Mathematical Social Sciences 66 (2013) 129–133

Contents lists available at SciVerse ScienceDirect

Mathematical Social Sciences

journal homepage: www.elsevier.com/locate/econbase

Agreeing to disagree with lexicographic prior beliefs

Christian W. Bach, Andrés Perea *

Department of Quantitative Economics & EpiCenter, School of Business and Economics, Maastricht University, 6200 MD Maastricht, The Netherlands

HIGHLIGHTS

- Two agents with almost identical priors can agree to completely disagree on their posterior beliefs.
- A slight perturbation of the common lexicographic prior can lead to common knowledge of completely opposed posterior beliefs.
- Agents can agree to disagree even if there is only a slight deviation from the common prior assumption.

ARTICLE INFO

Article history:
Received 14 August 2012
Received in revised form
19 March 2013
Accepted 21 March 2013
Available online 9 May 2013

ABSTRACT

The robustness of Aumann's seminal agreement theorem with respect to the common prior assumption is considered. More precisely, we show by means of an example that two Bayesian agents with almost identical prior beliefs can agree to completely disagree on their posterior beliefs. Besides, a more detailed agent model is introduced where posterior beliefs are formed on the basis of lexicographic prior beliefs. We then generalize Aumann's agreement theorem to lexicographic prior beliefs and show that only a slight perturbation of the common lexicographic prior assumption at some – even arbitrarily deep – level is already compatible with common knowledge of completely opposed posterior beliefs. Hence, agents can actually agree to disagree even if there is only a slight deviation from the common prior assumption.

1. Introduction

The impossibility of two agents to agree to disagree is established by Aumann's (1976) so-called agreement theorem. More precisely, it is shown that two Bayesian agents entertaining a common prior belief necessarily hold equal posterior beliefs in an event upon receiving private information in the case of their posterior beliefs being common knowledge. In other words, distinct posteriors cannot be common knowledge among Bayesian agents with a common prior. In this sense, agents cannot agree to disagree.

From an empirical as well as intuitive point of view the agreement theorem seems quite startling, since people frequently disagree on a variety of issues, while at the same time acknowledging their divergent opinions. It is thus natural to analyze whether Aumann's impossibility result still holds with weakened or slightly modified assumptions. In this spirit, Geanakoplos and Polemarchakis (1982) show that without assuming common knowledge of the posteriors, agents following a specific communication procedure can nevertheless not agree to disagree. Furthermore, Monderer and Samet (1989) replace common knowledge by the weaker concept of common p-belief and establish an agreement theorem with such an approximation of common knowledge. Indeed, they show that if the posteriors of Bayesian agents equipped with a common prior are common p-belief for large enough p, then these posteriors cannot differ significantly. Besides, a bounded rationality approach is taken by Samet (1990) who drops the implicit negative introspection assumption – which states that agents know what they do not know – and establishes that Aumann's agreement theorem remains valid with agents ignorant of their own ignorance. Yet a different generalization is provided by Bacharach (1985) who shows that if two agents follow a common decision procedure in line with the sure thing principle – which states that for every event and every partition of it, if each cell of the partition induces the same decision, then the event itself generates precisely this decision – and their particular decisions are common knowledge, then these decisions must coincide. In fact, Aumann's agreement theorem can be seen as a special case of Bacharach's result, since Bayesian updating from a common prior belief in some event constitutes a specific decision procedure satisfying the sure thing principle for determining a posterior belief in that event. Moreover, Bonanno and Nehring (1997) as well as Ménager (2012) provide rather comprehensive surveys on works on the agreement theorem. More recently, Hellman (2013) analyzes Aumann's impossibility result in a particular context of almost common priors and obtains a generalization of it. Besides, Bach and Cabessa (2012) derive a possibility result for agreeing to disagree in a topologically enriched epistemic model by replacing common knowledge of posteriors by limit knowledge – defined as the limit of iterated mutual

* Corresponding author. Tel.: +31 618829367.
E-mail addresses: c.bach@maastrichtuniversity.nl (C.W. Bach), a.perea@maastrichtuniversity.nl (A. Perea).

0165-4896/$ – see front matter
http://dx.doi.org/10.1016/j.mathsocsci.2013.03.004

knowledge – of posteriors. Further, Heifetz et al. (2012) provide an agreement theorem under unawareness.

The common prior assumption in economic theory in general and in game theory in particular is controversial and has been criticized, for example, by Morris (1995) and Gul (1998). With regard to Aumann's agreement theorem the question then arises to what extent the impossibility of agents to agree to disagree depends on their common prior.

Here, we first assume almost identical priors and show that agents can entertain completely opposed posteriors while at the same time satisfying common knowledge of these posteriors. In a more general context we then introduce an enriched and arguably more plausible model of lexicographically-minded agents, who form their posterior beliefs on the basis of lexicographic prior beliefs. Indeed, the relevant beliefs, i.e. the posteriors, remain single probability measures, yet agents can reason on the basis of a more detailed initial perception of the world. Note that lexicographic beliefs are typically used to model cautious reasoning and have also recently quite frequently been adopted in epistemic game theory. Here, we provide an agreement theorem for lexicographic beliefs. For this theorem to obtain, a strengthened common prior assumption is needed. More precisely, the agents' prior beliefs do not only have to be identical according to their primary perception of the state space but on all lexicographic levels. However, only slightly perturbing the common lexicographic prior assumption at some – even arbitrarily deep – level is already compatible with common knowledge of completely opposed posteriors. In this sense agents can actually agree to disagree.

2. Aumann's model

Before our possibility result on agreeing to disagree is formally presented, we briefly recall the required ingredients of Aumann's epistemic framework. A so-called Aumann structure $\mathcal{A} = (\Omega, (\mathcal{I}_i)_{i\in I}, p)$ consists of a finite set Ω of possible worlds, which are complete descriptions of the way the world might be, a finite set of agents I, a possibility partition $\mathcal{I}_i$ of Ω for each agent $i \in I$ representing his information, and a common prior belief function $p : \Omega \to [0, 1]$ 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 contains all worlds considered possible by i at world ω. In other words, agent i cannot distinguish between any two worlds ω and ω' that are in the same cell of his partition $\mathcal{I}_i$. Moreover, an event $E \subseteq \Omega$ is defined as a set of possible worlds. For instance, the event of it raining in London consists of all worlds in which it does rain in London.

Note that the common prior belief function p can naturally be extended to a common prior belief measure on the event space $p : \mathcal{P}(\Omega) \to [0, 1]$ by setting $p(E) = \sum_{\omega\in E} p(\omega)$. In this context, it is assumed that any information set has non-zero prior probability, i.e. $p(\mathcal{I}_i(\omega)) > 0$ for all $i \in I$ and $\omega \in \Omega$. Such a hypothesis seems plausible since it ensures no piece of information to be excluded a priori. Moreover, all agents are assumed to be Bayesians and to hence update the common prior belief given their private information according to Bayes's rule. More precisely, given some event E and some world ω, the posterior belief of agent i in E at ω is given by $p(E \mid \mathcal{I}_i(\omega)) = \frac{p(E\cap\mathcal{I}_i(\omega))}{p(\mathcal{I}_i(\omega))}$.

In Aumann's epistemic framework, knowledge is formalized in terms of events. The event of agent i knowing 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 world ω. Intuitively, i knows some event E if in all worlds he considers possible E holds. Naturally, the event $K(E) = \bigcap_{i\in I} K_i(E)$ then denotes mutual knowledge of E among the set I of agents. Letting $K^0(E) := E$, m-order mutual knowledge of the event E among the set I of agents is inductively defined by $K^m(E) := K(K^{m-1}(E))$ for all $m > 0$. Accordingly, mutual knowledge can also be denoted as 1-order mutual knowledge. Furthermore, an event is said to be common knowledge among a set I of agents whenever all m-order mutual knowledge of it simultaneously hold. It is then standard to define the event that E is common knowledge among the set I of agents as the infinite intersection of all higher-order mutual knowledge. Formally, the event E is common knowledge among the agents at some world ω if $\omega \in \bigcap_{m>0} K^m(E)$. Hence, the standard definition of common knowledge of some event E can be stated as $CK(E) := \bigcap_{m>0} K^m(E)$.

An alternative – yet equivalent – definition of common knowledge in terms of the meet of the agents' possibility partitions is proposed by Aumann (1976) and also used in his agreement theorem. Before the meet definition of common knowledge can be given some further set-theoretic notions have to be introduced. Given two partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ of a set S, partition $\mathcal{P}_1$ is called *finer* than 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 $\bigwedge_{i=1}^n(\mathcal{P}_i)(x)$. Now, according to the meet definition of common knowledge, an event E is said to be common knowledge at some world ω among the set I of agents, if E includes the member of the meet $\bigwedge_{i\in I} \mathcal{I}_i$ that contains ω. Formally, the meet definition of common knowledge of some event E can thus be stated as $CK(E) := \{\omega \in \Omega : (\bigwedge_{i\in I} \mathcal{I}_i)(\omega) \subseteq E\}$.

3. Motivating example

We now turn to the possibility of agents to agree to disagree. The common prior assumption is slightly perturbed in the sense of assuming arbitrarily close prior belief functions for the agents. Indeed, the following example shows that two Bayesian agents with almost identical prior beliefs can agree to completely disagree on their posterior beliefs.

Example 1. Consider $\Omega = \{\omega_1, \omega_2, \omega_3\}$, $\mathcal{I}_{Alice} = \mathcal{I}_{Bob} = \{\{\omega_1, \omega_2\}, \{\omega_3\}\}$ and $E = \{\omega_1\}$. Moreover, let $\epsilon > 0$ and $p_{Alice} : \Omega \to [0, 1]$ be *Alice*'s prior belief function such that $p_{Alice}(\{\omega_1\}) = \epsilon$, $p_{Alice}(\{\omega_2\}) = 0$, and $p_{Alice}(\{\omega_3\}) = 1 - \epsilon$. Also, let $p_{Bob} : \Omega \to [0, 1]$ be *Bob*'s prior belief function such that $p_{Bob}(\{\omega_1\}) = 0$, $p_{Bob}(\{\omega_2\}) = \epsilon$, and $p_{Bob}(\{\omega_3\}) = 1-\epsilon$. At ω_1 as well as at ω_2, *Alice*'s posterior belief in E is given by $p_{Alice}(E \mid \mathcal{I}_{Alice}(\omega_1)) = \frac{\epsilon}{\epsilon+0} = 1$, while *Bob*'s posterior belief in E is given by $p_{Bob}(E \mid \mathcal{I}_{Bob}(\omega_1)) = \frac{0}{0+\epsilon} = 0$. Suppose ω_1 to be the actual world. Note that it is common knowledge at ω_1 that $p_{Alice}(E \mid \mathcal{I}_{Alice}(\omega_1)) = 1$ and $p_{Bob}(E \mid \mathcal{I}_{Bob}(\omega_1)) = 0$. Hence, at world ω_1 the two agents' posterior beliefs are common knowledge, yet completely different. □

Accordingly, two agents can entertain absolutely opposing posterior beliefs, despite being equipped with arbitrarily close prior beliefs and their posterior beliefs being common knowledge.[1] Hence, agents can indeed agree to disagree. Moreover, the possible effects of a slight perturbation of the common prior assumption in Aumann's impossibility result show that the agreement theorem is not robust.

Note that in Example 1 the agents agree to disagree on an event that is considered unlikely to occur a priori. Yet, it would be fallacious to infer the irrelevance of an event from its improbability. For

[1] Note that our example does not contradict Hellman's (2013) result that whenever agents have almost common priors the extent of their disagreement is small and bounded. In fact, Hellman assumes that the meet of the players' partitions equals the set of all possible worlds—a condition which is violated in our example.

instance, in the context of dynamic games, precisely those events that are initially believed not to occur can have a crucial influence on what agents do later on in the game and whether their behavior conforms to particular reasoning patterns or solution concepts. In general, events that are surprising or deemed improbable can thus certainly be relevant and should as other, more probable, events be handled with the same care.

4. Lexicographic prior beliefs

In Aumann's original model the posterior beliefs are derived from a common prior which is given by a *single* probability measure on the state space. In order to have well-defined posteriors the prior must assign non-zero probability to every information set of every agent. However, this assumption may be quite restrictive. To relax this assumption we model the prior belief of an agent not by a single probability measure but by a *sequence* of probability measures. That is, we employ lexicographic beliefs in the sense of Blume et al. (1991) for the agents' priors. To have well-defined posterior beliefs it is now sufficient to require that every information set receives positive probability at *some* level – not necessarily the first – of the lexicographic prior belief. Yet, an agent's posterior is given by a single probability measure and the lexicographic nature of the prior is only used to generate this posterior. Hence, we maintain the classical view that the agents' relevant beliefs, i.e. the posteriors, are given by a single probability measure, while admitting agents to entertain a more fine-grained initial perception of the world. This approach is also taken by Asheim (2002) and Asheim and Perea (2005), in which the players' conditional beliefs in a dynamic game are derived from a lexicographic prior belief. Also in belief revision theory it is common to base conditional beliefs on lexicographic prior beliefs. In general, an agent equipped with a lexicographic prior belief can be interpreted as a cautious reasoner, who deems nothing impossible, but may consider some events infinitely more likely than other events.

Formally, the standard common prior belief is now replaced by subjective lexicographic prior beliefs for every agent in Aumann structures. Indeed, we call $\mathcal{A}_l = (\Omega, (\mathcal{I}_i)_{i\in I}, (b_i)_{i\in I})$ a lexicographic Aumann structure, where b_i is a lexicographic prior belief for all agents $i \in I$. More precisely, $b_i = (b_i^1, b_i^2, \ldots, b_i^K)$ for some $K \in \mathbb{N}$ is a finite sequence of prior belief functions $b_i^k : \Omega \to [0, 1]$ for all $k \in \{1, 2, \ldots, K\}$ such that

1. $\Sigma_{\omega\in\Omega} b_i^k(\omega) = 1$ for all $k \in \{1, 2, \ldots, K\}$,
2. for every $S \in \mathcal{I}_i$ there exists $k^* \in \{1, 2, \ldots, K\}$ such that $\text{supp}(b_i^{k^*}) \cap S \neq \emptyset$,
3. $\text{supp}(b_i^{k'}) \cap \text{supp}(b_i^{k''}) = \emptyset$ for all $k' \neq k''$.

Note that the first condition ensures that the agents' prior belief functions actually are probability distributions at every lexicographic level. Moreover, the second requirement guarantees that every possible information an agent may face receives positive probability at some lexicographic level by stipulating its intersection with the support of the probability distribution at some level to be non-empty. Further, according to the third condition any distinct lexicographic levels never allot positive probability to an identical world. This criterion seems natural as subsequent lexicographic levels exhibit differences in order of likelihood and hence a world being in the support of some lexicographic level should not reappear at any deeper lexicographic level. Hence by excluding overlapping supports, the class of lexicographic prior beliefs we consider corresponds to Blume et al. (1991)'s notion of *lexicographic conditional probability system*. Besides, note that a lexicographic Aumann structure generalizes the notion of Aumann structure, with the former being equivalent to the latter in the case of $K = 1$. From an interpretative point of view, the probability distributions at the various lexicographic levels can be considered as successive theories or hypotheses about the world ordered lexicographically with decreasing plausibility from the first level onwards.

Similar to the case of standard beliefs, an agent's lexicographic prior belief can naturally be extended to a lexicographic prior belief measure on the event space. Indeed, given an event $E \subseteq \Omega$, agent i's lexicographic prior belief in E is given by the sequence $b_i(E) = (b_i^1(E), b_i^2(E), \ldots, b_i^K(E)) = (\Sigma_{\omega\in E} b_i^1(\omega), \Sigma_{\omega\in E} b_i^2(\omega), \ldots, \Sigma_{\omega\in E} b_i^K(\omega))$. With lexicographic prior beliefs we define Bayesian updating as follows: given an event $E \subseteq \Omega$ and a world ω, the posterior belief $b_i(E|\mathcal{I}_i(\omega))$ is given by $\frac{b_i^{k^*}(E\cap\mathcal{I}_i(\omega))}{b_i^{k^*}(\mathcal{I}_i(\omega))}$ for the smallest $k^* \in \{1, 2, \ldots, K\}$ such that $\text{supp}(b_i^{k^*}) \cap \mathcal{I}_i(\omega) \neq \emptyset$.

Modeling Bayesian agents with lexicographic priors provides a very complete as well as plausible agent model. Before any information is received no possible piece of information is excluded while at the same time some worlds can be considered infinitely more likely than others, and after information is received the agents update the respectively relevant level of their lexicographic prior to form a unique posterior representing their relevant belief induced by subjective information. Note that a common lexicographic prior assumption requires identical prior belief functions at all lexicographic levels for the agents.

Moreover, the derivation of conditional beliefs from lexicographic beliefs in our model is in line with belief revision theory. Indeed, the way posterior beliefs are formed on the basis of lexicographic prior beliefs here can be seen as a probabilistic analogue to Grove's (1988) representation theorem, which connects the AGM belief revision axioms of Alchourrón et al. (1985) to some kind of plausibility orderings. More precisely, within Grove's model his system of spheres can be interpreted as a plausibility ordering over possible worlds: conditional on an event E an agent's conditional beliefs concentrate on those worlds in E that are deemed most plausible. In fact, within our model lexicographic beliefs induce a plausibility ordering over possible worlds in the sense of a possible world ω being deemed more plausible than another world ω' whenever ω receives positive probability at an earlier lexicographic level than ω'.

It is now shown that common knowledge of the agents' posterior beliefs together with a strengthened common lexicographic prior assumption ensures the impossibility of agents to agree to disagree.

Theorem 1. *Let $\mathcal{A}_l = (\Omega, (\mathcal{I}_i)_{i\in I}, (b_i)_{i\in I})$ be a lexicographic Aumann structure such that $b_i = b$ for all $i \in I$, let $\hat{b}_i \in \mathbb{R}$ for all $i \in I$, and let $E \subseteq \Omega$ be some event. If $CK(\bigcap_{i\in I}\{\omega' \in \Omega : b(E \mid \mathcal{I}_i(\omega')) = \hat{b}_i\}) \neq \emptyset$, then $\hat{b}_i = \hat{b}_j$ for all $i, j \in I$.*

Proof. Let $\omega \in \Omega$ such that $\omega \in CK(\bigcap_{i'\in I}\{\omega' \in \Omega : b(E \mid \mathcal{I}_{i'}(\omega')) = \hat{b}_{i'}\})$ and consider agent $i \in I$. First of all, since the meet is coarser than i's possibility partition, note that each cell of the meet can be written as the union of the cells of i's possibility partition that it includes. Hence, there exists a set $A_i \subseteq \Omega$ such that $(\bigwedge_{i'\in I} \mathcal{I}_{i'})(\omega) = \bigcup_{\omega''\in A_i} \mathcal{I}_i(\omega'')$ and for all $\omega_1, \omega_2 \in A_i$, if $\omega_1 \neq \omega_2$, then $\mathcal{I}_i(\omega_1) \neq \mathcal{I}_i(\omega_2)$. Furthermore, by the definition of common knowledge it follows that $(\bigwedge_{i'\in I} \mathcal{I}_{i'})(\omega) \subseteq \bigcap_{i'\in I}\{\omega' \in \Omega : b(E \mid \mathcal{I}_{i'}(\omega')) = \hat{b}_{i'}\}$ and thus $b(E \mid \mathcal{I}_i(\omega'')) = \hat{b}_i$ for all $\omega'' \in (\bigwedge_{i'\in I} \mathcal{I}_{i'})(\omega)$. Now, consider some world $\omega^* \in A_i$ and let $k \in \{1, 2, \ldots, K\}$ denote the smallest lexicographic level such that $\text{supp}(b^k) \cap (\bigwedge_{i'\in I} \mathcal{I}_{i'})(\omega^*) \neq \emptyset$. Then, either $\text{supp}(b^k) \cap \mathcal{I}_i(\omega^*) = \emptyset$, consequently $b^k(\mathcal{I}_i(\omega^*)) = 0$ and hence $b(E \mid \mathcal{I}_i(\omega^*)) \cdot b^k(\mathcal{I}_i(\omega^*)) = b^k(E \cap \mathcal{I}_i(\omega^*))$ is trivially satisfied, or $\text{supp}(b^k) \cap \mathcal{I}_i(\omega^*) \neq \emptyset$ and thus by definition of lexicographic Bayesian updating $b(E \mid \mathcal{I}_i(\omega^*)) \cdot b^k(\mathcal{I}_i(\omega^*)) = b^k(E \cap \mathcal{I}_i(\omega^*))$. Hence, in both cases $b(E \mid \mathcal{I}_i(\omega^*)) \cdot b^k(\mathcal{I}_i(\omega^*)) = b^k(E \cap \mathcal{I}_i(\omega^*))$ obtains. Since $\omega^* \in A_i \subseteq (\bigwedge_{i'\in I} \mathcal{I}_{i'})(\omega)$, it holds that $b(E \mid$

$\mathcal{I}_i(\omega^*)) = \hat{b}_i$. It then follows that $\hat{b}_i \cdot b^k(\mathcal{I}_i(\omega^*)) = b^k(E \cap \mathcal{I}_i(\omega^*))$. Summing over all worlds in A_i thus yields the following equation of sums $\sum_{\omega'' \in A_i} b^k(E \cap \mathcal{I}_i(\omega'')) = \hat{b}_i \cdot \sum_{\omega'' \in A_i} b^k(\mathcal{I}_i(\omega''))$. Observe that due to countable additivity it follows that $\sum_{\omega'' \in A_i} b^k(E \cap \mathcal{I}_i(\omega'')) = b^k(\bigcup_{\omega'' \in A_i}(E \cap \mathcal{I}_i(\omega''))) = b^k(E \cap \bigcup_{\omega'' \in A_i} \mathcal{I}_i(\omega'')) = b^k(E \cap (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))$ and $\sum_{\omega'' \in A_i} b^k(\mathcal{I}_i(\omega'')) = b^k(\bigcup_{\omega'' \in A_i} \mathcal{I}_i(\omega'')) = b^k((\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))$. Thus, the equation of sums can be written as $b^k(E \cap (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega)) = \hat{b}_i \cdot b^k((\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))$, thence $\hat{b}_i = \frac{b^k(E \cap (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))}{b^k((\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))}$. Since agent i has been chosen arbitrarily, $\hat{b}_1 = \hat{b}_2 = \cdots = \hat{b}_K = \frac{b^k(E \cap (\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))}{b^k((\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega))}$ obtains. □

From a lexicographic point of view Theorem 1 unveils a strong common prior assumption for the impossibility of agents to agree to disagree. Indeed, agents need to entertain absolutely identical priors at all lexicographic levels. Intuitively, the same complete perception of the state space has to be shared by all agents including the way they assign probabilities to worlds considered infinitely less likely than others. It seems demanding and somewhat implausible to require agents not only to exhibit an equal perception of the state space in line with their respective primary prior hypotheses but also in line with any deeper prior hypotheses they form.

Note that formally Theorem 1 follows as a corollary from Bacharach's (1985) generalization of Aumann's agreement theorem, since the way we define Bayesian updating of a lexicographic prior belief can be viewed as a decision rule that satisfies the sure-thing principle. However, our proof of Theorem 1, being self-contained and direct, is still instructive and useful.

We turn towards relaxing the common lexicographic prior assumption. In fact, it is now shown that assuming distinct priors only at *some* lexicographic level already enables agents to agree to disagree on their posteriors.

Theorem 2. *Let Ω be a set of possible worlds, let I be a set of agents, let $b_{i'}$ be a lexicographic prior belief on Ω for each agent $i' \in I$ such that $b_i \neq b_j$ for some agents $i \neq j$. Then, there exist a possibility partition $\mathcal{I}_{i'}$ for all agents $i' \in I$, some numbers $\hat{b}_{i'} \in \mathbb{R}$ for all agents $i' \in I$ with $\hat{b}_i \neq \hat{b}_j$ and some event $E \subseteq \Omega$ such that $CK(\bigcap_{i \in I}\{\omega' \in \Omega : b_i(E \mid \mathcal{I}_i(\omega')) = \hat{b}_i\}) \neq \emptyset$.*

Proof. Let $k \in \{1, 2, \ldots, K\}$ be the smallest lexicographic level such that $b_i^k \neq b_j^k$. Then, there exists a world $\omega \in \Omega$ such that $b_i^k(\omega) \neq b_j^k(\omega)$. Hence, $b_i^k(\omega) > 0$ or $b_j^k(\omega) > 0$. Without loss of generality assume that $b_i^k(\omega) > 0$ and let $\mathcal{I}_{i'} = \{\{\bigcup_{k' < k} \text{supp}(b_i^{k'})\}, \Omega \setminus \{\bigcup_{k' < k} \text{supp}(b_i^{k'})\}\}$ for all agents $i' \in I$. Note that $b_i^k(\Omega \setminus \{\bigcup_{k' < k} \text{supp}(b_i^{k'})\}) = 1$, since $\text{supp}(b_i^k) \subseteq \Omega \setminus \{\bigcup_{k' < k} \text{supp}(b_i^{k'})\}$. As $b_j^{k'} = b_i^{k'}$ for all $k' < k$ it also holds that $b_j^k(\Omega \setminus \{\bigcup_{k' < k} \text{supp}(b_i^{k'})\}) = 1$. Now consider event $E = \{\omega\}$. It follows that $b_i(E \mid \mathcal{I}_i(\omega)) = \frac{b_i^k(E \cap \mathcal{I}_i(\omega))}{b_i^k(\mathcal{I}_i(\omega))} = \frac{b_i^k(\omega)}{b_i^k(\Omega \setminus \{\bigcup_{k' < k} \text{supp}(b_i^{k'})\})} = \frac{b_i^k(\omega)}{1} \neq \frac{b_j^k(\omega)}{1} = \frac{b_j^k(\omega)}{b_j^k(\Omega \setminus \{\bigcup_{k' < k} \text{supp}(b_i^{k'})\})} = \frac{b_j^k(E \cap \mathcal{I}_j(\omega))}{b_j^k(\mathcal{I}_j(\omega))} = b_j(E \mid \mathcal{I}_j(\omega))$. Let $\hat{b}_{i'} = b_{i'}(E \mid \mathcal{I}_{i'}(\omega))$ for every agent $i' \in I$. Note that then $\hat{b}_i > 0$ and $\hat{b}_i \neq \hat{b}_j$. Moreover, since an agent's posterior belief in any event always remains constant throughout any of his possibility cells, and $\bigwedge_{i' \in I} \mathcal{I}_{i'} = \mathcal{I}_i$ for all agents $i \in I$, it follows that $(\bigwedge_{i' \in I} \mathcal{I}_{i'})(\omega) = \mathcal{I}_{i'}(\omega) \subseteq \bigcap_{i' \in I}\{\omega' \in \Omega : b_{i'}(E \mid \mathcal{I}_{i'}(\omega')) = \hat{b}_{i'}\}$. Therefore, $\omega \in CK(\bigcap_{i' \in I}\{\omega' \in \Omega : b_{i'}(E \mid \mathcal{I}_{i'}(\omega')) = \hat{b}_{i'}\}) \neq \emptyset$. □

Accordingly, it is already possible for agents to agree to disagree if only at some – arbitrarily deep – lexicographic level they entertain different prior beliefs, despite their perception of the state space being completely identical at all respectively lower lexicographic levels.

Note that Theorem 2 depends on the assumption that the intersection of the supports of any distinct lexicographic levels is empty. For example, consider $\Omega = \{\omega_1, \omega_2\}$, $I = \{Alice, Bob\}$, $b_{Alice} = (b_{Alice}^1, b_{Alice}^2)$, and $b_{Bob} = (b_{Bob}^1, b_{Bob}^2)$ such that $b_{Alice}^1(\omega_1) = 1$, $b_{Alice}^2(\omega_1) = \frac{1}{2}$, $b_{Alice}^2(\omega_2) = \frac{1}{2}$, $b_{Bob}^1(\omega_1) = 1$, and $b_{Bob}^2(\omega_2) = 1$. Then, for every event E and for all possibility partitions $\mathcal{I}_{Alice}$ and $\mathcal{I}_{Bob}$, there exists no world at which there is common knowledge of the posterior beliefs in E being unique but different. The conclusion of Theorem 2 thus no longer holds when admitting overlapping supports for the agents' lexicographic prior beliefs.

Finally, the robustness of agreeing to disagree with lexicographic beliefs is analyzed. Indeed, a lexicographic Aumann structure is constructed in which two agents entertain almost identical lexicographic prior beliefs, yet their posterior beliefs are completely opposed and at the same time common knowledge. Towards this purpose we now introduce the notion of *ϵ-close priors*, using the notion of maximum norm. For every vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, the maximum norm is defined as $\|x\|_\infty := \max_{k \in \{1,2,\ldots,n\}} |x_k|$ for some $n \in \mathbb{N}$.

Definition 1. Let $\mathcal{A}_l$ be a lexicographic Aumann structure. The lexicographic prior beliefs b_i and b_j of two agents $i, j \in I$ of lexicographic depth K are called ϵ-close, if $\max_{k \in \{1,2,\ldots,K\}} \|b_i^k - b_j^k\|_\infty \leq \epsilon$.

The non-robustness of the impossibility of lexicographic agreeing to disagree with regard to the strengthened common prior assumption is formally stated as follows.

Theorem 3. *For all $\epsilon > 0$ and for all $k^* > 0$, there exists a lexicographic Aumann structure $\mathcal{A}_l = (\Omega, (\mathcal{I}_i)_{i \in \{Alice, Bob\}}, (b_i)_{i \in \{Alice, Bob\}})$ and some event $E \subseteq \Omega$ such that $b_{Alice}^k = b_{Bob}^k$ for all $k < k^*$, b_{Alice} and b_{Bob} are ϵ-close, $CK(\bigcap_{i \in \{Alice, Bob\}}\{\omega' \in \Omega : b_i(E \mid \mathcal{I}_i(\omega')) = \hat{b}_i\}) \neq \emptyset$, $\hat{b}_{Alice} = 1$ but $\hat{b}_{Bob} = 0$.*

Proof. Let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{k^*}, \omega_{k^*+1}, \omega_{k^*+2}\}$, $\mathcal{I}_{Alice} = \mathcal{I}_{Bob} = \{\{\omega_1, \omega_2, \ldots, \omega_{k^*-1}\}, \{\omega_{k^*}, \omega_{k^*+1}\}, \{\omega_{k^*+2}\}\}$, as well as $b_{Alice} = (b_{Alice}^1, b_{Alice}^2, \ldots, b_{Alice}^{k^*})$ and $b_{Bob} = (b_{Bob}^1, b_{Bob}^2, \ldots, b_{Bob}^{k^*})$ that coincide for every lexicographic level $k < k^*$ and only differ at the last lexicographic level k^*. More precisely, let the agents' common lexicographic prior beliefs up to level k^*-1 be given by b^k such that $b^k(\omega_k) = 1$ for all $k \leq k^*-1$, and let the agents' lexicographic prior beliefs at level k^* be given by $b_{Alice}^{k^*}(\omega_{k^*}) = \epsilon$, $b_{Alice}^{k^*}(\omega_{k^*+1}) = 0$, and $b_{Alice}^{k^*}(\omega_{k^*+2}) = 1 - \epsilon$, as well as, $b_{Bob}^{k^*}(\omega_{k^*}) = 0$, $b_{Bob}^{k^*}(\omega_{k^*+1}) = \epsilon$, and $b_{Bob}^{k^*}(\omega_{k^*+2}) = 1 - \epsilon$, respectively. Then, b_{Alice} and b_{Bob} are ϵ-close. Let $E = \{\omega_{k^*}\}$ and note that $b_{Alice}(E \mid \mathcal{I}_{Alice}(\omega_{k^*})) = \frac{\epsilon}{\epsilon+0} = 1$, whereas $b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega_{k^*})) = \frac{0}{0+\epsilon} = 0$. Moreover, since an agent's posterior belief in any event always remains constant throughout any of his possibility cells and $\bigwedge_{i' \in \{Alice, Bob\}} \mathcal{I}_{i'} = \mathcal{I}_i$ for both agents $i \in \{Alice, Bob\}$, it follows that $(\bigwedge_{i' \in \{Alice, Bob\}} \mathcal{I}_{i'})(\omega_{k^*}) = \{\omega_{k^*}, \omega_{k^*+1}\} = \{\omega' \in \Omega : b_{Alice}(E \mid \mathcal{I}_{Alice}(\omega')) = 1\} \cap \{\omega' \in \Omega : b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega')) = 0\}$, and hence $\omega_{k^*} \in CK(\{\omega' \in \Omega : b_{Alice}(E \mid \mathcal{I}_{Alice}(\omega')) = 1\} \cap \{\omega' \in \Omega : b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega')) = 0\}) \neq \emptyset$. □

The preceding theorem illustrates that Aumann's impossibility result is also not robust with lexicographic prior beliefs. Indeed, only a slight perturbation of a common lexicographic prior at some – even arbitrarily deep – level can already yield completely opposed posteriors. A strong reliance of the impossibility of agents to agree to disagree on the common prior assumption is thus unveiled.

Note that Theorem 3 still holds when condition (2) of lexicographic Aumann structures $\mathcal{A}_l$ is strengthened to the following requirement: (2′) for every $\omega \in \Omega$ there exist $k^* \in \{1, 2, \ldots, K\}$ such that $\omega \in \mathrm{supp}(b_i^{k^*})$. Indeed, let $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{k^*}, \omega_{k^*+1}, \omega_{k^*+2}, \omega_{k^*+3}\}$, $\mathcal{I}_{Alice} = \mathcal{I}_{Bob} = \{\{\omega_1, \omega_2, \ldots, \omega_{k^*-1}\}, \{\omega_{k^*}, \omega_{k^*+1}\}, \{\omega_{k^*+2}\}\{\omega_{k^*+3}\}\}$, as well as $b_{Alice} = (b^1_{Alice}, b^2_{Alice}, \ldots, b^{k^*}_{Alice})$ and $b_{Bob} = (b^1_{Bob}, b^2_{Bob}, \ldots, b^{k^*}_{Bob})$ that coincide for every lexicographic level $k < k^*$ and only differ at the last two lexicographic levels k^* and $k^* + 1$. More precisely, let the agents' common lexicographic prior beliefs up to level $k^* - 1$ be given by b^k such that $b^k(\omega_k) = 1$ for all $k \leq k^* - 1$; let the agents' ϵ-close lexicographic prior beliefs at level k^* be given by $b^{k^*}_{Alice}(\omega_{k^*}) = \epsilon$, $b^{k^*}_{Alice}(\omega_{k^*+1}) = 0$, $b^{k^*}_{Alice}(\omega_{k^*+2}) = 1 - \epsilon$, as well as, $b^{k^*}_{Bob}(\omega_{k^*}) = 0$, $b^{k^*}_{Bob}(\omega_{k^*+1}) = \epsilon$, and $b^{k^*}_{Bob}(\omega_{k^*+2}) = 1 - \epsilon$, respectively; let the agents' ϵ-close lexicographic prior beliefs at level $k^* + 1$ be given by $b^{k^*+1}_{Alice}(\omega_{k^*+1}) = \epsilon$, $b^{k^*+1}_{Alice}(\omega_{k^*+3}) = 1 - \epsilon$, as well as, $b^{k^*+1}_{Bob}(\omega_{k^*}) = \epsilon$, and $b^{k^*+1}_{Bob}(\omega_{k^*+3}) = 1 - \epsilon$, respectively. Consider $E = \{\omega_{k^*}\}$ and observe that $b_{Alice}(E \mid \mathcal{I}_{Alice}(\omega_{k^*})) = \frac{\epsilon}{\epsilon+0} = 1$, whereas $b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega_{k^*})) = \frac{0}{0+\epsilon} = 0$. Moreover, since an agent's posterior belief in any event always remains constant throughout any of his possibility cells and $\bigwedge_{i' \in \{Alice,\ Bob\}} \mathcal{I}_{i'} = \mathcal{I}_{i'}$, it follows that $(\bigwedge_{i' \in \{Alice,\ Bob\}} \mathcal{I}_{i'})(\omega_{k^*}) = \{\omega_{k^*}, \omega_{k^*+1}\} = \{\omega' \in \Omega : b_{Alice}(E \mid \mathcal{I}_{Alice}(\omega')) = 1\} \cap \{\omega' \in \Omega : b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega')) = 0\}$, and hence $\omega_{k^*} \in CK(\{\omega' \in \Omega : b_{Alice}(E \mid \mathcal{I}_{Alice}(\omega')) = 1\} \cap \{\omega' \in \Omega : b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega')) = 0\}) \neq \emptyset$.

5. Conclusion

With regard to the controversial common prior assumption Aumann's agreement theorem has been shown not to be robust. Already a slight perturbation of the common prior is compatible with common knowledge of completely opposed posteriors. Moreover, the agent model has been extended from standard to lexicographic prior beliefs and a corresponding agreement theorem provided. However, the impossibility of agents to agree to disagree is also not robust in such an enriched lexicographic context. Indeed, only a slight difference of the agents' priors at some – even arbitrarily deep – lexicographic level may already yield completely opposed posteriors. These possibility results for slightly perturbed common priors unveil a strong reliance of the impossibility of agents to agree to disagree on the common prior assumption.

Further, note that we have assumed that agents hold lexicographic prior yet standard posterior beliefs. Our model can be modified such that agents also entertain lexicographic posteriors. However, from a conceptual point of view the simpler model with agents that are only lexicographically prior minded seems plausible: accordingly agents entertain a rich perception of the state space prior to receiving any information, while they subsequently use their information to form a unique posterior that then represents their relevant perception of the state space. Besides, lexicographic beliefs are used to generate conditional beliefs, which is in line with belief revision theory.

The implications of Aumann's impossibility theorem for speculative trade are studied by Milgrom and Stokey (1982). Reconsidering the consequences for speculation in the context of lexicographic agreeing to disagree constitutes an interesting problem for further research. Moreover, it would also be intriguing for future work to analyze the robustness of the converse of the agreement theorem. Indeed, the question can be addressed whether small disagreement can be compatible with large differences in priors.

Acknowledgments

This work was presented at the Ninth Conference on Logic and the Foundations of Game and Decision Theory (LOFT9), Toulouse, July 2010. We are grateful to the conference participants as well as to an anonymous Associate Editor and two anonymous referees at Mathematical Social Sciences for useful and constructive comments.

References

Alchourrón, C.E., Gärdenfors, P., Makinson, D., 1985. On the logic of theory change: partial meet contraction and revision functions. Journal of Symbolic Logic 50, 510–530.
Asheim, G.B., 2002. On the epistemic foundation for backward induction. Mathematical Social Sciences 44, 121–144.
Asheim, G.B., Perea, A., 2005. Sequential and quasi-perfect rationalizability in extensive games. Games and Economic Behavior 53, 15–42.
Aumann, R.J., 1976. Agreeing to disagree. Annals of Statistics 4, 1236–1239.
Bach, C.W., Cabessa, J., 2012. Limit-agreeing to disagree. Working Paper.
Bacharach, M., 1985. Some extensions of a claim of Aumann in an axiomatic model of knowledge. Journal of Economic Theory 37, 167–190.
Blume, L.E., Brandenburger, A., Dekel, E., 1991. Lexicographic probabilities and choice under uncertainty. Econometrica 59, 61–79.
Bonanno, G., Nehring, K., 1997. Agreeing to disagree: a survey. Mimeo, University of California.
Geanakoplos, J., Polemarchakis, M., 1982. We can't disagree forever. Journal of Economic Theory 26, 363–390.
Grove, A., 1988. Two modellings for theory change. Journal of Philosophical Logic 17, 157–170.
Gul, F., 1998. A comment on Aumann's Bayesian view. Econometrica 66, 923–928.
Heifetz, A., Meier, M., Schipper, B.C., 2012. Unawareness, beliefs and games. Working Paper.
Hellman, Z., 2013. Almost common priors. International Journal of Game Theory 42, 399–410.
Ménager, L., 2012. Agreeing to disagree: a review. Working Paper.
Milgrom, P., Stokey, N., 1982. Information, trade and common knowledge. Journal of Economic Theory 26, 17–27.
Monderer, D., Samet, D., 1989. Approximating common knowledge with common beliefs. Games and Economic Behavior 1, 170–190.
Morris, S., 1995. The common prior assumption in economic theory. Economics and Philosophy 11, 227–253.
Samet, D., 1990. Ignoring ignorance and agreeing to disagree. Journal of Economic Theory 52, 190–207.