Agreeing to Disagree with Lexicographic Prior Beliefs

Christian W. Bach and Andrés Perea

Department of Quantitative Economics
Maastricht University
6200 MD Maastricht
The Netherlands

c.bach@maastrichtuniversity.nl a.perea@maastrichtuniversity.nl

Abstract. The robustness of Aumann's seminal agreement theorem is considered. A more detailed agent model is introduced where posteriors are formed on the basis of lexicographic priors. 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 posteriors. Hence, contrary to the conclusions of Aumann's original impossibility result, agents can actually agree to disagree.

Keywords: agreeing to disagree, agreement theorems, common prior assumption, interactive epistemology, lexicographic beliefs.

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 posterior beliefs cannot be common knowledge among Bayesian agents with a common prior belief. 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. For instance, 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, and Samet (1990) establishes the agreement theorem in a weakened epistemic model without negative introspection. Moreover, Bonanno and Nehring

(1997) provide a rather comprehensive survey on further works on the agreement theorem.

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). 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 priors. Here, we slightly weaken this assumption and then analyze the robustness of Aumann's theorem in such a marginally perturbed context.

First of all, we 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. Moreover, we provide an agreement theorem for lexicographic beliefs. For this theorem to obtain, 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. The non-robustness of Aumann's agreement theorem as well as of its lexicographic generalization considerably weakens its conclusion of the impossibility of agreeing 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 $p : \Omega \rightarrow [0,1]$ 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 $p : \mathcal{P}(\Omega) \rightarrow [0,1]$ by setting $p(E) = \sum_{\omega \in E} p(\omega)$. In this context, it is supposed that each information set of each agent 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 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 $\bigwedge_{j=1}^{n} \mathcal{P}_j$. Moreover, given $x \in S$, the cell of the meet $\bigwedge_{j=1}^{n} \mathcal{P}_j$ containing x is denoted by $\bigwedge_{j=1}^{n} \mathcal{P}_j(x)$. 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 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 $p_{Alice} : \Omega \rightarrow [0, 1]$ 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 $p_{Bob} : \Omega \rightarrow [0, 1]$ 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.

The preceding example illustrates that two agents can entertain absolutely opposing posterior beliefs, despite being equipped with arbitrarily close prior beliefs and their posterior beliefs being common knowledge. Hence, agents can indeed agree to disagree. Moreover, the slight perturbation of the common prior assumption in Aumann's impossibility result shows that the agreement theorem is not robust. The agreement theorem itself is thus considerably weakened, since it already ceases to hold if one of its central assumptions is only marginally modified.

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

Now the robustness of the agreement theorem is addressed from a more general angle in terms of lexicographic beliefs. First of all, we introduce the notion of lexicographic prior beliefs and use it to replace standard prior beliefs in Aumann structures. Then, we generalize Aumann's theorem to lexicographic beliefs. This generalization requires a considerable strengthening of the common prior assumption to common lexicographical priors i.e. identical beliefs at all levels. Finally, it is shown that a small perturbation of completely identical lexicographic beliefs at some – even at a lexicographically very deep – level does already yield different posteriors that are common knowledge.

4.1 Extension to Lexicographic Prior Beliefs

Let $\mathcal{A}_l = (\Omega, (\mathcal{I})_{i \in I}, (b_i)_{i \in I})$ be called 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) $\Sigma_{\omega \in \Omega} b_i^k(\omega) = 1$,
(2) for every $\omega \in \Omega$ there exists $k^* \in \{1, 2, \ldots, K\}$ such that $\omega \in supp(b_i^{k^*})$,
(3) $supp(b_i^{k'}) \cap supp(b_i^{k''}) = \emptyset$ 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 world is assigned positive prior probability at some lexicographic level. Intuitively, no possible world is thus excluded a priori, while at the same time some worlds can be considered infinitely more likely than other worlds before any information is received. Farther, according to the third condition any distinct lexicographic levels never allot positive probability to a same world. This criterion seems natural as subsequent lexicographic levels exhibit differences in infinite likeliness and hence a world being in the support of some lexicographic level should not reappear at any deeper lexicographic level. Besides, observe that the second and third condition imply that for every world the lexicographic level k^* according to which it receives positive probability actually is unique. 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 $E \subseteq \mathit{\Omega}$, agent i's lexicographic prior belief in E is given by the sequence $b_i(E) = (b_i^1(E), b_i^2(E), \ldots, b_i^K(E)) = (\mathit{\Sigma}_{\omega \in E} b_i^1(\omega), \mathit{\Sigma}_{\omega \in E} b_i^2(\omega), \ldots, \mathit{\Sigma}_{\omega \in E} b_i^K(\omega))$. With lexicographic prior beliefs Bayesian updating is defined as follows: given an event $E \subseteq \mathit{\Omega}$ 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 $supp(b_i^{k^*}) \cap \mathcal{I}_i(\omega) \neq \emptyset$. Modelling Bayesian agents with lexicographic priors provides very complete as well as plausible agent model. Before any information is received no world is excluded while at the same time some worlds can be considered infinitely more likely than others by agents, and after information is received the agents update the respectively relevant level of their lexicographic prior to form a unique posterior. Note that a common lexicographic prior assumption requires identical prior belief functions at all lexicographic levels for the agents.

4.2 Aumann's Agreement Theorem for Lexicographic Prior Beliefs

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

Proof. Let $\omega' \in \mathit{\Omega}$ such that $\omega' \in CK(\bigcap_{i=1}^n \{\omega \in \mathit{\Omega} : b(E \mid \mathcal{I}_i(\omega)) = \hat{b}_i\})$ and consider agent $i \in I$. First of all, note that, since the meet is coarser than i's possibility partition, 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 $A_i \subseteq \mathit{\Omega}$ 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 $b(E \mid \mathcal{I}_i(\omega'')) = \hat{b}_i$ 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 $\mathit{supp}(b^k) \cap \bigwedge_{i \in I} \mathcal{I}_i(\omega^*) \neq \emptyset$. Then, $b(E \mid \mathcal{I}_i(\omega^*)) \cdot b^k(\mathcal{I}_i(\omega^*)) = b^k(E \cap \mathcal{I}_i(\omega^*))$. Since $b(E \mid \mathcal{I}_i(\omega)) = \hat{b}_i$, it 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 $\sum_{\omega'' \in A_i} b^k(E \cap \mathcal{I}_i(\omega'')) = \hat{b}_i \cdot \sum_{\omega'' \in A_i} b^k(\mathcal{I}_i(\omega''))$. Therefore, $\sum_{\omega'' \in A_i} b^k(E \cap \mathcal{I}_i(\omega'')) = b^{k^*}(\bigcup_{\omega'' \in A_i}(E \cap \mathcal{I}_i(\omega''))) = b^k(E \cap \bigcup_{\omega'' \in A_i} \mathcal{I}_i(\omega'')) = b^k(E \cap \bigwedge_{i=1}^n \mathcal{I}_i(\omega'))$ and $\sum_{\omega'' \in A_i} b^k(\mathcal{I}_i(\omega'')) = b^k(\bigcup_{\omega'' \in A_i} \mathcal{I}_i(\omega'')) = b^k\ (\bigwedge_{i=1}^n \mathcal{I}_i(\omega'))$. Thus, the equation of sums can be written as $b^k(E \cap \bigwedge_{i=1}^n \mathcal{I}_i(\omega')) = \hat{b}_i \cdot b^k\ (\ \bigwedge_{i=1}^n\ \mathcal{I}_i(\omega'))$, thence $\hat{b}_i = \frac{b^k(E \cap \bigwedge_{i=1}^n \mathcal{I}_i(\omega'))}{b^k(\bigwedge_{i=1}^n \mathcal{I}_i(\omega'))}$. Since agent i has also been arbitrarily chosen, $\hat{b}_1 = \hat{b}_2 = \ldots = \hat{b}_K = \frac{b^k(E \cap \bigwedge_{i=1}^n \mathcal{I}_i(\omega'))}{b^k(\bigwedge_{i=1}^n \mathcal{I}_i(\omega'))}$, which concludes the proof. □

From a lexicographic point of view Theorem 1 unveils a considerably 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 highly demanding and somewhat implausible to require agents not only to exhibit an equal perception on the state space in line with their respective primary prior hypotheses but also in line with any revised prior hypotheses they form.

4.3 Relaxing the Common Prior Assumption

We turn towards relaxing the common lexicographic prior assumption. Indeed, it is 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, and 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 $i \in I$, an event E, and a world $\omega \in \Omega$ such that $\omega \in CK(\bigcap_{i \in I}\{\omega' \in \Omega : b_i(E \mid \mathcal{I}_i(\omega')) = \hat{b}_i\})$ and $\hat{b}_i \neq \hat{b}_j$.*

Proof. Let $k \in \{1, 2, \ldots, K\}$ be the first 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 $\mathcal{I}_{i'} = \{\{\omega' \in \Omega : \omega' \in \bigcup_{k' \geq k} \mathit{supp}(b_i^{k'})\}, \{\omega' \in \Omega : \omega' \notin \bigcup_{k' < k} \mathit{supp}(b_i^{k'})\}\}$ for all agents $i' \in I$. Then, $b_i(E \mid \mathcal{I}_i(\omega)) = b_i^k(\omega)$ and $b_j(E \mid \mathcal{I}_j(\omega)) = b_j^k(\omega)$. Now consider event $E = \{\omega\}$ and observe that $b_i(E \mid \mathcal{I}_i(\omega)) = b_i(E \mid \{\omega' \in \Omega : \omega' \in \mathit{supp}(b_i^k)\}) \neq b_j(E \mid \{\omega' \in \Omega : \omega' \in \mathit{supp}(b_i^k)\} = b_j(E \mid \mathcal{I}_j(\omega))$. Let $\hat{b}_i$ denote the particular values of i's posterior belief 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 $\bigwedge_{i' \in I} \mathcal{I}_{i'} = \mathcal{I}_{i'}$, 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, $\omega \in CK(\bigcap_{i'\in I}\{\omega' \in \Omega : b_{i'}(E \mid \mathcal{I}_{i'}(\omega'))\})$, which concludes the proof. □

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

Next, the robustness of agreeing to disagree with lexicographic beliefs is scrutinized. 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.

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

Proof. Consider the set of all possible worlds $\Omega = \{\omega_1, \omega_2, \dots, \omega_{k^*}, \omega_{k^*+1}, \omega_{k^*+2}\}$, the event $E = \{\omega_{k^*+1}\}$, the possibility partitions $\mathcal{I}_{Alice} = \mathcal{I}_{Bob} = \{\ \{\ \omega_1,\ \omega_2,\ \dots,\ \omega_{k^*-1}\ \},\ \{\ \omega_{k^*},\ \omega_{k^*+1}\ \},\ \{\ \omega_{k^*+2}\ \}\ \}$, as well as two lexicographic prior belief functions $b_{Alice} = (b^1_{Alice}, b^2_{Alice}, \dots, b^{k^*}_{Alice})$ and $b_{Bob} = (b^1_{Bob}, b^2_{Bob}, \dots, b^{k^*}_{Bob})$ 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' ϵ-close lexicographic prior beliefs at level k^* be given by $b^{k^*}_{Alice}(\omega_{k^*}) = \epsilon$, $b^{k^*}_{Alice}(\omega_{k^*+1}) = 0$, and $b^{k^*}_{Alice}(\omega_{k^*+2}) = 1 - \epsilon$, 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. Recall $E = \{\omega_{k^*+1}\}$ 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 $\bigwedge_{i\in\{Alice,Bob\}}\mathcal{I}_i = \mathcal{I}_i$, it follows that $\bigwedge_{i\in\{1,2,\dots,n\}}\mathcal{I}_i(\omega_{k^*}) = \{\omega_{k^*}, \omega_{k^*+1}\} = \{\omega' \in \Omega : b_{Alice}(E \mid (\mathcal{I}_{Alice}(\omega')) = 1\} \cap \{\omega' \in \Omega : b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega')) = 0\}$, and hence $\omega_{k^*} \in CK(\{\omega' \in \Omega : b_{Alice}(E \mid (\mathcal{I}_{Alice}(\omega')) = 1\} \cap \{\omega' \in \Omega : b_{Bob}(E \mid \mathcal{I}_{Bob}(\omega')) = 0\})$, which concludes the proof. □

The preceding theorem illustrates that Aumann's impossibility result is also not robust with lexicographic 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.

Since the agreement theorem's consequences are not preserved at the limit, this non-robust result can be critically regarded. In other words, the possibility results for agreeing to disagree in line with Example 1, Theorem 2 and Theorem 3 can be interpreted as objections to Aumann's conclusion that it is impossible for agents to agree to disagree. Farther, in case of a more precise and arguably more natural agent model with various lexicographically ordered prior hypotheses on the state space, Theorem 1 shows that a considerable and somewhat implausible strengthening of the common prior to a common lexicographic prior assumption is needed to maintain the impossibility of agents to agree to disagree.

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 induce a critical stance towards Aumann's non-robust impossibility theorem.

The analysis of robustness of Aumann's agreement theorem given here could be applied to other assumptions of the theorem. For instance, a replacement of common knowledge of the posteriors by some notion of approximate common knowledge can be considered for future work. In a more general sense, the robustness of game-theoretic solution concepts that depend on the common prior assumption could be considered such as Aumann's (1987) Bayesian foundation for correlated equilibrium.

References

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

Aumann, R. J. (1987): Correlated Equilibrium as an Expression of Bayesian Rationality. *Econometrica* 55, 1–18.

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

Geanakoplos, J. and Polemarchakis, M. (1982): We Can't Disagree Forever. *Journal of Economic Theory* 26, 363–390.

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.