Rational Dialogues

John Geanakoplos, Herakles Polemarchakis

Rational Dialogues

John Geanakoplos*
Herakles Polemarchakis**

Any finite conversation, no matter how crazy it sounds, can be given context in which it is a rational dialogue.

DIALOGUES RATIONNELS

Toute conversation, aussi étrange qu'elle paraisse, est un dialogue rationnel dans un contexte approprié.

Keywords: dialogue, rationality

Mots clés : dialogue, rationalité

JEL Codes: D83.

INTRODUCTION

In the summer of 1977 Herakles came to John very excited about a paper of Bob Aumann's on common knowledge. We couldn't believe the paper, much less figure it out. Our adviser Kenneth Arrow couldn't either. Even the title "Agreeing to Disagree" seemed to say the opposite of the paper's conclusion.

There is nothing more tantalizing than a paradox. As Aumann has managed with other young students time and again, he had us hooked. His teaching style builds on paradoxes. He once described capitalism through a letter he had gotten from his son about life in the Kibbutz. In the morning, the son had written, I do something for my community and in the afternoon something for myself. Aumann said, wouldn't it be better if by doing something for himself he was at the same time doing something for his community? He described integration and

the fundamental theorem of calculus as showing that it is easier to solve many hard problems than a single hard problem. At a conference in India, he was asked by a reporter to say a word explaining game theory. Aumann replied that the question reminded him of Nikita Khrushchev's first press conference in front of foreign journalists. A reporter asked Khrushchev to say a word about the health of the Russian economy. Khrushchev said "Good." The reporter said he didn't literally mean one word, could Khrushchev say two words about the health of the Russian economy? Khrushchev replied "Not good." Aumann continued by saying that in one word game theory is about "interaction." In two words, it is about "rational interaction."

A parodox is something that sounds crazy, but looked at the right way makes sense. For Aumann, paradoxes abound. In honor of Bob Aumann, we prove here that paradoxes are ubiquitous. We show that any conversation, no matter how crazy the opinions and the rejoinders sound, can be explained as the first part of a dialogue between two perfectly rational interlocutors. Dialogues are interactions. In the right context they might all be rational interactions.

Turing famously suggested that one could distinguish a (non-thinking) machine from a man by engaging it in conversation and then letting a panel of judges review the transcript and vote man or machine. As is becoming clearer today with ChatGPT, and as our theorem suggests, it may not be as easy as Turing hoped.

Bob Aumann himself has often written that what is called irrational behavior by behavioral economists might one day be better understood as rational behavior in a complicated environment with constraints. Our theorem has a similar flavor. Perhaps the most comforting aspect of our theorem is that it provides some hope for our current troubled and polarized discourse.

Aumann [1976] defined common knowledge and proved that consensus is a necessary condition for common knowledge, that is, that people cannot agree to disagree about the probability of an event. A Bayesian dialogue is a sequential exchange of beliefs about the probability of an event. It is the prototype of a rational dialogue. One of two interlocutors states his belief, then the other responds with her belief, perhaps informed or influenced by his stated belief. He then responds, perhaps with a revision of his prior opinion (in view of her opinion), and then she responds again, and so on. The dialogue is said to terminate at a time *T* if neither agent changes his or her mind thereafter. In Geanakoplos and Polemarchakis [1982], we proved that Bayesian dialogues must always terminate, and that when they do, the agents are in agreement.[1]

We show here that a third party, with access only to the transcript of a dialogue, cannot be sure that any arbitrary finite sequence of alternating opinions is not part of a Bayesian dialogue. If the transcript were infinitely long, then it would necessarily terminate in agreement. We show that the available finite transcript of opinions can always be continued to reach an agreement in such a way that the whole dialogue from the beginning is rational.

Our argument covers the special case of a *didactic dialogue*, in which an expert is better informed than his interlocutor. The expert never changes his opinion, but the interlocutor follows an arbitrary path. Some of Plato's dialogues might

---

1. We allowed for an arbitrary but finite state space. Bacharach [1979] looked at Bayesian dialogues when information is normally distributed. Nielsen [1984] considered dialogues with an uncountable number of states.

be considered didactic dialogues in our sense. Socrates knows the right answer to which he leads his interlocutor. Plato perhaps understood our theorem in the sense that in some of his dialogues he has an interlocutor of Socrates, such as Protagoras, appear at first to move further away from the answer until eventually coming back to the right path.

Our theorem relies on one important premise. If an agent expresses absolute certainty in her opinion, then her interlocutor must immediately agree. Absolute certainty is tantamount to claiming a proof. If the interlocutor does not agree, then one or the other cannot be rational. She can be 99.9999% certain of one thing, and then 99.9999% certain of the opposite at the next stage; as long as neither she nor he is 100% certain, then whatever her interlocutor and she say can be rationalized.

Loosely speaking, one can consider common knowledge and agreement as an equilibrium, and the dialogue that leads to common knowledge as the adjustment path. We are arguing that along the adjustment path, anything goes. This bears an analogy with general competitive analysis. As follows from Debreu [1974], the Walrasian *tâtonnement* that leads to equilibrium, if it does, is arbitrary.

THE ARGUMENT

Bayesian Dialogues

A *Bayesian opinion framework* is defined by a finite probability space, a subset, two partitions, and an agent,

$$(\Omega, \pi, A, P, Q, i),$$

where Ω is a *finite* set of states and π is a strictly positive probability on Ω, and A is a subset of Ω. The probability π is the common prior of two agents p, q. P and Q are partitions of Ω, corresponding to the two agents p, q, defined by disjoint subsets or cells (P_c) and (Q_d), $c = 1, \ldots, C$ and $d = 1, \ldots, D$ respectively. For any $\omega \in \Omega$, $P(\omega)$ is defined as the unique cell P_c containing ω, and likewise for $Q(\omega)$. Finally, the agent $i \in \{p, q\}$.

The *Bayesian opinion* of agent $i = p$ about the likelihood of A, conditional on what p knows, is defined by the function $i_A = p_A : \Omega \to [0, 1]$

$$i_A(\omega) = p_A(\omega) = \frac{\pi(P(\omega) \cap A)}{\pi(P(\omega))},$$

and, likewise, when $i = q$,

$$i_A(\omega) = q_A(\omega) = \frac{\pi(Q(\omega) \cap A)}{\pi(Q(\omega))},$$

defines q's Bayesian opinion of the likelihood of A conditional on what q knows.

If $i = p$, then, after hearing p's Bayesian opinion p_A, q will revise her understanding of the world, replacing Q with $Q' = Q \vee p_A$ defined by

$$[Q \vee p_A](\omega) = Q(\omega) \cap \{\omega' : p_A(\omega') = p_A(\omega)\} \text{ for all } \omega \in \Omega.$$

Similarly, if $i = q$, then after hearing Q's Bayesian opinion q_A, p will revise his understanding of the world, replacing P with $P' = P \vee q_A$ defined by

$$[P \vee q_A](\omega) = P(\omega) \cap \{\omega' : q_A(\omega') = q_A(\omega)\} \text{ for all } \omega \in \Omega.$$

Thus the Bayesian opinion framework $(\Omega, \pi, A, P, Q, p)$ generates a unique successor $(\Omega, \pi, A, P', Q', -p) = (\Omega, \pi, A, P, [Q \vee p_A], q)$ and the Bayesian opinion framework $(\Omega, \pi, A, P, Q, q)$ generates a unique successor

$$(\Omega, \pi, A, P', Q', -q) = (\Omega, \pi, A, [P \vee q_A], Q, p).$$

It follows that any Bayesian opinion framework $(\Omega, \pi, A, P, Q, i)$ generates a uniquely defined infinite sequence of Bayesian opinion frameworks:

$$(\Omega, \pi, A, P, Q, i) = (\Omega, \pi, A, P_1, Q_1, i_1),$$

$$(\Omega, \pi, A, P_2, Q_2, i_2 = -i),$$

$$(\Omega, \pi, A, P_3, Q_3, i_3 = i),$$

$$(\Omega, \pi, A, P_4, Q_4, i_4 = -i),$$

$$. . .$$

in which, at each period, one agent i gives his opinion based on his partition at that time, and then in the next period the *other* agent $-i$ gives her opinion based on her previous partition revised in light of the previous opinion expressed by him. We call this whole infinite sequence a *Bayesian Dialogue* $(\Omega, \pi, A, P, Q, i)_\infty$.

Dialogues and Rational Dialogues

Bayesian dialogues contain many counterfactual statements, covering opinions conditional on all possible worlds $\omega \in \Omega$. In reality we typically only hear about a *finite* number of *actual* opinions. We define a *dialogue* as a finite sequence of opinions or beliefs $(b_1, b_2,\dots, b_T)$ with $b_t \in [0, 1]$ for all t. Once we specify a fixed state of the world $\omega^* \in \Omega$, every Bayesian dialogue generates an infinite sequence of beliefs $(\Omega, \pi, A, P, Q, i, \omega^*) \equiv (r_1, r_2,\dots)$ where $r_t = i_{t,A}(\omega^*)$ is the opinion expressed by the opining agent at time t for state ω^*. We call this infinite sequence of opinions a Bayesian dialogue at a fixed state. A *rational dialogue* is any finite sequence of beliefs $(r_1, r_2,\dots, r_T)$ that can be realized as the first part of a Bayesian dialogue at a fixed state.

Geanakoplos and Polemarchakis [1982] showed that in any Bayesian dialogue at a fixed state, $(r_1, r_2,\dots)$, there must be a finite time T by which consensus is reached, $r_1 = r_T$ for all $t \geq T$. Moreover, because Bayesian rational agents believe

in each other's rationality, if for some T, $r_T \in \{0, 1\}$, meaning one of the agents is absolutely certain and will never change his/her mind, then consensus must have already been reached by time T.

We say that the event $E \subset \Omega$ is common knowledge at $\omega^* \in E$ if $P(\omega) \cup Q(\omega) \subset E$ for all $\omega \in E$. It is evident that if $E \subset \Omega$ is common knowledge at ω^*, then the Bayesian dialogue at a fixed state $\omega^* (\Omega, \pi, A, P, Q, i_A, \omega^*) \equiv (r_1, r_2, \ldots)$ does not depend on any $P(\omega)$, $Q(\omega)$ or $\pi(\omega)$ for $\omega \notin E$.

Irrational Dialogues?

Are there dialogues $(b_1, \ldots, b_T)$ that look so crazy that they could not be the beginnings of a rational dialogue? Geanakoplos and Polemarchakis [1982] showed that given any positive integer n, there is a Bayesian dialogue $(c, d, c, d, \ldots, c, d, c, c, \ldots)$ in which one agent obstinately maintains the opinion c while the other maintains $d \neq c$ and then, suddenly, after n such alternations, consensus is reached at c.

In an unpublished paper Polemarchakis [2016] showed that any dialogue could be rational. Di Tillio, Lehrer and Samet [2022] extended the theorem to infinite dialogues with a countably infinite state space. The following theorem gives a similar result to Polemarchakis [2016] but in a slightly different setting and with a different proof.

There is one property that must hold for any rational dialogue, because rationality presumes both agents are rational and know that both are rational. If one of the agents is certain, then the other must immediately agree. Certainty is tantamount to claiming a proof, and if the other does not agree, one of the two interlocutors must not be rational.

> DEFINITION. *The dialogue* $(b_1, \ldots, b_T)$ *violates certainty acquiescence if for some $t < T$, $b_t \in \{0, 1\}$, yet $b_{t+1} \neq b_t$.*

Needless to say, if in the dialogue $(b_1, \ldots, b_T)$ nobody expresses absolute certainty, then the dialogue does not violate certainty acquiescence. The opinions could bounce around arbitrarily, as long as none hit 0 or 1.

> THEOREM. *Let* $(b_1, \ldots, b_T)$ *be an arbitrary dialogue that does not violate certainty acquiescence. Then* $(b_1, \ldots, b_T)$ *is a rational dialogue generated by some* $(\Omega, \pi, A, P, Q, p_A, \omega^*)_\infty$. *Moreover, consensus is reached at time T at* b_T.

Proof. The proof is by backward induction. Suppose $T = 1$. Suppose $0 < b_T < 1$. Let $\Omega = \{y, n\}$ and let $A = \{y\}$. Let $\pi(y) = b_T$, and $\pi(n) = 1 - b_T$. Let $P = Q = \{\{y, n\}\}$. Let $\omega^* = y$. Clearly consensus is reached at b_T because both agents have the same information.

If $b_T = 1$, delete the point n, and continue as above. If $b_T = 0$, let $\Omega = \{n\}$, and let $A = \phi$ and let $P = Q = \{\{\omega^* = n\}\}$ and let $\pi(n) = 1$. Clearly in all three cases the rational dialogue reaches consensus on the first step at b_T, no matter which agent is the first opiner, because both agents have the same information and reveal nothing when they opine.

Now suppose the theorem has been proved for all $T \leq n$ and let $(b_1, b_2, \ldots, b_T)$ be a given dialogue that does not violate certainty acquiescence for $T = n + 1$. By the induction hypothesis we can find a Bayesian dialogue at a fixed state

$(\Omega, \pi, A, P, Q, q, \omega^*)_\infty = (b_2, \ldots, b_T, \ldots)$, with consensus at b_T, in which q is the first speaker (with opinion b_2). As part of the induction hypothesis, we can assume that q reveals nothing with her initial opinion b_2. We shall now define

$$(\Omega^*, \pi^*, A^*, P^*, Q^*, p, \omega^{**})_\infty \equiv (r_1, \ldots, r_T, \ldots)$$

with consensus at time T at r_T, with $r_t = b_t$ for $t = 1, \ldots, T$ in which p is the first speaker. If $b_1 = 1$, then by certainty acquiescence, $b_t = 1$, for all t and we can rationalize that with the Bayesian dialogue in the second paragraph, and similarly if $b_1 = 0$.

So suppose $0 < b_1 < 1$. Define Ω^* and P^* by adding to Ω two extra points y_c, n_c for *each* partition cell P_c, so $P_c^* = P_c \cup \{y_c, n_c\}$. The partition Q^* adds two cells to those already in Q, namely Q_y^* consisting of all the y_c, and the other Q_n^* consisting of all the n_c. A^* extends A by including also all the y_c. This situation is depicted in Diagrams 1 and 2.

Diagram 1.

	Q_1	$\cdots$	Q_D
P_1	ω^*		
P_2			
$\cdots$			
P_C			

Diagram 2.

	Q_1^*	$\cdots$	Q_D^*	Q_y^*	Q_n^*
P_1^*	ω^*			y_1	n_1
P_2^*				y_2	n_2
$\cdots$					
P_C^*				y_C	n_C

The crucial step is to note that for *every* partition cell P_c, there exists numbers $0 < \pi(y_c) < 1$ and $0 < \pi(n_c) < 1$ such that

$$b_1 = \frac{\pi(A \cap P_c) + \pi(y_c)}{\pi(P_c) + \pi(y_c) + \pi(n_c)}.$$

This extends the probability measure π to a measure on all of Ω^*. Define π^* by rescaling the π (over all Ω^*) so that they add to 1. Observe that the rescaling in numerator and denominator cancel, so for *all* P_c,

$$b_1 = \frac{\pi(A \cap P_c) + \pi(y_c)}{\pi(P_c) + \pi(y_c) + \pi(n_c)} = \frac{\pi^*(A^* \cap P_c^*)}{\pi^*(P_c^*)}.$$

Take $\omega^{**} = \omega^*$. This completes the definition of the Bayesian dialogue announced above.

At the first step agent p announces

$$p_A(\omega^{**}) \equiv r_1 = \frac{\pi^*(A^* \cap P^*(\omega^{**}))}{\pi^*(P^*(\omega^{**}))} = b_1,$$

and reveals nothing, because, as noted, for every partition cell p would announce the same, hence at the next step $Q^{*\prime} = Q^*$. But $\omega^{**} = \omega^* \in \Omega$, hence by construction $Q^*(\omega^{**}) = Q^*(\omega^*) = Q(\omega^*) \subset \Omega$. Thus q then announces

$$\frac{\pi^*(A^* \cap Q^{*\prime}(\omega^{**}))}{\pi^*(Q^{*\prime}(\omega^{**}))} = \frac{\pi(A \cap Q(\omega^*))}{\pi(Q(\omega^*))} = b_2,$$

where the last equality follows from the induction hypothesis and the fact that π^* scales π. If $b_2 \in \{0, 1\}$, then this rational dialogue, like all rational dialogues, repeats b_2 thereafter, reproducing the given dialogue which, by certainty acquiescence would also have to repeat b_2 thereafter.

So suppose $0 < b_2 < 1$. Then, by the last part of the induction hypothesis, this announcement of b_2 reveals precisely that $\omega^{**} \in \Omega$, because had q seen partition cell Q_y^* or Q_n^*, she would have announced 1 or 0 instead of b_2. Thus the Bayesian dialogue at a fixed state $(\Omega^*, \pi^*, A^*, P^*, Q^*, p, \omega^{**})$ begins with b_1 and b_2, and then from step 2 onwards proceeds as the Bayesian dialogue with a fixed state $(\Omega, \pi, A, P, Q, q, \omega^*)$. ∎

REMARKS

An Example of the Construction. The constructive argument above can generate any dialogue, no matter how curious. For example, the two agents could agree with each other on say the probability 1/4 for many iterations, and then suddenly jump to consensus at 3/4.

We give the construction for the dialogue

$$\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{3}{4}, \frac{3}{4} \right)$$

in the matrix below, where each y corresponds to a different state in A and each n corresponds to a different state in $\Omega \backslash A$, and the numbers in the brackets are measures for the corresponding states. The probability measure that is the common prior of the agents is given by normalizing these measures to add to one, namely the numbers in brackets divided by $13\frac{2}{3}$. Observe that conditional probabilities are not affected by replacing a measure with any scalar multiple of the measure[2]. The partition of agent p consists of the rows of the matrix, and the partition of agent q corresponds to the columns of the matrix. The state of nature

---

2. In the example some of the states are assigned measure zero. The example could be modified to make all the measures strictly positive, at the cost of more complicated fractions.

ω^* is the y in the top left corner. Notice that the top left cell of the matrix is the only one containing two points.

$y[3/4], n[1/4]$	$y[0]$	$n[2]$	$y[0]$	$n[0]$
$y[0]$	$y[1/4]$	$n[3/4]$	$y[0]$	$n[0]$
$n[2]$	$y[2/3]$	$n[0]$	$y[0]$	$n[0]$
$y[0]$	$y[0]$	$y[1]$	$y[0]$	$n[3]$
$n[0]$	$n[11/4]$	$n[1/4]$	$y[1]$	$n[0]$

The reader can check that p will announce $1/4 = (3/4)/(3/4 + 1/4 + 2)$, revealing nothing since the conditional probability of A given any row is exactly $1/4$. Then q will announce $1/4$, since the conditional probability of A in the left most column (and also in the second and third columns) is $1/4$. That reveals precisely that ω^* is not in one of the last two columns, since they would have led to the announcements of 1 or 0. With this information, p still says $1/4$, since that is the conditional probability of A given the top row without its last two elements. This announcement reveals precisely that ω^* is not in the bottom two rows, since they would have led to the announcements of 1 or 0. Agent q responds to this by still saying $1/4$ since that is the probability of A given the first column without its last two elements. That reveals to p that ω^* is not the second or third columns, since they would have led to the announcements of 1 or 0. With this information, p finally says $3/4$. This reveals that ω^* is in the top left cell, and gets agreement from q at $3/4$.

The measure makes clear how the probabilities were constructed by backward induction. The top left cell is first in the construction. If that cell were common knowledge, p and q would agree on $3/4$, giving the last two opinions $(3/4, 3/4)$ in the dialogue. Next we add the second and third elements of the first column. The measures assigned to y and n induce q to assign conditional probability of $1/4$ to seeing this part of the first column. Thus we can generate the dialogue $(1/4, 3/4, 3/4)$.

$$y[3/4], n[1/4] \quad \rightarrow \quad \begin{matrix} y[3/4], n[1/4] \\ y[0] \\ n[2] \end{matrix} \quad \rightarrow$$

Next we added the second and third columns of the first three rows, assigning the measures to make sure that player p gives conditional probability of A of $1/4$ to each row so far constructed.

$y[3/4], n[1/4]$	$y[0]$	$n[2]$
$y[0]$	$y[1/4]$	$n[3/4]$	$\rightarrow$
$n[2]$	$y[2/3]$	$n[0]$

This gives us a dialogue $(1/4, 1/4, 3/4, 3/4)$. Next we move to add the fourth and fifth rows, as indicated below, so that q assigns the same probability $1/4$ to A in each column. This gives us a dialogue $(1/4, 1/4, 1/4, 3/4, 3/4)$. Finally we add the last two columns so that p gives conditional probability of A of $1/4$ to each row, giving us the whole dialogue $(1/4, 1/4, 1/4, 1/4, 3/4, 3/4)$.

y[3/4], n[1/4]	y[0]	n[2]	y[3/4], n[1/4]	y[0]	n[2]	y[0]	n[0]
y[0]	y[1/4]	n[3/4]	y[0]	y[1/4]	n[3/4]	y[0]	n[0]
n[2]	y[2/3]	n[0] →	n[2]	y[2/3]	n[0]	y[0]	n[0]
y[0]	y[0]	y[1]	y[0]	y[0]	y[1]	y[0]	n[3]
n[0]	n[11/4]	n[1/4]	n[0]	n[11/4]	n[1/4]	y[1]	n[0]

Experts. We described a didactic dialogue as one in which an expert leads a student through a conversation. At a state of the world, ω^*, individual 1 is an expert concerning the event A if no information in the join (coarsest refinement of the partitions of the individuals) would cause him to alter his beliefs.

A *didactic dialogue* is a dialogue

$$(\bar{q}, q^2,\ldots, \bar{q}, q^{2^t},\ldots, \bar{q}),\ \ 0 \leq \bar{q},\ldots,q^T \leq 1,$$

where the opinion at t odd is unchanging, $q^{2t+1} = \bar{q}$, and at t even, q^{2t}, is arbitrary but never 1 or 0.

COROLLARY. *Any didactic dialogue* $(\bar{q}, q^2,\ldots, \bar{q}, q^{2^t},\ldots, \bar{q})$ *is a rational dialogue.*

The following matrix of states and measures displays a Bayesian opinion framework, with the row player opining first. At the fixed state given by y in the to left, this gives a Bayesian dialogue with the opinions

$$\left(\frac{3}{4},\frac{1}{4},\frac{3}{4},\frac{1}{4},\frac{3}{4},\frac{3}{4},\ldots\right)$$

y[3/4], n[1/4]	y[0]	n[0]	y[0]	n[0]
y[0]	y[3/4]	n[1/4]	y[0]	n[0]
n[2]	y[6]	n[0]	y[0]	n[0]
y[0]	y[0]	y[1/12]	y[0]	n[1/36]
n[0]	n[81/4]	n[0]	y[243/4]	n[0]

Silence. Here, a dialogue is an alternating sequence of opinions. Formally, an interlocutor cannot remain silent when it is her turn to speak. Herbert Scarf, a mentor of ours at Yale, said that for a teacher the most important thing is where to put the silences. One can interpret silence by an interlocutor at t as the repetition of her opinion at $t-2$: that is, $b_t = b_{t-2}$. Thus if our tape contains only the opinion of one agent that is changing over time, we can interpret it as a conversation with an expert who constantly repeats the same opinion.

REFERENCES

Aumann, R. J. [1976]. "Agreeing to Disagree," *Annals of Statistics*, 4 (6): 1236–1239.

Bacharach, M. [1979]. "Normal Bayesian Dialogues," *Journal of the American Statistical Association*, 74 (368): 837–846.

Debreu, G. [1974]. "Excess Demand Functions," *Journal of Mathematical Economics*, 1 (1): 15–21.

Di Tillio, A., Lehrer, E. and Samet, D. [2022]. "Monologues, Dialogues and Commom Priors," *Theoretical Economics*, 17 (2): 587–615.

Geanakoplos, J. D. and Polemarchakis, H. M. [1982]. "We Can't Disagree Forever," *Journal of Economic Theory*, 28 (1): 192–200.

Nielsen, L. T. [1984]. "Common Knowledge, Communication, and Convergence of Beliefs," *Mathematical Social Sciences*, 8 (1): 1–14.

Polemarchakis, H. [2016]. "Rational Dialogs," *CRETA Discussion Paper*, 19.