# Bayesian monologues and dialogues [1]

A. Di Tillio [2] E. Lehrer [3] H. Polemarchakis [4]

D. Samet [5]

March 4, 2019

Preliminary and incomplete; do not quote or circulate

[1]Lehrer acknowledges the support of ISF Grant #963/15 and NSFC-ISF Grant #2510/17. Polemarchakis wants to thank John Geanakoplos for earlier joint work; and Christophe Chamley, Aviad Heifetz and Christina Pawlowitsch for discussions and comments. Samet acknowledges the support of ISF Grant #722/18.

[2]Bocconi University; alfredo.ditillio@unibocconi.it

[3]University of Tel Aviv; lehrer@post.tau.ac.il

[4]University of Warwick; h.polemarchakis@warwick.ac.uk

[5]University of Tel Aviv; samet@post.tau.ac.il

## Abstract

At each stage of a bayesian dialogue, each of two interlocutors states his beliefs formed after the revision prompted by the beliefs stated by the other at the previous stage. A third party, with access only to the transcript of a dialogue, cannot distinguish a bayesian dialogue from an arbitrary sequence of pairs of utterances. Equivalently, two rational individuals who learn form each other can hold different, even divergent beliefs for any number of rounds of communication.

The knowledge and belief of two agents is modelled here, as in Aumann (1976), by two partitions of a probability space. Starting with an agent's partition, a learning process for the agent is a sequence of partitions, each is a refinement of its predecessor. The probabilities ascribed by an agent along the process to a fixed event form a sequence of numbers called a *monologue* concerning this event. A *dialogue* is a pair of two monologues, one for each agent, concerning the same event such that the refinements are achieved by publicly announcing the probabilities in each stage. We introduce a new measure of fluctuation of sequences called *relative variation* and characterize the sequences of probabilities that are monologues in terms of this measure. We provide a necessary and sufficient condition for two sequences of probabilities to make a dialogue, in terms of their joint fluctuation, showing that two monologues do not necessarily make a dialogue.

**Key words**: dialogue; rationality; agreement.

**JEL classification**: D83.

*Two monologues do not make a dialogue.*

Noel De Nevers

## Introduction

A bayesian *dialogue* is a sequential exchange of beliefs; it is the prototype of a rational dialogue. At each stage, each of two interlocutors states his beliefs formed after the revision prompted by the beliefs stated by the other at the previous stage.

Aumann (1976) defined *common knowledge* and proved that consensus is a necessary condition for common knowledge. Geanakoplos and Polemarchakis (1982) proved that finite bayesian dialogues terminate in consensus and common knowledge [1].

Convergence and a terminal condition of bounded vari-

[1]Bacharach (1979) looked at bayesian dialogues when information is normally distributed. Nielsen (1984) proved convergence for infinite dialogues.

ation characterize bayesian dialogues; eventual consensus characterizes a finite bayesian dialogue that terminates when nothing is left to be said. The argument extends to the special case of a *didactic dialogue*, in which an expert is better informed than his interlocutor.

A third party, with access only to the transcript of a dialogue, cannot distinguish a bayesian dialogue from an arbitrary sequence of pairs of utterances. Equivalently, two rational individuals who learn form each other and will eventually agree, can hold different and divergent beliefs for any number of rounds of communication [2].

[2] Loosely speaking, one can consider common knowledge and agreement as an equilibrium and the dialogue that leads to common knowledge as the adjustment path; it follows that, across fundamentals, rationality is a refutable claim at equilibrium, while, along the adjustment path, it is not. Which bears an analogy with general competitive analysis: as follows from Debreu (1974), the Walrasian tâtonnement, the adjustment path, is arbitrary; nevertheless, equilibrium prices and quantities are not arbitrary, in Brown and Matzkin (1996), and, furthermore, in Chiappori, Ekeland, Kubler, and Polemarchakis (2004), they identify the fundamentals.

**Exogenous learning and monologues.** A learning process in which an agent obtains more and more information over time is classically modeled by a sequence of partitions of a probability space, where each partition refines its predecessor. The partitions' elements that contain a given state form a decreasing sequence of events that reflect the agent's increasing information at the state. The learning process is exogenous in the sense that the partitions are given and are not generated endogenously in the probability space.

A Bayesian agent updates her beliefs along the learning process by conditioning the probability at each stage on the corresponding partition's element. This results in a sequence of probability distributions at each state. We fix an event $E$ and in each stage of the learning process we assign to each state the probability of $E$ at the state. The sequence of functions thus defined form a martingale. We call the sequence of the probabilities of $E$ in a state a *monologue* concerning $E$. We first characterize the sequences of probability numbers that can emerge as a monologue. Such a characterization deems some monologues as being incompatible with Bayesian updating.

Observe that when the agent is certain about the event or its complement at some point, she never changes her mind thereafter. This means that when a monologue hits a point on the boundaries, 0 or 1, it remains there forever. It is easy to construct for such a sequence a finite probability space and a learning process with respect to which the sequence is a monologue.[3] We therefore restrict our study to sequences contained in the open unit interval, which we call *internal* sequences.

We assume that the partitions' elements have all positive probability and hence the conditional probability

[3]This is true for any sequence that becomes constant at some point.

is well defined. However, in the limit the probabilities of these elements can either vanish or be positive. In the latter case we say that the monologue is *sound* and in the former case that it is *unsound*. When we do not require that the monologue be sound, every sequence qualifies. We show:

*Any internal sequence can be a monologue.*

Thus, Bayesian updating does not limit monologues as long as we allow an agent to err. That is, to assign zero probability to the real state of the world.

The situation is different when the monologue is required to be sound. The first easy observation we make is that a sound monologue must converge. Thus, convergence is a necessary condition for an internal sequence to be a monologue. Moreover, convergence which is monotonic from some point is also a sufficient condition for being a monologue. It is the ups and downs of the sequence, its fluctuations, that may prevent a sequence from being a monologue. A Bayesian updater cannot be too jumpy.

Before we describe the condition that limits the fluctuation of a sequence, we note that monologues come in dual pairs. If $(p_k)$ is a monologue concerning some event, then its dual $(\bar{p}_k)$, where $\bar{p}_k = 1 - p_k$, is a monologue concerning the complement of the event. The changes of $(p_k)$ and its dual at stage $k$ are $p_{k+1} - p_k$ and $\bar{p}_{k+1} - \bar{p}_k$, which are of the same magnitude but opposite signs. We measure these changes relative to the actual probability, namely, we consider $(p_{k+1} - p_k)/p_k$ and $(\bar{p}_{k+1} - \bar{p}_k)/\bar{p}_k$ which are typically of different magnitude, but of opposite signs. We refer to the positive one as the *relative change* at $n$, and to the sum of the relative changes as the *relative variation.*

Obviously, a sequence and its dual have the same relative variation.

For sound monologues we show the following.

*An internal sequence is a sound monologue if and only if its relative variation is finite.*

We note that this condition implies bounded variation of $(p_k)$, which means that the sum of the stage variations $|p_{k+1} - p_k|$ is finite. This, in turn, implies that the sequence converges. However, if the limit of the sequence is not 0 or 1, then the two conditions are equivalent. We discuss the relation of the almost surly bounded variation of martingales in the literature review at the end of the introduction.

**Endogenous learning and dialogues.** We next study learning processes of two agents that are endogenously determined by their interaction. The two learning process are described in the same probability space. At the beginning, the knowledge of each agent is given by a partition of the probability space, as in the model in Aumann (1976). The posterior probability of an agent is formed at each state by conditioning on her partition. Since for forming their posteriors the agents start in the first stage with the same probability, this probability is called a *common prior*.

Fix an event $E$ and consider its two posteriors. We use these posteriors, which are endogenously given in the model, to increase the information of the agents by letting each agent know the posterior of the other agent.[4] This is done by publicly announcing the two posteriors, which means that they become commonly known. The knowledge of each agent after the announcement is described by a partition which refines her original one.[5] This ends

[4]Since the posterior of the agent is fixed on each element of her partition, she knows her own posterior.

[5]The assignment of two posteriors to each state induces a partition of the space into events on which this assignment is fixed. The new partition of an agent is the common refinement of her partition with the partition into posteriors. In the partition thus defined, the posterior probabilities of $E$ in the previous stage are commonly known.

the first stage of the two learning processes.

The learning processes of the two agents continue by repeatedly constructing pairs of partitions. In each stage the pair of posteriors of $E$ with respect to the pair of partitions is publicly announced. The result is a new pair of partitions that describes the knowledge of the players after the announcement.

We call the sequence of pairs of numbers announced along the processes a *dialogue*. Since the sequence of partitions of each agent is a learning process, the sequence of the agent's posteriors of $E$ is a monologue. It is easy to show that the two monologues are sound (undsound) if and only if one of them is sound (unsound). We call a dialogue *sound* if both monologues are sound and unsound if both are unsound.

We address now the question of which pair of probability sequences form a dialogue. When the dialogue is not required to be sound there is no restriction on the pair. We show by construction:

*Any pair of internal sequences is a dialogue.*

When we require that the two sequences form a sound dialogue, each of the sequences should be a sound monologue and must satisfy the required condition for such monologues. However, the two sequences must satisfy a stronger condition that we describe next. Given two sequences, a *joint sequence* is any sequence whose $n$-th element is the $n$-th element of one of the sequences, for every $n$. Note that each of the two given sequences is in particular a joint sequence. We show:

*Two strictly internal sequences form a sound dialogue, if and only if every joint sequence of these sequences has a bounded variation. In this case, all joint sequences converge to the same limit.*

We show that this condition can be simplified by noting that instead of checking the boundedness of *all* infinitely many joint sequence it is enough to check the boundedness of only three sequences.

The previous results help us to provide a necessary and sufficient condition for two monologues to from a dialogue:

> *Two strictly internal sequences* $(p_k)$ *and* $(q_k)$ *that are monologues form a dialogue if and only if* $\sum_k |p_k - q_k| < \infty$.

Note that all the conditions on sequences and pair of sequences here are tail conditions and do not depend on any finitely many elements of the sequences. In particular, any finite sequence or pair of finite sequences can be completed to infinite ones that satisfy these conditions.

**Eventually constant learning processes.** These are processes in which from some stage on the partitions remain the same. Monologues and dialogues in such processes are eventually constant sequences. Obviously, any learning process in a finite space is eventually constant. Thus, monologues and dialogues in such spaces are necessarily eventually constant. Moreover, in the case of dialogues, the posteriors of the given event become the same and commonly known in finite time.

Observe that any eventually constant sequence satisfies all the conditions for monologues. Also, every pair of eventually constant sequences, with the same constant, satisfies all the conditions for dialogues. It is easy to see that any such pair can be a dialogue in a finite space, and thus the constant becomes common knowledge in finite time. However, as our examples show, it can also be a dialogue in an infinite space where common knowledge is not attained in finite time. Thus, the fact that the agents have the same constant posterior in finite time does not imply that it is commonly known.

**Literature review.** The first mentioning of a dialogue in the sense used here appears in Aumann (1976). In the last paragraph of this paper, Aumann delineates a simultaneous dialogue concerning the probability of a coin falling on H after each of the agents made a number of observations known only to him.[6] In light of the agreement theorem, proved in this paper, common knowledge of the posteriors of an event imply that the two posteriors coincide. Aumann therefore concludes that the dialogue must end with the same posterior. By a simultaneous dialogue we mean that in each period the posteriors of both agents are revealed to them simultaneously, as in

[6] A dialogue is simultaneous when in each period the posteriors of both agents are revealed to them simultaneously, as in our paper.

our paper.

Geanakoplos and Polemarchakis (1982) proved that any serial dialogue must end with the same probability ascribed by both agent to the given event.[7] They showed, moreover, that in all but the last period the agents can repeat each the same probability, and only in the last period an agreement is reached which is commonly known. Hart and Taumann (2004) showed in a similar model, where communication is replaced by observation of the market, how behavior in the market can remain constant for several periods, and then the market crashes. The analysis in these papers is made locally. That is, a state is fixed and the updating of the knowledge of the players is followed in this state. All these papers assumed finite partitions which guarantees that common knowledge of the posterior probability of the event is reached in finite

[7] A dialogue is serial when in each period only one of the agents informs the other her posterior.

time.

Nielsen (1984) extended both papers by allowing knowledge structures given by sigma algebras rather than finite partitions. He formulated and proved Aumann's agreement theorem for such knowledge structures and showed that dialogues, simultaneous and serial, which may be infinite, converge almost surely to the same probability. His analysis, like ours, is global: In each period the knowledge of the agents is described in *all* states by specifying a knowledge structure in each period.

Cave (1983) extended Aumann (1976) by considering a function $f$ on events with the property that if $f(X) = f(Y)$ then $f(X \cup Y) = f(X)$, which he calls *union consistency*. He studies learning in which the values of $f$ on the elements of the partitions of the agents are publicly announced. Cave did not describe a learning process, but considered partitions for which the learning process does not change the partition, which he called an equilibrium. He showed that in equilibrium the values of $f$ are the same and they are commonly known. Note that Aumann (1976) is the special case where $f(X)$ is the conditional probability of a fixed event $E$ given $X$. Observe also that infinite learning processes cannot be described in this general set up without endowing the range of $f$ with some metric that enables the values of $f$ to converge. Parikh and P.Krasucki (1990) studied a condition on the function $f$ that guarantees that communication in pairs in a finite space leads to common knowledge.

We generalize Geanakoplos and Polemarchakis (1984) in two ways. First, we study also *infinite* sequences. Second, while they point out to somewhat surprising pairs that can form a dialogue we characterize *all* such pairs.

The sequence of probabilities of one agent in a dialogue form a monologue, which is simply the result of a learning process. The literature on such processes dealt with such sequences. Burkholder (1966) showed that and $L^1$-bounded martingale sequence is of bounded variation almost surely on every atom of the basic probability space. A simpler proof was given in Tsuchikura and Yamasaki (1976). We proved a stronger result. For our martingales the sequence must be of *relative* bounded variation. Moreover, we show that every sequence can be realized when the prior of a state is 0. Recently, Shaiderman (2018) has shown that any $L^2$-bounded martingale, when conditioned on a positive probability event, has a bounded variation.This is typically incorrect when the martingale is only $L^1$-bounded.

## The Model

**Learning and monologues** Consider a measurable state space $(\Omega, \Sigma)$ with a countable knowledge-partition $\Pi$ of an agent and her type function $t\colon \Omega \to \Delta(\Omega)$. The probability distribution $t(\omega)$ describe the agent's belief at $\omega$. We assume that $t$ is measurable with respect to $\Pi$, and for each $\omega$, $t(\omega)(\Pi(\omega)) = 1$, where $\Pi(\omega)$ is the element of $\Pi$ containing $\omega$.

A *learning process* is a sequence of countable knowledge-partitions $\Pi_1, \Pi_1, \Pi_2, \ldots$ such that $\Pi_1 = \Pi$, and $\Pi_{k+1}$ refines $\Pi_k$. We assume that for each $k$ and $\omega$, $t(\omega)(\Pi_k(\omega)) > 0$. We associate with each knowledge partition $\Pi_k$ a type function $t_k$ for this partition, where for each $E$, $t_k(\omega)(E) = t(\omega)(E \mid \Pi_k(\omega))$ is the posterior of $E$ in stage $k$ at $\omega$.

Fix an event $E$. The sequence $t_k(\omega)(E)$ is called the *monologue concerning $E$ at $\omega$*. We call the monologue *sound* if $t(\omega)(\cap_k \Pi_k(\omega)) > 0$. A sequence $p_k$ is called a (sound) *monologue* if it is a (sound) monologue concerning an event $E$ at some state $\omega$ in some learning process.

A sequence in the open interval $(0,1)$ is called *internal*. A sequence is an interval $(\xi, 1-\xi)$ for $\xi > 0$ is called *strictly internal*. Observe that if $p_k$ is a monologue and for some $n$, $p_n$ is either 0 or 1, then for all $k > n$, $p_k = p_n$.

Note that a sound monologue must converge. Indeed, $p_k = t(\omega)(E \cap \Pi_k(\omega))/t(\omega)(\Pi_k(\omega))$. As $\Pi_k(\omega)$ converges monotonically to the event $\cap_i \Pi_i(\omega)$ it follows by the continuity of $t(\omega)$ that $p_k$ converge to $t(\omega)(E \cap \cap_k \Pi_k(\omega))/t(\omega)(\cap_k \Pi_k(\omega))$.

The monologue concerning $E$ at $\omega$ depends only on the decreasing sequence of events $\Pi_k(\omega)$. Thus, when talking about monologues we sometime omit the partitions and refer only to a decreasing sequences of events that contain $\omega$.

**Joint learning and dialogues** Consider two agents, 1 and 2, with knowledge partitions $\Pi^1$ and $\Pi^2$ on $\Omega$ and type functions $t^1$ and $t^2$. A probability $\mu$ on $(\Omega, \Sigma)$ is a *common prior* if for each $\omega$, $t^i(\omega)(\cdot) = \mu(\cdot \mid \Pi^i(\omega))$ for $i = 1, 2$.

Let $(\Pi^1_k)$ and $(\Pi^2_k)$ be two learning process of 1 and 2 and $t^1_k$ ,$t^2_k$ the corresponding type sequences. We say that the pair of learning processes is a *joint learning process generated by $E$*, if the partitions are generated endogenously as follows. Define by $\Pi_k$ the partition of $\Omega$ into the events that are the inverse image of the function $\omega \to (t^1_k(\omega)(E), t^2_k(\omega)(E))$. Then $\Pi^1_{k+1}$ is the common refinement of $\Pi^1_k$ and $\Pi_k$ and similarly for agent 2. Note that $\Pi_k$ is a coarsening of the meet of the the partitions $\Pi^1_{k+1}$ and $\Pi^2_{k+1}$.

Thus, we can think of the joint learning process as being the result of making the pair $(t^1_k(\omega)(E), t^2_k(\omega)(E))$ common knowledge at $\omega$ at each stage $k$. We call the pair of two monologues concerning $E$ at $\omega$, $t^1_k(\omega)(E)$ and $t^1_k(\omega)(E)$ a *dialogue concerning $E$ at $\omega$*. If there is a common prior on the state space the pair of monologues is called a *dialogue with a common prior*. A pair of sequences $p_k^1$ and $p_k^2$ is a dialogue (with a common prior) if there exists a state space (with a common prior), event $E$ and state $\omega$ such that the two sequences are a dialogue (with a common prior) concerning $E$ and $\omega$.

## The main results

### Monologues

The characterization of sequences as monologues concerns restrictions on the way these sequences fluctuate, namely restrictions on the differences $p_{k+1} - p_k$. Note that monologues come in pairs. A monologue $(p_k)$ concerning an event $E$ is coupled with a dual monologue $(\bar{p}_k)$, where $\bar{p}_k = 1 - p_k$ concerning $\bar{E}$, the complement of $E$. Thus, if some restriction is applied to a sequences $(p_k)$ to be a monologue, then the same restriction should be applied to the sequence $(\bar{p}_k)$. Our first restriction is on the following valuation of the sequence $(p_k)$.

**Definition** (Variation:)**.** The *variation* of a sequence $p_k$ is $\sum_k |p_{k+1} - p_k|$. It is *bounded* if the sum is finite.

Note that the variation of $(\bar{p}_k)$ is the same as the variation of $(p_k)$, as $|p_{k+1} - p_k| = |\bar{p}_{k+1} - \bar{p}_k|$. Thus the requirement of boundedness of variation is the same for $(p_k)$ and $(\bar{p}_k)$.

Next, we consider the difference $p_{k+1} - p_k$ relative to $p_k$ and similarly for the dual sequence, $\bar{p}_{k+1} - \bar{p}_k$ relative to $\bar{p}_k$. That is, we consider $(p_{k+1} - p_k)/p_k$ and $(\bar{p}_{k+1} - \bar{p}_k)/\bar{p}_k$. Observe that these two numbers are either both 0 or of opposite sign. We choose the positive one, which is the maximal one, and sum over $k$.

**Definition** (Relative variation:)**.** The *relative varia-*

*tion* of a sequence $(p_k)$ is

$$\sum_k \max\left\{\frac{p_{k+1}-p_k}{p_k}, \frac{\bar{p}_{k+1}-\bar{p}_k}{\bar{p}_k}\right\}.$$

It is *bounded* if the sum is finite.

Note that the relative variation of $(p_k)$ can be written as

$$\sum_k |p_{k+1}-p_k|/r_k, \tag{1}$$

where $r_k = p_k$ when $p_k \geq p_{k+1}$ and $r_k = \bar{p}_k$ when $p_{k+1} \geq p_k$.

The following two observations are straightforward:

*Observation* 1. If a sequence has a bounded variation then it converges.

*Observation* 2. If a sequence has a bounded relative variation then it has a bounded variation.

The first observation follows since bounded variation implies that the sequence is Cauchy and therefore converges. The second observation follows since in Eq. (1), $r_k \leq 1$.

The implications in the opposite direction do not hold. The sequence $p_n = \sum_{k=1}^{n}(-1)^k/n$ converges, but its variation is $2\sum_{k=1}^{\infty} 1/n$ which is unbounded.

To show that bounded variation does not imply bounded relative variation, consider a sequence $x, y, x/2, y/2, \ldots,$ $x/2^n, y/2^n, x/2^{n+1} \ldots,$ where $x$ and $y$ are positive and $x < y$. The variation of this sequence is $\sum_{n=1}^{\infty} |x/2^n - y/2^n| + \sum_{n=1}^{\infty} |y/2^n - x/2^{n+1}|$. Each of the two sums is a geometric series with quotient $1/2$ and therefore converge. Thus the variation is bounded. Note that for each $n$, $x/2^n < y/2^n$. Therefore, the relative variation contains the sum $\sum_n |x/2^n - y/2^n|/(x/2^n)$. However, each term in this sum is $|1 - y/x| > 0$ and thus the sum is unbounded, and so is the relative variation.

There are cases in which bounded variation does imply bounded relative variation:

*Observation* 3. Bounded variation of a sequence $p_k$ implies bounded relative variation in the following cases:

1. $p_k$ is strictly internal;

2. $p_k$ converges to a number $p$ in $(0, 1)$;

3. $p_k$ is a decreasing sequence that converges to 0;

4. $p_k$ is an increasing sequence that converges to 1.

*Proof.* For (1) and (2), note that since $(p_k)$ is of bounded variation, it converges, and thus the conditions in (1) and (2) are equivalent. Now, if $p_k$ is in $(\xi, 1-\xi)$ for some $\xi > 0$, then $|p_{k+1} - p_k|/r_k < |p_{k+1} - p_k|/\xi$. For (3), we can assume that $p_k \leq 1/2$ for all $k$. The relative variation of $p_k$ is $\sum_k(\bar{p}_{k+1} - \bar{p}_k)/\bar{p}_k \leq 2\sum_k(\bar{p}_{k+1} - \bar{p}_k)$. For (4), we can assume that $p_k \geq 1/2$ for all $k$. The relative variation of $p_k$ is $\sum_k(p_{k+1} - p_k)/p_k \leq 2\sum_k(p_{k+1} - p_k)$. □

**Theorem 1.** *(i) Every internal sequence is a monologue.*

*(ii) An internal sequence is a sound monologue if and only if it has a bounded relative variation.*

*(iii) A sequence which is either strictly internal or monotonic is a sound monologue if and only if it has a bounded variation.*

*Proof.* Let $(p_k)$ be an internal sequence which is a monologue concerning $E$ at $\omega$. Denote $Q_k = \Pi^k(\omega)$, $\mu = t(\omega)$, and $a_k = \mu(Q_k)$.

**Claim 1.** *For each $k$,*

$$\frac{a_k}{a_{k+1}} \geq \max\left\{\frac{p_{k+1}}{p_k}, \frac{\bar{p}_{k+1}}{\bar{p}_k}\right\}. \tag{2}$$

*Proof.* As $\mu(E \cap (Q_k \setminus Q_{k+1})) = \mu(E \cap Q_k) - \mu(E \cap Q_{k+1}) = p_k a_k - p_{k+1} a_{k+1}$, it follows that

$$0 \leq p_k a_k - p_{k+1} a_{k+1} \leq a_k - a_{k+1}, \tag{3}$$

which is equivalent to (2). □

To prove $(i)$ we use (2) to construct a learning process that yields $p_k$. Fix $a^1$ in $(0,1)$, and define by induction a sequence $a_k$ as follows. When $a_k$ is defined, we set $a_{k+1}$ to be a number such that $a_k/a_{k+1}$ satisfies (2), and in addition $a_k/a_{k+1} \geq 2$, which guarantees that the sequence$(a_k)$ is decreasing and converging to 0. Thus, we can choose a state space with a state $\omega$ and a learning process such that $\mu(Q_k) = a_k$ and $\mu(\cap Q_k) = 0$. By Eq. (3) we can choose for each $k$ an event $E_k$ in $Q_k \setminus Q_{k+1}$ such that $\mu(E_k) = p_k a_k - p_{k+1}a_{k+1}$. Finally, let $E = \cup_k E_k$. As $\mu(\cap Q_k) = 0$, $\mu(E \cap Q_k) = \sum_{i \geq k} \mu(E \mid Q_k \setminus Q_{k+1}) = \sum_{i \geq k} p_i a_i - p_{i+1}a_{i+1} = p_k a_k$. Hence, $\mu(E \mid Q_k) = p_k a_k / a_k = p_k$, as required.

Next we formulate the conditions of soundness and bounded relative variation in terms of infinite products.

**Claim 2.** *Let the internal sequence $(p_k)$ be a dialogue* *concerning $E$ at $\omega$. Then, the dialogue is sound if and only if*

$$\lim_n \prod_{k=1}^{n} \frac{a_k}{a_{k+1}} < \infty. \tag{4}$$

*Proof.* The finite products in (4 are $a_1/a_{n+1}$. They converge if and only if $\lim_n a_n > 0$, that is if and only if the dialogue is sound. □

**Claim 3.** *An internal sequence $(p_k)$ has a bounded relative variation if and only if*

$$\lim_n \prod_{k=1}^{n} \max\left\{\frac{p_{k+1}}{p_k}, \frac{\bar{p}_{k+1}}{\bar{p}_k}\right\} < \infty \tag{5}$$

*Proof.* Note that

$$\begin{aligned} \max\left\{\frac{p_{k+1}}{p_k}, \frac{\bar{p}_{k+1}}{\bar{p}_k}\right\} &= \max\left\{1 + \frac{p_{k+1} - p_k}{p_k}, 1 + \frac{\bar{p}_{k+1} - \bar{p}_k}{\bar{p}_k}\right\} \\ &= 1 + \max\left\{\frac{p_{k+1} - p_k}{p_k}, \frac{\bar{p}_{k+1} - \bar{p}_k}{\bar{p}_k}\right\}. \end{aligned} \tag{6}$$

To relate the infinite product in (5) to the infinite sum in the definition of relative variation we use the following lemma.

**Lemma 1.** *Let* $(\varepsilon_k)$ *be a non-negative sequence. Then,* $\prod(1+\varepsilon_k) < \infty$ *if and only if* $\sum \varepsilon_k < \infty$.

*Proof.* For one direction observe that $\prod_{i=1}^{n}(1+\varepsilon_k) \geq 1+\sum_{i=1}^{n} \varepsilon_k$. For the other direction, note that $\prod(1+\varepsilon_k) < \infty$ if and only if $\sum \ln(1+\varepsilon_k) < \infty$, and $\sum_{i=1}^{n} \varepsilon_k \geq \sum \ln(1+\varepsilon_k)$. □

Setting $\varepsilon_k = \max\{(p_{k+1}-p_k)/p_k, (\bar{p}_{k+1}-\bar{p}_k)/\bar{p}_k)\}$, it follows by Lemma 1 and Eq. (6), that Eq. (5) holds if and only if $\sum_k \varepsilon_k < \infty$, that is if and only if $(p_k)$ has a bounded relative variation. □

To prove $(ii)$, suppose that $(p_k)$ is a sound monologue. Then by Claim 2 Eq. (4) holds. Therefore, by Eq. (2), Eq. (5) holds. Thus, by Claim 3, $(p_k)$ has a bounded relative variation.

For the converse direction we use the following observation . If Eq. (2) holds with equality for each $k$, and $(p_k)$ has a bounded relative variation, then Eq. (5) holds, and by the equality in Eqs. (2), (4) holds. Thus, by Claim 2 the monologue is sound.

Now, suppose that $(p_k)$ is a sequence with a bounded relative variation, then by Observations 1 and 2 it converges. Let $p = \lim p_k$. We construct a learning process as in $(i)$, only that in stage $k$ we set $a_{k+1}$ to satisfy Eq. (2) with equality. Thus, as we observed before, the monologue is sound, that is $\lim a_n = a > 0$. The events $E_k$ are defined as in the construction in $(i)$. In addition we define $E_\infty$ to be an event in $\cap Q_k$ such that $\mu(E_\infty) = pa$. We set $E = E_\infty \cup (\cup E_k)$. Thus, $\mu(E \cap Q_k) = (\sum_{i \geq k} p_i a_i - p_{i+1} a_{i+1}) + pa$. As $p_i a_i$ converges to $pa$, this sum is $p_k a_k$, implying $\mu(E \mid Q_k) = p_k$, as desired.

$(iii)$ If $p_k$ is a sound monologue concerning $E$ at $\omega$, then it has, by $(ii)$, a bounded relative variation which implies, by Observation 2, that it has a bounded variation. Conversely, suppose that $p_k$ is bounded or monotonic and it has a bounded variation. By Observation 1 $p_k$ converges, and by Observation 3 it has a bounded relative variation. By $(ii)$, $p_k$ is a sound monologue in some learning process. □

#### Dialogues

When we do not restrict ourselves to *sound* monologues then every internal sequence is a monologue as stated in Theorem 1. A similar result holds for dialogues that are not necessarily sound.

**Theorem 2.** *Any pair of internal sequences is a dialogue with a common prior.*

By giving up the common prior in the consequence of Theorem 2 we can gain soundness, as stated in the following theorem.

**Theorem 3.** *Any pair of internal sequences is a sound dialogue.*

Theorems 2 and 3 are proved in Section .

Next, we study dialogues with common prior and soundness. For simplicity we assume that the sequences involved are strictly internal.

A *mixture* of two sequences $(p^1_k)$ and $(p^2_k)$ is a sequence $p^{i(1)}_1, p^{i(2)}_2, p^{i(3)}_3, \ldots$, where $i(k)$ is a sequences of names of the players, that is, $i(k) \in \{1, 2\}$ for each $k \geq 0$. Note, that $p^1_k$ and $p^2_k$ are in particular mixture sequences for the constant sequences $i(k)$.

**Theorem 4.** *A pair of strictly internal sequences is a sound dialogue with a common prior if and only if their mixtures have a bounded variation. In this case they all converge to the same limit.*

Theorem 4 is proved in Section .

The condition in Theorem 4 involves the infinitely many sums that are associated with to the mixture sequences. However, it can be simplified by requiring bounded variation of only three sequences. Consider the five sums:

(a) $\sum_k |p^1_{k+1} - p^1_k|$,

(b) $\sum_k |p^2_{k+1} - p^2_k|$,

(c) $\sum_k |p^1_{k+1} - p^2_k|$,

(d) $\sum_k |p^2_{k+1} - p^1_k|$

(e) $\sum_k |p^1_k - p^2_k|$.

The first four are associated with the mixture sequences with the orders of names $(1, 1, 1, \dots)$, $(2, 2, 2, \dots)$, $(1, 2, 1, 2, \dots)$, and $(2, 1, 2, 1, \dots)$ correspondingly. The sum (e) is not associated with a mixture sequence.

**Proposition 1.** *The variations of all mixtures of* $(p^1_k)$ *and* $(p^2_k)$ *are bounded if and only if the variations of* *the sums (a), (b), and one of (c),(d), and (e) are bounded.*

*Proof.* Let $(p_i^{i(k)})$ be a mixture sequence. Then each summand $|p_{k+1}^{i(k+1)} - p_k^{i(k)}|$ appears in either (a), (b), (c), or (d). Thus, all mixture sequences have bounded variation if and only if (a), (b), (c), and (d) are bounded. Suppose that (a), (b), and (e) are bounded. Since for each $k$, $|p_{k+1}^1 - p_k^2| \leq |p_{k+1}^1 - p_k^1| + |p_k^1 - p_k^2|$, it follows that (c) is bounded, and similarly, (d) is also bounded. Suppose that (a), (b), and (c) are bounded. Since for each $k$, $|p_k^1 - p_k^2| \leq |p_k^1 - p_{k+1}^1| + |p_{k+1}^1 - p_k^2|$ it follows that (e) is bounded and therefore also (d) is bounded. The case of bounded (a), (b), and (d) is similar. □

We can now state a simple condition for two strictly internal sound monologues to be a sound dialogue.

**Theorem 5.** *Let $(p_k^1)$ and $(p_k^2)$ be two strictly internal sequences. Then the pair $(p_k^1)$ and $(p_k^2)$ is a sound* *dialogue with a common prior if and only if the two sequences are sound monologues and* $\sum_k |p_k^1 - p_k^2| < \infty$.

*Proof.* Suppose that the sequences $(p_k^1)$ and $(p_k^2)$ form a sound dialogue with a common prior, then by Theorem 4 and Proposition 1, (a), (b), and (c) are all bounded. The boundedness of (a) and (b) implies by Theorem 1 and Observation 3 that each of the sequences is a sound monologue. The boundedness of (e) is the condition in the theorem.

Conversely, suppose that $(p_k^1)$ and $(p_k^2)$ are sound monologues for which (e) is bounded. Then by Theorem 1 and Observation 3, (a) and (b) are bounded. Thus, by Theorem 4 and Proposition 1, the pair of sequences is a sound dialogue. □

## Proofs

*Proof of Theorem* 2. Let $(p_k^1)$ and $(p_k^2)$ be two internal sequences. We construct a type space with a common prior $\mu$ and an event $E$ such that these two sequences are the dialogue concerning $E$ in a state $\omega$. Let $\Omega = \{\omega_{i,j}, \eta_{i,j} \mid i, j \in [1, \ldots, \infty]\}$. We call the events $C_{i,j} = \{\omega_{i,j}, \eta_{i,j}\}$ *cells*. The part of cells which are in $E$ are $E_{i,j} = \{\eta_{i,j}\}$ and thus $E = \cup_{i,j} E_{i,j}$. We denote $\text{Row}(m, \vec{k}) = \cup_{j \geq k} C_{m,j}$ and call it the *k-truncated m-row*. Similarly, the *k-truncated m-column* is $\text{Col}(\vec{k}, m) = \cup_{i \geq k} C_{i,m}$. Let $\omega = \omega_{\infty,\infty}$.

The initial partition of player 1, $\Pi_1^1$, consists of all the 1-truncated rows, $\text{Row}(i, \vec{1})$ for $i \in [1, \ldots \infty]$. For player 2, $\Pi_1^2$ consists of all the 1-truncated columns $\text{Col}(\vec{1}, j)$ for $j \in [1, \ldots \infty]$. We will define a common prior $\mu$ on

$(\Omega, \Sigma)$ such that for all $i, j \in [k+1, \infty]$,

$$\mu(E \mid \text{Row}(i, \vec{k})) = p_k^1 \quad \text{and} \quad \mu(E \mid \text{Col}(\vec{k}, j))) = p_k^2 \tag{7}$$

whlie

$$\mu(E \mid \text{Row}(k, \vec{k})) \neq p_k^1 \quad \text{and} \quad \mu(E \mid \text{Col}(\vec{k}, k))) \neq p_k^2. \tag{8}$$

Note that by (7), for a fixed row $i$, $\mu(E \mid \text{Row}(i, \vec{k})) = p_k$ for all $k \leq i-1$ and similarly for agent 2.

By (7) and (8), the event that agents 1 and 2 ascribe to $E$ probability $p_1^1$ and $p_1^2$, correspondingly, is the event $\cup_{i \geq 2, j \geq 2} C_{i,j}$. We make this event commonly known by describing the agents's knowledge in stage 2 by a pair of partitions $\Pi_2^1$ and $\Pi_2^2$, where $\Pi_2^1$ includes the 2-truncated $i$-rows for $i \geq 2$, and $\Pi_2^1$ include the 2-truncated $j$-columns for $j \geq 2$. Using (7) and (8) repeatedly, the joint learning process goes on and in stage $k$, $(\Pi_k^1)$ includes the $k$-truncated $i$-rows, $\text{Row}(i, \vec{k})$ for $i \geq k$, and

$\Pi_k^2$ includes $k$-truncated $j$-columns, $\mathrm{Col}(\vec{k}, j)$ for $j \geq k$. The resulting dialogue concerning $E$ at $\omega$ will be the pair of sequences $(p_k)$ and $(q_k)$.

It remains to construct the probability $\mu$. In this construction we use the term *center* to denote the set of diagonal cells $C_{i,i}$ and the cells adjacent to the diagonal. Thus the *off center* cells are those $C_{i,j}$ with $i, j$ in $[0, \infty]$ such that either $i > j + 1$ or $j > i + 1$. The construction of $\mu$ is carried out in four steps.

In step 1, $\mu(C_{i,j})$ will be defined for all cells $C_{i,j}$. We define in step 2, $\mu(E_{i,j})$ for all cells off center. In Step 3, we fix $n < \infty$ and define a probability $\mu^n$ that agrees with $\mu$ off center and generates a dialogue that agrees with the given dialogue in the first $n$ stages. In step 4 we define $\mu$ to be limit of the probabilities $\mu^n$.

**Step 1**: defining $\mu$ on cells. For each $k < \infty$ let

$$e_k = \left(\max\{1/p_k^1, 1/\bar{p}_k^1, 1/p_k^2, 1/\bar{p}_k^2\}\right)^{-1},$$

and $\varepsilon_k = \min\{e_0, \ldots, e_k\}/3$. Define for each $i, j < \infty$, $\mu(C_{i,j}) = W\varepsilon_j^j\varepsilon_i^i$, $\mu(C_{\infty,j}) = W\varepsilon_j^{2j}$, $\mu(C_{i,\infty}) = W\varepsilon_i^{2i}$, and finally, $\mu(C_{\infty,\infty}) = 0$. The constant $W$ is chosen to normalize the sum of these numbers.

The only property of $\mu$ which is needed in the following steps is described in the next claim.

**Claim 4.** *Let $a_k^i = \mu(\mathrm{Row}(i, \vec{k}))$ and $b_k^j = \mu(\mathrm{Col}(j, \vec{k}))$. Then, for all $i$, $j$ in $[1, \infty]$ and $k < \infty$ such that $i, j \geq k + 1$,*

$$\frac{a_k^i}{a_{k+1}^i} \geq \max\left\{\frac{1}{p_k^1}, \frac{1}{\bar{p}_k^1}\right\} \text{ and } \frac{b_k^j}{b_{k+1}^j} \geq \max\left\{\frac{1}{p_k^2}, \frac{1}{\bar{p}_k^2}\right\}. \tag{9}$$

To prove this claim, consider first a pair $(i, k)$ where $k + 1 \leq i < \infty$, The ratio $a_k^i/a_{k+1}^i$ is $(\sum_{j \geq k} \varepsilon_i^i\varepsilon_j^j + \varepsilon_i^{2i})/(\sum_{j \geq k+1} \varepsilon_i^i\varepsilon_j^j + \varepsilon_i^{2i})$. After cancelling $\varepsilon_i^i$ this ratio becomes $(\sum_{j \geq k} \varepsilon_j^j + \varepsilon_i^i)/(\sum_{j \geq k+1} \varepsilon_j^j + \varepsilon_i^i)$. The nominator exceeds $\varepsilon_k^k$. We increase the denominator by replacing each $\varepsilon_j^j$ in the infinite sum by $\varepsilon_k^j$ which results in a geometric series smaller than $2\varepsilon_k^{k+1}$, and by decreasing both the index and the power of $\varepsilon_i^i$ to $\varepsilon_k^{k+1}$. Thus the ratio is bigger than $\varepsilon_k^k/(3\varepsilon_k^{k+1}) = 1/e_k \geq \max\{1/p_k^1, 1/\bar{p}_k^1\}$. This proves the first part of (9) for $i < \infty$.

Next, consider the pair $(\infty, k)$. The ratio $a_\infty^k/a_\infty^{k+1}$ is $(\sum_{j\geq k}\varepsilon_j^{2j}/\sum_{j\geq k+1}\varepsilon_j^{2j}$. The nominator exceeds $\varepsilon_k^{2k}$. We increase the denominator by replacing each $\varepsilon_j^{2j}$ by $\varepsilon_k^{2j}$, getting $\sum_{j=k+1}^{\infty}\varepsilon_k^{2j} = \varepsilon_k^{2k+2}(1-\varepsilon_k^2) \leq 2\varepsilon_k^{2k+2}$. Thus the ratio is more than $1/(2\varepsilon_k^2) \geq 1/(2\varepsilon_k) \geq 1/((2/3)e_k) \geq \max\{1/p_k^1, 1/\bar{p}_k^1\}$. This proves the first part of (9) for $i = \infty$. The proof for agent 2 is similar.

**Step 2**: defining $\mu(E_{i,j})$ off center. For $k < \infty$ and $i \geq k+2$ we let $\mu(E_{i,k}) = p_k^1 a_k^i - p_{k+1}^1 a_{k+1}^i$. To justify this definition we need to show that this difference falls between 0 and $\mu(C_{i,k})$, note that $\max\{1/p_k^1, 1/\bar{p}_k^1\} \geq \max\{p_{k+1}^1/p_k^1, \bar{p}_{k+1}^1/\bar{p}_k^1\}$. Thus (9) and the equivalence of (2) and (3) imply that $\mu(E_{i,j})$ falls in the required range. Similarly, for $k < \infty$ and $j \geq k+2$ we let $\mu(E_{k,j}) = p_k^2 b_k^i - p_{k+1}^2 b_{k+1}^i$.

Observe that if for $i > k$, $\mu(E \mid \text{Row}(i, \vec{k})) = p_k$, then by the definition of $\mu(E_{i,k-1})$ it follows that $\mu(E \mid \text{Row}(i, \overrightarrow{k-1})) = p_{k-1}^1$. Thus, in order to show that (7) holds in row $i$ it is enough to show that $\mu(E \mid \text{Row}(i, \overrightarrow{i-1})) = p_{i-1}^1$, and similarly for the second agent. This is done in the next step.

**Step 3**: constructing probabilities in the center. It remains to define $\mu(E_{i,j})$ for the cells in the center. This is done as follows. For a fixed $n > 1$ we define $\mu(E_{i,j})$ for center cells such that (7) and (8) hold for all $i, j \leq n$. We denote the resulting probability $\mu_n$. Obviously, all measures $\mu_n$ coincide off center. The construction of $\mu^n$ is carried out by induction on $k = n+1, \ldots, 1$.

For $k = n+1$ we define arbitrarily $\mu^n(E_{n+1,n+1})$,

$\mu^n(E_{n+2,n+1})$, and $\mu^n(E_{n+1,n+2})$. Suppose the construction was carried out for $k+1$. We construct $\mu^n(E_{k,k})$, $\mu^n(E_{k+1,k})$, and $\mu^n(E_{k,k+1})$.

We start with $\mu(E_{k+1,k})$. Denote $p = \mu(E \mid \text{Row}(k+1, \overrightarrow{k+1}))$ and define $\mu(E_{k+1,k}) = p^1_k a^{k+1}_k - p a^{k+1}_{k+1}$. As, $\max\{1/p^1_k, 1/\bar{p}^1_k\} \geq \max\{p/p^1_k, \bar{p}/\bar{p}^1_k\}$, it follows from (9) and the equivalence of (2) and (3) that $\mu(E_{i,j})$ falls between 0 and $\mu(E_{k+1,k})$, and thus the definition is valid. Moreover, $\mu(E \mid \text{Row}(k+1, \vec{k}) = p^1_k$. Thus, (7) holds in row $k+1$. We similarly define $\mu(E_{k,k+1})$.

We need to define $\mu^n(E_{k,k})$ such that (8) is satisfied. Since we want to keep the inequality in the limit of $\mu^n$ we need $\hat{p}^1_k$ and $\hat{p}^1_k$ to be bounded away from $p^1_k$ and $p^2_k$, respectively, uniformly for all $n$. Let $M_1 = \mu^n(E \cap \text{Row}(k, \overrightarrow{k+1}))$ and $K_1 = \mu^n(\text{Row}(k, \vec{k}))$. We similarly define $M_2$ and $K_2$ for agent 2. If we set $\mu^n(E_{k,k}) = 0$, then $\mu^n(E \mid \text{Row}(k, \vec{k})) = M_1/K_1$ and $\mu^n(E \mid \text{Col}(\vec{k}, k)) =$

$M_2/K_2$. If we set $\mu^n(E_{k,k}) = \mu^n(C_{k,k}) = \varepsilon_k^{2k}$, then $\mu^n(E \mid \text{Row}(k, \overrightarrow{k+1})) = (\varepsilon_k^{2(k)} + M_1)/K_1$ and $\mu^n(E \mid \text{Col}(\overrightarrow{k+1}, k)) = (\varepsilon_k^{2k} + M_2)/K_2$. Thus, we can choose the pair $(\hat{p}_k^1, \hat{p}_k^2)$, in the plain $\mathbb{R}^2$, on the interval with endpoints $(M_1/K_1, M_2/K_2)$ and $((\varepsilon_k^{2k} + M_1)/K_1, (\varepsilon_k^{2k} + M_2)/K_2)$. The difference between these endpoints is $(\varepsilon_k^{2k}/K_1, \varepsilon_k^{2k}/K_2)$ which depends only on the definition of $\mu(C_{i,j})$ and neither on the definition of $\mu(E_{i,j})$, nor on $n$. Thus, we can find $\rho_k > 0$ small enough and a pair $(\hat{p}_k^1, \hat{p}_k^2)$ on the said interval, such that $|\hat{p}_k^1 - p_k^1| \geq \rho_k$ and $|\hat{p}_k^2 - p_k^2| \geq \rho_k$. The choice of the pair $\hat{p}_k^1, \hat{p}_k^2$ may depend on $n$, but $\rho_k$ will be the same for all $n$.

**Step 4**: taking the limit. Let $I$ be the set of indices of center cells. The set $[0,1]^I$ with the product topology is a compact set. For each $n$, $(\mu^n(E_{i,j}))_{(i,j)\in I}$ is an element of this set. Thus, there exists a limit point $x_{i,j}$ of this sequence. Obviously, for each $(i,j) \in I$, $0 \leq x_{i,j} \leq$

$\mu(C_{i,j})$. Thus we can extend $\mu$ to $E_{i,j}$ in the center by defining $\mu(E_{i,j}) = x_{i,j}$.

We need to show that $\mu$ satisfies (7) and (8). Each equation for $i$ and $k$ in (7) is a linear equation in the three numbers $\mu^n(E_{i,i-1})$, $\mu^n(E_{i,i})$, and $\mu^n(E_{i,i+1})$, where the coefficients are the same for all $n$. Thus, the equation holds also in the limit, that is, for $\mu$.

For (8), $\mu^n(\text{Row}(k, \vec{k}))$ is a linear expression in $\mu^n(E_{i,i})$, and $\mu^n(E_{i,i+1})$ where the coefficients are independent of $n$. Thus, $\mu^n(\text{Row}(k, \vec{k})) \to_n \mu(\text{Row}(k, \vec{k}))$. As $|\mu^n(\text{Row}(k, \vec{k})) - p_k^1| \geq \rho_k > 0$, for each $n$, it follows that $|\mu^n(\text{Row}(k, \vec{k})) - p_k^1| \geq \rho_k > 0$, and thus $\mu(\text{Row}(k, \vec{k})) \neq p_k^1$. The argument for agent 2 are the same. □

*Proof of Theorem 4.* Let the pair $(p_k^1)$ and $(p_k^2)$ be a sound dialogue concerning $E$ at $\omega$ with a common prior $\mu$, and consider a sequence $i(k)$ of names in $\{1, 2\}$. We define a decreasing sequence of events, $Q_k$ such that $Q_k \subseteq$

$\Pi_k(\omega)$, the element of the meet of $\Pi_k^1$ and $\Pi_k^2$ that contains $\omega$. The event $Q_k$ is the union of all the elements of the partition $\Pi_k^{i(k)}$ in which the posterior of $E$ is $p_k^{i(k)}$. Since $\Pi_{k+1} \subseteq Q_k$, it follows that the sequence $Q_k$ is decreasing. Thus, the sequence $p_k^{i(k)}$ is the monologue concerning $E$ at $\omega$ generated by the sequence $Q_k$ and the common prior $\mu$. By construction, $\cap_k \Pi_{k+1} \subseteq \cap_k Q_k$. By the soundness of the dialogue $\mu(\cap_k \Pi_{k+1}) > 0$. Thus the monologue $p_k^{i(k)}$ is sound. By the second part of Theorem 1 and Observation 3, the sequence $p_k^{i(k)}$ has a bounded variation.

Since the variations of $p_k^1$ and $p_k^2$ are bounded they converge. Moreover, because of the soundness they both converge to $\mu(E \mid \cap_k \Pi_k)$. Obviously, any mixture converges to the same limit.

For the converse, we construct a state space in which the two sequences form a sound dialogue.

**A sketch of the construction.** We show that two strictly internal monologues, $p^1_k$ and $p^2_k$, that satisfy $\sum_{k=0}^{\infty} |p^1_k - p^2_k| < \infty$ form a sound dialogue concerning an event $E$ at a state $\omega$ whose probability is positive in a type space with CP $\mu$.

As in the previous proof we construct the probabilities backward. Define,

$$H_k := \cup_{1 \leq i,j \leq \infty} C_{i,j} = \cup_{j \geq k} \operatorname{Col}(\overrightarrow{k}, j). \qquad (10)$$

$H_k$ is the south-eastern $k$-th corner. Suppose that all the probabilities (of $C_{i,j}$ and $E_{i,j}$) in $H_{k+1}$ have been defined. We want to extend the definition to $H_k$. We define first the probabilities on the $k$-th column, $\operatorname{Col}(\overrightarrow{k+1}, k)$. By doing it, we ensure that the conditional probabilities of $E$ on every row $\operatorname{Row}(i, \overrightarrow{k})$ is $p^1_k$, $i \geq k+1$.

A main issue in the construction is to control the size of $\operatorname{Col}(\overrightarrow{k+1}, k)$. It should not grow too fast. The reason is that otherwise, the normalized probabilities at the end of the process might vanish. How large the column $\text{Col}(\overrightarrow{k+1}, k)$ could be? Lemma 3 shows that its size is of the magnitude of $H_{k+1}$ times $|p_k^1 - p_{k+1}^1| + |p_{k+1}^1 - p_{k+1}^2|$. So, the total weight accumulated after defining $\text{Col}(\overrightarrow{k+1}, k)$ is that of $H_{k+1}$ times $(1 + |(p_k^1 - p_{k+1}^1| + |(p_{k+1}^1 - p_{k+1}^2|)$.

Next we define the $k$-th row, $\text{Row}(k, \overrightarrow{k+1})$. The purpose of this definition is to make the conditional probability of $E$ on each column $\text{Col}(\overrightarrow{k}, j), j \geq k+1$ equal to $p_k^2$. Following the reasoning above, the total weight after the extension to the $k$-th row and column is that of $H_{k+1}$ times $(1 + |p_k^1 - p_{k+1}^1| + |p_{k+1}^1, p_{k+1}^2|)(1 + |p_k^2 - p_{k+1}^2|)$. Continuing in this fashion until $k = 1$, we get that the total size is of $H_1$ is that of $H_{k+1}$ times

$$\prod_{i=1}^{k}(1 + |p_i^1 - p_{i+1}^1| + |(p_{i+1}^1, p_{i+1}^2|)(1 + |p_i^2 - p_{i+1}^2|).$$

However, due to the fact that the two sequences are monologues and due their closeness, this product is bounded (see Lemma 1).

One additional point should be clarified. Notice that when defining the diagonal cell, $C_{k,k}$, the probabilities of $E$ conditional on the row $\text{Row}(k, \overrightarrow{k})$ should be different from $p^1_k$. This is the reason why the definition of the probabilities (total and conditional) on the diagonal cell is different from that on the other cells. But there is another difference that adds to the complication. The diagonal cell determines also the conditional probability of $E$ on the $k$-th column, $\text{Col}(\overrightarrow{k}, k)$, which has to be different from $p^2_k$.

Before we proceed to the proof we need a few notations. For every $a, b \in (0,1)$ we denote,

$$\delta(a,b) = \max\left\{\frac{b-a}{a}, \frac{\bar{b}-\bar{a}}{\bar{a}}\right\}.$$

**Lemma 2.** *Let $x, y \in (0,1)$, $(\Omega, \mu)$ be a measurable space, $A, B \subseteq \Omega$ two disjoint events, and an event*

*$E \subseteq \Omega$ such that **(a)** $\mu(A) > 0$, **(b)** $\mu(B) = 2\varphi(x,y)\mu(A)$; **(c)** $\mu(E \mid A) = y$; and **(d)** $\mu(E \mid B) = (x + 1\!\!1_{x>y})\varphi(x,y)\mu(A)$. Then, $\mu(E \mid A \cup B) = x$.*

*Proof.* Suppose first that $x > y$. Then,

$$\begin{aligned}\mu(E \mid A \cup B) &= \frac{\mu(A)(x\varphi(x,y) + \varphi(x,y) + y)}{\mu(A)(2\varphi(x,y)+1)} \\ &= \frac{(x-y)(1+x) + y(1-x)}{2(x-y)+1-x} = \frac{x(x-2y+1)}{x-2y+1} = x.\end{aligned}$$

Suppose now that $x \leq y$. Then,

$$\mu(E \mid A \cup B) = \frac{\mu(A)(x\varphi(x,y) + y)}{\mu(A)(2\varphi(x,y)+1)} = \frac{(y-x+y)x}{2(y-x)+x} = x.$$

□

Lemma 2 is of a great importance for the constructive proofs that follow. Suppose that the probability of an event $A$, say $\mu(A)$, and the conditional $\mu(E \mid A)$ have already been defined. We would like to add another event $B$, disjoint of $A$, such that (a) the probability of $B$ is $\mu(B) = 2\varphi(x,y)\mu(A)$; and (b) the conditional probability of $E$ given the union $A \cup B$ is equal to $x$. Is it possible, and if so, what should be the probability of $E$ within $B$? Lemma 2 states that this is possible. Furthermore, if $\mu(E \mid B) = (x + 1\!\!1_{x>y})2\varphi(x,y)\mu(A)$, then $\mu(E \mid A \cup B) = x$. In words, $E$ occupies $x$ of a half of $B$, and on the other half $E$ occupies all of it if $x > y$ and nothing, otherwise. This is the role of $1\!\!1_{x>y}$ in the formula.

*Remark* 1. When we define $\mu(B)$ as $2\varphi(x,y)\mu(A)$ and $\mu(E \mid B)$ as $(x+1\!\!1_{x>y})\varphi(x,y)\mu(A)$, we say that we apply $(x;\mu(A),y)$**-scheme** on $B$.
Suppose that $\mu(A) = y$. If we apply $(x;\mu(A),y)$-scheme on $B$, Lemma 2 states that $\mu(E \mid B \cap A) = x$.

We use the same notations as in the proof of Theorem 2. As opposed to this proof, here the probabilities of $C_{i,j}$ are not defined once and for all; they are defined alongside with the conditional probability of $E$.

The sequences $(p_k^1)$ and $(p_k^2)$ are strictly internal and thus one can find $0 < \xi$ such that $\xi < p_k^i < 1 - \xi$ for every $k$ and $i = 1, 2$.

For every integer $\ell$ we are going to define the measure $\mu_\ell$. $\mu$ will then be defined as a limit of $\mu_\ell$. Fix an integer $\ell$. During the construction we are going to define a few arrays of weights, not necessarily probabilities, and conditional probabilities:

- $c_{i,j}^\ell := \mu_\ell(C_{i,j}) = \mu_\ell(\omega_{i,j}, \eta_{i,j})$;
- $\alpha_{i,j}^\ell := \mu_\ell(E|C_{i,j}) = \mu_\ell(\eta_{i,j})/\mu_\ell(\omega_{i,j}, \eta_{i,j})$;
- $d_{i,j}^\ell := \mu_\ell(\mathrm{Col}(\overrightarrow{i}, j))$;
- $\gamma_{i,j}^\ell := \mu_\ell(E|\,\mathrm{Col}(\overrightarrow{i}, j))$;
- $r_{i,j}^\ell := \mu_\ell(\mathrm{Row}(i, \overrightarrow{j}))$;
- $\rho_{i,j}^\ell := \mu_\ell(E|\,\mathrm{Row}(i, \overrightarrow{j}))$;
- $t_{i,j}^\ell := \sum_{k \geq i} r_{k,j}^\ell$.

During the construction, for every $i, j < \infty$, we will care to keep $c^{\ell}_{i,j}$ (across $\ell$) away from 0. The reason is that a converging subsequence (as $\ell$ goes to infinity) will define $(c_{i,j})$ and we want to make sure that the latter does not vanish.

**Step 0: Defining the last column and row.** We start with the weights on the last column, $\mathrm{Col}(\overrightarrow{1}, \infty)$: $(\alpha^{\ell}_{i,\infty})$ and $(c^{\ell}_{i,\infty})$, $i = 1, ..., \infty$. By assumption, $(p^2_k)$ is a sound monologue. That is, one can find a sequence of decreasing events $(Q_k)$, an event $E$ and a measure $\nu_2$ such that $\nu_2(\cap_k Q_k) > 0$ and $(p^2_k)$ is a monologue concerning $E$ with respect to $(Q_k)$.

Define $c_{\infty,\infty} = \nu_2(\cap_k Q_k)$ and $\alpha^{\ell}(\eta_{\infty,\infty}) = \frac{\mu_\ell(\eta_{\infty,\infty})}{\mu_\ell(\eta_{\infty,\infty},\omega_{\infty,\infty})} = \lim_k p^2_k$. This takes care of the limit cell in which the conditional probability is $\lim_k p^2_k$. Next define, $c_{i,\infty} = \nu_2(Q_k \setminus Q_{k+1})$ and $\alpha^{\ell}(i, \infty) = \frac{\mu_\ell(\eta_{i,\infty})}{c_{i,\infty}} = \nu_2(E \mid Q_k \setminus Q_{k+1})$.

We turn to the last row, $\text{Row}(\infty, \overrightarrow{1})$: $(\alpha^{\ell}_{\infty,j})$ and $(c^{\ell}_{\infty,j})$. As $(p^1_k)$ is a sound monologue, we can find a sequence of decreasing events $(Q'_k)$, an event $E'$ and a measure $\nu_1$ such that $\nu_1(\cap_k Q'_k) > 0$ and $(p^1_k)$ is a monologue concerning $E'$ with respect to $(Q'_k)$. Without loss of generality $\nu_1(\cap_k Q'_k) = \nu_2(\cap_k Q_k)$. We now define the last row in a manner similar to that of the right margin. These definitions guarantee Eq. (7) for every $i, j = \infty$ and every $n$.

Note that the wights on the last row and column do not depend on $\ell$. Denote by $M$ their total size. That is, $M := r^{\ell}_{\infty,1} + d^{\ell}_{1,\infty} - c^{\ell}_{\infty,\infty}$.

We proceed with the other cells. For any $\ell + 1 \leq i, j < \infty$, set $c^{\ell}_{i,j} = \alpha^{\ell}_{i,j} = 0$. That is, the bottom-right corner $H_\ell$ is 0. For the following definitions recall the $(x; \mu(A), y)$-scheme in Notation 1.

**Step 1: Defining a column.** We now define the weights on column Col$(\ell+1,\ell)$. Fix $i \geq \ell+1$. Apply the $(p^1_\ell; r^\ell_{i,\ell+1}, \rho^\ell_{i,\ell+1})$-scheme on $C_{i,\ell}$. By Lemma 2, $\delta^\ell_{i+1,\ell} = p^2_n$. What is the total size of the $\ell$-th column just added? The scheme dictates that the size of $C_{i,\ell}$ is $2\delta(p^1_\ell, p^1_{\ell+1}) d^\ell_{i,\ell+1}$, which is bounded by $2d^\ell_{i+1,\ell}/\xi$ (recall, $\xi < p^1_\ell < 1-\xi$). Thus, the total weight added is bounded by $2M/\xi$.

**Step 2: Defining a row.** Next we define the terms on the row Row$(\ell,\ell+1)$. Fix $j \geq \ell+1$ and apply the $(p^2_n; d^\ell_{\ell+1,j}, \gamma^\ell_{\ell+1,j})$-scheme on the cell $C_{\ell,j}$. Again, the total added weight is bounded by $2M/\xi$.

**Step 3: Defining a diagonal cell.** The diagonal cell requires a special treatment. We have to define $c_{\ell,\ell}$ in a way that satisfies Eq. (7) for $n=\ell$. Stated differently, $\rho^\ell_{\ell,\ell}$ is bounded away from $p^1_\ell$ and $\gamma^\ell_{\ell,\ell}$ is bounded away from $p^2_\ell$. For this purpose, let $(\varepsilon_\ell)$ be a sequence of positive numbers such that $\varepsilon_\ell < 2\xi$ and $\sum_\ell \varepsilon_\ell < \infty$. We choose $\tilde{p}_\ell^\ell \in (p_\ell^2 + \varepsilon_\ell, p_\ell^2 + 2\varepsilon_\ell)$ ($\varepsilon_\ell < 2\xi$ implies $\tilde{p}_\ell^2 < 1$) and apply the $(\tilde{p}_\ell^\ell; d_{\ell+1,\ell}^\ell, \gamma_{\ell+1,\ell}^\ell)$-scheme on $C_{\ell,\ell}$. This implies that $\gamma_{\ell,\ell}^\ell$ is bounded away from $p_\ell^2$ (specifically, $\gamma_{\ell,\ell}^\ell > p_\ell^2 + \varepsilon_\ell$). It might be then that the choice of $\tilde{q}_\ell^\ell$ creates an equality $\rho_{\ell,\ell}^\ell = p_\ell^1$. We argue that it is possible to choose $\tilde{q}_\ell^\ell$ so that $\rho_{\ell,\ell}^\ell$ is bounded away from $p_\ell^1$. Indeed, $\rho_{\ell,\ell}^\ell$ is increasing (almost linearly) in $\tilde{q}_\ell^\ell$. Thus, there is only one choice where equality is created. The term $\tilde{q}_\ell^\ell$ is chosen in the interval $[p_\ell^2 + \varepsilon_\ell, p_\ell^2 + 2\varepsilon_\ell]$ so that $\rho_{\ell,\ell}^\ell$ is the farthest possible from $p_\ell^1$. This way we make sure that $\rho_{\ell,\ell}^\ell$ is bounded away from $p_\ell^1$, that is $|\rho_{\ell,\ell}^\ell - p_\ell^1|$ does not get close to 0 as $\ell$ changes.

What is the size of the total weight in the rows added? By a similar argument to the one used above, it is $4M/\xi^2$. Thus, the total size of the cells defined so far is

$$M_1 := M + 2M/\xi + 4M/\xi^2. \tag{11}$$

To summarize what we have on $H_\ell$ is that the conditionals $\rho^\ell_{i,\ell} = p^1_\ell$ for $i \geq \ell + 1$, $\gamma_{\ell,j} = p^2_\ell$ for $j \geq \ell + 1$, $\gamma^\ell_{\ell,\ell} = \tilde{p}^\ell_\ell > p^2_\ell + \varepsilon_\ell$. Finally, $\rho^\ell(\ell, \ell)$ is bounded away from $p^1_\ell$. The conditional probability of $E$ given $H_\ell$ is a convex combination of $p^2_\ell$ and $\tilde{q}^\ell_\ell$ and we denote it by $\bar{q}^\ell_\ell$.

We now continue the definitions of $\mu_\ell$ on all other cells by a backward induction. Suppose that all $c^\ell_{i,j}$ and $\alpha^\ell_{i,j}$, $k+1 \leq i, j$, have been defined. The inductive step has three step:

Step 1: Define the $k$-th column. For every $i \geq k+1$ apply the $(p^1_k; r^\ell_{i,k}, \rho^\ell_{i,k})$-scheme on $C_{i,k}$;

Step 2: Define the $k$-th row. For every $j \geq k+1$ apply the $(p^2_k; d^\ell_{k,j}, \gamma^\ell_{k,j})$-scheme on $C_{k,j}$;

Step 3: Define the $k$-diagonal cell. Choose $\tilde{q}^\ell_k \in [p^2_k + \varepsilon_k, p^2_k + 2\varepsilon_k]$ and apply the $(\tilde{p}^\ell_k; d^\ell_{k+1,k}, \gamma^\ell_{k+1,k})$-scheme on $C_{k,k}$. The choice of $\tilde{q}^\ell_k$ should keep $\rho^\ell_{k,k}$ bounded away from

$p_k^1$.

The conditional probability of $E$ given $H_k$ is a convex combination of $p_k^2$ and $\tilde{q}_k^\ell$ and we denote it by $\bar{q}_k^\ell$. By how much the total weight has increased? We use the following lemma in order to estimate it.

**Lemma 3.** *Let $D = D_1 \cup D_2$ and $E$ be events, where $D_i \cap D_2 = \emptyset$ and $\mu$ be a measure. Suppose that $\mu(E \mid D_1) = p$, $\mu(E \mid D_2) = x$ and $\mu(E \mid D) = q$. Then, for every $p'$,*

$$\mu(D_1 \mid D)|p-p'|+\mu(D_2 \mid D)|x-p'| \leq |p-p'|+|q-p|.$$

*Proof.* By assumption, $q = \mu(D_1 \mid D)p + \mu(D_2 \mid D)x = (1 - \mu(D_2 \mid D))p + \mu(D_2 \mid D)x$. Thus,

$$\mu(D_1 \mid D)|p-p'| + \mu(D_2 \mid D)|x-p'| =$$
$$\mu(D_1 \mid D)|p-p'| + |q + \mu(D_2 \mid D)p - p - \mu(D_2 \mid D)p'| \leq$$
$$\mu(D_1 \mid D)|p-p'| + |q-p| + \mu(D_2 \mid D)|p-p'| = |p-p'| + |q-p|$$

□

We start by estimating the size increase due to Step 1. For this end, we use Lemma 3 with $D = H_{k+1}$, $D_1 = H_{k+2} \cup \mathrm{Col}(\overrightarrow{k+2}, k+1)$, $D_2 = \mathrm{Row}(k+1, \overrightarrow{k+1})$, the $k+1$-th row. Also let $p = p^1_{\ell+1}, p' = p^1_k$, $x = \rho^\ell_{k+1,k+1}$ and $q = \bar{q}^\ell_{k+1}$. We obtain,

$$\mu_\ell(D_1 \mid D)|p^1_{k+1} - p^1_k| + \mu_\ell(D_2 \mid D)|\rho^\ell_{k+1,k+1} - p^1_k| \leq |p^1_{k+1} - p^1_k| + |\bar{q}^\ell_{k+1} - p^1_k \tag{12}$$

When using the scheme described in Step 1, the size of $k$-th column is the sum of two components: (a) $2\mu_\ell(D_1)\varphi(p^1_k, p^1_{k+1})$, the size of the cells on the left hand side of $D_1$; plus (b) $2\mu_\ell(D_2)\varphi(p^1_k, \rho^\ell_{k+1,k+1})$, which is the weight $\mu_\ell(C_{k+1,k})$, assigned to the cell just left $D_2$. We use here the fact that the sequences $(p^1_k)$ and $(p^2_k)$ are in $(\xi, 1-\xi)$ to obtain,

$$2\mu_\ell(D_1)\varphi(p^1_k, p^1_{k+1}) + 2\mu_\ell(D_2)\varphi(p^1_k, \rho^\ell_{k+1,k+1}) \leq$$
$$2\mu_\ell(H_{k+1})\Big(\mu_\ell(D_1 \mid H_{k+1})|p^1_{k+1} - p^1_k|/\xi + \mu_\ell(D_2 \mid H_{k+1})|\rho^\ell_{k+1,k+1} -$$
$$2\mu_\ell(H_{k+1})\left(|p^1_{k+1} - p^1_k| + |\bar{q}^\ell_{k+1} - p^1_{k+1}|\right)/\xi,$$

where the second inequality is derived from Eq. (12). To summarize, the total weight defined in Step 1 is bounded from above by, $2\mu_\ell(H_{k+1})\left(|p^1_{k+1} - p^1_k| + |\bar{q}^\ell_{k+1} - p^1_{k+1}|\right)/\xi$. It implies that the total size of the cells defined up to, and including, Step 1 (excluding the last row and column) is bounded by

$$\mu_\ell(H_{k+1}) \cdot \left[1 + 2\left(|p^1_{k+1} - p^1_k| + |\bar{q}^\ell_{k+1} - p^1_{k+1}|\right)/\xi\right]. \quad (13)$$

The increase in the size of weights due to Step 2 and Step 3 is bounded by $\max(2\varphi(p^2_k, p^2_{k+1}), 2\varphi(p^2_k, \tilde{q}^\ell_{k+1})) \leq 2\max(|p^2_k - p^2_{k+1}|, |p^2_k - \tilde{q}^\ell_{k+1}|)/\xi$. The total size after defining the probabilities on the $\ell$-th row and column is therefore bounded by (see Eq. (13)),

$$\mu_\ell(H_{k+1}) \cdot \left[1+2\left(|p^1_{k+1}-p^1_k|+|\bar{q}^\ell_{k+1}-p^1_{k+1}|\right)/\xi\right]\left[1+2\max\left(|p^2_k-p^2_{k+1}|, |p\right.\right.$$

Continuing backward until $k = 1$, we get that the total size is bounded by (recall $M_1$ at Eq. 11),

$$M_1 \cdot \prod_{k=1}^{\ell}\left[1+2\left(|p^1_{k+1}-p^1_k|+|\bar{q}^\ell_{k+1}-p^1_{k+1}|\right)/\xi\right]\left[1+2\max\left(|p^2_k-p^2_{k+1}|, |p^2_k-\right.\right.$$

Note that $|\bar{q}^{\ell}_{k+1} - p^1_{k+1}| \leq |\bar{q}^{\ell}_{k+1} - p^1_{k+1}| + |p^1_{k+1} - p^2_{k+1}| < |p^1_{k+1} - p^2_{k+1}| + 2\varepsilon_k$. Furthermore, $|\tilde{q}^{\ell}_{k+1} - p^2_{k+1}| < 2\varepsilon_k$. Thus, all the sequences involved have a bounded variation. Due to Lemma 1, the product is uniformly (for every $\ell$) bounded, say by $M_2$. We conclude that the total size of all cells defined at the $\ell$-th step is bounded my $M_1 \cdot M_2$. One can therefore normalize the weights to obtain probabilities and thereby a type space in which Eqs. (7) and (8) are satisfied for any $k \leq \ell$.

As in the previous proof we take now a converging sequence of the arrays $(c^{\ell}_{i,j})$ and $(\alpha^{\ell}_{i,j})$, as $\ell \to \infty$. The limit, denoted $(c_{i,j})$ and $(\alpha_{i,j})$, is such that $c_{i,j}, \alpha_{i,j}$ are strictly positive for every $i, j$. The reason is that for any $i, j$ and $\ell$, $c^{\ell}_{i,j}$ and $\alpha^{\ell}_{i,j}$ were defined so as to keep them bounded away from 0. Moreover the normalization factor is uniformly bounded (had it diverged to infinity, the normalized figures would have converged to zero).

Therefore, the normalized figures are also bounded away from zero. The implication is that all the equalities is Eqs. (7) are well defined (the probabilities are positive and therefore the conditional probabilities are well defined) and satisfied. Moreover, during the construction the conditional probabilities on the diagonal cells were kept bounded away from their respective probabilities, and therefore the inequalities in Eqs. (8) are also satisfied in the limit. This completes the proof. □

*The Proof of Theorem* 3. We use Theorem 5 twice. We use the same space, but we assign different probabilities to each cell. We first construct a dialogue in which both agents have the sequence $p^1_k$. That is, we construct a dialogue generating $p^1_k$ and $\widetilde{p}^2_k$, where $\widetilde{p}^2_k = p^1_k$. Here, the two sequences coincide. Obviously, these sequences satisfy the conditions of Theorem 5. The measure of this space is defined to be the prior of Agent 1. In a similar way, by constructing a dialogue generating the sequences $\widetilde{p}^1_k = p^2_k$ and $p^2_k$, we define the prior of agent 2. $\square$

## References


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

M. Bacharach. Normal bayesian dialogues. *Journal of the American Statistical Association*, 74:837–846, 1979.

D. Brown and R. Matzkin. Testable restrictions on the equilibrium manifold. *Econometrica*, 64:1249–1262, 1996.

D. L. Burkholder. Martingale transforms. *Annals of Mathematical Statistics*, 37:1494–1504, 1966.

J. Cave. Learning to agree. *Economics Letters*, 12:147–152, 1983.

P.-A. Chiappori, I. Ekeland, F. Kubler, and H. Polemarchakis. Testable implications of general equilibrium theory: a differentiable approach. *Journal of Mathematical Economics*, 40:105–119, 2004. URL http://www.polemarchakis.org/a64-tge.pdf.

G. Debreu. Excess demand functions. *Journal of Mathematical Economics*, 1:000–000, 1974.

J. D. Geanakoplos and H. Polemarchakis. We cannot disagree forever. *Journal of Economic Theory*, 28: 192–200, 1982. URL http://www.polemarchakis.org/a16-cdf.pdf.

J. D. Geanakoplos and H. Polemarchakis. Intertemporally separable overlapping generation economies. *Journal of Economic Theory*, 34:207–215, 1984. URL http://www.polemarchakis.org/a27-iso.pdf.

S. Hart and Y. Taumann. Market crashes without external shocks. *Journal of Business*, 77:1–8, 2004.

L. T. Nielsen. Common knowledge, communication, and convergence of beliefs. *Mathematical Social Sciences*, 8:1–14, 1984.

R Parikh and P.Krasucki. Consensus, and knowledge, *J*, 52, 178-189. *Journal of . of Economic Theory*, 52 (178-189), 1990.

D. Shaiderman. Variation of the conditional means of a martingale. Unpublished manyscript, 2018.

T. Tsuchikura and M. Yamasaki. Martingale sequence of bounded variation 123-127. *Tohoku Mathatics Journal*, pages 123–127, 1976.