Strategic Manipulation in Bayesian Dialogues*

Christina Pawlowitsch†

June 23, 2021

Abstract

In a *Bayesian dialogue* two individuals report their Bayesian updated belief about a certain event back and forth, at each step taking into account the additional information contained in the updated belief announced by the other at the previous step. Such a process, which operates through a reduction of the set of possible states of the world, converges to a commonly known posterior belief, which can be interpreted as a dynamic foundation for Aumann's agreement result. Certainly, if two individuals have diverging interests, truthfully reporting one's Bayesian updated belief at every step might not be optimal. This observation could lead to the intuition that always truthfully reporting one's Bayesian updated belief were the best that two individuals could do if they had perfectly coinciding interests and these were in line with coming to know the truth. This article provides an example which shows this intuition to be wrong. In this example, at some step of the process, one individual has an incentive to deviate from truthfully reporting his Bayesian updated belief. However, not in order to hide the truth, but to help it come out at the end: to prevent the process from settling into a commonly known belief—the "Aumann conditions"—on a certain subset of the set of possible states of the world (in which the process then would be blocked), and this way make it converge to a subset of the set of possible states of the world on which it will be commonly known whether the event in question has occurred or not. The strategic movement described in this example is similar to a *conversational implicature*: the correct interpretation of the deviation from truthfully reporting the Bayesian updated belief thrives on it being common knowledge that the announced probability cannot possibly be the speaker's Bayesian updated belief at this step. Finally, the argument is embedded in a game-theoretic model.

Keywords: Common knowledge – common belief – convergence of beliefs – Bayesian implicatures.

*I would like to thank Robert Aumann and Herakles Polemarchakis for critical comments and discussions.

†LEMMA–Laboratoire d'Économie Mathématique et de Microéconomie Appliquée, Université Panthéon-Assas, Paris II; christina.pawlowitsch@u-paris2.fr

1 Introduction

In a "Bayesian dialogue" (Bacharach, 1979; Geanakoplos and Polemarchakis, 1982), two individuals, who assign the same prior distribution to some random variable and then receive private information about that distribution, communicate their posterior distribution back and forth—thereby successively updating their posterior distribution in a Bayesian-rational way—until the process has reached an absorbing state, in which the two individuals will have reached consensus about the posterior distribution. Aumann (1976) suggests that such a process, which he illustrates by an example, can be seen as a dynamic foundation for his "agreement result."

Aumann's (1976) "agreement result" says the following: If two individuals impose the same prior probability on the set of all possible states of the world Ω and if, after realization of the true state of the world (in virtue of the common knowledge of the prior probability and individuals' information partitions), it is common knowledge that the posterior probability that individual 1 attributes to a certain event A is q_1 and that the posterior probability that individual 2 attributes to A is q_2, then: $q_1 = q_2$, that is, the two posteriors have to be equal.

Both Bacharach (1979) as well as Geanakoplos and Polemarchakis (1982) provide dynamic foundations for Aumman's result, in the sense that the result holds in a nonvacuous way at the absorbing state of the respective process. While Bacharach works with the more general framework of conditional distributions over the set of all possible states of the world, Geanakoplos and Polemarchakis study a dynamic process of belief revision that stays closer to Aumman's original model: in their model, two individuals who receive private information in the form of a finite information partition communicate their Bayesian posterior belief about a certain event back and forth (thereby making their posterior at that step common knowledge) until their respective posteriors at the next step will already be common knowledge (without the announcement needed), and as a consequence—this is Aumann's result—will be equal. Geanakoplos and Polemarchakis refer to such a process as one of *indirect communication*, as opposed to a process of *direct communication*, by which they refer to a scenario in which the two individuals directly reveal to each other the information that they have received according to their information partition. Nielsen (1984) extends on Aumann's result and the dynamic scenario studied by Geankoplos and Polemarchakis by using the more general framework of sigma algebras, instead of partitions, for modeling knowledge. There is continued interest in Bayesian dialogues as, for instance, recent work by Polemarchakis (2016) and Di Tillio, Lehrer, and Samet (2021) demonstrates.[1]

[1]Formal models of belief formation based on Bayesian updates first have been introduced by Dalkey (1969) and DeGroot (1974). The idea, in its rudimentary form, dates back much further into the history of ideas. It can be found, somewhat implicitly, in the kind of riddles often referred to as *three-hat problems*, such as the riddle of the three prisoners (mentioned by Lacan, 1945; see also Billot, 2007, and Cléro, 2008) or the riddle of the three ladies with dirty faces (mentioned by Littlewood 1953, 3–4), which in an oral tradition probably have been around for

1.1 Always truthfully speak Bayesian?

In a Bayesian dialogue, individuals *by assumption* always truthfully report their Bayesian updated belief. In the literature on Bayesian dialogues, this assumption is not explicitly grounded in some other theory or model.

It is easy to see that truthfully reporting one's Bayesian updated belief might *not* be the best to do at every step of an exchange of probability assessments in case that the two individuals have divergent interests. The Covid-19 crisis provides numerous real-life examples of this. Think, for instance, of the question of vaccine effectiveness. Some might have private information that leads to an updated belief that a given vaccine is $q\%$ effective. But the holder of this belief might have an interest in not making it known that they do hold this belief, knowing that this information would lead others to update their belief in a certain way, which might not be in their interest.[2]

From this observation one could easily gain the intuition that always truthfully reporting one's Bayesian updated belief would be the best that "rational" individuals could do, or ought to to, if they had coinciding interests and these were in line with coming to know the truth about the event in question. This article provides an example which shows this intuition to be wrong.

1.2 The key to understanding a Bayesian dialogue

What is observed in a Bayesian dialogue is the sequence of announced probabilities. However, in the background—or more precisely in the theorized mind of the parties engaged in the conversation—a Bayesian dialogue operates through a *reduction process of the set of all possible states of the world* Ω. At each step, with the announcement of the Bayesian updated belief of one of the two parties, it becomes *common knowledge* that a certain subset of the set of possible states of the world Ω *cannot* contain the true state of the world, and this subset of Ω hence can be deleted from Ω *in common knowledge*.

The key to understanding a Bayesian dialogue is in how this reduction process relates to the static conditions that characterize Aumann's result; it is specifically in the understanding that:

(1) the conditions that characterize Aumann's result are what defines an absorbing state of this process, and that

(2) there is *not* necessarily a unique subset of Ω on which these conditions are satisfied, but potentially many and that which of these will be reached depends on who of the two parties

much longer. The term "Bayesian dialogue" is due to Bacharach (1979). On the formal treatment of processes of belief revision based on Bayesian updates see also Fagin et al. (1995 [1988]).

[2]See, for example, the article in the *The New York Times* from March 23, 2021 (online edition, consulted March 24, 2021), "U.S. Health Officials Question AstraZeneca Vaccine Trial Results," which reports: "While AstraZeneca said on Monday that its vaccine appeared to be 79 percent effective at preventing Covid-19, the panel of independent experts said the actual number may have been between 69 percent and 74 percent."

starts the process.

On these different subsets of Ω on which the Aumann conditions hold, potentially different commonly known Bayesian posteriors might arise. This has a number of consequences, among others that the outcome of a Bayesian dialogue can depend on the order of communication and possibly also on whether some information from outside is injected into the process somewhere along the way (both of these phenomena recently have been addressed by Polemarchakis 2016). But more generally, it makes a Bayesian dialogue vulnerable to strategic manipulation. And this not only when the two individuals have diverging interests, but also when they have perfectly common interests, for of course it can be that the two jointly prefer the commonly known Bayesian posterior that arises on some subset of Ω on which the Aumann conditions hold over the commonly known Bayesian posterior that arises on some other subset of Ω on which the Aumann conditions hold as well. This is what the example presented here exploits.

1.3 The incentive to deviate

In the example presented, at some step of the process, one of the individuals has an incentive to deviate from truthfully reporting his Bayesian updated belief. However, not in order to deceive the other, but in order to "keep the process going," that is, to prevent the process from settling into the Aumann conditions on a certain subset of Ω, on which they will not have certainty about the event in question, and thereby make it converge to another subset of Ω on which the Aumann conditions hold as well but on which they will have certainty about the event in question.

1.4 Bayesian Implicatures

The specific deviation from truthfully stating one's Bayesian updated belief that is investigated in this example thrives on the assumption that under usual conditions individuals do truthfully state their Bayesian updated belief. The announcement of the individual who deviates from truthfully stating his Bayesian updated belief is interpreted as such—and indeed only can be interpreted as such—because it is *commonly known* between the two individuals that the announced probability *cannot* possibly be the speaker's truthful Bayesian update at this step. This is similar to a *conversational implicature* (Grice, 1975), which operates on the principle that the meaning of a speech act is generated by the flouting of certain conversational maxims (such as: Try to make your contribution one that is true. Be relevant. Be perspicuous.). This observation suggests a particular line of defense of the Bayesian paradigm in the study of dialogues: the importance of the assumption that individuals do truthfully state their Bayesian updated belief is not so much in that they always would have an interest to do so, much less would do so, but in that the assumption *serves to correctly interpret deviations from that rule.*

1.5 Game-theoretic model

Finally, the argument is embedded in a formal game-theoretic model. This step is essential, because it makes it necessary to define explicitly what are the potential alternatives to the behavioral strategy of always truthfully reporting one's Bayesian updated belief and how deviations from that rule shall be evaluated. A simple model with two behavioral types, *Bayesian* and *strategically Bayesian*, is presented. The *strategically Bayesian* differs from the *Bayesian* in the following way: When the process has boiled down to a subset of Ω on which the Aumann conditions hold but there is no certainty about the event in question (the commonly known posterior is some number strictly between 0 and 1), then, instead of announcing his Bayesian posterior at the current step (which would bring no new information because this posterior is already common knowledge), the *strategically Bayesian* announces the Bayesian posterior that he had in the first place, that is, given the original Ω before any states have been deleted; and, accordingly, when hearing the other announce a probability of which it is commonly known that it cannot be the speaker's Bayesian posterior at the current step, the *strategically Bayesian* interprets it as the other's Bayesian posterior given the original Ω. In this game, the strategy profile in with both individuals act according to the *Bayesian* type is a Nash equilibrium. However, as will be shown, it fails to be *strategically stable* (Kohlberg and Mertens, 1986).

2 Preliminaries: Aumann's result and dialogues

Let $(\Omega, \mathscr{B}, p)$ be a probability space: Ω the set of possible states of the world, $\mathscr{B}$ a σ-algebra on Ω, and p the prior probability distribution defined on $(\Omega, \mathscr{B})$. Furthermore, let there be two individuals, 1 and 2, who, on the one hand, impute the same prior probability, which is given by p, to the events in $\mathscr{B}$, but who, on the other hand, have access to private information regarding the true state of the world. For each individual $i \in \{1, 2\}$, his or her private information is given by a finite partition $\mathscr{P}_i$ of Ω, that is, a finite set

$$\mathscr{P}_i = \{P_{i1}, P_{i2}, \ldots, P_{ik}, \ldots, P_{iK_i}\}$$

of nonempty subsets of Ω, the *classes* of the partition, such that:

(a) each pair $(P_{ik}, P_{ik'})$, $k \neq k'$, is disjoint and

(b) $\bigcup_k P_{ik} = \Omega$.

The partition $\mathscr{P}_i$ models individual i's information in the following sense: when $\omega \in \Omega$ is the true state, the individual characterized by $\mathscr{P}_i$ will learn that one of the states that belong to the class of the partition $\mathscr{P}_i$ to which belongs ω, which shall be denoted by $P_i(\omega)$, has materialized. In order to guarantee that the classes P_{ik} of the partition $\mathscr{P}_i$ are measured by p, we suppose, of

course, that they belong to $\mathscr{B}$. With this interpretation, if ω is the true state and $P_i(\omega) \subset A$, that is, $P_i(\omega)$ *implies* A, then individual i (at state ω) "knows" that event A has happened. Following Aumann (1976), we assume that the prior p defined on $(\Omega, \mathscr{B})$ as well as the information partitions of the two individuals, $\mathscr{P}_i$, $i \in I = \{1, 2\}$, are *common knowledge* between the two individuals.

More generally, if individual i is Bayesian rational, then for any event A that belongs to the σ-algebra defined on Ω, after realization of the true state of the world, i can calculate the posterior probability of A given the information provided by the partition $\mathscr{P}_i$, that is, the conditional probability of A given that the true state belongs to $P_i(\omega)$:

$$q_i = p(A \mid P_i(\omega)) = \frac{p(A \cap P_i(\omega))}{p(P_i(\omega))}.$$

The probability attributed to an event is also called a *belief*. In this terminology, $p(A)$ is the *prior belief of* A, which by assumption is common knowledge between the two individuals, and $p(A \mid P_i(\omega))$ the *posterior* or *updated belief* that i attributes to A given the information received through his partition.

2.1 Aumann's result

An event is *common knowledge* between two individuals if not only both know it but also both know that the other knows it and that both know that the other knows that they both know it, ad infinitum (Lewis, 1969). To capture this notion within a set-theoretic framework that relies on the notion of a state of the world, it turns out to be useful—and having established this is one of Aumann's achievements—to consider the *meet* of the two partitions.

Definition 1 ("meet") Let $\mathscr{P}_1$ and $\mathscr{P}_2$ be two finite partitions of Ω. The *meet* of $\mathscr{P}_1$ and $\mathscr{P}_2$, denoted by $\hat{\mathscr{P}} = \mathscr{P}_1 \wedge \mathscr{P}_2$, is the *finest common coarsening* of $\mathscr{P}_1$ and $\mathscr{P}_2$, that is, the finest partition of Ω such that, for each state $\omega \in \Omega$,

$$P_i(\omega) \subset \hat{P}(\omega), \quad \forall i \in I = \{1, 2\},$$

where $\hat{P}(\omega)$ is the class of the meet to which belongs ω.

Lemma (Aumann, 1976) An event $A \subset \Omega$, at state ω, is *common knowledge* between individuals 1 and 2 in the sense of the recursive definition (Lewis, 1969) if and only if $\hat{P}(\omega) \subset A$.

Note that the fact that individual i attributes to a certain event A a certain posterior $q_i \in [0, 1]$ is itself an event: it corresponds to the union of all information classes of i's partition that lead to this posterior $q_i \in [0, 1]$.

Proposition (Aumann, 1976) Let $(\Omega, \mathscr{B}, p)$ a probability space, $\mathscr{P}_1$ and $\mathscr{P}_2$ two finite partitions of Ω, measurable with respect to $\mathscr{B}$, which represent the information accessible to individual

1 and 2, all of this being common knowledge between the two individuals. Let furthermore $A \in \mathscr{B}$ be an event. If at state ω (in virtue of the common knowledge of the prior probability p and the information partitions) the posteriors q_1 and q_2 that the individuals attribute to A are common knowledge, then they have to be equal: $q_1 = q_2$.

2.2 The Aumann conditions

It is worthwhile to review the proof of Aumann's result because it contains an important characterization that is crucial for understanding how the result relates to Bayesian dialogues.

The proof of the theorem can be understood as being composed of three steps. Step 1, which is conceptually the most important one, consists in establishing that common knowledge of q_i implies that for any information class of $\mathscr{P}_i$ that is contained in the information class of the meet to which belongs the true state, $P(\omega)$, the conditional probability of A has to be equal to q_i:

$$q_i = \frac{p(A \cap P_i(\omega))}{p(P_i(\omega))} \quad = \quad \frac{p(A \cap P_{ik})}{p(P_{ik})}, \quad \forall\, P_{ik} \subset \hat{P}(\omega). \tag{1}$$

The condition has to be true, for otherwise there would be some level of knowledge at which q_i would not be known, and so common knowledge of q_i would break down.

Step 2: From (1) and the fact that the classes of i's partition induce a partition of $\hat{P}(\omega)$, one obtains that:

$$q_i = \frac{p(A \cap \hat{P}(\omega))}{p(\hat{P}(\omega))}, \tag{2}$$

that is, the posterior attributed to A given $P_i(\omega)$, which is denoted by q_i, has to be equal to the posterior probability of A given $\hat{P}(\omega)$, that is, the element of the meet to which belongs ω. To see why (2) holds, note that from (1), one gets

$$q_i\, p(P_{ik}) \quad = \quad p(A \cap P_{ik}), \qquad \forall\, P_{ik} \subset \hat{P}(\omega),$$

which, by summing over all $P_{ik} \subset \hat{P}(\omega)$, gives

$$q_i \sum_{P_{ik} \subset \hat{P}(\omega)} p(P_{ik}) \quad = \quad \sum_{P_{ik} \subset \hat{P}(\omega)} p(A \cap P_{ik}).$$

Since the elements P_{ik} of i's partition are disjoint and the union over all P_{ik}-s that are contained in $\hat{P}(\omega)$ gives $\hat{P}(\omega)$, by the σ–additivity of p, one gets

$$q_i\, p(\hat{P}(\omega)) \quad = \quad p(A \cap \hat{P}(\omega)),$$

which, by rearranging terms, gives (2).

Step 3: From the fact that (1) and (2) have to hold for each i, one obtains

$$q_1 = \frac{p(A \cap \hat{P}(\omega))}{p(\hat{P}(\omega))} = q_2, \tag{3}$$

that is, $q_1 = q_2$, which concludes the proof.[3]

For later reference it is useful to isolate the following condition contained in the proof. Combining (1) and (2), one gets:

$$q_i = \frac{p(A \cap P_i(\omega))}{p(P_i(\omega))} \quad = \quad \frac{p(A \cap P_{ik})}{p(P_{ik})} = \frac{p(A \cap \hat{P}(\omega))}{p(\hat{P}(\omega))} \quad \forall P_{ik} \subset \hat{P}(\omega), \quad \forall\, i \in I \tag{4}$$

That is: for each i, the posterior attributed to A given $P_i(\omega)$ has to be equal to:

(1) the posterior probability of A given *any* of the classes P_{ik} of i's partition that are contained in the class of the meet to which belongs the true state of the world $\hat{P}(\omega)$, and has to be equal to

(2) the posterior probability of A given $\hat{P}(\omega)$.

I refer to equation (4) as the *Aumann conditions*.

Example 1: Aumann's result

This is an example in which Aumann's result hold in a nonvacuous way. Let $\Omega = \{a, b, c, d, e, f\}$ with uniform prior, and

$$\begin{aligned} \mathscr{P}_1 &= \{\{a, b\}, \{c, d\}, \{e\}, \{f\}\}, \\ \mathscr{P}_2 &= \{\{a, c\}, \{b, d\}, \{e, f\}\}. \end{aligned}$$

Suppose that $A = \{b, c\}$ is the event of interest, and $\omega^\star = a$ the true state of the world. Then:

$$\begin{aligned} q_1 &= \frac{p(A \cap P_1(a))}{p(P_1(a))} = \frac{p(\{b, c\} \cap \{a, b\})}{p(\{a, b\})} = \frac{p(\{b\})}{p(\{a, b\})} = \frac{1}{2}, \\ q_2 &= \frac{p(A \cap P_2(a))}{p(P_2(a))} = \frac{p(\{b, c\} \cap \{a, c\})}{p(\{a, c\})} = \frac{p(\{c\})}{p(\{a, c\})} = \frac{1}{2}. \end{aligned}$$

The meet of the two partitions is $\hat{\mathscr{P}} = \{\{a, b, c, d\}, \{e, f\}\}$. Hence, the class of the meet that contains the true state of the world is $\hat{P}(a) = \{a, b, c, d\}$. Here, each individual thinks it possible that the other has received as information any of the classes in the other's partition that are included in the class of the meet that contains the true state of the world. However, for both

[3] Note that step 2 relies on the more general fact that if A_k is a sequence of disjoint subsets of Ω and $p(B \mid A_k) = q$ for all k, then $p(B \mid \cup A_k) = q$, which is a simple consequence of the Kolmogorov Axioms. Geanakoplos (1992, 66) points out that this property mimics the *sure-thing principle*. In probability theory, in its more general form, namely that if $\mathscr{H}$ is a sub-σ algebra of $\mathscr{G}$, then

$$\mathbf{E}[\mathbf{E}(X \mid \mathscr{G}) \mid \mathscr{H}] = \mathbf{E}[X \mid \mathscr{H}],$$

this property is sometimes referred to as the *Tower Property* (see, for instance, Williams, 1991, p. 88). I would like to thank Mathias Beiglböck and Daniel Toneian for having pointed this out to me.

individuals, for any of those classes, the conditional probability of $A = \{b, c\}$ leads to the same posterior, because it is also true that

$$\begin{aligned} \frac{p(\{b,c\} \cap \{c,d\})}{p(\{c,d\})} &= \frac{p(\{c\})}{p(\{c,d\})} = \frac{1}{2}, \\ \frac{p(\{b,c\} \cap \{b,d\})}{p(\{b,d\})} &= \frac{p(\{b\})}{p(\{b,d\})} = \frac{1}{2}, \end{aligned}$$

which implies that the posterior beliefs of the two individuals are common knowledge between them. One easily verifies that in this example—as it should be according to the Aumann conditions—one also has:

$$p(\{b,c\} \mid \hat{P}(a)) = \frac{p(\{b,c\} \cap \{a,b,c,d\})}{p(\{a,b,c,d\})} = \frac{p(\{b,c\})}{p(\{a,b,c,d\})} = \frac{1}{2}.$$

Figure 1 illustrates this situation.

$$\begin{aligned} \mathscr{P}_1 &= \{ \overbrace{\underbrace{\{a,\, b\}}_{p(A|\{a,b\})=\frac{1}{2}} \quad \underbrace{\{c,\, d\}}_{p(A|\{c,d\})=\frac{1}{2}}}^{p(A|\{a,b,c,d\})=\frac{1}{2}} , \{e\}, \{f\}\} \\ \mathscr{P}_2 &= \{ \underbrace{\underbrace{\{a,\, c\}}_{p(A|\{a,c\})=\frac{1}{2}} \quad \underbrace{\{b,\, d\}}_{p(A|\{b,d\})=\frac{1}{2}}}_{p(A|\{a,b,c,d\})=\frac{1}{2}} , \{e, f\}\} \end{aligned}$$

Figure 1. The Aumann conditions.

2.3 Geanakoplos et Polemarchakis's scenario of indirect communication—a Bayesian dialogue

In a Bayesian dialogue in the form of Geanakoplos and Polemarchakis's (1982) scenario of indirect communication, two individuals, after having received private information about the true state of the world (as given by a finite partition), turn in turn, communicate their Bayesian updated belief about a certain event back and forth. That is: one individual starts by announcing his posterior given the information received through his information partition; then, the other individual re-calculates his posterior, given the information received through his information partition and in addition to that also taking into account the information that can be extracted from the announcement of the other at the previous step (given that they have common knowledge of the other's information partition); etc. Who of the two individuals starts the process is given exogenously.

At each step, the information contained in the announcement of the updated belief of the other consists in that a certain subset of Ω *cannot* contain the true state of the world: namely the union of all those information classes of the partition of the other that *do not* lead to the updated

probability announced at this step. Since the information partitions are not just knowledge but common knowledge between the two individuals, this subset of Ω can then be discarded from Ω *in common knowledge*.

A Bayesian dialogue hence can be understood as operating through a successive reduction of the set of all possible states of the world Ω in the following way:

- Step 1: The process starts by *discarding all states that are not in the information class of the meet to which belongs the true state of the world*, $\hat{P}(\omega^\star)$.

  Of course, because simply by receiving information through their partitions, and thanks to the common knowledge of these partitions, it becomes common knowledge between the two individuals that any state that is not in the class of the meet to which belongs the true state of the world cannot be the true state of the world. (See Example 1 for an illustration: If, for instance, a is the true state of the world, just by receiving their private information and thanks to the common knowledge of their information partitions, without anything being said, it is immediately common knowledge between the two individuals that the true state of the world can neither be e nor f.)

- Step t: Then, at each step t, given that the information partitions are common knowledge between the two individuals, with the announcement of the Bayesian updated posterior of the individual whose turn it is to talk at this step, it becomes common knowledge between the two individuals that the union of all those partition classes of the individual who has just announced his posterior that *do not lead to that posterior* can be discarded from Ω *in common knowledge*.

The condition that terminates this process—and this is important—is not that the two individuals announce the same posterior, but that a subset of Ω is reached such that the announcement of the posterior of any of the two individuals does not allow them to discard any more states—that is, a subset of Ω on which the posteriors are already common knowledge (by force of the common knowledge of the information partitions induced by that subset of Ω), and hence, by Aumann's result, will be equal.

More formally: Let $\Omega_0 = \Omega$. Then:

- Step 1: $\Omega_1 = \hat{P}(\omega^\star)$, where $\omega^\star$ is the true state of the world.
- Step t: $\Omega_t = \Omega_{t-1} \backslash \bar{\mathscr{P}}_{i(t-1),t-1}$, where

$$\bar{\mathscr{P}}_{i(t),t} = \bigcup_{i(t),k} P_{i(t),k}, \text{ such that } P_{i(t),k} \in \mathscr{P}_{i(t)} \text{ and } \frac{p(A \cap P_{i(t),k} \cap \Omega_t)}{p(P_{i(t),k} \cap \Omega_t)} \neq q_{i(t),t},$$

$$q_{i(t),t} = \frac{p(A \cap P_i(\omega) \cap \Omega_t)}{p(P_i(\omega) \cap \Omega_t)}$$

with $i(t)$ given by the sequence $1, 2, 1, 2, \ldots$ if individual 1 starts, and by $2, 1, 2, 1 \ldots$ if individual 2 starts.[4]

It can be shown that this process, in a finite number of steps, converges to a situation in which the posteriors are common knowledge and hence—by Aumann's result—identical (Geanakoplos and Polemarchakis, 1982). In that sense, such a process can be interpreted as a dynamic foundation of Aumann's result.

The set Ω_t has a meaningful interpretation: it is the union of all those states of which it is *not commonly known at step n that they cannot be the true state of the world*; in other words, the set of states that *cannot be excluded in common knowledge*. In echoing Genakopolos and Polemarchakis (1982, p. 196), I shall refer to this set as the *fund of common knowledge* at step t.

The fund of common knowledge at step t, Ω_t, plays a similar role here as the *common ground* in the philosophy and study of language. As Stalnaker (2002, 704), for instance, says: "The common ground is just what a speaker presupposes to be common or mutual belief. The common beliefs of the parties of a conversation are the beliefs they share, and that they recognize that they share." In Stalnaker's account, similarly as to what happens in a Bayesian dialogue, the common ground also evolves during a conversation.

2.4 Properties of a Bayesian dialogue

A Bayesian dialogue is stopped by the Aumann conditions. Of course, once a Bayesian dialogue has settled into the Aumann conditions, it will stay there forever, because the posteriors will be common knowledge—solely in virtue of the common knowledge of the information partitions on the reduced set of possible states of the world (the fund of common knowledge) at the given step of the process. One observation is immediate: If the Aumann conditions are already satisfied on the original Ω, then the process stops immediately at step 1, or to say it more correctly, will have reached its absorbing state at step 1. In Example 1, for instance, imagine that individual 1 starts a Bayesian dialogue by announcing that her posterior is $1/2$. Individual 2 cannot learn

[4]This way of defining the reduction process that operates in the back of a Bayesian dialogue differs slightly from the definition originally given by Geanakoplos and Polemarchakis (1982). The difference is in the first step: Geanakoplos and Polemarchakis *do not* start the process by first restricting the set of possible states of the world to the element of the meet to which belongs the true state of the world $\hat{P}(\omega^\star)$. Instead, they carry from one step to the next the union of all those classes of the partition of the individual who talks at step t that remain compatible with the announcement made by that individual at step t (including those that are not in the class of the meet to which belongs the true state). Note however that this set possibly contains states of which it has already become common knowledge that they cannot be the true state of the world, namely, the intersection of this set with the set of all those states that do not belong to the element of the meet that contains the true state of the world. It seems to me therefore more coherent, in terms of the interpretation, to restrict the set Ω first to the class of the meet in which lies the true state of the world. In terms of the generated dialogue—the sequence of announced probabilities—this difference in the definition is irrelevant.

anything from that announcement, because for any class of individual 1's information partition of which individual 2 thinks it possible that the true state lies in it, individual 1 would come up with the same posterior. Individual 2 therefore cannot eliminate any state from the set of possible states of the world. And importantly, the fact that this is so is common knowledge between the two individuals. Similarly when the roles of individual 1 and 2 are reversed.

There is no regularity in the announced probabilities (Polemarchakis, 2016). The visible trace of a Bayesian dialogue is the sequence of announced posteriors. It can be that at that level "nothing happens" for a certain number of rounds, in the sense that each individual, on his side, repeats the same posterior (of which it is not commonly known that this is the posterior of that individual at that step), while in the background, nevertheless, the two individuals—in common knowledge—successively discard possible states of the world, until Ω has been reduced to a set on which they suddenly agree on a commonly known posterior. Geanakoplos and Polemarchakis (1982, p. 179) demonstrate this property by a parametric example, which they attribute to Aumann. Polemarchakis (2016) has recently addressed the more general question whether there is any pattern in the sequence of announced probabilities that stems from a Bayesian dialogue. And he shows that there isn't: he shows that for any sequence of numbers strictly between 0 and 1, one can find a set Ω of possible states of the world and two partitions such that the given sequence is the visible trace of a Bayesian, or as Polemarchakis says in this context, "rational dialogue."

A Bayesian dialogue depends on the order in which the two individuals announce their updated Bayesian posteriors (Polemarchakis, 2016). Depending on whether it is individual 1 or individual 2 who starts the process by announcing his or her posterior (understanding that from then on they will do that in an alternating manner), the process can end with two different subsets of Ω, on each of which the Aumann conditions hold. But, on these two different subsets of Ω (reached as a function of which individual starts the process), different commonly known posteriors attributed to A might arise. Example 2, which builds on an example given by Polemarchakis, provides an illustration.

Example 2: Bayesian dialogues

Let $\Omega = \{a, b, c, d, e, f, g, h, i, j, k\}$ the set of possible states of the world, endowed with uniform prior probability, that is, $p(\omega) = 1/11$ for all possible states of the world and

$$\begin{aligned} \mathscr{P}_1 &= \{\{a, b, c, d, e, f\}, \{g, h, i, j, k\}\}, \\ \mathscr{P}_2 &= \{\{a, b, g, h\}, \{c, d, i, j\}, \{e, f, k\}\}. \end{aligned}$$

Suppose that $A = \{a, b, i, j, k\}$ is the event of interest, and $\omega^\star = a$ the true state of the world.[5]

[5]The example is derived from an example given by Polemarchakis (2016, 12) by transposing the latter into a

In matrix representation (see Appendix), this example looks as follows:

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\{e, f\}$	$\frac{1}{3}$
$\{g, h\}$	$\{\mathbf{i}, \mathbf{j}\}$	$\{\mathbf{k}\}$	$\frac{3}{5}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	

In the matrix above, rows represent the information classes of individual 1; columns that of individual 2; states belonging to A are indicated in bold face; and the number at the end of each row respectively bottom of each column indicates the conditional probability of A given that row respectively column.

In this example, the meet (the finest common coarsening) of the two partitions is the coarsest partition: $\mathscr{P}_1 \wedge \mathscr{P}_2 = \{\Omega\}$. The join (the coarsest common refinement, see Appendix) is $\mathscr{P}_1 \vee \mathscr{P}_2 = \{\{a, b\}, \{c, d\}, \{e, f\}, \{g, h\}, \{i, j\}, \{k\}\}$. The information class of the meet that contains the true state of the world therefore is Ω. In this example, as in Polemarchakis's original one, the outcome of a Bayesian dialogue depends on the order in which the two individuals report their posteriors.

- *If individual 1 starts*:

  Step 1: $\Omega_0 = \{a, b, c, d, e, f, g, h, i, j, k\}$, $\mathscr{P}_{1,\Omega_0} = \{\{a, b, c, d, e, f\}, \{g, h, i, j, k\}\}$

  $$q_1 \quad = \quad \frac{p(\{a, b, i, j, k\} \cap \{a, b, c, d, e, f\})}{p(\{a, b, c, d, e, f\})} = \frac{p(\{a, b\})}{p(\{a, b, c, d, e, f\})} = \frac{1}{3}$$

  If individual 1 announces 1/3, then it will become common knowledge between the two individuals that the true state cannot belong to the set $\{g, h, i, j, k\}$, and therefore this set should be deleted from what remains in the *fund of common knowledge* at this step, which becomes $\Omega_1 = \{a, b, c, d, e, f\}$. In matrix form:

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\{e, f\}$	$\frac{1}{3}$
1	0	0	

  Step 2: $\Omega_1 = \{a, b, c, d, e, f\}$, $\mathscr{P}_{2,\Omega_1} = \{\{a, b\}, \{c, d\}, \{e, f\}\}$

  $$q_2 \quad = \quad \frac{p(\{a, b\} \cap \{a, b\})}{p(\{a, b\})} = \frac{p(\{a, b\})}{p(\{a, b\})} = 1.$$

  If individual 2 announces 1, then it will be common knowledge between the two individuals that the true state of the world cannot be in $\{c, d, e, f\}$, and hence this set should be deleted

model with a uniform prior on the possible states of the world, which is achieved by replicating some of the states in Polemarchakis's model.

in common knowledge. Certainly, for no matter if 2 had received the information that the true state was in $\{c, d\}$ or if he had received the information that the true state was in $\{e, f\}$, in both cases, he would have announced a posterior of 0, which is not the posterior that he has effectively announced. The matrix hence becomes:

$\{\mathbf{a}^\star, \mathbf{b}\}$	1
1	

Step 3: $\Omega_2 = \{a, b\}$, $\mathscr{P}_{1,\Omega_2} = \{\{a, b\}\}$. Individual 1 announces also "1," and the process has reached its absorbing state. Note that on the set of states that are still alive at step 3, $\Omega_2 = \{a, b\}$, the Aumann conditions are trivially satisfied because the information partitions of the two individuals induced by $\Omega_2 = \{a, b\}$ are identical: $\mathscr{P}_{1,\Omega_2} = \{\{a, b\}\} = \mathscr{P}_{2,\Omega_2}$.

It shall be noted that in this example, the element of the join to which belongs the true state of the world, that is, the information that can be attained in common knowledge if the two individuals directly communicate to each other the information that each one of them has received individually, is also $\{a, b\}$. Direct communication will therefore also lead to a posterior of 1 attributed to A. In other words, if individual 1 starts, they get through indirect communication exactly what can be attained through direct communication.

- *If individual 2 starts*:

  Step 1: $\Omega_0 = \{a, b, c, d, e, f, g, h, i, j, k\}$, $\mathscr{P}_{2,\Omega_0} = \{\{a, b, g, h\}, \{c, d, i, j\}, \{e, f, k\}\}$

  $$q_1 = \frac{p(\{a, b, i, j, k\} \cap \{a, b, g, h\})}{p(\{a, b, g, h\})} = \frac{p(\{a, b\})}{p(\{a, b, g, h\})} = \frac{1}{2}$$

  As a consequence, $\{e, f, k\}$ can be deleted in common knowledge, and $\Omega_1 = \{a, b, c, d, g, h, i, j\}$. But then the matrix is:

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\frac{1}{2}$
$\{g, h\}$	$\{\mathbf{i}, \mathbf{j}\}$	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	

  And the process of deletion ends here, with each of them announcing 1/2 from this moment on, forever. Certainly, the Aumann conditions are satisfied on what remains in the fund of common knowledge, $\Omega_1 = \{a, b, c, d, g, h, i, j\}$.

Figure 2 summarizes by putting the two processes, under the two different orders, next to each other.

If individual 1 starts:

Step 1:

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\{e, f\}$	$\frac{1}{3}$
$\{g, h\}$	$\{\mathbf{i}, \mathbf{j}\}$	$\{\mathbf{k}\}$	$\frac{3}{5}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	

Step 2

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\{e, f\}$	$\frac{1}{3}$
1	0	0	

Step 3

$\{\mathbf{a}^\star, \mathbf{b}\}$	1
1	

If individual 2 starts:

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\{e, f\}$	$\frac{1}{3}$
$\{g, h\}$	$\{\mathbf{i}, \mathbf{j}\}$	$\{\mathbf{k}\}$	$\frac{3}{5}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	

$\{\mathbf{a}^\star, \mathbf{b}\}$	$\{c, d\}$	$\frac{1}{2}$
$\{g, h\}$	$\{\mathbf{i}, \mathbf{j}\}$	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	

Figure 2

The Aumann conditions may hold on several subsets of Ω: the door to strategic manipulability. The order-dependency of a Bayesian dialogue stems from a more general property of the underlying static model, namely that the Aumann conditions can hold on different subsets of Ω on which different commonly known posteriors may arise. But this, over and above its consequences for the relevance of order, more generally makes a Bayesian dialogue fragile with respect to strategic deviations, namely not only when the parties engaged in a dialogue have diverging interests but potentially also *when they have perfectly coinciding interests*. The reason for this, at first sight counterintuitive, phenomenon is simple: If on different subsets of Ω on which the Aumann conditions are satisfied different commonly known posteriors arise, it could be that the two individuals jointly prefer one of these possible posteriors over another. Deviating from truthfully reporting their Bayesian posterior at some step of the process then might allow them to navigate "around" the Aumann conditions: to *prevent the process from settling into the Aumann conditions on a certain subset of* Ω *in order to make it settle into the Aumann conditions on another subset of* Ω on which a commonly known posterior arises that they jointly have a preference for. This is what will be explored in the following section.

3 Embedding the example in a "story": turning it into a game

Suppose that the two individuals appearing in Example 2 are two professional chess players who have been thrown into prison. The director of the prison calls on both of them and announces:

> "Here is Ω, here are your information partitions, here is the prior p and the event A that we are interested in (all as defined above in the presentation of Example 2). One state of the world will materialize and each of you will receive information according to his or her information partition. Then, I will ask one of you, in front of the other: *What is the probability that you attribute to the event A?* After his answer, I will ask the other: *What is the probability that you attribute to the event A?* After his answer, I will turn to the first again and ask: *Now, has the event A occurred or not?* If his answer is correct, then both of you will get free. If it is not correct, both of you will sit for the rest of your lives."

Suppose, as above, that the state of the world that materializes is $\omega^\star = a$, or is picked by the director (whether one or the other does not matter for the purpose of the current investigation). The director hence sends out the information to individual 1 that the true state belongs to the set $\{a, b, c, d, e, f\}$, and to individual 2 the information that the true state belongs to the set $\{a, b, g, h\}$. The director first calls on individual 2, in front of individual 1, and asks his posterior. Individual 2 truthfully says "1/2." The director then calls on individual 1 to step forward and to report her posterior. Individual 1 says ...

Well, what do you think that she says? Realize that after individual 2's announcement (see right-hand panel in Figure 2, step 2), what is left of Ω and the two information partitions induced by that set is common knowledge between the two individuals, which is to say that it is common knowledge between the two individuals that from this moment onward, if they were to state their posteriors truthfully, they would forever be stuck with the answers "1/2" (even if they had as many more rounds to go as they wanted). However, at a second thought, there is a way out of that situation.

Since the original information partitions as well as the announcement of individual 2, at the first step, are common knowledge between the two individuals, it is also common knowledge between the two individuals:

(1) that it is the very announcement of individual 2, at the first step, that has brought them into this situation in which the current updated belief of individual 1, whose turn it is now, has no informational value anymore,

(2) but also that before that—before individual 2's announcement of "1/2" and the ensuing

reduction of the fund of common knowledge—the posterior that individual 1 had then did have some informational value, because it was either 1/3 or 1/5 (see right-hand panel in Figure 2, step 1).

Imagine that you are individual 2 (the one who the director has asked first and who will also be asked at the third step) and that in response to the director's question to individual 1, at the second step, you hear individual 1 say: "1/3." You know that this *cannot* possibly be the truthful Bayesian update of individual 1 after you have made your announcement at the previous step, because you know that individual 1's Bayesian update at this step is 1/2, and in fact this is common knowledge between the two of you. But it is also common knowledge between the two of you that before your announcement at the first step, 1/3 was a possible truthful Bayesian response of individual 1, which corresponds to the fact that individual 1 has received the information that the true state of the world belongs to the set $\{a, b, c, d, e, f\}$ and hence cannot be in $\{g, h, i, j, k\}$ (see right-hand panel in Figure 2, step 1). Now, knowing that individual 2 is highly rational (and does not say "1/3" because she made a mistake in determining her Bayesian posterior after your announcement), you will probably infer that this—that at step 1, her Bayesian posterior was 1/3—is precisely what individual 1 wants to tell you by her announcement of 1/3. What you do, so to say, is to look for some kind of repair, some way to reconcile what you observe (which is commonly known *not* to be a truthful Bayesian response) with strategically rational behavior, and you understand that the states in $\{g, h, i, j, k\}$ can be discarded from the set of possible states of the world, which leaves you with $\{a, b, c, d\}$ as the fund of common knowledge at this step of your "conversation." Combining that with your own information, which is that the true state belongs to the set $\{a, b, g, h\}$, you understand that the true state of the world belongs to $\{a, b\}$ and that the event $A = \{a, b, i, j, k\}$ has hence surely occurred. You announce "1," and both of you get free. Now that was a thought experiment of individual 2 at the last step. If you are individual 1, you understand that this is the way that individual 2 will reason. Anticipating this, you as individual 1, at the second step, announce "1/3."

3.1 A linguistic interpretation: a "Bayesian implicature"

In the story above, the profitable deviation from truthfully stating one's Bayesian updated belief thrives on the fact that *by doing the deviation*, it will become *common knowledge* that the announced probability cannot possibly be the speaker's Bayesian updated belief that this step; in other words, that she has deviated from the rule of truthfully stating her Bayesian updated belief at that step.

Philosophers of language and linguists might recognize in this movement a *conversational implicature* (Grice, 1975): the phenomenon that the meaning of a speech act, here the announced

probability, will be implied by a deviation from some predefined convention how to talk under normal conditions—what Grice calls the "flouting" of a conversational maxim.

Under the name of the "Cooperative Principle," Grice (pp. 45–46) isolates four main conversational maxims, "supermaxims," as he says:

- The maxim of Quantity: "1. Make your contribution as informative as required (for the current purpose of the exchange). 2. Do not make your contribution more informative than is required."
- The maxim of Quality: "Try to make your contribution one that is true."
- The maxim of Relation: "Be relevant."
- The maxim of Manner: "Be perspicuous."

Under the category of quality, Grice places two submaxims: "1. Do not say what you believe to be false." "2. Do not say that for which you do lack adequate evidence." Under the category of manner, Grice places: "1. Avoid obscurity of expression." "2. Avoid ambiguity." "3. Be brief (avoid unnecessary prolixity)." "4. Be orderly."

In the example above, the maxim flouted can be said to be that of quality, which here takes the specific form that one ought to truthfully announce one's Bayesian updated belief at the current state of a conversation. In other words, the maxim to truthfully report one's Bayesian updated belief at the current state of a conversation can be seen as a submaxim of the supermaxim of quality. The implicature comes from it being common knowledge that a deviation from that maxim has occurred because the probability stated cannot possibly be the speaker's truthful Bayesian updated belief at that step. To "flout" a maxim, as Grice (p. 49) explains, is to "blatantly fail to fulfill it." But what can be a more blatantly committed offense than one that is committed *in the face of common knowledge*? We have here indeed a mathematically precise manifestation of a communicative implicature. I propose to call such an implicature that thrives on it being common knowledge that an expressed belief cannot possibly be the truthful Bayesian updated belief of the speaker (at the current state of the conversation), a *Bayesian implicature*.

In addition to that one could bring to bear that the implicature observed in the example above is triggered by a *clash* of maxims (Grice 1975, p. 49). In a situation in which the Bayesian updated beliefs of the two individuals are already common knowledge (given the fund of common knowledge Ω_t at the current step t), reporting one's Bayesian updated belief would amount to making a perfectly *irrelevant* speech act. In other words, there is a clash between the maxim of *quality* and that of *relation*. In the example above, individual 1, when at step 2 she announces her original posterior, can be said to sacrifice *quality* in order to save *relevance*.

4 A game-theoretic analysis

With the story above, the iterated exchange of probabilities silently has been embedded in a "game." By force of the context of the narrative (the prison, getting free or not), the individuals silently have been attributed preferences over the end result of their interaction. Moreover, being asked to announce "the probability that they attribute to event A" does not oblige—much less enable—them to state their Bayesian updated belief as calculated according to the rules of a Bayesian dialogue; in other words, it leaves them some strategic choice.

Still, the story above is not the complete description of a game in the game-theoretic sense, because it has not been explicitly defined what the alternatives to truthfully reporting one's Bayesian updated belief at every step of the process are, and importantly, how to react if the other has reported a probability that cannot possibly be his truthful Bayesian update at this step. To speak game-theoretically: it has not been defined what the strategy sets of players are. This is indispensable though, because without that one cannot fully evaluate deviations from the rule of always truthfully stating one's Bayesian updated belief much less characterize them in terms of solution concepts for games, like Nash equilibrium and refinements thereof.

In this section, I propose a grounding of the above story in a complete game-theoretic model. There are numerous ways of doing this. The model that I suggest is a minimal model, in which players' strategy sets are derived from two behavioral types, *Bayesian* and *strategically Bayesian*.

4.1 The game

Players and rules. The players of this game are the two prisoners. The role of the director of the prison is not as a player but in that his instructions provide the framework for setting up the rules of the game. The rules of the game are: A state of the world materializes. The players, who ex ante do not know which state has materialized, receive private information about the state of the world through their individual information partitions. Then, one player is asked, in front of the other, to state a number between 0 and 1 (which refers to the probability attributed to event A). Then, the other player is asked, in front of the other, to state a number between 0 and 1. Then, the player who has been asked first has to make a binary choice between 0 and 1, corresponding to the answers "A has not occurred" and "A has occurred."

Payoffs. If the answer of the player who is asked at the third step corresponds to the truth value of the proposition at the state of the world that has materialized, then both players will have a payoff of 1 (they get free); if not, they both will have a payoff of 0 (they sit for the rest of their lives).

Strategy sets. Players' strategies are derived from behavioral types that govern how a player

plays this game, that is, according to which rule he or she determines respectively interprets the probabilities stated at each step and how he or she makes her choice between 0 and 1 at the third step:

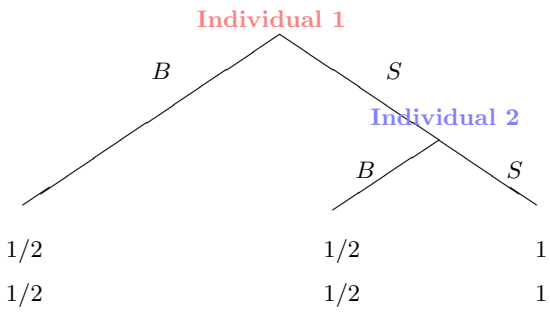
- For step 1, no strategic choice is considered: the player who acts at step 1 is considered to always truthfully report his Bayesian posterior as given by his information partition and the original set Ω.
- The player who acts at step 2, is considered to have two different choices, according to the following behavioral types:
  - "Act as a Bayesian" (B): state your truthful Bayesian updated belief on the assumption that the other has done so at the first step (that is, given your information partition induced by the set Ω from which all those information classes of the other have been subtracted that do not lead to the posterior that the other has stated at step 1).
  - "Act as a "strategically Bayesian" (S): state your truthful Bayesian updated belief on the assumption that the other has done so at the first step—unless, with the assumption that it is common knowledge that the other has truthfully stated his Bayesian posterior at step 1, on what remains of Ω at the current step, the Aumann conditions are satisfied (that is, the Bayesian posterior of each of you is already common knowledge and hence the two are equal). If the latter is the case, then state your original Bayesian posterior, that is, the Bayesian posterior that you had given the original Ω and the information that you have received through your partition.
- For the player who acts at step 3 (who is the one who has already been called on at the first step, and as assumed, there has truthfully stated his Bayesian posterior), two positions, namely conditional on whether the other player, at step 2, has acted as a *Bayesian* or as a *strategically Bayesian*, have to be distinguished. However, a strategic choice shall be considered only for the second: If what the other player has stated at step 2 was a possible truthful Bayesian updated belief of this player at this step (under the assumption that it is common knowledge that what the other has said at step 1 was his truthful Bayesian posterior), then for the player who acts at step 3 no strategic choice shall be considered: he is supposed to act on his truthful Bayesian updated belief at this step. Only if what the other player has stated at step 2 was *not* a possible truthful Bayesian updated belief of this player, at this step, then two strategic possibilities for the player who acts at step 3 are considered:
  - "Act as a Bayesian" (B): act on the posterior that you had at step 1.

– "Act as a "strategically Bayesian" (S): check if the probability announced by the other at the step before was a possible truthful Bayesian posterior for him at step 1. If yes, exploit the information accordingly to reduce Ω (that is, delete from Ω any subset of Ω corresponding to an element in the other individual's information partition that does not lead to the probability that the other has announced at the previous step) and act on your updated belief given your information partition induced by what remains of Ω after that reduction.

Certainly, at the third step, to "act on one's posterior" is to maximize expected utility: that is, state 0 if the posterior that one attributes to A is below $1/2$, and state 1 if the posterior that one attributes to A is above $1/2$. If the posterior attributed to A is equal to $1/2$, then the following tie-breaking rule shall be considered: randomize between 0 and 1, both with probability $1/2$.

4.2 The game tree and the payoff matrix

Under the assumption that the director calls on individual 2 first and considering only those actions of the two individuals that corresponds to strategic choices of the players in the game, this boils down to the following game in extensive form:



To see why this is the case, one has to "play through" all possible sequences of moves, for all possible states of the world for which players have strategic choices. Note first that for the states e, f, and g players have no strategic choice. For any of the other states, the various combinations of strategies always lead to the payoffs indicated at the end nodes of the tree shown above, which is why the game can be represented by this simple extensive-form game:

- When the true state is either a or b:
  - (B, B): If both are "Bayesian," the outcome will just be what can be seen in the right-hand panel of Figure 2: individual 2, at step 3, will attribute to A a probability of $1/2$.

According to the rule assumed, he therefore will state with a probability of 1/2 that A has occurred, and hence both will get free with a probability of 1/2, which gives both of them an expected payoff of 1/2.

- (B, S): If individual 1 is "Bayesian" and individual 2 is "strategically Bayesian," the same sequence of actions as above will be acted out (individual 2, at step 3 would be ready to make the right inference if individual 1 had announced her initial posterior, but because she didn't, but did instead announce her updated posterior at step 2, individual 2 will act on his updated posterior at step 3).
- (S, B) If individual 1 is "strategically Bayesian" and individual 2 is "Bayesian," then individual 1, at step 2, will state her initial posterior of 1/3, but since this is not a possibly truthful Bayesian posterior at this step, individual 2, at step 3, will revert to his initial posterior and will attribute to A a probability of 1/2, which again gives both of them an expected payoff of 1/2.
- (S, S) If both are "strategically Bayesian," what will happen is what is described in the story in the previous section: individual 1, at step 2, will state her initial Bayesian posterior, 1/3. Individual 2, at step 3, will correctly interpret this and take from it the information that the true state cannot belong to $\{g, h, i, j, k\}$, and, combining this with his own information, will know that the true state belongs to $\{a, b\}$ and that event A hence has occurred for sure. He will announce this to the director. Both will get free and will have a payoff of 1.

- When the true state is either i or j: as for the case that the true state is either a or b only that in case of the strategy profiles (S, B) and (S, S), the numerical value of the initial Bayesian posterior of individual 1 will be 3/5 (instead of 1/3).
- When the true state is either c or d: as for the case that the true state is either a or b only that in case of the strategy profile (S, S), individual 2 will announce "0" in the end.
- When the true state is either g or h: as for the case that the true state is either a or b only that in case of the strategy profiles (S, B) and (S, S), the numerical value of the initial Bayesian posterior of individual 1 will be 3/5 (instead of 1/3), and in case of the strategy profile (S, S), individual 2 will announce "0" in the end.

In the form of a payoff matrix the game looks as follows:

		Individual 2	
		B	S
Individual 1	B	$\frac{1}{2},\ \frac{1}{2}$	$\frac{1}{2},\ \frac{1}{2}$
	S	$\frac{1}{2},\ \frac{1}{2}$	$1,\ 1$

4.3 Equilibrium analysis

In this game, both the strategy profile (B, B) and the strategy profile (S, S) are Nash equilibria. However, only the second is *strategically stable*.

Strategy stability is an equilibrium refinement criterion introduced by Kohlberg and Mertens (1986). It is based on the observation that a Nash equilibrium is not necessarily self-enforcing, namely then not if a player by deviating from the strategy that he is supposed to take in the equilibrium under study can make it known to the other player that he is deviating to some other strategy and this way force the other player to change strategy as well, namely taking the strategy that is a best response to that deviation.

The first equilibrium, (B, B), is in particular not stable against the kind of forward-induction test as proposed by Kohlberg and Mertens (1986), which for the game considered here works out like this: Suppose that (B, B) is the convention how to play this game. If individual 2 finds that individual 1 cannot possibly have played according to the strategy "Bayesian" (B) at the previous step (because what she has said was not compatible with that strategy), he will, if he is rational, understand that he can do better by deviating from the strategy that he, by convention, initially was supposed to adopt, namely to speak truthfully Bayesian (B), and will also adopt the strategy "strategically Bayesian" (S). Individual 1, being rational and counting on the rationality of individual 2, can anticipate this, and hence will push individual 2 into that situation by acting as a "strategically Bayesian." In fact, forward induction here boils down to a simple back-ward induction argument (Selten, 1975), which Kohlberg and Mertens retain as a necessary (but not sufficient) condition for strategic stability.

4.4 The game model and the narrative argument

Rephrasing the "story" from the previous section formally as a game, forces the model builder to define what are the strategic alternatives over truthfully stating one's Bayesian updated belief, and most importantly, also how deviations from the principle of always truthfully reporting one's Bayesian update are interpreted (in short, what are the strategy sets of players). But certainly, the solution of the so-generated game will depend on the strategic alternatives that one allows the players in the game to have. It is interesting to see that in the presence of the *strategically Bayesian* type, the movement away from the regular *Bayesian* type (the one who always truthfully states his Bayesian updated belief) does not correspond to a failure of the profile in which both players act according to the *Bayesian* type to constitute a Nash equilibrium, but a failure of this equilibrium to satisfy certain equilibrium refinement criteria. From a game-theoretic point of view this is not surprising. It rather reflects the contingency that being supported by Nash-equilibrium conditions is not enough for the outcome of a game to qualify as self-enforcing. Game theorists

have developed equilibrium refinement concepts in the very aim of sorting out equilibria that do not correspond to ways of playing the game that one would expect rational players to engage in.

On the other hand, there is a part of the narrative argument that is not made visible by the game model: namely *how come* that individual 1's departure from the principle of always truthfully stating one's Bayesian updated belief should entail what it does for the *strategically Bayesian*, namely that what individual 1 has announced instead is her Bayesian posterior on the original Ω. This, however, is not a limitation of the particular game model presented; it rather reflects a limitation of game-theoretic models in general. In a game-theoretic model, strategy sets have to be defined ex ante. A game-theoretic model therefore cannot explain *why* a player would come up with the possibility of using a particular strategy—here: *why* he would make such an inference. A game-theoretic model can only investigate if, in the presence of a given set of possible alternatives, *using* a particular strategy—making such an inference—can be sustained by equilibrium conditions and which further refinements such an equilibrium satisfies (or if it satisfies any other solution concept for games that one finds apt). This is where any game-theoretic model has to rely on input from outside—intuition coming from prospective applications or insight from other disciplines.

5 Interpretation

This study is not meant as a critique of the assumption that individuals do correctly *evaluate* their Bayesian updated belief. It is assumed throughout that individuals always correctly *evaluate* their Bayesian updated belief. What is pointed out here is that when an exchange of probabilities is embedded into a game, the players of this game—even if they have perfectly coinciding interests and these are in line with wanting to learn the truth about a certain event—might not always have an interest to truthfully *report* (reveal) their Bayesian updated belief at every step of a conversation.

The example presented here is meant to make us aware of what *cannot serve*, or at least *is not sufficient*, as a theoretical foundation for the assumption that people always do truthfully state their Bayesian updated belief as required in a Bayesian dialogue: namely that individuals have coinciding interests and that what they want in the end is just to gain certainty about a certain event.

The assumption that the parties engaged in the dialogue always do truthfully state their Bayesian updated belief provides an important theoretical reference point. It serves as a benchmark but it derives its centrality also from the fact that it is quite difficult to define in general what could be another behavioral rule that is applicable for any dialogue of any length. The observation made in the example that the movement away from always truthfully reporting one's Bayesian updated

belief is similar to a conversational implicature suggests a further line of defense of the Bayesian paradigm in the study of dialogues: that the importance of the assumption that individuals always do truthfully state their Bayesian posterior is not so much in the fact that they always would have an interest to do so but in that the assumption serves to correctly interpret deviations from that rule. In other words, that it is not just a theoretical but also a practical reference point that is functional in the practice of ordinary speech.

Appendix

A.1. The join of two partitions

Definition 2 Let $\mathscr{P}_1$ and $\mathscr{P}_2$ two partitions of Ω. The *join* of $\mathscr{P}_1$ and $\mathscr{P}_2$, denoted by $\check{\mathscr{P}} = \mathscr{P}_1 \vee \mathscr{P}_2$, is the *coarsest common refinement* of $\mathscr{P}_1$ and $\mathscr{P}_2$, that is, the coarsest partition of Ω such that, for each $\omega \in \Omega$,

$$\check{P}(\omega) \subset P_i(\omega), \quad \forall i \in I = \{1, 2\},$$

where $\check{P}(\omega)$ is the class of the join to which belongs ω.

The classes of $\check{\mathscr{P}} = \mathscr{P}_1 \vee \mathscr{P}_2$ are obtained by taking for each class of one partition its pairwise intersection with the classes of the other partition (see, for instance, Barbut 1968).[6]
In Example 1, for instance:

$$\check{\mathscr{P}} = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \{f\}\}.$$

The join of two information partitions has an immediate interpretation: it represents the information that the two individuals can have when they directly exchange the information that each one of them has received through his partition (see, for example, Geanakoplos and Polemarchakis 1982). Of course, if 1 says to 2: "I have received the information that the true state of the world belongs to $\{a, b\}$," and 2 says to 1: "I have received the information that the true state of the world belongs to $\{a, c\}$," it will become commonly known between them that the true state of the world must belong to the intersection of the two sets $\{a, b\} \cap \{a, c\} = \{a\}$.

A.2. Matrix representation of the two partitions

Any two finite partitions can be written in the form of a matrix such that

[6] Barbut uses $\mathscr{P}_1 \vee \mathscr{P}_2$ for the meet and $\mathscr{P}_1 \wedge \mathscr{P}_2$ for the join. In view of how these objects are obtained from the individual partitions, it seems to me that this is the better notation. I nevertheless use $\mathscr{P}_1 \vee \mathscr{P}_2$ for the join and $\mathscr{P}_1 \wedge \mathscr{P}_2$ for the meet, because this is the notation employed in Aumann's article and the literature that builds on it.

- the elements of the matrix are occupied by the elements of the join of the two partitions, with possibly some elements of the matrix empty but without any rows or columns completely empty, and
- the information classes of one individual correspond to the rows of the matrix and that of the other individual to the columns of the matrix (see, for instance, Barbut, 1968).

In such a matrix, the classes of the meet of the two partitions appear as the unions of those elements of the join that have the same empty elements along rows as well as columns.

In Example 1:

$\{a^\star\}$	$\{\mathbf{b}\}$		$\frac{1}{2}$
$\{\mathbf{c}\}$	$\{d\}$		$\frac{1}{2}$
		$\{e\}$	0
		$\{f\}$	0
$\frac{1}{2}$	$\frac{1}{2}$	0	

In the figure above, states belonging to A are indicated by boldface (b and c here) and the true state of the world is indicated by a star, $a^\star$. Furthermore, for each row (information class of individual 1), to the right of the vertical line, appears the conditional probability of A given that row; and, for each column (information class of individual 2), below the horizontal line, appears the conditional probability of A given the column.

References

[1] Aumann, R. J. (1976). Agreeing to disagree, *The Annals of Statistics* 4, 1236–1239.

[2] Bacharach, M. (1979). Normal bayesian dialogues. *Journal of the American Statistical Association* 74, 837–846.

[3] Barbut, M. (1968). Partitions d'un ensemble fini: leur treillis (cosimplexe) et leur représentation géométrique. *Mathématiques et Sciences Humaines* 22, 5–22.

[4] Billot, A. (2007). Le raisonnement stratégique (Ce qu'Aumann doit à Lacan). In: B. Walliser (Ed.) *Economie et Cognition*. Paris: Editions Ophrys (pp. 115–146).

[5] Cléro, J.-P. (2008). Lacan et les probabilités. *Revue de Synthèse* 129, 297–319.

[6] DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association* 69, 118–121.

[7] Di Tillio, A., Lehrer, E. & Samet, D. (2021). Monologues, dialogues, and common priors. Working paper (February 9).

[8] Dalkey, N. C. (1969). The Delphi method: an experimental study of group opinion. *United States Air Force Project Rand* RM-5888-PR, 1–79.

[9] Fagin, R., Halpern, J. Y., Moses, Y. & Vardi, M. Y. (1995). *Reasoning about Knowledge*. Cambridge, MA: MIT Press. (First published as *Reasoning about Knowledge*, *Mimeo, IBM*, San Jose, California, 1988.)

[10] Geanakoplos, J. (1992). Common knowledge. *Journal of Economic Perspectives* 6 (4), 53–82.

[11] Geanakoplos, J. & Polemarchakis, H. M. (1982). We can't disagree forever. *Journal of Economic Theory* 28, 192–200.

[12] Grice, H. P. (1975). Logic and conversation. In: P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics*, Vol. 3, Speech Acts. New York: Academic Press (pp. 41–58).

[13] Kohlberg, E. & Mertens, J. F. (1986). On the strategic stability of equilibria. *Econometrica* 54(5), 1003–1037.

[14] Lacan, J. (1945). Le temps logique et l'assertion de certitude anticipée: un nouveau sophisme. *Cahiers d'art, 1940–1944*, 32–42; reprinted in: *Écrits* (1966). Paris: Seuil.

[15] Lewis, D. K. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Cambridge University Press.

[16] Littlewood, J. E. (1953). *A Mathematician's Miscellany*. London: Methuen & Co.

[17] Nielsen, L. T. (1984). Common knowledge, communication, and convergence of beliefs. *Mathematical Social Sciences* 8, 1–14.

[18] Polemarchakis, H. (2016). Rational dialogs. Working Paper (February 2016).

[19] Selten, R. (1975). Reexamination of the perfectness concept for equilibrium points in extensive games. *International Journal of Game Theory*, 4(1), 25–55.

[20] Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy* 25, 701–721.

[21] Williams, D. (1991). *Probability with Martingales*. Cambridge, UK: Cambridge University Press.