = \rho$ for all $j \in I$.

3 Weak Agreement

Since the agents hold levels of posterior beliefs, agreement becomes a multifarious notion. Identical beliefs can obtain (or not) at different lexicographic layers. In fact, it is now shown that common knowledge of lexicographic posteriors ensures the agents' first level posterior beliefs to coincide.

Theorem 1 (WAT). *Let $\mathcal{A}_{LCP}$ be a lexicographic Aumann structure with a common prior, $E \subseteq \Omega$ be some event, and $\omega \in \Omega$ be some world. If*

$$CK\Big(\bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}\big(E \mid \mathcal{I}_i(\omega')\big) = p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big)\big\}\Big) \neq \emptyset,$$

then

$$p^{m_1}\big(E \mid \mathcal{I}_i(\omega)\big) = p^{m_1}\big(E \mid \mathcal{I}_j(\omega)\big)$$

for all $i, j \in I$.

Proof. Let $j \in I$ be some agent, $A_j \subseteq \Omega$ be some set such that $(\bigwedge_{i \in I} \mathcal{I}_i)(\omega) = \bigcup_{\omega' \in A_j} \mathcal{I}_j(\omega')$ and $\mathcal{I}_j(\omega_1) \cap \mathcal{I}_j(\omega_2) = \emptyset$ for all $\omega_1, \omega_2 \in A_j$. Moreover, let $m \in \{1, \dots, M\}$ be the first lexicographic level such that $p^m\big((\bigwedge_{i \in I} \mathcal{I}_i)(\omega)\big) > 0$. Consider some world $\bar{\omega} \in A_j$. If $p^m\big(\mathcal{I}_j(\bar{\omega})\big) > 0$, then $p^{m_1}\big(\cdot \mid \mathcal{I}_j(\bar{\omega})\big) = p^m\big(\cdot \mid \mathcal{I}_j(\bar{\omega})\big)$, and by Bayesian updating,

$$p^{m_1}\big(E \mid \mathcal{I}_j(\bar{\omega})\big) \cdot p^m\big(\mathcal{I}_j(\bar{\omega})\big) = p^m\big(E \cap \mathcal{I}_j(\bar{\omega})\big)$$

holds. Alternatively, if $p^m\big(\mathcal{I}_j(\bar{\omega})\big) = 0$, then $p^m\big(E \cap \mathcal{I}_j(\bar{\omega})\big) = 0$. Since $p^{m_1}\big(\cdot \mid \mathcal{I}_j(\bar{\omega})\big)$ is well-defined,

$$p^{m_1}\big(E \mid \mathcal{I}_j(\bar{\omega})\big) \cdot p^m\big(\mathcal{I}_j(\bar{\omega})\big) = p^m\big(E \cap \mathcal{I}_j(\bar{\omega})\big)$$

holds trivially. Therefore,

$$p^{m_1}\big(E \mid \mathcal{I}_j(\omega')\big) \cdot p^m\big(\mathcal{I}_j(\omega')\big) = p^m\big(E \cap \mathcal{I}_j(\omega')\big)$$

obtains for all $\omega' \in A_j$.

As

$$A_j \subseteq (\bigwedge_{i \in I} \mathcal{I}_i)(\omega) \subseteq CK\Big(\bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}\big(E \mid \mathcal{I}_i(\omega')\big) = p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big)\big\}\Big)$$

$$\subseteq \bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}\big(E \mid \mathcal{I}_i(\omega')\big) = p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big)\big\},$$

it is the case that $p^{m_l}(E \mid \mathcal{I}_i(\omega')) = p^{m_l}(E \mid \mathcal{I}_i(\omega))$, for all $i \in I$ for all $l \in \{1, \dots, L\}$ and for all $\omega' \in A_j$. In particular, $p^{m_1}(E \mid \mathcal{I}_j(\omega')) = p^{m_1}(E \mid \mathcal{I}_j(\omega))$ holds for all $\omega' \in A_j$. It follows that

$$p^{m_1}(E \mid \mathcal{I}_j(\omega)) \cdot p^m\big(\mathcal{I}_j(\omega')\big) = p^m\big(E \cap \mathcal{I}_j(\omega')\big)$$

holds for all $\omega' \in A_j$. Summing over all $\omega' \in A_j$ and using countable additivity yields

$$p^{m_1}\big(E \mid \mathcal{I}_j(\omega)\big) = \frac{p^m\big(E \cap (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)\big)}{p^m\big((\bigwedge_{i \in I} \mathcal{I}_i)(\omega)\big)}.$$

Since j has been chosen arbitrarily, it can be concluded that

$$p^{m_1}\big(E \mid \mathcal{I}_i(\omega)\big) = p^{m_1}\big(E \mid \mathcal{I}_j(\omega)\big)$$

for all $i, j \in I$. ■

Agents can thus not agree to disagree on their first level posterior beliefs. The preceding result remains silent though on any lexicographic level deeper than level one. In this sense, **WAT** establishes a form of weak agreement within the lexicographic framework.

Note that it is not possible to establish **WAT** by simply truncating the lexicographic Aumann structure at the first prior level and then applying Aumann's proof of his original agreement theorem to this simpler structure. This is because the first level prior may not assign positive probability to some agent's information cell, which in turn implies that a deeper level prior needs to be invoked to compute his first level posterior. Such possibilities need to be accommodated by the proof of weak agreement theorem.

For the special case of exclusively admitting the first level posteriors – formally, only considering $p_i^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega)\big)$ for all $\omega \in \Omega$ and for all $i \in I$ – our framework of lexicographic Aumann structures becomes essentially equivalent to Bach and Perea (2013)'s model, which only employs a lexicographic common prior but unique posteriors. Their non-overlapping support condition across lexicographic prior levels is not assumed in our framework though. Thus, **WAT** can be seen as a generalization of Bach and Perea (2013, Theorem 1). If not only the posteriors but also the common prior are restricted to a single probability measure, i.e. $M = 1$, then Aumann (1976)'s model can be recovered and **WAT** becomes the original agreement theorem.

4 Disagreement

Attention is now focussed on the deeper lexicographic levels. It turns out that agents can agree to disagree on posteriors beyond the first lexicographic level.

Proposition 1 (DIS). *There exist a lexicographic Aumann structure $\mathcal{A}_{LCP}$ with a common prior, some event $E \subseteq \Omega$, and some world $\omega \in \Omega$, such that*

$$CK\Big(\bigcap_{i \in I} \bigcap_{l \in \{1,\ldots,L\}} \big\{\omega' \in \Omega : p^{m_l}(E \mid \mathcal{I}_i(\omega')) = p^{m_l}(E \mid \mathcal{I}_i(\omega))\big\}\Big) \neq \varnothing$$

and

$$p^{m_{l^*}}(E \mid \mathcal{I}_i(\omega)) \neq p^{m_{l^*}}(E \mid \mathcal{I}_j(\omega))$$

for some $i, j \in I$ and for some $l^* \in \{2, \ldots, L\}$.

Proof. Let $\mathcal{A}_{LCP} = \big(\Omega, I, (\mathcal{I}_i)_{i \in I}, \rho\big)$ be a lexicographic Aumann structure with a common prior, where

- $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$,
- $I = \{Alice, Bob\}$,
- $\mathcal{I}_{Alice} = \big\{\{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}\big\}$,

– $\mathcal{I}_{Bob} = \{\Omega\}$,
– and $\rho = (p^1, p^2, p^3)$ with $p^1(\omega_1) = 1$, $p^2(\omega_2) = \frac{1}{3}$, $p^2(\omega_3) = \frac{2}{3}$, $p^3(\omega_4) = 1$.

Consider the event $E = \{\omega_1, \omega_3\}$. Observe that

$$p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = p^1\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = 1$$

for all $\omega \in \{\omega_1, \omega_2\}$, and

$$p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = p^2\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = 1$$

for all $\omega \in \{\omega_3, \omega_4\}$.[10] Consequently, $p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = 1$ obtains at every world $\omega \in \Omega$. Also, observe that

$$p^{m_1}\big(E \mid \mathcal{I}_{Bob}(\omega)\big) = p^1\big(E \mid \mathcal{I}_{Bob}(\omega)\big) = 1$$

for all $\omega \in \Omega$. Therefore, *Alice*'s and *Bob*'s first level posterior beliefs of E coincide.

Moreover, it is the case that

$$p^{m_2}\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = p^2\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = 0$$

for all $\omega \in \{\omega_1, \omega_2\}$, and

$$p^{m_2}\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = p^3\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = 0$$

for all $\omega \in \{\omega_3, \omega_4\}$. Hence, $p^{m_2}\big(E \mid \mathcal{I}_{Alice}(\omega)\big) = 0$ obtains at every world $\omega \in \Omega$. Also,

$$p^{m_2}\big(E \mid \mathcal{I}_{Bob}(\omega)\big) = p^2\big(E \mid \mathcal{I}_{Bob}(\omega)\big) = \frac{2}{3}$$

holds at every world $\omega \in \Omega$. Therefore, *Alice*'s and *Bob*'s second level posterior beliefs of E do not coincide.

Taking $\omega = \omega_1$ guarantees that

$$CK\Big(\bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}(E \mid \mathcal{I}_i(\omega')) = p^{m_l}(E \mid \mathcal{I}_i(\omega))\big\}\Big) = CK(\Omega) = \Omega \neq \varnothing,$$

while

$$p^{m_2}(E \mid \mathcal{I}_{Alice}(\omega)) = 0 \neq \frac{2}{3} = p^{m_2}(E \mid \mathcal{I}_{Bob}(\omega))$$

obtains at the second lexicographic level m_2. ■

A possibility result on agreeing to disagree thus emerges with lexicographic probability systems. Common knowledge of the agents' lexicographic posteriors does manifestly not suffice to establish agreement at all lexicographic levels. The agents can entertain

[10] Recall that in the expressions $p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega_1)\big)$ and $p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega_3)\big)$, index m_1 is a shortcut notation for the two different indices $m_{Alice,\omega_1,1}$ and $m_{Alice,\omega_3,1}$, respectively. Hence, equalities $p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega_1)\big) = p^1\big(E \mid \mathcal{I}_{Alice}(\omega_1)\big)$ and $p^{m_1}\big(E \mid \mathcal{I}_{Alice}(\omega_3)\big) = p^2\big(E \mid \mathcal{I}_{Alice}(\omega_3)\big)$ imply that $m_{Alice,\omega_1,1} = 1$ and $m_{Alice,\omega_3,1} = 2$, respectively.

distinct posteriors at lexicographic levels beyond one, and at the same time acknowledge this divergence. This result is somewhat surprising as it lexicographically counters Aumann's impossibility theorem. Besides, note that **DIS** would still apply and the same proof would remain valid, if a disjoint support condition were to be imposed on the lexicographic level priors.

Conceptually, **DIS** raises the question as to what drives the disagreement in a lexicographically enriched set-up. From Aumann's agreement theorem, it is typically concluded that asymmetric information does not suffice to explain heterogeneity in posterior beliefs of Bayesian agents with a common prior. Consequently, disagreement can be reached by either weakening the common knowledge assumption or the common prior assumption. Such a conclusion does no longer apply in our lexicographic framework, since by **DIS** heterogeneous posteriors can obtain despite common knowledge of posteriors as well as the common prior remaining intact. In contrast to Aumann's original set-up with standard beliefs, the lexicographic beliefs in our framework are capable of capturing hypothetical reasoning. The conceptual conclusion of Aumann's impossibility result with regard to disagreement is thus refined by **DIS** which detects hypothetical reasoning as a third source for heterogeneity in posterior beliefs.

5 Strong Agreement

The impossibility theorem of **WAT** is weak in the sense that it only affects the first lexicographic posterior level and agreement can already fall apart at the second level as **DIS** shows. Further assumptions about the agents' like-mindedness are thus needed for a stronger result yielding equal posteriors at every lexicographic level. For this purpose an adaptation of absolute mutual absolute continuity from probability theory is introduced.

Definition 5. *Let $\mathcal{A}_{LCP}$ be a lexicographic Aumann structure with a common prior and $\omega \in \Omega$ be some world. The common prior ρ is* mutually absolutely continuous*, whenever*

$$p^m\big(\mathcal{I}_i(\omega)\big) = 0, \text{ if and only if, } p^m\big(\mathcal{I}_j(\omega)\big) = 0$$

for all $\omega \in \Omega$, for all $i, j \in I$, and for all $m \in \{1, \dots, M\}$.

Mutual absolute continuity ensures that at every lexicographic level the corresponding common prior handles the agents' information in synchrony. In any conceivable contingency, either the received private information at a world is deemed possible for all agents or it is excluded for everyone. Mutual absolute continuity can thus be viewed as a kind of lexicographic "same-excluding" condition.

The interpretation of the common prior assumption in the original Aumann structures with standard beliefs as agent like-mindedness can be adapted to our framework with lexicographic beliefs. The lexicographic common prior adds a contingent form of like-mindedness that also covers the different layers of hypothetical reasoning a priori. In this sense a lexicographic common prior that is mutually absolutely continuous constitutes an *intensified like-mindedness* assumption, where the players' hypothetical reasoning conditional on their information is aligned. In fact, this condition ensures that for every posterior level the agents' conditional beliefs are computed with the same level prior. If the agents violate intensified like-mindedness, then it can happen that at some posterior level they base their updated beliefs on distinct level priors. In other words, the lexicographic like-mindedness a priori gets lost in the process of Bayesian updating. The lexicographic Aumann structure constructed in the proof of **DIS** illustrates this phenomenon.

Formally, our mutual absolute continuity condition imposed on the common prior is closely related to the standard notion in probability theory which concerns two probability measures. Let μ and ν be measures on some set Ω, and define $\mu << \nu$, if $\nu(F) = 0$ implies $\mu(F) = 0$ for all $F \in 2^{\Omega}$. Let the two measures μ and ν be called *standard mutually absolutely continuous*, whenever $\mu << \nu$ and $\nu << \mu$.[11] Observe that the common prior ρ induces for every level $m \in \{1, \ldots, M\}$ and for every player $i \in I$ a measure $\mu_i^m : 2^{\Omega} \to [0,1]$ given by

$$\mu_i^m(F) := \begin{cases} 0 & \text{if } F = \varnothing \\ \sum_{\omega \in F} \frac{p^m\left(\mathcal{I}_i(\omega)\right)}{|\mathcal{I}_i(\omega)|} & \text{otherwise,} \end{cases}$$

for all $F \in 2^{\Omega}$. Now, if $\mu_i^m(F) = \sum_{\omega \in F} \frac{p^m\left(\mathcal{I}_i(\omega)\right)}{|\mathcal{I}_i(\omega)|} > 0$ for some $F \in 2^{\Omega}$, then there exists $\omega' \in F$ such that $p^m\left(\mathcal{I}_i(\omega')\right) > 0$. By the mutual absolute continuity condition of Definition 5, $p^m\left(\mathcal{I}_j(\omega')\right) > 0$ thus holds too, and consequently $\mu_j^m(F) = \sum_{\omega \in F} \frac{p^m\left(\mathcal{I}_j(\omega)\right)}{|\mathcal{I}_j(\omega)|} > 0$. Conversely, if $p^m\left(\mathcal{I}_i(\omega)\right) > 0$ for some $\omega \in \Omega$, then $\mu_i^m(\{\omega\}) > 0$. By standard mutual absolute continuity, $\mu_j^m(\{\omega\}) > 0$ hence also obtains, and consequently $p^m\left(\mathcal{I}_j(\omega)\right) > 0$. Therefore, the following formal characterization our mutual absolute continuity adaptation in terms of standard mutual absolute continuity from probability theory ensues.

Remark 1. Let $\mathcal{A}_{LCP}$ be a lexicographic Aumann structure with a common prior. The common prior ρ is mutually absolutely continuous, if and only if, μ_i^m and μ_j^m are standard mutually absolutely continuous for all $i, j \in I$ and for all $m \in \{1, \ldots, M\}$.

Mutual absolute continuity in line with Definition 5 can thus be viewed as a variant of standard mutual absolute continuity from probability theory.

In fact, our condition of Definition 5 is also similar to Stuart (1997)'s use of mutual absolute continuity.[12] Accordingly, if some agent's belief assigns a positive probability to a state (which essentially corresponds to a possible world in our framework), then so do all the other agents. Even though Stuart (1997) does not impose any priors, an agent's belief in his model can be viewed as a posterior. While the underlying idea of Stuart's (1997) mutual absolute continuity and ours is the same – some form of synchronicity in both consideration and omission – his version concerns posterior beliefs and possible worlds, whereas ours refers to prior beliefs and information.

[11] In probability theory, two mutually absolutely continuous measures are sometimes also called equivalent.

[12] In Stuart (1997), mutual absolute continuity plays an important role in establishing all period defection in the normal-form model of the finitely repeated prisoners' dilemma.

It turns out that mutual absolute continuity together with the common prior assumption and common knowledge of posteriors implies that the agents' posterior beliefs coincide at all lexicographic levels.

Theorem 2 (SAT). *Let $\mathcal{A}_{LCP}$ be a lexicographic Aumann structure with a common prior, $E \subseteq \Omega$ be some event, and $\omega \in \Omega$ be some world. If ρ is mutually absolutely continuous and*

$$CK\Big(\bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}\big(E \mid \mathcal{I}_i(\omega')\big) = p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big)\big\}\Big) \neq \emptyset,$$

then

$$p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big) = p^{m_l}\big(E \mid \mathcal{I}_j(\omega)\big)$$

for all $i, j \in I$ and for all $l \in \{1, \dots, L\}$.

Proof. We first show that if ρ is mutually absolutely continuous, then the lexicographic indices of the ω'-conditional i-LPS $\rho_i^{\omega'}$ are the same for all $\omega' \in (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)$ and for all $i \in I$. Let $j \in I$, $\omega' \in (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)$ as well as $(m_l)_{l=1}^L$ and $(m'_l)_{l=1}^L$ be the indices of ρ_j^{ω} and $\rho_j^{\omega'}$, respectively. Since $\omega' \in (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)$, the world ω' is doxastically reachable from ω, i.e., there exists a sequence $(P^k)_{k=1}^N$ of information cells such that $\omega \in P^1$, $\omega' \in P^N$, and $P^k \cap P^{k+1} \neq \emptyset$ for all $1 \leqslant k < N$. Since ρ is mutually absolutely continuous, it is the case that, $p^m(P^k) = 0$ if and only if $p^m(P^{k+1}) = 0$ for all $m \in \{1, \dots, M\}$ and for all $1 \leqslant k < N$. Thus, $p^m(P^1) = 0$ if and only if $p^m(P^N) = 0$ for all $m \in \{1, \dots, M\}$. Since $\omega \in \mathcal{I}_j(\omega) \cap P^1$, $\omega' \in \mathcal{I}_j(\omega') \cap P^N$ and ρ is mutually absolutely continuous, it follows that $p^m\big(\mathcal{I}_j(\omega)\big) = 0$ if and only if $p^m(P^1) = 0$ and $p^m\big(\mathcal{I}_j(\omega')\big) = 0$ if and only if $p^m(P^N) = 0$, and thus $p^m\big(\mathcal{I}_j(\omega)\big) = 0$ if and only if $p^m\big(\mathcal{I}_j(\omega')\big) = 0$, for all m $\in \{1, \dots, M\}$. Consequently, $(m_l)_{l=1}^L = (m'_l)_{l=1}^L$. Now, towards a contradiction, suppose that there exist $j' \in I$ and $l \in \{1, \dots, L\}$ such that $m'_l \neq m''_l$, where $(m''_l)_{l=1}^L$ are the indices of $\rho_{j'}^{\omega'}$. Without loss of generality, suppose that l is the least such index. Then, either $m'_l < m''_l$, in which case, $p^{m'_l}\big(\mathcal{I}_j(\omega')\big) > 0$ and $p^{m'_l}\big(\mathcal{I}_{j'}(\omega')\big) = 0$, or $m'_l > m''_l$, in which case, $p^{m''_l}\big(\mathcal{I}_j(\omega')\big) = 0$ and $p^{m''_l}\big(\mathcal{I}_{j'}(\omega')\big) > 0$. In both cases, a contradiction with the mutual absolute continuity of ρ obtains. Consequently, $(m_l)_{l=1}^L = (m'_l)_{l=1}^L = (m''_l)_{l=1}^L =: (\bar{m}_l)_{l=1}^L$. The ω'-conditional i-LPS can then be written as

$$\rho_i^{\omega'} = \rho(\cdot \mid \mathcal{I}_i(\omega')) = \Big(p^{\bar{m}_1}\big(\cdot \mid \mathcal{I}_i(\omega')\big), \dots, p^{\bar{m}_L}\big(\cdot \mid \mathcal{I}_i(\omega')\big)\Big)$$

for all $i \in I$ and for all $\omega' \in (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)$.

We are now ready to derive agreement in posteriors. Let $j' \in I$ and $A_{j'} \subseteq \Omega$ such that $(\bigwedge_{i \in I} \mathcal{I}_i)(\omega) = \bigcup_{\omega' \in A_{j'}} \mathcal{I}_{j'}(\omega')$ and $\mathcal{I}_{j'}(\omega_1) \cap \mathcal{I}_{j'}(\omega_2) = \emptyset$ for all $\omega_1, \omega_2 \in A_{j'}$. Note that

$$A_{j'} \subseteq (\bigwedge_{i \in I} \mathcal{I}_i)(\omega) \subseteq CK\Big(\bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}\big(E \mid \mathcal{I}_i(\omega')\big) = p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big)\big\}\Big)$$

$$\subseteq \bigcap_{i \in I} \bigcap_{l \in \{1,\dots,L\}} \big\{\omega' \in \Omega : p^{m_l}\big(E \mid \mathcal{I}_i(\omega')\big) = p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big)\big\}.$$

Consider some $l' \in \{1, \dots, L\}$. It follows that

$$p^{m_{l'}}\big(E \mid \mathcal{I}_{j'}(\omega)\big) = p^{m_{l'}}\big(E \mid \mathcal{I}_{j'}(\omega')\big) = \frac{p^{m_{l'}}\big(E \cap \mathcal{I}_{j'}(\omega')\big)}{p^{m_{l'}}\big(\mathcal{I}_{j'}(\omega')\big)} = \frac{p^{\bar{m}_{l'}}\big(E \cap \mathcal{I}_{j'}(\omega')\big)}{p^{\bar{m}_{l'}}\big(\mathcal{I}_{j'}(\omega')\big)}$$

for all $\omega' \in A_{j'}$. Consequently,

$$p^{m_{l'}}\big(E \mid \mathcal{I}_{j'}(\omega)\big) \cdot p^{\bar{m}_{l'}}\big(\mathcal{I}_{j'}(\omega')\big) = p^{\bar{m}_{l'}}\big(E \cap \mathcal{I}_{j'}(\omega')\big),$$

for all $\omega' \in A_{j'}$. Summing over all $\omega' \in A_{j'}$ and using countable additivity yields

$$p^{m_{l'}}\big(E \mid \mathcal{I}_{j'}(\omega)\big) = \frac{p^{\bar{m}_{l'}}\big(E \cap (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)\big)}{p^{\bar{m}_{l'}}\big((\bigwedge_{i \in I} \mathcal{I}_i)(\omega)\big)} = p^{\bar{m}_{l'}}\big(E \mid (\bigwedge_{i \in I} \mathcal{I}_i)(\omega)\big).$$

Since j' and l' have been chosen arbitrarily, it can be concluded that

$$p^{m_l}\big(E \mid \mathcal{I}_i(\omega)\big) = p^{m_l}\big(E \mid \mathcal{I}_j(\omega)\big)$$

for all $i, j \in I$ and for all $l \in \{1, \dots, L\}$. ■

It is thus impossible for lexicographically-minded agents to agree to disagree whenever mutual absolute continuity is satisfied. In contrast to **WAT**, which only ensures a weak form of agreement at the first posterior level, **SAT** establishes strong agreement at all lexicographic posterior levels.

From a conceptual perspective, agreement is only ensured in the lexicographically enriched framework by a substantial strengthening of the agents' like-mindedness. It does not suffice to require a common prior at all reasoning levels. On top of that, each of these priors also has to synchronically consider or synchronically neglect the agents' information in order to reconcile their updating. Together with common knowledge of posteriors, the assumption of intensified like-mindedness drives the homogeneity of the posteriors in our lexicographic framework.

The particular lexicographic Aumann structure constructed in the proof of **DIS** suggests that **SAT** qualifies as tight with respect to the mutual absolute continuity condition.[13] There the other two key assumptions, i.e. common prior as well common knowledge of posteriors, but not mutual absolute continuity hold, while the consequent, i.e. lexicographically identical posterior beliefs, fails.

Continuity in agreeing to lexicographically disagree follows from **SAT** in the sense that equal prior beliefs up to some lexicographic prior level imply equal posterior beliefs up to a corresponding lexicographic posterior level. Suppose that the common prior assumption is weakened such that the agents' priors coincide up to some level $\bar{M} < M$, and modify the initial lexicographic Aumann structure by truncating the agent's lexicographic priors at $\bar{M}$, which is equivalent to imposing a common prior $\rho = (p^1, \dots, p^{\bar{M}})$. By **SAT** it follows that common knowledge of lexicographic posteriors at some world $\omega \in \Omega$ implies equal posterior measures for every level $l \in \big\{1, \dots, \min\{L_{i,\omega} \in \mathbb{N} : i \in I\}\big\}$ in the truncated structure, and hence also up to level $\min\{L_{i,\omega} \in \mathbb{N} : i \in I\}$ in the initial lexicographic Aumann structure. In this sense, the lexicographic impossibility result of **SAT** is continuous.

[13] Tightness is interpreted in the style of Aumann and Brandenburger (1995), i.e. whether dropping only one assumption of a result were to already break its conclusion.

6 Perfect Equilibrium

Next, we turn to game theory where some of our results on lexicographic agreeing to disagree are employed for an epistemic analysis of tremble equilibria. In game theory, strategic interaction of multiple agents is modelled, and possible outcomes are predicted based on different assumptions. Static games with complete information constitute the most elementary analytical framework. Formally, such games are represented by a tuple

$$\Gamma = \big(I, (S_i)_{i\in I}, (U_i)_{i\in I}\big)$$

consisting of a finite set I of players and finite non-empty strategy sets S_i as well as real-valued utility functions U_i with domain $\times_{j\in I} S_j$ for every player $i \in I$. In terms of notation, the set $S_{-i} := \times_{j\in I\setminus\{i\}} S_j$ refers to the product set of the i's opponents' strategy combinations. The tuple $\Gamma = \big(I, (S_i)_{i\in I}, (U_i)_{i\in I}\big)$ is often also referred to as normal form. As background hypotheses it is stipulated that all players choose their strategies simultaneously and that the ingredients of the game, i.e. the normal form, is common knowledge among the players. Solution concepts propose plausibility criteria or decision rules in line with which the players are supposed to act. Formally, a solution concept defines a subset $SC \subseteq \times_{i\in I} S_i$ of the set of all strategy combinations as possible outcomes of the game.

The solution concept of Nash equilibrium – due to Nash (1950) and (1951) – requires players to choose utility maximizing against fixed strategies of the opponents. In order to ensure existence of an equilibrium point in any game, also randomizations over strategies are admitted. The set of choice objects for every player $i \in I$ is thus enlarged from S_i to $\Delta(S_i)$, where a typical element σ_i of $\Delta(S_i)$ is called a mixed strategy of player i. The utility functions U_i are extended from $\times_{j\in I} S_j$ to $\times_{j\in I} \big(\Delta(S_j)\big)$ for every player $i \in I$ by an expected utility computation. A tuple of mixed strategies $\sigma = (\sigma_j)_{j\in I}$ constitutes a *Nash equilibrium*, whenever

$$s_i^* \in \arg\max_{s_i\in S_i}\Big\{ \sum_{s_{-i}\in S_{-i}} \big(\bigotimes_{j\in I\setminus\{i\}} \sigma_j\big)(s_{-i}) \cdot U_i(s_i, s_{-i})\Big\} \tag{1}$$

for all $s_i^* \in \text{supp}(\sigma_i)$ and for all $i \in I$.[14] If equation (1) holds, s_i^* is called a best response to σ_{-i}, where $\sigma_{-i} := (\sigma_j)_{j\in I\setminus\{i\}}$. Player i is said to strictly prefer a strategy s_i to some other strategy s_i' given σ_{-i}, whenever $\sum_{s_{-i}\in S_{-i}}(\bigotimes_{j\in I\setminus\{i\}} \sigma_j)(s_{-i}) \cdot U_i(s_i, s_{-i}) > \sum_{s_{-i}\in S_{-i}}(\bigotimes_{j\in I\setminus\{i\}} \sigma_j)(s_{-i}) \cdot U_i(s_i', s_{-i})$ holds.

In classical game theory, the multiplicity of Nash equilibria in many games has been deemed unsatisfactory and refinements have thus been sought. A particular class of equilibrium refinements is based on the idea that players can make mistakes with small probability. Phrased in more vivid terms: players possibly tremble when implementing their strategies. In line with this intuition, various tremble equilibria have been

[14] Given a probability measure $p \in \Delta(X)$ on some set X its support is defined as $\text{supp}(p) := \{x \in X : p(x) > 0\}$. Fixing $K \in \mathbb{N}$ and probability measures p_k on sets X_k for all $k \in \{1, \dots, K\}$, $\bigotimes_{k\in\{1,\dots,K\}} p_k$ denotes the product measure on the set $\times_{k\in\{1,\dots,K\}} X_k$.

proposed in the literature. The most basic such solution concept is Selten's (1975) perfect equilibrium.[15] Essentially, attention is restricted to Nash equilibria that obtain as limits of sequences of perturbed strategy combinations. While originally introduced by Selten (1975, Section 8) as a solution concept for dynamic games, perfect equilibrium has also been widely used in static games. A formal definition of perfect equilibrium for the class of static games ensues as follows.

Definition 6. *Let Γ be a game and $\sigma = (\sigma_i)_{i \in I} \in \times_{i \in I} \Delta(S_i)$ be a tuple of mixed strategies. The tuple σ constitutes a* perfect equilibrium *of Γ, if there exists a sequence of tuples of mixed strategies $(\sigma^k)_{k \in \mathbb{N}} = \big((\sigma_i^k)_{i \in I}\big)_{k \in \mathbb{N}} \in \big(\times_{i \in I} \Delta(S_i)\big)^{\mathbb{N}}$ such that*

(i) $\lim_{k \to \infty} \sigma^k = \sigma$;
(ii) for all $i \in I$ and for all $k \in \mathbb{N}$, it is the case that $\mathrm{supp}(\sigma_i^k) = S_i$;
(iii) for all $i \in I$ and for all $k \in \mathbb{N}$, if $s_i \in \mathrm{supp}(\sigma_i)$, then s_i is a best response to σ_{-i}^k.

A perfect equilibrium thus always coincides with the limit of a sequence of trembles. Moreover, for every player, his perfect equilibrium mixed strategy only assigns positive probability to strategies that are best responses to any of the opponents' tremble combinations. It can be shown that a perfect equilibrium must be a Nash equilibrium (Selten, 1975, Lemma 9). This result essentially rests on the fact that the expected utilities are continuous in mixed strategy profiles. Conversely, Nash equilibrium does not imply perfect equilibrium. The latter solution concept thus is stronger than the former. In classical parlance, perfect equilibrium constitutes a refinement of Nash equilibrium.

The following example illustrates these two solution concepts.

Example 1. Consider the two player game depicted in Figure 1 with players *Alice* and *Bob*, where *Alice* chooses a "row" (a or b) and *Bob* picks a "column" (y or z). The mixed strategy tuple $\sigma = (\sigma_{Alice}, \sigma_{Bob})$, where $\sigma_{Alice}(a) = 1$ and $\sigma_{Bob}(y) = 1$, forms a Nash equilibrium, as a is a best response to σ_{Bob} and y is a best response to σ_{Alice}. To see that σ also constitutes a perfect equilibrium, construct a sequence of tuples of mixed strategies $(\sigma^k)_{k \in \mathbb{N}\backslash\{0\}} = \big((\sigma_{Alice}^k, \sigma_{Bob}^k)\big)_{k \in \mathbb{N}\backslash\{0\}}$ by setting $\sigma_{Alice}^k(a) = 1 - \frac{1}{k+1}$, $\sigma_{Alice}^k(b) = 0 + \frac{1}{k+1}$, $\sigma_{Bob}^k(y) = 1 - \frac{1}{k+1}$ and $\sigma_{Bob}^k(z) = 0 + \frac{1}{k+1}$ for all $k \in \mathbb{N}\backslash\{0\}$. Observe that $\lim_{k \to \infty} \sigma^k = \sigma$ as well as $\mathrm{supp}(\sigma_{Alice}^k) = S_{Alice}$ and $\mathrm{supp}(\sigma_{Bob}^k) = S_{Bob}$ for all $k \in \mathbb{N}\backslash\{0\}$. Moreover, a is a best response to σ_{Bob}^k for all $k \in \mathbb{N}\backslash\{0\}$ and y is a best response to σ_{Alice}^k for all $k \in \mathbb{N}\backslash\{0\}$. It follows that σ is a perfect equilibrium.

	y	z
a	1, 1	0, 0
b	0, 0	0, 0

Fig. 2.

[15] Other tremble equilibria are, for instance, Myerson's (1978) proper equilibrium, van Damme's (1984) quasi-perfect equilibrium, as well as Selten and Harsanyi's (1988) uniformly perfect equilibrium.

The mixed strategy tuple $\sigma' = (\sigma'_{Alice}, \sigma'_{Bob})$, where $\sigma'_{Alice}(b) = 1$ and $\sigma'_{Bob}(z) = 1$ also constitutes a Nash equilibrium, since b is a best response to σ_{Bob} and z is a best response to σ_{Alice}. However, it does not form a perfect equilibrium. Suppose that there exists a sequence of full support mixed strategy tuples $(\sigma^k_{Alice}, \sigma^k_{Bob})_{k \in \mathbb{N}\setminus\{0\}} \in \big(\Delta(S_{Alice}) \times \Delta(S_{Bob})\big)^{\mathbb{N}\setminus\{0\}}$ with limit point σ'. Then, b cannot be a best response to σ^k_{Bob} for any $k \in \mathbb{N}\setminus\{0\}$. Indeed, as soon as y receives positive probability, only a can be a best reponse for *Alice*. It follows that σ' is not a perfect equilibrium. ♣

7 Lexicographic Characterization

It is known that tremble equilibria with their sequences of full support mixed strategy tuples can be characterized in terms of lexicographic conjectures. The latter can be modelled as lexicographic probability systems in which for every player the set of opponents' choice combinations defines the basic space of uncertainty. Perfect equilibrium and proper equilibrium have been reformulated with lexicographic conjectures by Blume et al. (1991b) and shown to be equivalent to their notion of lexicographic Nash equilibrium plus further restrictions, respectively. In this section we define lexicographic perfect equilibrium and lexicographic semi-perfect equilibrium. While these two solution concepts phrased in terms of lexicographic conjectures essentially correspond to variants of Blume et al.'s (199b) lexicographic Nash equilibrium, our definitions are aligned with our formal framework and formulated in a direct way.

Some further concepts and notation need to be introduced. Let Γ be a game and $i \in I$ be some player. A sequence $\beta_i = (b_i^1, \ldots, b_i^L) \in \big(\Delta(S_{-i})\big)^L$ of probability measures, for some $L \in \mathbb{N}$, is called player i's *lexicographic conjecture*. For the sake of simplicity we assume the same number L of levels for all $i \in I$. A lexicographic conjecture β_i is *cautious*, whenever for all $j \in I\setminus\{i\}$ and for all $s_j \in S_j$, there exists some lexicographic level $l^* \in \{1, \ldots, L\}$ such that $\text{marg}_{S_j} b_i^{l^*}(s_j) > 0$, where $\text{marg}_{S_j} b_i^{l^*}(s_j) := \sum_{s_{-(i,j)} \in S_{-(i,j)}} b_i^{l^*}(s_{-(i,j)}, s_j)$ for all $s_j \in S_j$. Given a strategy $s_i \in S_i$ and a lexicographic conjecture $\beta_i = (b_i^1, \ldots, b_i^L) \in \big(\Delta(S_{-i})\big)^L$,

$$u_i^l(s_i, \beta_i) := \sum_{s_{-i} \in S_{-i}} b_i^l(s_{-i}) \cdot U_i(s_i, s_{-i})$$

is player i's *level-l expected utility* for all $l \in \{1, \ldots, L\}$. Equipped with a lexicographic conjecture $\beta_i \in \big(\Delta(S_{-i})\big)^L$, player i *strictly lex-prefers* a strategy $s_i \in S_i$ to some other strategy $s_i' \in S_i$, whenever there exists a lexicographic level $l^* \in \{1, \ldots, L\}$ such that

$$u_i^{l^*}(s_i, \beta_i) > u_i^{l^*}(s_i', \beta_i) \text{ and } u_i^l(s_i, \beta_i) = u_i^l(s_i', \beta_i)$$

for all $l < l^*$. A strategy $s_i^* \in S_i$ is called *lex-optimal* given β_i, if there exists no strategy $s_i \in S_i$ such that i strictly lex-prefers s_i to s_i^*. Similarly, player i is said to be *lex-indifferent* between s_i and s_i', whenever $u_i^l(s_i, \beta_i) = u_i^l(s_i', \beta_i)$ for all $l \in \{1, \ldots, L\}$. Player i *weakly lex-prefers* s_i to s_i', if he strictly lex-prefers the former to the latter or feels lex-indifferent. A lexicographic conjecture β_i is called *lexicographic product* *conjecture*, if $b_i^l = \bigotimes_{j \in I \setminus \{i\}} \text{marg}_{S_j} b_i^l$ holds for all $l \in \{1, \ldots, L\}$, and is formally written as

$$\beta_i := \bigotimes_{j \in I \setminus \{i\}} \text{marg}_{S_j} \beta_i := \big(\bigotimes_{j \in I \setminus \{i\}} \text{marg}_{S_j} b_i^1, \ldots, \bigotimes_{j \in I \setminus \{i\}} \text{marg}_{S_j} b_i^L \big).$$

Conceptually, a player with a lexicographic product conjecture treats his opponents' choices as uncorrelated.[16]

Selten's (1975) solution concept of perfect equilibrium can be expressed in terms of lexicographic conjectures.

Definition 7. *Let Γ be a finite game, $\sigma = (\sigma_i)_{i \in I} \in \times_{i \in I} \big(\Delta(S_i)\big)$ be a tuple of mixed strategies, and $L \in \mathbb{N}$. The tuple σ constitutes a* lexicographic perfect equilibrium *of Γ, if there exist a tuple $\beta = (\beta_i)_{i \in I} \in \Big(\big(\Delta(S_{-i})\big)^L\Big)_{i \in I}$ of lexicographic conjectures and a lexicographic product measure $\pi = (\pi^1, \ldots, \pi^L) \in \big(\Delta(\times_{i \in I} S_i)\big)^L$ such that for all $i \in I$, the following properties hold:*

(a) $\beta_i = (b_i^1, \ldots, b_i^L)$ *is cautious;*
(b) $\sigma_i = \text{marg}_{S_i} b_j^1$ *for all* $j \in I \setminus \{i\}$*;*
(c) if $s_i \in \text{supp}(\sigma_i)$*, then* s_i *is lex-optimal given* β_i*;*
(d) $\beta_i = \bigotimes_{j \in I \setminus \{i\}} \text{marg}_{S_j} \beta_i$*;*
(e) $\text{marg}_{S_{-i}} \pi = \beta_i$.

A lexicographic formulation of perfect equilibrium thus builds on an interpretation of mixed strategies as conjectures. In this regard, condition (b) blocks any doxastic ambiguity by requiring that for a given player all opponents share the same belief about his choice. The trembles of the classical definition are mimicked via condition (a) which requires the lexicographic conjectures to be cautious. The best response property of the perfect equilibrium tuple is ensured by condition (c) according to which only choices supported by the player's lexicographic conjecture receive positive probability. Epistemic independence is built in via condition (d) postulating that the players' lexicographic conjectures are the product of their marginals. Each of the lexicographic beliefs are required by condition (e) to stem from a joint source. In essence, lexicographic perfect equilibrium corresponds to Blume et al.'s (1991b) lexicographic Nash equilibrium plus full support at all lexicographic levels, a common prior, and some independence condition.

The classical and the lexicographic versions of perfect equilibrium are equivalent.

Lemma 1. *Let Γ be a finite game and $\sigma \in \times_{i \in I} \big(\Delta(S_i)\big)$ be a tuple of mixed strategies. The tuple σ constitutes a perfect equilibrium of Γ, if and only if, σ constitutes a lexicographic perfect equilibrium of Γ.*

Proof. See Appendix.

[16] While players by assumption do choose independently of course, it is well known that this does not preclude the possibility that beliefs about opponents' choices violate statistical independence. Essentially, the reason lies in two distinct forms of independence – causal and epistemic – which do not imply each other.

The classical formulation (Definition 6) and the lexicographic variant (Definition 7) of perfect equilibrium can thus be used interchangeably. Lemma 1 is by and large equivalent to Blume et al. (1991b, Proposition 7), where classical perfect equilibrium is characterized in terms of their notion of lexicographic Nash equilibrium plus some additional assumptions. For the sake of completeness and self-containedness we explicitly show the equivalence. However, since Lemma 1 lies outside the focus of this paper its proof is deferred to the Appendix.

A possibly meaningful weakening of lexicographic perfect equilibrium would obtain, if conditions (d) and (e) of Definition 7 were to be dropped.

Definition 8. *Let Γ be a finite game, $\sigma = (\sigma_i)_{i \in I} \in \bigtimes_{i \in I} \big(\Delta(S_i)\big)$ be a tuple of mixed strategies, and $L \in \mathbb{N}$. The tuple σ constitutes a* lexicographic semi-perfect equilibrium *of Γ, if there exists a tuple $\beta = (\beta_i)_{i \in I} \in \Big(\big(\Delta(S_{-i})\big)^L\Big)_{i \in I}$ of lexicographic conjectures such that for all $i \in I$, the following properties hold:*

(a) $\beta_i = (b_i^1, \ldots, b_i^L)$ is cautious;
(b) $\sigma_i = \mathrm{marg}_{S_i} b_j^1$ for all $j \in I \backslash \{i\}$;
(c) if $s_i \in \mathrm{supp}(\sigma_i)$, then s_i is lex-optimal given β_i.

A lexicographic semi-perfect equilibrium does admit a player's lexicographic conjecture about his opponents' choices to not be independent. Accordingly, he may deem it lexicographically possible for some opponents' choices to be correlated. Note that correlated beliefs at some level do not imply the belief that players do not choose independently from each other. Even though the actions of any two players in a static game are entirely autonomous, the reasoning leading to these actions might be related in a way that makes them correlated from the perspective of a third player. Also, in contrast to perfect equilibrium, more flexibility about the lexicographic conjectures is permitted by Definition 8, as they no longer need to be projections of a joint source. The solution concept of lexicographic semi-perfect equilibrium basically coincides with Blume et al.'s (1991b) notion of lexicographic Nash equilibrium plus some full support property.

It is clear that perfect equilibrium implies semi-perfect equilibrium, as the latter requires two properties less than the former. The following example shows that the converse does not hold though.

Example 2. Consider the three player game depicted in Figure 2 with players *Alice*, *Bob*, and *Claire*, where *Alice* chooses a "row" (a or b), *Bob* picks a "column" (y or z), and *Claire* selects a "matrix" (*left*, *middle*, or *right*).

	y	z
a	$1,1,2$	$0,0,0$
b	$0,0,0$	$1,1,0$

left

	y	z
a	$1,1,2$	$0,0,0$
b	$0,0,0$	$1,1,2$

middle

	y	z
a	$1,1,0$	$0,0,0$
b	$0,0,0$	$1,1,2$

right

Fig. 3.

It is first shown that the mixed strategy tuple $\sigma = (\sigma_{Alice}, \sigma_{Bob}, \sigma_{Claire})$, where $\sigma_{Alice}(a) = \sigma_{Alice}(b) = 0.5$, $\sigma_{Bob}(y) = \sigma_{Bob}(z) = 0.5$, and $\sigma_{Claire}(middle) = 1$ forms a lexicographic semi-perfect equilibrium. Define conjectures $\beta_{Alice} = (b^1_{Alice}, b^2_{Alice})$, $\beta_{Bob} = (b^1_{Bob}, b^2_{Bob})$, and $\beta_{Claire} = (b^1_{Claire}, b^2_{Claire})$ such that

$$\begin{aligned}
b^1_{Alice} &= 0.5 \cdot (y, middle) + 0.5 \cdot (z, middle),\\
b^2_{Alice} &= 0.5 \cdot (y, left) + 0.5 \cdot (z, right)\big),\\
b^1_{Bob} &= 0.5 \cdot (a, middle) + 0.5 \cdot (b, middle),\\
b^2_{Bob} &= 0.5 \cdot (a, left) + 0.5 \cdot (b, right)\big),\\
b^1_{Claire} &= 0.5 \cdot (a, y) + 0.5 \cdot (b, z),\\
b^2_{Claire} &= 1 \cdot (a, y).
\end{aligned}$$

Each of the three conjectures is cautious, as all choices of all respective opponents' receive positive probability at some lexicographic level. Moreover,

$$\begin{aligned}
\text{marg}_{S_{Alice}} b^1_{Bob} &= \text{marg}_{S_{Alice}} b^1_{Claire} = 0.5 \cdot a + 0.5 \cdot b = \sigma_{Alice},\\
\text{marg}_{S_{Bob}} b^1_{Alice} &= \text{marg}_{S_{Bob}} b^1_{Claire} = 0.5 \cdot y + 0.5 \cdot z = \sigma_{Bob},\\
\text{marg}_{S_{Claire}} b^1_{Alice} &= \text{marg}_{S_{Claire}} b^1_{Bob} = middle = \sigma_{Claire}.
\end{aligned}$$

Observe that a and b are lex-optimal given β_{Alice}, y and z are lex-optimal given β_{Bob}, as well as *middle* is lex-optimal given β_{Claire}. Consequently, σ constitutes a lexicographic semi-perfect equilibrium. However, σ is not lexicographic perfect, as b^1_{Claire}'s probability measure violates independence and property (d) of Definition 7 is thus not satisfied. ♣

It could be interesting to explore new solution concepts based on various weakenings of lexicographic perfect equilibrium such as lexicographic semi-perfect equilibrium. Another possibility would be to also admit conjectures that violate the projection property on the first lexicographic level. A corresponding perfect equilibrium variant could then be defined directly in terms of lexicographic conjectures and be required to satisfy the conditions (a) and (c) of Definition 7. We leave such thoughts for further research.

8 Epistemic Characterization

We now explore the interactive reasoning assumptions of perfect equilibrium and thereby extend the work of Blume et al. (1991b). While Blume et al. (1991b) characterize perfect equilibrium in terms of lexicographic conjectures, they do not perform any epistemic analysis involving higher-order beliefs to unveil the interactive thinking that drive players to choose in line with this solution concept. The latter is precisely the focus of this section. A key role will be played by our results on lexicographic agreeing to disagree. In particular, the weak agreement theorem (**WAT**) as well as the strong agreement theorem (**SAT**) turn into essential ingredients to establish an epistemic foundation for perfect equilibrium.

In game theory, reasoning is captured by means of epistemic structures that are added to the formal framework. Different patterns or assumptions about reasoning can then be expressed in the form of epistemic hypotheses. Classical solution concepts can be characterized in terms of reasoning by epistemic conditions. In this way, the interactive thinking a solution concept requires on behalf of the players so that they act in line with its prediction is made explicit.

Before we turn to reasoning foundations, some more formal structure and notions have to be fixed. First of all, the basic framework of games as embodied by Γ needs to be enlarged by an epistemic dimension. To this end we introduce the notion of a lexicographic Aumann model.

Definition 9. *Let Γ be a finite game. A* lexicographic Aumann model *of Γ is a tuple*

$$\mathcal{A}^{\Gamma}_{LCP} = \big(\Omega, \rho, I, (\mathcal{I}_i, \hat{s}_i)_{i \in I}\big)$$

where

- *Ω is a set of possible worlds,*
- *$\rho = (p^1, \dots, p^M)$ is a common prior,*
- *I is the set of players from Γ,*
- *$\mathcal{I}_i \subseteq 2^{\Omega}$ is player i's possibility partition of Ω for all $i \in I$,*
- *$\hat{s}_i : \Omega \to S_i$ is player i's choice function that is $\mathcal{I}_i$-measurable for all $i \in I$, i.e., $\hat{s}_i(w') = \hat{s}_i(w)$ for all $w, w' \in \Omega$ such that $w' \in \mathcal{I}_i(\omega)$,*
- *for every player $i \in I$ and for every world $\omega \in \Omega$, there exists a level $m \in \{1, \dots, M\}$ such that $p^m\big(\mathcal{I}_i(\omega)\big) > 0$.*

A lexicographic Aumann models thus corresponds to a lexicographic Aumann structure (Definition 2) supplemented by choice functions for every player that connect the interactive epistemology to games. It then becomes possible to express game-theoretic events and interactive beliefs as well as knowledge about these.

The event that player i chooses strategy $s_i \in S_i$ is formalized as

$$[s_i] := \{\omega \in \Omega : \hat{s}_i(\omega) = s_i\}$$

and the event that i's opponents choose $s_{-i} \in S_{-i}$ is given by

$$[s_{-i}] := \bigcap_{j \in I \setminus \{i\}} [s_j].$$

Note that the $\mathcal{I}_i$-measurability of $\hat{s}_i$ ensures that either $\mathcal{I}_i(\omega) \subseteq [s_i]$ or $\mathcal{I}_i(\omega) \subseteq [s_i]^{\complement}$. A lexicographic conjecture function can be defined as $\hat{\beta}_i : \Omega \to \big(\Delta(S_{-i})\big)^L$, where

$$\begin{aligned}\hat{\beta}_i(\omega)(s_{-i}) &= \big(\hat{b}^1_i(\omega)(s_{-i}), \dots, \hat{b}^L_i(\omega)(s_{-i})\big) \\ &:= \rho\big([s_{-i}] \mid \mathcal{I}_i(\omega)\big) = \Big(p^{m_1}\big([s_{-i}] \mid \mathcal{I}_i(\omega)\big), \dots, p^{m_L}\big([s_{-i}] \mid \mathcal{I}_i(\omega)\big)\Big)\end{aligned}$$

for all $\omega \in \Omega$ and for all $s_{-i} \in S_{-i}$. From the $\mathcal{I}_i$-measurability of the level posteriors it follows that $\hat{\beta}_i$ is $\mathcal{I}_i$-measurable too, i.e. $\hat{\beta}_i(\omega') = \hat{\beta}_i(\omega)$ for all $\omega' \in \mathcal{I}_i(\omega)$. Hence, for every lexicographic conjecture β_i of player i, the lexicographic conjecture function induces a coarsening of $\mathcal{I}_i$ of the form

$$[\beta_i] := \{\omega \in \Omega : \hat{\beta}_i(\omega) = \beta_i\}.$$

As $\hat{b}_i^l(\omega)(s_{-i}) = p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega)\big)$, it is the case that

$$\text{marg}_{S_j} \hat{b}_i^l(\omega)(s_j) = p^{m_l}\big([s_j] \mid \mathcal{I}_i(\omega)\big)$$

for all $\omega \in \Omega$, for all $l = 1, \ldots, L$, for all $s_j \in S_j$, and for all $j \in I \backslash \{i\}$.

Epistemic hypotheses can be formalized by means of events. Some assumptions that will be used for the purpose of describing the interactive thinking underlying perfect equilibrium are now spelled out. The set

$$T_i := \{\omega \in \Omega : \hat{\beta}_i(\omega) \text{ is cautious}\}$$

denotes the event that *player i is cautious* and the event that *all players are cautious* is given by

$$T := \bigcap_{i \in I} T_i.$$

The set

$$R_i := \{\omega \in \Omega : \hat{s}_i(\omega) \text{ is lex-optimal given } \hat{\beta}_i(\omega)\}$$

constitutes the event that *player i is rational* and the event that *all players are rational* is denoted by

$$R := \bigcap_{i \in I} R_i.$$

Given some event $E \subseteq \Omega$, the set

$$PB_i(E) := \{\omega \in \Omega : p^{m_1}\big(E \mid \mathcal{I}_i(\omega)\big) = 1\}$$

represents the event that *player i primarily believes in E* and the event that all players *primarily believe in E* is given by

$$PB := \bigcap_{i \in I} PB_i.$$

Note that primary belief concerns the first lexicographic *posterior* level $l = m_1$ which may differ from the first lexicographic *prior* level $m = 1$.

As a preliminary observation we provide an epistemic foundation for perfect equilibrium in the special case of two player games.

Proposition 2. *Let Γ be a finite game with two players i and j, $\mathcal{A}^{\Gamma}_{LCP}$ be some lexicographic Aumann model of Γ, and $\omega^* \in \Omega$ be some world. If $\omega^* \in PB(T) \cap PB(R) \cap K\big([\hat{\beta}_i(\omega^*)] \cap [\hat{\beta}_j(\omega^*)]\big)$, then there exists a pair of mixed strategies $(\sigma_i, \sigma_j) \in \Delta(S_i) \times \Delta(S_j)$ such that*

(i) $\sigma_i = \hat{b}_j^1(\omega^*)$ and $\sigma_j = \hat{b}_i^1(\omega^*)$;
(ii) the pair of mixed strategies (σ_i, σ_j) constitutes a perfect equilibrium of Γ.

Proof. (i) Define $\beta_i := \hat{\beta}_i(\omega^*)$ and $\beta_j := \hat{\beta}_j(\omega^*)$ as well as $\sigma_i := b^1_j$ and $\sigma_j := b^1_i$. Then, $\sigma_i = \hat{b}^1_j(\omega^*)$ and $\sigma_j = \hat{b}^1_i(\omega^*)$ directly obtains.

(ii) Let $k \in \{i, j\}$ be one of the two players and $-k$ be his opponent. As $\omega^* \in K\big([\hat{\beta}_i(\omega^*)] \cap [\hat{\beta}_j(\omega^*)]\big) \subseteq K_{-k}\big([\hat{\beta}_k](\omega^*)\big)$, it follows that $\mathcal{I}_{-k}(\omega^*) \subseteq [\hat{\beta}_k(\omega^*)]$ and consequently $\hat{\beta}_k(\omega) = \hat{\beta}_k(\omega^*)$ for all $\omega \in \mathcal{I}_{-k}(\omega^*)$. As $\omega^* \in PB(T) \subseteq PB_{-k}(T_k)$, it is the case that

$$p^{m_1}\big(T_k \mid \mathcal{I}_{-k}(\omega^*)\big) = \frac{p^{m_1}\big(T_k \cap \mathcal{I}_{-k}(\omega^*)\big)}{p^{m_1}\big(\mathcal{I}_{-k}(\omega^*)\big)} = 1$$

and thus there exists $\omega' \in T_k \cap \mathcal{I}_{-k}(\omega^*)$. Then, $\hat{\beta}_k(\omega')$ is cautious and $\hat{\beta}_k(\omega') = \hat{\beta}_k(\omega^*) = \beta_k$. It follows that β_k is cautious too. Since k has been chosen arbitrarily, property (a) of Definition 7 obtains. In addition, $\sigma_k = b^1_{-k} = \mathrm{marg}_{S_k} b^1_{-k}$ ensures that property (b) of Definition 7 is satisfied. Next consider some strategy $s_k \in \mathrm{supp}(\sigma_k) = \mathrm{supp}(\hat{b}^1_{-k}(\omega^*))$. Then, $\hat{b}^1_{-k}(\omega^*)(s_k) = p^{m_1}\big([s_k] \mid \mathcal{I}_{-k}(\omega^*)\big) > 0$, and thus there exists $\omega' \in [s_k] \cap \mathrm{supp}\big(p^{m_1}\big(\cdot \mid \mathcal{I}_{-k}(\omega^*)\big)\big) \subseteq [s_k] \cap \mathcal{I}_{-k}(\omega^*)$. Consequently, $\hat{s}_k(\omega') = s_k$ and $\hat{\beta}_k(\omega') = \hat{\beta}_k(\omega^*)$. Also, as $\omega^* \in PB(R) \subseteq PB_{-k}(R_k)$, it is the case that $p^{m_1}\big(R_k \mid \mathcal{I}_{-k}(\omega^*)\big) = 1$ and thus $\mathrm{supp}\big(p^{m_1}\big(\cdot \mid \mathcal{I}_{-k}(\omega^*)\big)\big) \subseteq R_k$. Hence, $\omega' \in R_k$, i.e. $\hat{s}_k(\omega') = s_k$ is lex-optimal given $\hat{\beta}_k(\omega') = \hat{\beta}_k(\omega^*) = \beta_k$. This establishes property (c) of Definition 7. Besides, note that $\beta_k = \mathrm{marg}_{S_{-k}} \beta_k = \bigotimes_{j \in I \setminus \{k\}} \mathrm{marg}_{S_j} \beta_k$ holds trivially as there is only one opponent for each player, which establishes property (d) of Definition 7. Finally, define $\pi := \beta_i \bigotimes \beta_j$. Then, $\mathrm{marg}_{S_{-k}} \pi = \beta_k$ directly follows, and property (e) of Definition 7 is satisfied. ∎

The reasoning assumptions underlying perfect equilibrium, if attention is restricted to two players thus consist of mutual primary belief in caution, mutual primary belief in rationality, and mutual knowledge of conjectures.

In order to tame the complications arising once more than two players are admitted, the epistemic conditions need to be tightened. The problem of projection can be tackled by strengthening mutual knowledge of conjectures to common knowledge. By the aid of **WAT**, an epistemic foundation then ensues for the notion of lexicographic semi-perfect equilibrium.

Lemma 2. *Let Γ be a finite game, $\mathcal{A}^{\Gamma}_{LCP}$ be some lexicographic Aumann model of Γ, and $\omega^* \in \Omega$ be some world. If $\omega^* \in PB(T) \cap PB(R) \cap CK\big(\bigcap_{i \in I} [\hat{\beta}_i(\omega^*)]\big)$, then there exists a tuple of mixed strategies $(\sigma^*_i)_{i \in I} \in \times_{i \in I} \big(\Delta(S_i)\big)$ such that*

(i) $\sigma^*_i = \mathrm{marg}_{S_i} \hat{b}^1_j(\omega^*)$ for all $i \in I$ and for all $j \in I \setminus \{i\}$;
(ii) the tuple of mixed strategies $(\sigma^*_i)_{i \in I}$ constitutes a lexicographic semi-perfect equilibrium of Γ.

Proof. (i) Consider the tuple of lexicographic conjectures $\big(\hat{\beta}_i(\omega^*)\big)_{i \in I}$ at world ω^*. Let $i \in I$ be some player. Observe that $[\hat{\beta}_j(\omega^*)] \subseteq [\hat{b}^l_j(\omega^*)] \subseteq [\mathrm{marg}_{S_i}\big(\hat{b}^l_j(\omega^*)\big)]$ for all $l = 1, \ldots, L$ and for all $j \in I \setminus \{i\}$. Then, by monotonicity of common knowledge,

$$CK\Big(\bigcap_{j \in I \setminus \{i\}} [\hat{\beta}_j(\omega^*)]\Big) \subseteq CK\Big(\bigcap_{j \in I \setminus \{i\}} \bigcap_{l \in \{1, \ldots, L\}} \big[\mathrm{marg}_{S_i} \hat{b}^l_j(\omega^*)\big]\Big) \neq \varnothing.$$

As $\operatorname{marg}_{S_i}\hat{b}^l_j(\omega)(s_i) = p^{m_l}\big([s_i] \mid \mathcal{I}_j(\omega)\big)$ for all $\omega \in \Omega$, for all $j \in I\backslash\{i\}$, for all $l \in \{1,\ldots,L\}$, and for all $s_i \in S_i$,

$$CK\Big(\bigcap_{j\in I\backslash\{i\}}\ \bigcap_{l\in\{1,\ldots,L\}}\big\{\omega\in\Omega : p^{m_l}\big([s_i]\mid\mathcal{I}_j(\omega)\big) = p^{m_l}\big([s_i]\mid\mathcal{I}_j(\omega^*)\big)\big\}\Big) \neq \varnothing$$

holds for all $s_i \in S_i$. By Theorem 1, it follows that

$$p^{m_1}\big([s_i]\mid\mathcal{I}_j(\omega^*)\big) = p^{m_1}\big([s_i]\mid\mathcal{I}_k(\omega^*)\big)$$

for all $s_i \in S_i$ as well as for all $j,k \in I\backslash\{i\}$, and thus

$$\operatorname{marg}_{S_i}\hat{b}^1_j(\omega^*) = \operatorname{marg}_{S_i}\hat{b}^1_k(\omega^*)$$

for all $j,k \in I\backslash\{i\}$. For every player $i \in I$, define $\sigma^*_i := \operatorname{marg}_{S_i}\hat{b}^1_{i'}(\omega^*)$ for some $i' \in I\backslash\{i\}$. Then, $\sigma^*_i = \operatorname{marg}_{S_i}\hat{b}^1_j(\omega^*)$ holds for all $i \in I$ and for all $j \in I\backslash\{i\}$.

(ii) Consider the tuple of lexicographic conjectures $\big(\hat{\beta}_i(\omega^*)\big)_{i\in I}$, where $\hat{\beta}_i(\omega^*) = \big(\hat{b}^1_i(\omega^*),\ldots,b^L_i(\omega^*)\big)$ for all $i \in I$. Let $i,j \in I$ be two players such that $i \neq j$. Since $\omega^* \in CK\big(\bigcap_{i\in I}[\hat{\beta}_i(\omega^*)]\big) \subseteq K_j\big([\hat{\beta}_i(\omega^*)]\big)$, it follows that $\mathcal{I}_j(\omega^*) \subseteq [\hat{\beta}_i(\omega^*)]$, and thus $\hat{\beta}_i(\omega) = \hat{\beta}_i(\omega^*)$ for all $\omega \in \mathcal{I}_j(\omega^*)$. Note that $\operatorname{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_j(\omega^*)\big)\Big) \subseteq \mathcal{I}_j(\omega^*)$. Moreover, as $\omega^* \in PB_j(T_i)$, the equation $p^{m_1}\big(T_i \mid \mathcal{I}_j(\omega^*)\big) = 1$ holds, thus $\operatorname{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_j(\omega^*)\big)\Big) \subseteq T_i$. Now, consider $\omega' \in \operatorname{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_j(\omega^*)\big)\Big)$. Then, $\omega' \in \mathcal{I}_j(\omega^*) \cap T_i$. Consequently, on the one hand $\hat{\beta}_i(\omega') = \hat{\beta}_i(\omega^*)$ and on the other hand $\hat{\beta}_i(\omega')$ is cautious. Therefore, $\hat{\beta}_i(\omega^*)$ is cautious, which establishes property (a) of Definition 8.

By part (i), the property $\sigma^*_i = \operatorname{marg}_{S_i}\hat{b}^1_j(\omega^*)$ holds for all $i \in I$ and for all $j \in I\backslash\{i\}$. Thus property (b) of Definition 8 obtains.

Let $i,j \in I$ such that $i \neq j$ and consider some $s_i \in \operatorname{supp}(\sigma^*_i) = \operatorname{supp}\big(\operatorname{marg}_{S_i}\hat{b}^1_j(\omega^*)\big)$. Thus, $\operatorname{marg}_{S_i}\hat{b}^1_j(\omega^*)(s_i) = p^{m_1}\big([s_i] \mid \mathcal{I}_j(\omega^*)\big) > 0$. Hence, there exists $\omega^\circ \in \operatorname{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_j(\omega^*)\big)\Big) \subseteq \mathcal{I}_j(\omega^*)$ such that $\hat{s}_i(\omega^\circ) = s_i$. As shown above, it is also the case that $\hat{\beta}_i(\omega) = \hat{\beta}_i(\omega^*)$ for all $\omega \in \mathcal{I}_j(\omega^*)$. Consequently, $\hat{\beta}_i(\omega^\circ) = \hat{\beta}_i(\omega^*)$. Since $\omega^* \in PB(R) \subseteq PB_j(R_i)$, it holds that $p^{m_1}(R_i \mid \mathcal{I}_j(\omega^*)) = 1$, i.e. $\omega' \in R_i$ for all $\omega' \in \operatorname{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_j(\omega^*)\big)\Big)$. Thus, $\omega^\circ \in R_i$, i.e. $\hat{s}_i(\omega^\circ)$ is lex-optimal given $\hat{\beta}_i(\omega^\circ)$. As $\hat{s}_i(\omega^\circ) = s_i$ and $\hat{\beta}_i(\omega^\circ) = \hat{\beta}_i(\omega^*)$, it follows that s_i is lex-optimal given $\hat{\beta}_i(\omega^*)$, which establishes property (c) of Definition 8.

Therefore, $(\sigma^*_i)_{i\in I}$ constitutes a lexicographic semi-perfect equilibrium of Γ. ■

The weak agreement theorem (**WAT**) plays a major role in the preceding result, as it ensures that players always agree on their marginal conjectures about any common opponent they face in the game. The possibility that any two players entertain distinct beliefs about a third player's choice is thereby blocked and the problem of projection solved. Formally, condition (i) of Theorem 2 and property (b) of Definition 8 are driven by **WAT**.

Yet additional armoury has to be invoked to establish an epistemic foundation for perfect equilibrium in the general set-up of many player games. Requiring the common prior to be mutually absolutely continuous enables the application of **SAT**, which can be used in turn to resolve the problem of independence.

Theorem 3. *Let Γ be a finite game, $\mathcal{A}^{\Gamma}_{LCP}$ be some lexicographic Aumann model of Γ such that the common prior ρ is mutually absolutely continuous, and $\omega^* \in \Omega$ be some world. If $\omega^* \in PB(T) \cap PB(R) \cap CK\big(\bigcap_{i \in I}[\hat{\beta}_i(\omega^*)]\big)$, then there exists a tuple of mixed strategies $(\sigma^*_i)_{i \in I} \in \times_{i \in I}\big(\Delta(S_i)\big)$ such that*

(i) $\sigma^*_i = \mathrm{marg}_{S_i} \hat{b}^1_j(\omega^*)$ for all $i \in I$ and for all $j \in I \backslash \{i\}$;
(ii) the tuple of mixed strategies $(\sigma^*_i)_{i \in I}$ constitutes a perfect equilibrium of Γ.

Proof. (i) Consider the tuple of lexicographic conjectures $\big(\hat{\beta}_i(\omega^*)\big)_{i \in I}$ at world ω^*. For every player $i \in I$, define $\sigma^*_i := \mathrm{marg}_{S_i} \hat{b}^1_{i'}(\omega^*)$ for some $i' \in I \backslash \{i\}$. Part (i) of Theorem 2 ensures that $\sigma^*_i = \mathrm{marg}_{S_i} \hat{b}^1_j(\omega^*)$ for all $i \in I$ and for all $j \in I \backslash \{i\}$.

(ii) By Lemma 2, properties (a), (b), and (c) of Definition 7 hold. Let $i \in I$ be some player and $l \in \{1, \ldots, L\}$ be some lexicographic level. Since $CK\Big(\bigcap_{j \in I}\big[\hat{\beta}_j(\omega^*)\big]\Big) \neq \emptyset$, it is the case that

$$CK\Big(\big[\mathrm{marg}_{S_{-i}} \hat{b}^l_i(\omega^*)\big]\Big) \neq \emptyset$$
$$CK\Big(\big[\mathrm{marg}_{S_{i+1}} \hat{b}^l_i(\omega^*)\big]\Big) \neq \emptyset$$
$$CK\Big(\bigcap_{j \in \{i,i+1\}} \big[\mathrm{marg}_{S_{-(i,i+1)}} \hat{b}^l_j(\omega^*)\big]\Big) \neq \emptyset.$$

Consider some opponents' strategy combination $s_{-i} \in S_{-i}$. As $\hat{b}^l_i(\omega^*)(\cdot) = p^{m_l}\big([\cdot] \mid \mathcal{I}_i(\omega^*)\big)$, it follows that

$$CK\Big(\big\{\omega \in \Omega : p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega)\big) = p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega^*)\big)\big\}\Big) \neq \emptyset$$
$$CK\Big(\big\{\omega \in \Omega : p^{m_l}\big([s_{i+1}] \mid \mathcal{I}_i(\omega)\big) = p^{m_l}\big([s_{i+1}] \mid \mathcal{I}_i(\omega^*)\big)\big\}\Big) \neq \emptyset$$
$$CK\Big(\bigcap_{j \in \{i,i+1\}} \big\{\omega \in \Omega : p^{m_l}\big([s_{-(i,i+1)}] \mid \mathcal{I}_j(\omega)\big) = p^{m_l}\big([s_{-(i,i+1)}] \mid \mathcal{I}_j(\omega^*)\big)\big\}\Big) \neq \emptyset$$

By the proof of Theorem 2, there exist some indices α_l, β_l and γ_l independent from i, $i+1$ and ω such that

$$p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega)\big) = p^{\alpha_l}\big([s_{-i}] \mid (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega^*)\big)$$
$$p^{m_l}\big([s_{i+1}] \mid \mathcal{I}_i(\omega)\big) = p^{\beta_l}\big([s_{i+1}] \mid (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega^*)\big)$$
$$p^{m_l}\big([s_{-(i,i+1)}] \mid \mathcal{I}_i(\omega)\big) = p^{m_l}\big([s_{-(i,i+1)}] \mid \mathcal{I}_{i+1}(\omega)\big) = p^{\gamma_l}\big([s_{-(i,i+1)}] \mid (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega^*)\big)$$

for all $\omega \in (\bigwedge_{i\in I} \mathcal{I}_i)(\omega^*)$. Since ρ is mutually absolutely continuous, the first part of the proof of Theorem 2 ensures that the lexicographic levels of $\rho(\cdot \mid \mathcal{I}_i(\omega))$ are the same for all $\omega \in (\bigwedge_{i\in I} \mathcal{I}_i)(\omega^*)$, and thus $\alpha_l = \beta_l = \gamma_l := \bar{m}_l$. Moreover, since either $\mathcal{I}_i(\omega) \subseteq [s_i]$ or $\mathcal{I}_i(\omega) \subseteq [s_i]^\complement$, the following property holds

$$p\big(E \cap [s_i] \mid \mathcal{I}_i(\omega)\big) = p\big(E \mid \mathcal{I}_i(\omega)\big) \cdot p\big([s_i] \mid \mathcal{I}_i(\omega)\big)$$

for all probability measures $p \in \Delta(\Omega)$, for all $E \subseteq \Omega$ and for all $i \in I$. Let $\mathcal{P} := \{P_{i+1} \in \mathcal{I}_{i+1} : P_{i+1} \subseteq (\bigwedge_{i\in I} \mathcal{I}_i)(\omega^*)\}$ be the possibility cells of player $i+1$ included in the meet cell of ω^*. By using the above properties together with the law of total probability, it follows that

$$\begin{aligned}
&p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega^*)\big)\\
&= p^{\bar{m}_l}\big([s_{-i}] \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big)\\
&= \sum_{P_{i+1}\in\mathcal{P}} p^{\bar{m}_l}\big([s_{-i}] \mid P_{i+1}\big) \cdot p^{\bar{m}_l}\big(P_{i+1} \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big)\\
&= \sum_{P_{i+1}\in\mathcal{P}} p^{\bar{m}_l}\big([s_{-(i,i+1)}] \mid P_{i+1}\big) \cdot p^{\bar{m}_l}\big([s_{i+1}] \mid P_{i+1}\big) \cdot p^{\bar{m}_l}\big(P_{i+1} \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big)\\
&= \sum_{P_{i+1}\in\mathcal{P}} p^{\bar{m}_l}\big([s_{-(i,i+1)}] \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big) \cdot p^{\bar{m}_l}\big([s_{i+1}] \mid P_{i+1}\big) \cdot p^{\bar{m}_l}\big(P_{i+1} \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big)\\
&= p^{\bar{m}_l}\big([s_{-(i,i+1)}] \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big) \cdot \sum_{P_{i+1}\in\mathcal{P}} p^{\bar{m}_l}\big([s_{i+1}] \mid P_{i+1}\big) \cdot p^{\bar{m}_l}\big(P_{i+1} \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big)\\
&= p^{\bar{m}_l}\big([s_{-(i,i+1)}] \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big) \cdot p^{\bar{m}_l}\big([s_{i+1}] \mid (\bigwedge_{j\in I} \mathcal{I}_j)(\omega^*)\big)\\
&= p^{\bar{m}_l}\big([s_{-(i,i+1)}] \mid \mathcal{I}_{i+1}(\omega^*)\big) \cdot p^{\bar{m}_l}\big([s_{i+1}] \mid \mathcal{I}_i(\omega^*)\big).
\end{aligned}$$

Analogously,

$$p^{\bar{m}_l}\big([s_{-(i,i+1)}] \mid \mathcal{I}_{i+1}(\omega^*)\big) = p^{\bar{m}_l}\big([s_{-(i,i+1,i+2)}] \mid \mathcal{I}_{i+2}(\omega^*)\big) \cdot p^{\bar{m}_l}\big([s_{i+2}] \mid \mathcal{I}_{i+1}(\omega^*)\big)$$

ensues, and thus

$$\begin{aligned}
p^{\bar{m}_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega^*)\big) = {}& p^{\bar{m}_l}\big([s_{-(i,i+1,i+2)}] \mid \mathcal{I}_{i+2}(\omega^*)\big) \cdot p^{\bar{m}_l}\big([s_{i+2}] \mid \mathcal{I}_{i+1}(\omega^*)\big)\\
&\cdot\ p^{\bar{m}_l}\big([s_{i+1}] \mid \mathcal{I}_i(\omega^*)\big).
\end{aligned}$$

By induction, it follows that

$$p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega^*)\big) = \prod_{j\in I\setminus\{i-1\}} p^{\bar{m}_l}\big([s_{j+1}] \mid \mathcal{I}_j(\omega^*)\big).$$

Consequently,

$$\begin{aligned}
\hat{b}_i^l(\omega^*)(s_{-i}) &= p^{m_l}\big([s_{-i}] \mid \mathcal{I}_i(\omega^*)\big)\\
&= \prod_{j\in I\setminus\{i-1\}} p^{\bar{m}_l}\big([s_{j+1}] \mid \mathcal{I}_j(\omega^*)\big) = \prod_{j\in I\setminus\{i-1\}} \mathrm{marg}_{S_{j+1}} \hat{b}_j^l(\omega^*)(s_{j+1})\\
&= \prod_{j\in I\setminus\{i-1\}} \mathrm{marg}_{S_{j+1}} \hat{b}_i^l(\omega^*)(s_{j+1}) = \prod_{j\in I\setminus\{i\}} \mathrm{marg}_{S_j} \hat{b}_i^l(\omega^*)(s_j).
\end{aligned}$$

Therefore, $\hat{\beta}_i(\omega^*) = \bigotimes_{j \in I \setminus \{i\}} \mathrm{marg}_{S_j} \hat{\beta}_i(\omega^*)$, which establishes property (d) of Definition 7.

Furthermore, let $i \in I$ and $j, j' \in I \setminus \{i\}$ be some players, $s_i \in S_i$ be some strategy for player i, and $l \in \{1, \ldots, L\}$ be some lexicographic level. Observe that

$$\begin{aligned} \mathrm{marg}_{S_i} \hat{b}_j^l(\omega^*)(s_i) = p^{m_l}\big([s_i] \mid \mathcal{I}_j(\omega^*)\big) &= p^{\bar{m}_l}\big([s_i] \mid (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega^*)\big) \\ &= p^{m_l}\big([s_i] \mid \mathcal{I}_{j'}(\omega^*)\big) = \mathrm{marg}_{S_i} \hat{b}_{j'}^l(\omega^*)(s_i) \end{aligned}$$

and therefore, $\mathrm{marg}_{S_i} \hat{\beta}_j(\omega^*) = \mathrm{marg}_{S_i} \hat{\beta}_{j'}(\omega^*)$ for all $i \in I$ and for all $j, j' \in I \setminus \{i\}$. Now, take $i, i' \in I$ such that $i \neq i'$ and define the lexicographic product measure

$$\pi := \hat{\beta}_i(\omega^*) \otimes \mathrm{marg}_{S_i} \hat{\beta}_{i'}(\omega^*) = \big(\bigotimes_{j \in I \setminus \{i\}} \mathrm{marg}_{S_j} \hat{\beta}_i(\omega^*)\big) \otimes \mathrm{marg}_{S_i} \hat{\beta}_{i'}(\omega^*).$$

We show that $\mathrm{marg}_{S_{-k}} \pi = \hat{\beta}_k(\omega^*)$ for all $k \in I$. First, the definition of π combined with property (d) of Definition 7 ensures that

$$\mathrm{marg}_{S_{-i}} \pi = \bigotimes_{j \in I \setminus \{i\}} \mathrm{marg}_{S_j} \hat{\beta}_i(\omega^*) = \hat{\beta}_i(\omega^*).$$

If $k \in I \setminus \{i\}$, then the equality of the marginal conjectures established above together with property (d) of Definition 7 implies that

$$\begin{aligned} \mathrm{marg}_{S_{-k}} \pi &= \big(\bigotimes_{j \in I \setminus \{i,k\}} \mathrm{marg}_{S_j} \hat{\beta}_i(\omega^*)\big) \otimes \mathrm{marg}_{S_i} \hat{\beta}_{i'}(\omega^*) \\ &= \big(\bigotimes_{j \in I \setminus \{i,k\}} \mathrm{marg}_{S_j} \hat{\beta}_k(\omega^*)\big) \otimes \mathrm{marg}_{S_i} \hat{\beta}_k(\omega^*) \\ &= \bigotimes_{j \in I \setminus \{k\}} \mathrm{marg}_{S_j} \hat{\beta}_k(\omega^*) = \hat{\beta}_k(\omega^*). \end{aligned}$$

Consequently, π and $\big(\hat{\beta}_i(\omega^*)\big)_{i \in I}$ satisfy property (e) of Definition 7.

Therefore, $(\sigma_i^*)_{i \in I}$ forms a lexicographic perfect equilibrium of Γ, and thus, by Lemma 1, a perfect equilibrium of Γ. ■

The property that a player's belief about his opponents' strategies is independent poses a rather intricate matter in the proof of Theorem 3 and its accomplishment is assisted by our strong agreement theorem (**SAT**). The effective application of the two lexicographic agreement theorems (**WAT** and **SAT**) in establishing epistemic conditions for perfect equilibrium once again underlines the power that Aumann's seminal impossibility result on agreeing to disagree is capable of unfolding.

The following result addresses the converse direction by ensuring that the epistemic conditions of Theorem 3 always exist and can be aligned with any perfect equilibrium.

Theorem 4. *Let Γ be a finite game and $\sigma = (\sigma_i)_{i \in I} \in \times_{i \in I} (\Delta(S_i))$ be a tuple of mixed strategies that constitutes a perfect equilibrium of Γ. Then, there exists a lexicographic Aumann model $\mathcal{A}_{LCP}^{\Gamma}$ of Γ with a world $\omega^* \in \Omega$ such that the common prior ρ is mutually absolutely continuous, $\omega^* \in PB(T) \cap PB(R) \cap CK\big(\bigcap_{i \in I} [\hat{\beta}_i(\omega^*)]\big)$, as well as $\sigma_i = \mathrm{marg}_{S_i} \hat{b}_j^1(\omega^*)$ for all $i \in I$ and for all $j \in I \setminus \{i\}$.*

Proof. By Lemma 1, σ forms a lexicographic perfect equilibrium and there exist a tuple $\beta = (\beta_i)_{i \in I} \in \Big(\big(\Delta(S_{-i}) \big)^L \Big)_{i \in I}$ of lexicographic conjectures and a lexicographic product measure $\pi = (\pi^1, \ldots, \pi^L) \in \big(\Delta(\times_{i \in I} S_i) \big)^L$ in line with the properties (a) to (e) of Definition 7. Construct the lexicographic Aumann model $\mathcal{A}^{\Gamma}_{LCP} = \big(\Omega, \rho, I, (\mathcal{I}_i, \hat{s}_i)_{i \in I} \big)$ of Γ, where

- $\Omega = \{ \omega^s : s = (s_i)_{i \in I} \in \times_{i \in I} S_i \}$,
- $p^m \in \Delta(\Omega)$ is defined by $p^m(\omega^s) = \pi^m(s)$ for all $\omega^s \in \Omega$ and for all $m \in \{1, \ldots, M\}$, with $M = L$,
- $\mathcal{I}_i(\omega^s) = \Omega$ for all $\omega^s \in \Omega$ and for all $i \in I$,
- $\hat{s}_i : \Omega \to \times_{i \in I} S_i$ is defined by $\hat{s}_i(\omega^s) = s_i$, for all $\omega^s \in \Omega$ and for all $i \in I$.

As $\mathcal{I}_i(\omega^s) = \Omega$ for all $\omega^s \in \Omega$ and for all $i \in I$, it directly follows that $p^m\big(\mathcal{I}_i(\omega^s)\big) = p^m\big(\mathcal{I}_j(\omega^s)\big) = 1$, and thus $p^l\big(\mathcal{I}_i(\omega^s)\big) = 0$ if and only if $p^l\big(\mathcal{I}_j(\omega^s)\big) = 0$, for all $\omega^s \in \Omega$, for all $m \in \{1, \ldots, M\}$, and for all $i, j \in I$. Therefore, ρ is mutually absolutely continuous.

Since $p^m\big(\mathcal{I}_i(\omega^s)\big) = 1$ for all $\omega^s \in \Omega$, for all $m \in \{1, \ldots, M\}$, and for all $i \in I$, Definition 3 ensures that $m_l = l$ for all $l \in \{1, \ldots, L\}$. Consider some player $i \in I$, some world $\omega^s \in \Omega$, and some lexicographic level $l \in \{1, \ldots, L\}$. It follows that

$$\begin{aligned} \hat{b}^l_i(\omega^s)(s'_{-i}) = p^{m_l}\big([s'_{-i}] \mid \mathcal{I}_i(\omega^s)\big) &= \frac{p^l\big(\mathcal{I}_i(\omega^s) \cap [s'_{-i}]\big)}{p^l\big(\mathcal{I}_i(\omega^s)\big)} = \frac{p^l(\Omega \cap [s'_{-i}])}{p^l(\Omega)} = \frac{p^l([s'_{-i}])}{1} \\ &= \pi^l(\{s \in \bigtimes_{i \in I} S_i : s_{-i} = s'_{-i}\}) = \sum_{s_i \in S_i} \pi^l(s_i, s'_{-i}) = \mathrm{marg}_{S_{-i}} \pi^l(s'_{-i}) = b^l_i(s'_{-i}) \end{aligned}$$

for all $s'_{-i} \in S_{-i}$, where the last equality is due to property (e) of Definition 7. Consequently, $\hat{\beta}_i(\omega^s) = \beta_i$ for all $\omega^s \in \Omega$ and for all $i \in I$. Hence, $[\hat{\beta}_i(\omega^s)] = \{\omega^{s'} \in \Omega : \hat{\beta}_i(\omega^{s'}) = \hat{\beta}_i(\omega^s)\} = \{\omega^{s'} \in \Omega : \hat{\beta}_i(\omega^{s'}) = \beta_i\} = \Omega$ for all $\omega^s \in \Omega$ as well as for all $i \in I$, and thus $CK\big(\bigcap_{i \in I} [\hat{\beta}_i(\omega^s)]\big) = CK(\Omega) = \Omega$.

Next consider some world $\omega^s \in \Omega$ and some player $i \in I$. Since $\hat{\beta}_i(\omega^s) = \beta_i$, property (a) of lexicographic perfect equilibrium ensures that $\hat{\beta}_i(\omega^s)$ is cautious, i.e. $\omega^s \in T_i$. It follows that $T_i = \Omega$, and thus $T = \bigcap_{j \in I} T_j = \Omega$. Consequently, $\mathrm{supp}\big(p^{m_1}(\cdot \mid \mathcal{I}_i(\omega^s))\big) \subseteq T$ and hence $p^{m_1}(T \mid \mathcal{I}_i(\omega^s)) = 1$, i.e. $\omega^s \in PB_i(T)$. Also, by properties (b) and (e) of Definition 7, it follows that

$$\begin{aligned} p^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega^s)\big) &= p^1(\cdot \mid \Omega) = p^1 = \pi^1 = \bigotimes_{j \in I} \mathrm{marg}_{S_j} \pi^1 \\ &= \bigotimes_{j \in I} \mathrm{marg}_{S_j} \mathrm{marg}_{S_{-(j+1)}} \pi^1 = \bigotimes_{j \in I} \mathrm{marg}_{S_j} b^1_{j+1} = \bigotimes_{j \in I} \sigma_j \end{aligned}$$

Let $\omega^{s'} \in \mathrm{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega^s)\big)\Big)$. Then, $s' \in \mathrm{supp}(\bigotimes_{j \in I} \sigma_j)$, i.e. $s'_j \in \mathrm{supp}(\sigma_j)$ for all $j \in I$. By property (c) of Definition 7, s'_j is lex-optimal given β_j, and hence $\hat{s}_j(\omega^{s'})$ is lex-optimal given $\hat{\beta}_j(\omega^{s'})$, i.e. $\omega^{s'} \in R_j$ for all $j \in I$. Thus $\omega^{s'} \in \bigcap_{j \in I} R_j = R$. Hence,

$\text{supp}\Big(p^{m_1}\big(\cdot \mid \mathcal{I}_i(\omega^s)\big)\Big) \subseteq R$. Thus, $p^{m_1}\big(R \mid \mathcal{I}_i(\omega^s)\big) = 1$, i.e. $\omega^s \in PB_i(R)$. Since i has been chosen arbitrarily, $\omega^s \in \bigcap_{i\in I} PB_i(T) \cap \bigcap_{i\in I} PB_i(R) = PB(T) \cap PB(R)$. As ω^s has been picked arbitrarily too, $PB(T) \cap PB(R) = \Omega$ obtains.

Finally, let $\omega^* \in \Omega$ be some world and $i \in I$ be some player. Then, $\omega^* \in PB(T) \cap PB(R) \cap CK\big(\bigcap_{i\in I}[\hat{\beta}_i(\omega^*)]\big)$. Furthermore, property (b) of Definition 7 guarantees that $\sigma_i = \text{marg}_{S_i} b_j^1 = \text{marg}_{S_i} \hat{b}_j^1(\omega^*)$ for all $j \in I\backslash\{i\}$. Since i has been chosen arbitrarily, $\sigma_i = \text{marg}_{S_i} \hat{b}_j^1(\omega^*)$ for all $i \in I$ and for all $j \in I\backslash\{i\}$. ∎

Accordingly, the sufficient conditions for perfect equilibrium put forth by Theorem 3 are not too strong in the sense that every perfect equilibrium is attainable with them. The conjunction of Theorems 3 and 4 constitutes an epistemic characterization of perfect equilibrium in terms of mutual primary belief in caution, mutual primary belief in rationality, and common knowledge of conjectures.

The epistemic program in game theory has shed light on the reasoning assumptions underlying Nash equilibrium.[17] The decisive – yet conceptually not unproblematic – implicit property of Nash equilibrium lies in some correct beliefs assumption. By requiring common knowledge of conjectures, Theorems 3 and 4 show that a significant dose of doxastic inerrancy also underlies the more general solution concept of perfect equilibrium. In contrast, common knowledge of rationality is not required in terms of reasoning: it is not even needed at the first lexicographic level. A central conceptual insight due to Aumann and Brandenburger (1995) for Nash equilibrium – interactive beliefs in rationality do not enter the picture but only an interactive correct beliefs condition does – is thus fortified by Theorems 3 and 4 in the more general context of perfect equilibrium.[18] Both Nash equilibrium and perfect equilibrium hence only require iterated – and thus truly interactive – beliefs about conjectures and not about rationality or anything else. Consequently, some correct beliefs property constitutes the essence of these solution concepts. Nonetheless, the reasoning foundation for perfect equilibrium stretches beyond the one for Nash equilibrium. Indeed, some notion of caution is needed in order to reflect the inherent trembles property of perfect equilibrium, which is absent from Nash equilibrium though.

9 Conclusion

When interactive epistemology is enriched by lexicographic probability systems, three results on agreeing to disagree obtain. If the agents' posteriors are common knowledge, the weak agreement theorem ensures the first lexicographic level posteriors to coincide. Somewhat unexpectedly, however, disagreement cannot be excluded without further

[17] For instance, Brandenburger, 1992b; Aumann and Brandenburger, 1995; Perea, 2007; Barelli, 2009; Bach and Tsakas, 2014; Bonanno, 2018; Bach and Perea, 2019

[18] As Aumann and Brandenburger (1995) as well as Brandenburger (1992b) highlight, *common knowledge enters the picture in an unexpected way* for Nash equilibrium to ensue: *what is needed is common knowledge of the players' conjectures* but *not of the players' rationality* (cf. Aumann and Brandenburger, 1995, p. 1163), and *then only in games with more than two players* (cf. Brandenburger, 1992b, p. 96).

assumptions on the deeper lexicographic levels. In line with our disagreement result, agreement can already fail on the second lexicographic level. Imposing mutual absolute continuity on top of common knowledge of posteriors, the strong agreement rules out posterior disagreement at any lexicographic level.

The impossibility of lexicographic agreeing to disagree becomes an essential tool to shed light on interactive reasoning in games. Epistemic conditions are provided for the classical solution concept of perfect equilibrium. In particular, the weak agreement theorem and the strong agreement theorem fundamentally assist in overcoming the challenges that arise with more than two players. The reasoning assumptions underlying perfect equilibrium are identified in our lexicographic framework by mutual primary belief in caution, mutual primary belief in rationality, and common knowledge of conjectures. The solution concept's key epistemic ingredient thus lies in an interactive correct beliefs assumption, while caution as well as rationality only appear in a non-iterated doxastic way on the first lexicographic level.

From a conceptual perspective, our results on the (im)possibility of lexicographic agreeing to disagree are relevant for situations when reasoning about ordered layers of contingencies is considered. Notably the original conclusion of Aumann's agreement theorem breaks down. Agreeing to disagree becomes conceivable once hypothetical contingencies enter the picture. This could have intriguing consequences for economic applications such as the possibility of trade. We leave such considerations for future research.

Appendix

The proof of Lemma 1 requires some additional results that are laid out first.

Given a game Γ, some player $i \in I$, some strategy $s_i \in S_i$ of player i, and some general – not necessarily product – probability measure $q \in \Delta(S_{-i})$, player i's expected utility of strategy s_i is defined as

$$u_i(s_i, q) := \sum_{s_{-i} \in S_{-i}} q(s_{-i}) \cdot U_i(s_i, s_{-i}).$$

Let X be a finite space, let $L > 0$ be an integer, let $\alpha = (a^1, \dots, a^L) \in \big(\Delta(X)\big)^L$ be a tuple of probability measures and let $r = (r^1, \dots, r^{L-1}) \in (0,1)^{L-1}$ be a tuple of real numbers. Let $r \square \alpha$ be defined by

$$r \square \alpha := \begin{cases} a^1 & \text{if } L = 1 \\ (1-r^1)\cdot a^1 + r^1 \cdot (1-r^2) \cdot a^2 + r^1 \cdot r^2 \cdot (1-r^3) \cdot a^3 + \dots + & \\ r^1 \cdot r^2 \cdot \ldots \cdot r^{L-2} \cdot (1-r^{L-1}) \cdot a^{L-1} + r^1 \cdot r^2 \cdot \ldots \cdot r^{L-1} \cdot a^L & \text{if } L > 1 \end{cases}$$

Observe that $r \square \alpha \in \Delta(X)$, since

$$\sum_{x \in X} (r \square \alpha)(x) = (1-r^1) + r^1 \cdot (1-r^2) + r^1 \cdot r^2 \cdot (1-r^3) + \dots + r^1 \cdot r^2 \cdot \ldots \cdot r^{L-2} \cdot (1-r^{L-1}) + r^1 \cdot r^2 \cdot \ldots \cdot r^{L-1} = 1.$$

Lemma A.1. *Let Γ be a game, $i \in I$ be a player, $s_i', s_i'' \in S_i$ be two strategies of player i, $\beta_i = (b_i^1, \ldots, b_i^L) \in \big(\Delta(S_{-i})\big)^L$ be a lexicographic conjecture of player i, and $(r_n)_{n\in\mathbb{N}} = \big((r_n^1, \ldots, r_n^{L-1})\big)_{n\in\mathbb{N}} \in \big[(0,1)^{L-1}\big]^{\mathbb{N}}$ be a sequence such that $\lim_{n\to\infty} r_n = \mathbf{0} \in \mathbb{R}^{L-1}$. Then, the following properties hold:*

(i) If $u_i(s_i', r_n \square \beta_i) > u_i(s_i'', r_n \square \beta_i)$ for all $n \in \mathbb{N}$, then i strictly lex-prefers s_i' to s_i''.
(ii) If $u_i(s_i', r_n \square \beta_i) \geqslant u_i(s_i, r_n \square \beta_i)$ for all $n \in \mathbb{N}$ and for all $s_i \in S_i$, then s_i' is lex-optimal given β_i.
(iii) If s_i' is lex-optimal given β_i, then there exist a subsequence $(r_{n_k})_{k\in\mathbb{N}}$ of $(r_n)_{n\in\mathbb{N}}$ and an index $K \in \mathbb{N}$ such that $u_i(s_i', r_{n_k} \square \beta_i) \geqslant u_i(s_i, r_{n_k} \square \beta_i)$ for all $k \geqslant K$ and for all $s_i \in S_i$.

Proof. (i) Observe that $\lim_{n\to\infty} r_n = \mathbf{0}$ implies

$$\lim_{n\to\infty} r_n \square \beta_i = \lim_{n\to\infty} \big[(1 - r_n^1) \cdot b_i^1 + r_n^1 \cdot (1 - r_n^2) \cdot b_i^2 + \ldots + r_n^1 \cdot r_n^2 \cdot \ldots \cdot r_n^{L-1} \cdot b_i^L\big] = b_i^1.$$

In addition, for each $l \in \{1, \ldots, L\}$, define

$$\Delta^l := u_i^l(s_i', \beta_i) - u_i^l(s_i'', \beta_i) = \sum_{s_{-i}\in S_{-i}} b_i^l(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big].$$

Suppose that $u_i(s_i', r_n \square \beta_i) > u_i(s_i'', r_n \square \beta_i)$ for all $n \in \mathbb{N}$. It follows that

$$\begin{aligned} &\sum_{s_{-i}\in S_{-i}} (r_n \square \beta_i)(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= (1 - r_n^1) \cdot \Delta^1 + r_n^1 \cdot (1 - r_n^2) \cdot \Delta^2 + \cdot \ldots \cdot + r_n^1 \cdot r_n^2 \cdot \ldots \cdot r_n^{L-1} \cdot \Delta^L > 0 \end{aligned} \tag{2}$$

for all $n \in \mathbb{N}$. Consequently,

$$\begin{aligned} 0 &\leqslant \lim_{n\to\infty} \sum_{s_{-i}\in S_{-i}} (r_n \square \beta_i)(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= \sum_{s_{-i}\in S_{-i}} \lim_{n\to\infty} (r_n \square \beta_i)(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= \sum_{s_{-i}\in S_{-i}} b_i^1(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] = \Delta^1. \end{aligned}$$

If $\Delta^1 > 0$, then $u_i^1(s_i', \beta_i) > u_i^1(s_i'', \beta_i)$ and thus i strictly lex-prefers s_i' to s_i''. If $\Delta^1 = 0$, then define the truncated tuples $\beta_i^{(2)} := (b_i^2, \ldots, b_i^L)$ and $(r_n^{(2)})_{n\in\mathbb{N}} := \big((r_n^2, \ldots, r_n^{L-1})\big)_{n\in\mathbb{N}}$. Property (2) together with the fact that $\Delta^1 = 0$ ensures that

$$\begin{aligned} 0 &< \sum_{s_{-i}\in S_{-i}} (r_n \square \beta_i)(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= (1 - r_n^1) \cdot \Delta^1 + r_n^1 \cdot \sum_{s_{-i}\in S_{-i}} (r_n^{(2)} \square \beta_i^{(2)})(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= r_n^1 \cdot \sum_{s_{-i}\in S_{-i}} (r_n^{(2)} \square \beta_i^{(2)})(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \end{aligned}$$

for all $n \in \mathbb{N}$, and thus

$$\sum_{s_{-i} \in S_{-i}} (r_n^{(2)} \square \beta_i^{(2)})(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] > 0$$

for all $n \in \mathbb{N}$. Consequently,

$$\begin{aligned} 0 &\leqslant \lim_{n\to\infty} \sum_{s_{-i} \in S_{-i}} (r_n^{(2)} \square \beta_i^{(2)})(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= \sum_{s_{-i} \in S_{-i}} \lim_{n\to\infty} (r_n^{(2)} \square \beta_i^{(2)})(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] \\ &= \sum_{s_{-i} \in S_{-i}} b_i^2(s_{-i}) \cdot \big[U_i(s_i', s_{-i}) - U_i(s_i'', s_{-i})\big] = \Delta^2. \end{aligned}$$

If $\Delta^2 > 0$, then $u_i^2(s_i', \beta_i) > u_i^2(s_i'', \beta_i)$ and $u_i^1(s_i', \beta_i) = u_i^1(s_i'', \beta_i)$, and thus i strictly lex-prefers s_i' to s_i''. If $\Delta^2 = 0$, then by continuing in this fashion for $l \geqslant 3$, property (2) ensures that eventually there exists $l^* \in \{1, \ldots, L\}$ such that $\Delta^{l^*} > 0$ and $\Delta^l = 0$ for all $0 < l < l^*$. Equivalently, $u_i^{l^*}(s_i', \beta_i) > u_i^{l^*}(s_i'', \beta_i)$ and $u_i^l(s_i', \beta_i) = u_i^l(s_i'', \beta_i)$ for all $0 < l < l^*$. Therefore, i strictly lex-prefers s_i' to s_i''.

(ii) Let $s_i \in S_i$. Suppose that $u_i(s_i', r_n \square \beta_i) \geqslant u_i(s_i, r_n \square \beta_i)$ for all $n \in \mathbb{N}$. If $u_i(s_i', r_n \square \beta_i) = u_i(s_i, r_n \square \beta_i)$ for all $n \in \mathbb{N}$, then by similar arguments as in the proof of Lemma A.1 (i), it follows that $\Delta^l = 0$ for all $l \in \{1, \ldots, L\}$. Consequently, i weakly lex-prefers s_i' to s_i. If $u_i(s_i', r_n \square \beta_i) > u_i(s_i, r_n \square \beta_i)$ for some $n^* \in \mathbb{N}$, then again by similar arguments as in the proof of Lemma A.1 (i), there exists $l^* \in \{1, \ldots, L\}$ such that $\Delta^{l^*} > 0$ and $\Delta^l = 0$ for all $0 < l < l^*$. Hence, i weakly lex-prefers s_i' to s_i and, as s_i has been chosen arbitrarily, s_i' is thus lex-optimal given β_i.

(iii) Consider a subsequence $(r_{n_k})_{k\in\mathbb{N}}$ of $(r_n)_{n\in\mathbb{N}}$ that satisfies the following property: for every $s_i \in S_i$, if $u_i(s_i, r_{n_k} \square \beta_i) > u_i(s_i', r_{n_k} \square \beta_i)$ for infinitely many indices $k \in \mathbb{N}$, then $u_i(s_i, r_{n_k} \square \beta_i) > u_i(s_i', r_{n_k} \square \beta_i)$ all $k \in \mathbb{N}$. Since $\lim_{n\to\infty} r_n = \mathbf{0}$, it is the case that $\lim_{k\to\infty} r_{n_k} = \mathbf{0}$. Suppose that s_i' is lex-optimal given β_i. By the contraposition of Lemma A.1 (i), for all $s_i \in S_i$, it is not the case that $u_i(s_i, r_{n_k} \square \beta_i) > u_i(s_i', r_{n_k} \square \beta_i)$ for all $k \in \mathbb{N}$. The contraposition of the property of the sequence $(r_{n_k})_{k\in\mathbb{N}}$ then ensures that, for all $s_i \in S_i$, it is not the case that $u_i(s_i, r_{n_k} \square \beta_i) > u_i(s_i', r_k \square \beta_i)$ for infinitely many indices $k \in \mathbb{N}$. Equivalently, for all $s_i \in S_i$, there exists $K(s_i) \in \mathbb{N}$ such that $u_i(s_i', r_{n_k} \square \beta_i) \geqslant u_i(s_i, r_{n_k} \square \beta_i)$ for all $k \geqslant K(s_i)$. Consequently, $u_i(s_i', r_{n_k} \square \beta_i) \geqslant u_i(s_i, r_{n_k} \square \beta_i)$ for all $k \geqslant \max\{K(s_i) : s_i \in S_i\}$ and for all $s_i \in S_i$. ■

Lemma A.2. *Let Γ be a game and $\psi : \Delta(\times_{i\in I} S_i) \times \Delta(\times_{i\in I} S_i) \to \mathbb{R}$ be the function defined by*

$$\psi(\sigma, \tilde{\sigma}) := \sup\big\{r \in \mathbb{R} : \sigma(s) - r \cdot \tilde{\sigma}(s) \geqslant 0, \textit{ for all } s \in \bigtimes_{i\in I} S_i\big\}.$$

Then, ψ satisfies the following properties:

(1) $\psi(\sigma, \tilde{\sigma}) = 1$, if and only if, $\sigma = \tilde{\sigma}$.
(2) If $\operatorname{supp}(\tilde{\sigma}) \subseteq \operatorname{supp}(\sigma)$, then $\sigma(s) - \psi(\sigma, \tilde{\sigma}) \cdot \tilde{\sigma}(s) = 0$ for some $s \in \operatorname{supp}(\sigma)$.

(3) The function $\psi(\cdot,\tilde{\sigma}) : \Delta(\times_{i\in I} S_i) \to \mathbb{R}$ *is continuous, for all* $\tilde{\sigma} \in \Delta(\times_{i\in I} S_i)$.

Proof. (1) Suppose that $\psi(\sigma,\tilde{\sigma}) = 1$. Then $\sigma - 1\cdot\tilde{\sigma} \geqslant 0$ and thus $\sigma \geqslant \tilde{\sigma}$. If $\sigma(s') > \tilde{\sigma}(s')$ for some $s' \in \times_{i\in I} S_i$, then $1 = \sum_{s\in\times_{i\in I} S_i} \sigma(s) > \sum_{s\in\times_{i\in I} S_i} \tilde{\sigma}(s) = 1$, which is a contradiction. Therefore $\sigma = \tilde{\sigma}$. Conversely, suppose that $\sigma = \tilde{\sigma}$. Define $\Psi_r := \{r \in \mathbb{R} : \sigma(s) - r\cdot\tilde{\sigma}(s) \geqslant 0, \text{ for all } s \in \times_{i\in I} S_i\}$. Since $\sigma - 1\cdot\tilde{\sigma} = 0$, then $1 \in \Psi_r$. Let $\epsilon > 0$ and let $s \in \operatorname{supp}(\sigma) = \operatorname{supp}(\tilde{\sigma})$. Then $\sigma(s) - (1+\epsilon)\cdot\tilde{\sigma}(s) = -\epsilon\cdot\tilde{\sigma}(s) < 0$. Hence, $(1+\epsilon) \notin \Psi_r$ for all $\epsilon > 0$. Therefore, $\psi(\sigma,\tilde{\sigma}) = \sup_{r\in\mathbb{R}} \Psi_r = 1$.

(2) Towards a contradiction, suppose that $\operatorname{supp}(\tilde{\sigma}) \subseteq \operatorname{supp}(\sigma)$ and $\sigma(s) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) > 0$ for all $s \in \operatorname{supp}(\sigma)$. Let $\bar{s} \in \arg\min\{\sigma(s) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) : s \in \operatorname{supp}(\tilde{\sigma})\}$ and define $r := \frac{(\sigma(\bar{s}) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(\bar{s}))}{\tilde{\sigma}(\bar{s})}$ and $\psi'(\sigma,\tilde{\sigma}) := \psi(\sigma,\tilde{\sigma}) + r$. Since $\operatorname{supp}(\tilde{\sigma})$ is finite, $\bar{s}$ is well defined. Moreover, as $\bar{s} \in \operatorname{supp}(\tilde{\sigma}) \subseteq \operatorname{supp}(\sigma)$, then $\sigma(\bar{s}) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(\bar{s}) > 0$, hence $r > 0$, and thus $\psi'(\sigma,\tilde{\sigma}) > \psi(\sigma,\tilde{\sigma})$. Let $s \in \times_{i\in I} S_i$. If $s \in (\times_{i\in I} S_i)\backslash\operatorname{supp}(\sigma)$, then $\sigma(s) = \tilde{\sigma}(s) = 0$, and thus $\sigma(s) - \psi'(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) = 0$. If $s \in \operatorname{supp}(\sigma)\backslash\operatorname{supp}(\tilde{\sigma})$, then $\sigma(s) > 0$ and $\tilde{\sigma}(s) = 0$, and thus $\sigma(s) - \psi'(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) > 0$. If $s \in \operatorname{supp}(\tilde{\sigma})$, then $\sigma(s) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) \geqslant \sigma(\bar{s}) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(\bar{s}) > \sigma(\bar{s}) - \psi'(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(\bar{s}) = \sigma(\bar{s}) - [\psi(\sigma,\tilde{\sigma}) + r]\cdot\tilde{\sigma}(\bar{s}) = \sigma(\bar{s}) - \psi(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) - r\cdot\tilde{\sigma}(\bar{s}) = 0$. Consequently, $\sigma(s) - \psi'(\sigma,\tilde{\sigma})\cdot\tilde{\sigma}(s) \geqslant 0$ for all $s \in \times_{i\in I} S_i$ and $\psi'(\sigma,\tilde{\sigma}) > \psi(\sigma,\tilde{\sigma})$, which contradicts the supremacy of $\psi(\sigma,\tilde{\sigma})$.

(3) Let $\tilde{\sigma} \in \Delta(\times_{i\in I} S_i)$ and let $(\sigma^k)_{k\in\mathbb{N}}$ be a sequence such that $\lim_{k\to\infty} \sigma^k = \sigma$. Then, $\lim_{k\to\infty} \psi(\sigma^k,\tilde{\sigma}) = \psi(\lim_{k\to\infty}\sigma^k,\tilde{\sigma}) = \psi(\sigma,\tilde{\sigma})$, and thus $\psi(\cdot,\tilde{\sigma})$ is continuous. ■

Lemma A.3. *Let* $(\sigma^k)_{k\in\mathbb{N}} \in \big(\Delta(\times_{i\in I} S_i)\big)^{\mathbb{N}}$ *be a sequence of mixed strategy profiles. Then, there exist a lexicographic probability measure* $\pi = (\pi^1,\ldots,\pi^L) \in \big(\Delta(\times_{i\in I} S_i)\big)^L$ *and a sequence* $(r_n)_{n\in\mathbb{N}} = \big((r_n^1,\ldots,r_n^{L-1})\big)_{n\in\mathbb{N}} \in \big[(0,1)^{L-1}\big]^{\mathbb{N}}$ *with* $\lim_{n\to\infty} r_n = \mathbf{0}$ *such that a subsequence* $(\sigma^{k_n})_{n\in\mathbb{N}}$ *of* $(\sigma^k)_{k\in\mathbb{N}}$ *satisfies* $\sigma^{k_n} = r_n \,\square\, \pi$ *for all* $n \in \mathbb{N}$.

Proof. Consider a subsequence $(\sigma^{k_n})_{n\in\mathbb{N}}$ of $(\sigma^k)_{k\in\mathbb{N}}$ that satisfies the following property: for every $s \in \times_{i\in I} S_i$, if $\sigma^{k_n}(s) = 0$ for infinitely many indices $n \in \mathbb{N}$, then $\sigma^{k_n}(s) = 0$ for all $n \in \mathbb{N}$. Then, there exists some index $N \in \mathbb{N}$ such that the subsequence $(\sigma^{k_n})_{n\geqslant N}$ of $(\sigma^{k_n})_{n\in\mathbb{N}}$ satisfies the following property: for every $s \in \times_{i\in I} S_i$, if $\sigma^{k_N}(s) = 0$, then $\sigma^{k_n}(s) = 0$ for all $n \geqslant N$. By the Bolzano–Weierstrass Theorem, there exists some convergent subsequence of $(\sigma^{k_n})_{n\geqslant N}$, denoted by $(\sigma^k)_{k\in\mathbb{N}}$ for the sake of simplicity, with limit $\pi^1 := \lim_{k\to\infty}\sigma^k$.

Either $\sigma^k = \pi^1$ infinitely often or $\sigma^k = \pi^1$ finitely often. Suppose that $\sigma^k = \pi^1$ infinitely often. Let $(\sigma^{k_n})_{n\in\mathbb{N}}$ be a subsequence of $(\sigma^k)_{k\in\mathbb{N}}$ such that $\sigma^{k_n} = \pi^1$ for all $n \in \mathbb{N}$, let $(r_n)_{n\in\mathbb{N}}$ be the empty sequence, and let $\pi = (\pi^1)$. It follows that $\sigma^{k_n} = \pi^1 = r_n \,\square\, \pi$ for all $n \in \mathbb{N}$, which completes the proof in this case.

Otherwise, suppose that $\sigma^k = \pi^1$ finitely often. Then, there exists $N \in \mathbb{N}$ such that $\sigma^k \neq \pi^1$ for all $k \geqslant N$. Let $(\sigma^{k_n})_{n\in\mathbb{N}}$ be a subsequence of $(\sigma^k)_{k\in\mathbb{N}}$ such that $\sigma^{k_n} \neq \pi^1$ for all $n \in \mathbb{N}$. This subsequence is denoted by $(\sigma^k)_{k\in\mathbb{N}}$ for the sake of simplicity. By Lemma 9 (1), $\psi(\sigma^k,\pi^1) \neq 1$ for all $k \in \mathbb{N}$. Consider the then well-defined sequence $(\pi_k^2)_{k\in\mathbb{N}}$ given by

$$\pi_k^2 := \frac{\sigma^k - \psi(\sigma^k,\pi^1)\cdot\pi^1}{1 - \psi(\sigma^k,\pi^1)} \tag{3}$$

for all $k \in \mathbb{N}$. Note that for every $s \in \times_{i\in I} S_i$ and for each $k \in \mathbb{N}$, if $\sigma^k(s) = 0$, then $\pi^1(s) = 0$ and thus $\pi^2_k(s) = 0$. It follows that $\text{supp}(\pi^2_k) \subseteq \text{supp}(\sigma^k)$ for all $k \in \mathbb{N}$. In addition, Lemma 9 (2) ensures that for every $k \in \mathbb{N}$, there exists $s \in \text{supp}(\sigma^k)$ such that $\sigma^k(s) - \psi(\sigma^k, \pi^1)\cdot\pi^1(s) = 0$, and thus $s \notin \text{supp}(\pi^2_k)$. Consequently, $\text{supp}(\pi^2_k) \subsetneq \text{supp}(\sigma^k)$ for all $k \in \mathbb{N}$.

Equation (3) can be rewritten as

$$\sigma^k = \psi(\sigma^k, \pi^1) \cdot \pi^1 + \big[1 - \psi(\sigma^k, \pi^1)\big] \cdot \pi^2_k \tag{4}$$

for all $k \in \mathbb{N}$, where $\psi(\sigma^k, \pi^1) \in (0, 1)$. Lemma 9 (3) and Lemma 9 (1) ensure that $\lim_{k\to\infty} \psi(\sigma^k, \pi^1) = \psi(\lim_{k\to\infty} \sigma^k, \pi^1) = \psi(\pi^1, \pi^1) = 1$. Consider the sequence $(r^1_k)_{k\in\mathbb{N}}$ defined by

$$r^1_k := 1 - \psi(\sigma^k, \pi^1) \tag{5}$$

for all $k \in \mathbb{N}$, where $\lim_{k\to\infty} r^1_k = 1 - \lim_{k\to\infty} \psi(\sigma^k, \pi^1) = 0$. Equations (4) and (5) imply that

$$\sigma^k = (1 - r^1_k) \cdot \pi^1 + r^1_k \cdot \pi^2_k \tag{6}$$

for all $k \in \mathbb{N}$.

By similar reasoning applied to the sequence $(\pi^2_k)_{k\in\mathbb{N}}$, it follows that there exists a convergent subsequence $(\pi^2_{k_n})_{n\in\mathbb{N}}$ of $(\pi^2_k)_{k\in\mathbb{N}}$, also denoted as $(\pi^2_k)_{k\in\mathbb{N}}$ for the sake of simplicity, with limit $\pi_2 := \lim_{k\to\infty} \pi^2_k$. Either $\pi^2_k = \pi^2$ infinitely often or $\pi^2_k = \pi^2$ finitely often.

Suppose that $\pi^2_k = \pi^2$ infinitely often. Let $(\pi^2_{k_n})_{n\in\mathbb{N}}$ be a subsequence of $(\pi^2_k)_{k\in\mathbb{N}}$ such that $\pi^2_{k_n} = \pi^2$ for all $n \in \mathbb{N}$, let $(r_n)_{n\in\mathbb{N}} = \big((r^1_{k_n})\big)_{n\in\mathbb{N}}$ and let $\pi = (\pi^1, \pi^2)$. Equation (6) ensures that

$$\sigma^{k_n} = (1 - r^1_{k_n}) \cdot \pi^1 + r^1_{k_n} \cdot \pi^2 = r_n \,\square\, \pi$$

for all $n \in \mathbb{N}$, which completes the proof in this case.

Otherwise, suppose that $\pi^2_k = \pi^2$ finitely often. There exist sequences $(\pi^3_k)_{k\in\mathbb{N}}$ and $(r^2_k)_{k\in\mathbb{N}}$ such that the following properties hold:

$$\begin{aligned} \pi^2_k &= (1 - r^2_k) \cdot \pi^2 + r^2_k \cdot \pi^3_k \\ \pi^3_k &:= \frac{\pi^2_k - \psi(\pi^2_k, \pi^2) \cdot \pi^2}{1 - \psi(\pi^2_k, \pi^2)} \text{ and } \text{supp}(\pi^3_k) \subsetneq \text{supp}(\pi^2_k) \text{ for all } k \in \mathbb{N} \\ r^2_k &:= 1 - \psi(\pi^2_k, \pi^2) \text{ for all } k \in \mathbb{N} \text{ and } \lim_{k\to\infty} r^2_k = 0. \end{aligned} \tag{7}$$

Equations (6) and (7) imply that

$$\sigma^k = (1 - r^1_k) \cdot \pi^1 + r^1_k \cdot \big[(1 - r^2_k) \cdot \pi^2 + r^2_k \cdot \pi^3_k\big]. \tag{8}$$

Iterating the same reasoning for the sequences $(\pi^l_k)_{k\in\mathbb{N}}$ for $l \geqslant 3$ guarantees that there exist a lexicographic level $L \in \mathbb{N}$, $\pi = (\pi^1, \dots, \pi^L) \in \big(\Delta(\times_{i\in I} S_i)\big)^L$, and $(r_n)_{n\in\mathbb{N}} \in \big[(0, 1)^{L-1}\big]^{\mathbb{N}}$ such that $\lim_{n\to\infty} r_n = \mathbf{0}$ and $\sigma^{k_n} = r_n \,\square\, \pi$ for all $n \in \mathbb{N}$. Note that the iterative process necessarily terminates after finitely many rounds, since the set $\times_{i\in I} S_i$ is finite and $\text{supp}(\sigma^k) \supsetneq \text{supp}(\pi^2_k) \supsetneq \text{supp}(\pi^3_k) \supsetneq \dots$ for all $k \in \mathbb{N}$. ■

Equipped with Lemmas A.1, A.2, and A.3, we can now proceed to formally establish Lemma 1.

Proof. (⇒): Suppose that σ constitutes a perfect equilibrium of Γ. Then, there exists a sequence of tuples of mixed strategies $(\sigma^k)_{k\in\mathbb{N}}$ such that properties (i), (ii), and (iii) of Definition 6 hold. By Lemma 9, there exists a lexicographic probability measure $\pi = (\pi^1, \dots, \pi^L) \in \big(\Delta(\times_{i\in I} S_i)\big)^L$ and a sequence $(r_n)_{n\in\mathbb{N}} = \big((r_n^1, \dots, r_n^{L-1})\big)_{n\in\mathbb{N}} \in \big[(0,1)^{L-1}\big]^{\mathbb{N}}$ with $\lim_{n\to\infty} r_n = \mathbf{0}$ such that some subsequence $(\sigma^{k_n})_{n\in\mathbb{N}}$ of $(\sigma^k)_{k\in\mathbb{N}}$ can be expressed as $\sigma^{k_n} = r_n \square \pi$ for all $n \in \mathbb{N}$. For every $i \in I$, define the lexicographic conjecture $\beta_i := \text{marg}_{S_{-i}}\pi$. We show that $(\sigma^k)_{k\in\mathbb{N}}$, $(\beta_i)_{i\in I}$, and π satisfy properties (a), (b), (c), (d), and (e) of Definition 7.

First, note that property (e) of Definition 7 is directly satisfied. Since $\sigma^{k_n} = r_n \square \pi$ is a product measure for all $n \in \mathbb{N}$, it follows that π is a tuple of product measures. Consequently,

$$\beta_i = \text{marg}_{S_{-i}}\pi = \bigotimes_{j\in I\setminus\{i\}} \text{marg}_{S_j}\pi = \bigotimes_{j\in I\setminus\{i\}} \text{marg}_{S_j}\text{marg}_{S_{-i}}\pi = \bigotimes_{j\in I\setminus\{i\}} \text{marg}_{S_j}\beta_i$$

for all $i \in I$, which yields property (d) of Definition 7. Moreover, property (i) ensures that $\sigma = \lim_{n\to\infty} \sigma^{k_n} = \lim_{n\to\infty}(r_n \square \pi) = \pi^1$. Hence,

$$\sigma_i = \text{marg}_{S_i}\sigma = \text{marg}_{S_i}\pi^1 = \text{marg}_{S_i}\text{marg}_{S_{-j}}\pi^1 = \text{marg}_{S_i}b_j^1$$

for all $i \in I$ and all $j \in I\setminus\{i\}$, which establishes property (b) of Definition 7. Furthermore, property (ii) guarantees that $\sigma_i^{k_n}$ has full support for all $i \in I$ and for all $n \in \mathbb{N}$. Thus, π and hence β_i, is cautious for all $i \in I$, which establishes property (a) of Definition 7. Finally, let $s_i \in \text{supp}(\sigma_i)$. By property (iii), s_i is a best response to $\sigma_{-i}^{k_n} = \text{marg}_{S_{-i}}(r_n \square \pi) = r_n \square \beta_i$ for all $n \in \mathbb{N}$. By Lemma 9 (ii), s_i is lex-optimal given β_i, which corresponds to property (c) of Definition 7. Therefore, $\sigma = (\sigma_i)_{i\in I}$ constitutes a lexicographic perfect equilibrium of Γ.

(⇐): Suppose that σ constitutes a lexicographic perfect equilibrium of Γ. Then, there exists a tuple of lexicographic conjectures $\beta = (\beta_i)_{i\in I}$ and a lexicographic product measure $\pi = (\pi^1, \dots, \pi^L)$ satisfying properties (a), (b), (c), (d), and (e) of Definition 7. Consider the sequence $(r_n)_{n\in\mathbb{N}} = \big((\frac{1}{n+1}, \dots, \frac{1}{n+1})\big)_{n\in\mathbb{N}} \in \big[(0,1)^{L-1}\big]^{\mathbb{N}}$. Note that $\lim_{n\to\infty} r_n = \mathbf{0}$. For every $i \in I$ and for every $n \in \mathbb{N}$, define $\sigma_i^n := \text{marg}_{S_i}(r_n \square \pi)$ and $\sigma^n := (\sigma_i^n)_{i\in I}$. We show that there exists a subsequence of $(\sigma^n)_{n\in\mathbb{N}}$ satisfying properties (i), (ii), (iii) of Definition 6.

Let $i \in I$ be some player. Since $r_n \square \pi$ is a product measure and properties (e) and (b) hold,

$$\begin{aligned}\lim_{n\to\infty} \sigma_i^n &= \lim_{n\to\infty} \text{marg}_{S_i}(r_n \square \pi) = \lim_{n\to\infty} \text{marg}_{S_i}\text{marg}_{S_{-j}}(r_n \square \pi)\\ &= \lim_{n\to\infty} \text{marg}_{S_i}(r_n \square \text{marg}_{S_{-j}}\pi) = \lim_{n\to\infty} \text{marg}_{S_i}(r_n \square \beta_j)\\ &= \text{marg}_{S_i} \lim_{n\to\infty}(r_n \square \beta_j) = \text{marg}_{S_i} b_j^1 = \sigma_i\end{aligned}$$

for all $j \in I$ such that $i \neq j$. This establishes property (i) of Definition 6. In addition, let $j \in I\setminus\{i\}$, $s_j \in S_j$, and $n \in \mathbb{N}$. Property (a) ensures that there exists a level

$l^* \in \{1, \ldots, L\}$ such that $\text{marg}_{S_j} b_i^{l^*}(s_j) > 0$. It follows that

$$\begin{aligned}\sigma_j^n(s_j) &= \text{marg}_{S_j}(r_n \square \pi)(s_j) = \text{marg}_{S_j}\text{marg}_{S_{-i}}(r_n \square \pi)(s_j) \\ &= \text{marg}_{S_j}(r_n \square \text{marg}_{S_{-i}}\pi)(s_j) = \text{marg}_{S_j}(r_n \square \beta_i)(s_j) > 0.\end{aligned}$$

Hence, $\text{supp}(\sigma_j^n) = S_j$, which yields property (ii) of Definition 6. Besides, let $s_i \in \text{supp}(\sigma_i)$. Property (c) ensures that s_i is lex-optimal given β_i. By Lemma 9 (iii), there exists some subsequence $(r_{n_k})_{k \in \mathbb{N}}$ of $(r_n)_{n \in \mathbb{N}}$ and some index $K \in \mathbb{N}$ such that $u_i(s_i, r_{n_k} \square \beta_i) \geqslant u_i(s_i', r_{n_k} \square \beta_i)$ for all $k \geqslant K$ and for all $s_i' \in S_i$. Property (e) guarantees that

$$\begin{aligned}r_{n_k} \square \beta_i &= (r_{n_k} \square \text{marg}_{S_{-i}}\pi) = \text{marg}_{S_{-i}}(r_{n_k} \square \pi) \\ &= \bigotimes_{j \in I \setminus \{i\}} \text{marg}_{S_j}(r_{n_k} \square \pi) = \bigotimes_{j \in I \setminus \{i\}} \sigma_j^{n_k}.\end{aligned}$$

Hence, s_i is a best response to $\sigma_{-i}^{n_k}$ for all $k \geqslant K$, i.e. the subsequence $(\sigma_{-i}^{n_k})_{k \geqslant K}$ satisfies property (iii) of Definition 6. Consequently, the subsequence $(\sigma^{n_k})_{k \geqslant N}$ satisfies properties (i), (ii), (iii) of Definition 6. Therefore, σ constitutes a perfect equilibrium of Γ. ■

References

Asheim, G. B. (2001): Proper Rationalizability in Lexicographic Beliefs. *International Journal of Game Theory* 30, 453–478.

Asheim, G. B. (2002): On the Epistemic Foundation for Backward Induction. *Mathematical Social Sciences* 44, 121–144.

Asheim, G. B. and Perea, A. (2005): Sequential and Quasi-Perfect Rationalizability in Extensive Games. *Games and Economic Behavior* 53, 15–42.

Aumann, R. J. (1974): Subjectivity and Correlation in Randomized Strategies. *Journal of Mathematical Economics* 1, 67–96.

Aumann, R. J. (1976): Agreeing to Disagree. *Annals of Statistics* 4, 1236–1239.

Aumann, R. J. and Brandenburger, A. (1995): Epistemic Conditions for Nash Equilibrium. *Econometrica* 63, 1161–1180.

Bach, C. W. and Cabessa, J. (2017): Limit-Agreeing to Disagree. *Journal of Logic and Computation* 27, 1169–1187.

Bach, C. W. and Perea, A. (2013): Agreeing to Disagree with Lexicographic Prior Beliefs. *Mathematical Social Sciences* 66, 129–133.

Bach, C. W. and Perea, A. (2020): Generalized Nash Equilibrium without Common Belief in Rationality. *Economics Letters* 186, 1–6.

Bach, C. W. and Tsakas, E. (2014): Pairwise Epistemic Conditions for Nash Equilibrium. *Games and Economic Behavior* 85, 48–59.

Barelli, P. (2009): Consistency of Beliefs and Epistemic Conditions for Nash and Correlated Equilibria. *Games and Economic Behavior* 67, 363–375.

Battigalli, P. and Siniscalchi, M. (1999): Hierarchies of Conditional Beliefs and Interactive Epistemology in Dynamic Games. *Journal of Economic Theory* 88, 188–230.

Battigalli, P. and Siniscalchi, M. (2002): Strong Belief and Forward Induction Reasoning. *Journal of Economic Theory* 106, 356–391.

Blume, L.; Brandenburger, A. and Dekel, E. (1991a): Lexicographic Probabilities and Choice under Uncertainty. *Econometrica* 59, 61–79.

Blume, L.; Brandenburger, A. and Dekel, E. (1991b): Lexicographic Probabilities and Equilibrium Refinements. *Econometrica* 59, 81–98.

Bonanno, G. (2018): Behavior and Deliberation in Perfect-Information Games: Nash Equilibrium and Backward Induction. *International Journal of Game Theory* 47, 1001–1032.

Bonanno, G. and Nehring, K. (1997): Agreeing to Disagree: A Survey. Mimeo.

Börgers, T. (1994): Weak Dominance and Approximate Common Knowledge. *Journal of Economic Theory* 64, 265–276.

Brandenburger, A. (1992a): Lexicographic Probabilities and Iterated Admissibility. In *Economic Analysis of Markets and Games*, ed. P. Dasgupta et al., MIT Press, 282–290.

Brandenburger, A. (1992b): Knowledge and Equilibrium in Games. *The Journal of Economic Perspectives* 6, 83–101.

Brandenburger, A. and Dekel, E. (1987): Common Knowledge with Probability 1. *Journal of Mathematical Economics* 16, 237–245.

Brandenburger, A.; Friedenberg, A. and Keisler, J. (2008): Admissibility in Games. *Econometrica* 76, 307–352.

Catonini, E. and De Vito, N. (2018): Cautious Belief and Iterated Admissibility. Mimeo

Catonini, E. and De Vito, N. (2020): Weak Belief and Permissibility. *Games and Economic Behavior* 120, 154–179.

Chen, Y.-C.; Lehrer, E.; Li, J.; Samet, D. and Shmaya, E. (2015): Agreeing to Disagree and Dutch Books. *Games and Economic Behavior* 93, 108–116.

Contreras-Tejada, P.; Scarpa, G.; Kubicki, A. M.; Brandenburger, A. and La Mura, P. (2021): Observers of Quantum Systems cannot Agree to Disagree. *Nature Communications* 12, article number: 7021.

Dégremont, C. and Roy, O. (2012): Agreement Theorems in Dynamic-Epistemic Logic. *Journal of Philosophical Logic* 41, 735–764.

Dekel, E.; Friedenberg, A. and Siniscalchi, M. (2016): Lexicographic Beliefs and Assumption. *Journal of Economic Theory* 163, 955–985.

Demey, L. (2014): Agreeing to Disagree in Probabilistic Dynamic Epistemic Logic. *Synthese* 191, 409–438.

Dominiak, A. and Lefort, J.-P. (2015): Agreeing to Disagree Type Results Under Ambiguity. *Journal of Mathematical Economics* 61, 119–129.

Gale, D. (1953): A Theory of n-Person Games with Perfect Information. *Proceedings of the National Academy of Sciences* 39, 496–501.

Gizatulina, A. and Hellman, Z. (2019): No Trade and Yes Trade Theorems for Heterogeneous Priors. *Journal of Economic Theory* 182, 161–184.

Halpern, J. (2010): Lexicographic Probability, Conditional Probability, and Nonstandard Probability. *Games and Economic Behavior* 68, 155–179.

Hammond, P. (1994): Elementary non-Archimedean Representation of Probability for Decision Theory and Games. In *Patrick Suppes: Scientific Philosopher, Vol. 1: Probability and Probabilistic Causality*, ed. P. Humphrey, Kluwer, 25–61.

Harsanyi, J. C. (1967-68): Games of Incomplete Information played by "Bayesian Players". Part I, II, III. *Management Science* 14, 159–182, 320–334, 486–502.

Harsanyi, J. (1973): Games with Randomly Disturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points. *International Journal of Game Theory* 2, 1–23.

Heifetz, A.; Meier, M. and Schipper, B. C. (2013): Unawareness, Beliefs and Speculative Trade. *Games and Economic Behavior* 77, 100–121.

Hellman, Z. (2013): Almost Common Priors. *International Journal of Game Theory* 42, 399–410.

Hellman, Z. and Samet, D. (2012): How Common are Common Priors? *Games and Economic Behavior* 74, 517–525.

KREPS, D. AND RAMEY, G. (1987): Structural Consistency, Consistency, and Sequential Rationality. *Econometrica* 55, 1331–1348.

KREPS, D. AND WILSON, R. (1982): Sequential Equilibria. *Econometrica* 50, 863–894.

LEE, B. S. (2016): Admissibility and Assumption. *Journal of Economic Theory* 163, 42–72.

LEHRER, E. AND SAMET, D. (2014): Belief Consistency and Trade Consistency. *Games and Economic Behavior* 83, 165–177.

LIU, Y. (2019): Two Tales of Epistemic Models. *Thought: A Journal of Philosophy* 8, 291–302 .

MAILATH, G. J.; SAMUELSON, L. AND SWINKELS, J. M. (1997): How Proper is Sequential Equilibrium? *Games and Economic Behavior* 18, 193–218.

MÉNAGER, L. (2012): Agreeing to Disagree: A Review. Mimeo.

MILGROM, P. AND STOKEY, N. (1982): Information, Trade and Common Knowledge. *Journal of Economic Theory* 26, 17–27.

MYERSON, R. B. (1978): Refinements of the Nash Equilibrium Concept. *International Journal of Game Theory* 7, 73–80.

NASH, J. (1950): Equilibrium Points in N-Person Games. *Proceedings of the National Academy of Sciences* 36, 48–49.

NASH, J. (1951): Non-Cooperative Games. *The Annals of Mathematics* 54, 286–295.

PACUIT, E. (2018): Agreement Theorems with Qualitative Conditional Probability. Mimeo.

PEREA, A. (2007): A One-Person Doxastic Characterization of Nash Strategies. *Synthese* 158, 1251–1271.

PEREA, A. (2012): *Epistemic Game Theory: Reasoning and Choice*. Cambridge University Press.

RÉNYI, A. (1995): On a New Axiomatic Theory of Probability. *Acta Mathematica Academiae Scientiarum Hungarica* 6, 285–335.

SAMUELSON, L. (1992): Dominated Strategies and Common Knowledge. *Games and Economic Behavior* 4, 284–313.

SELTEN, R. (1975): Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games. *International Journal of Game Theory* 4, 25–55.

Stahl, D. O. (1995): Lexicographic Rationalizability and Iterated Admissibility. *Economics Letters* 47, 155–159.

Stuart, H. W. (1997): Common Belief of Rationality in the Finitely Repeated Prisoners' Dilemma. *Games and Economic Behavior* 19, 133–143.

Tarbush, B. (2016): Counterfactuals in "Agreeing to Disagree" Type Results. *Mathematical Social Sciences* 84, 125–133.

Tsakas, E. (2014): Epistemic Equivalence of Extended Belief Hierarchies. *Games and Economic Behavior* 86, 126–144 .

Tsakas, E. (2018): Agreeing to Disagree with Conditional Probability Systems. *The B.E. Journal of Theoretical Economics* 18, 1–7 .

Yang, C. (2015): Weak Assumption and Iterated Admissibility. *Journal of Economic Theory* 158, 87–101.