Comment on Leitgeb's Stability Theory of Belief

Hanti Lin Kevin T. Kelly
Carnegie Mellon University
{hantil, kk3n}@andrew.cmu.edu

Hannes Leitgeb's stability theory of belief provides three *synchronic* constraints on an idealized agent's degrees of belief and the propositions she believes. The theory requires that, for each instant of time, an idealized agent satisfies the following three *synchronic* conditions:

(P1) The set of one's beliefs is consistent and closed under deduction.

(P2) One's degrees of belief satisfy the axioms of probability.

(P3) *Lockean Thesis*: there exists a threshold r with $1/2 < r \leq 1$ such that one's beliefs turn out to be exactly the propositions that have degree of belief at least r.

Given P1 and P2, Leitgeb shows that the Lockean thesis P3 is equivalent to the following condition: every proposition with probability one is believed, and the strongest proposition B_W that one believes is probabilistically *stable* with respect to one's current probability measure P in the sense that, for every proposition E consistent with B_W for which $P(E) > 0$, conditional probability $P(B_W|E)$ is greater than 1/2.

Given P, there may be more than one stable proposition, and the theory itself does not specify which to believe. Indeed, from what has been said so far, Leitgeb's synchronic theory is compatible with the *generalized odds-threshold* acceptance rules recommended in Lin & Kelly (2012).[1] Here is a simple geometrical recipe for constructing an odds-threshold rule that satisfies P1-P3 in the case of a ternary partition $\{w_1, w_2, w_3\}$. Recall Leitgeb's figure 2, which we reproduce as figure 1.a below. For each vertex V_i of the triangle, construct the line from V_i that divides the triangle in half. Now mark an arbitrary point P_i on the line that lies above Leitgeb's 1/2 threshold for $\{w_i\}$ (figure 1.b). Draw the straight line that passes through P_i and V_j for each distinct i, j (figure 1.b). The lines so constructed partition the triangle into regions. Label each region with the strongest proposition accepted by a Bayesian credal state in that region in the manner depicted (figure 1.c).[2] The resulting acceptance rule jointly satisfies Leitgeb's P1-P3 along with all of our requirements.[3]

Generalized odds threshold rules have a specific geometrical shape that is not mandated by Leitgeb's synchronic principles P1-P3. The motivation for that shape is based essentially

[1] Leitgeb's synchronic theory is also compatible with, say, the defeasiblity condition proposed by Douven (2002).

[2] Credal states that fall exactly on the boundary of a region may all fall to one side or may all fal to the other.

[3] In Lin & Kelly (2012), we likewise assume P1 and P2, but we do not assume P3 unless the underlying partition is binary, so we view the Lockean thesis not as a general norm of rationality, but as a framing effect that arises when one focuses on the partition ~~Q_1 vs. $\neg Q_1$~~ rather than on the partition ~~Q_1 vs. Q_2 vs. Q_3~~.

Just for consistency of notation.

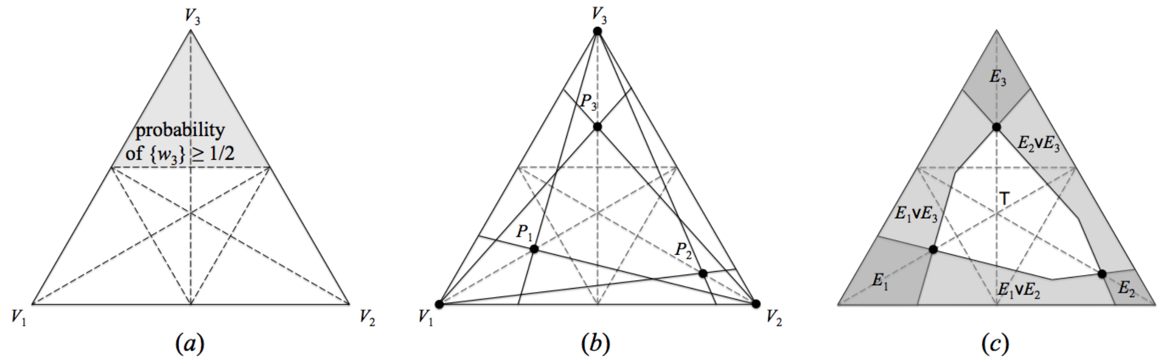

Figure 1: Compatibility of the odds threshold rule with Leitgeb's axioms.

upon *diachronic* considerations, and that is where we begin to disagree with Leitgeb. While we prefer to drop a key principle of the AGM belief revision theory (Alchourrón et al. 1985), Leitgeb expresses a desire to retain it in section 3 of his paper. In this note, we show that, if AGM belief revision is incorporated into Leitgeb's theory, a plausible, diachronic norm of case reasoning must be sacrificed.[4]

Leitgeb interprets his theory as providing nothing more than simultaneous constraints on one's beliefs and degrees of belief. To make that idea explicit, let relation $\mathsf{R}(P, Bel)$ mean that it is (synchronically) *permitted* for an agent with probabilistic credal state P to have belief set Bel. Then Leitgeb's thesis reads:

> (**Leitgeb's Thesis**) For every P and every Bel, if $\mathsf{R}(P, Bel)$, then P and Bel jointly satisfy conditions P1-P3.

Synchronic permission leads naturally to a diachronic constraint on belief revision. Suppose that $\mathsf{R}(P, Bel)$ and that, upon receipt of new information E such that $P(E) > 0$, the agent is obligated to update P by conditioning[5] to obtain P_E and revises Bel by a propositional belief revision operator $*$, to obtain new belief state $Bel * E$. Then say that $*$ is *diachronically* admissible for P and Bel if and only if $\mathsf{R}(P_E, Bel * E)$, for all E.[6] The most fundamental assumption of the AGM theory of belief revision is *accretiveness*, which requires that, whenever Bel and E are compatible, $Bel*E$ be the set of the logical consequences of $Bel \cup \{E\}$, written $Bel + E$. Since the accretiveness condition holds for all belief revision operators in AGM theory, the assumption that AGM theory is *obligatory* together with the requirement that

[4]Our argument against AGM in Lin & Kelly (2012) assumes that the agent must possess a comprehensive *policy* for accepting propositions and for revising them in light of new information—an assumption that Leitgeb does not subscribe to. The impossibility argument in this note is carried out in Leitgeb's much weaker setting.

[5]$P_E(X)$ is defined to be $P(X \wedge E)/P(E)$ if $P(E) > 0$; otherwise it is undefined.

[6]In Lin & Kelly (2012), we require that $*$ be a belief revision *policy* defined for all pairs Bel, E, rather than just for a fixed Bel and all E.

belief revision be diachronically admissible entails the following requirement, which Leitgeb endorses in section 3 of his paper:

(**Diachronic Admissibility of Accretive Belief Revision**) For every Bel, every P, and every E such that E is consistent with Bel and that $P(E) > 0$, if $\mathsf{R}(P, Bel)$ then $\mathsf{R}(P_E, Bel{+}E)$.

Unfortunately, the two principles Leitgeb endorses—Leitgeb's Thesis and Diachronic Admissibility of Accretive Belief Revision—are jointly incompatible with another plausible, diachronic principle relating acceptance to belief revision. Let $\mathsf{R}(P, \Diamond(A))$ say that it is *permitted* for an agent with credal state P to believe A in the following sense:

$$\mathsf{R}(P, \Diamond(A)) \quad \text{iff} \quad \exists Bel : \mathsf{R}(P, Bel) \wedge Bel(A).$$

Similarly, let $\mathsf{R}(P, \Box(A))$ say that it is *required* for an agent with credal state P to believe A in the following sense:

$$\mathsf{R}(P, \Box(A)) \quad \text{iff} \quad \forall Bel : \mathsf{R}(P, Bel) \rightarrow Bel(A).$$

The following is a form of reasoning by cases, the case of learning E and the case of learning its negation:

(**Diachronic Admissibility of Case Reasoning**) For every probability measure P as an agent's current credal state, every proposition E with $P(E) > 0$, and every proposition A, if $\mathsf{R}(P_E, \Box(A))$ and $\mathsf{R}(P_{\neg E}, \Diamond(A))$, then $\mathsf{R}(P, \Diamond(A))$.

In words, if it required for an agent with probabilistic credal state P to be believe A upon learning E, and if it is permitted for her to believe A upon learning $\neg E$, then it is already permitted for her to believe A prior to learning any new information. For example, for a couch potato who never ventures outdoors, it is obligatory for her to believe that she will not contract poison ivy given that she has genetic immunity against it and it is at least admissible for her to believe that she will not get it otherwise, since she has not traveled outdoors. So it should be admissible for her to believe that she won't get it. Then we have the following impossibility result:

Theorem. *Suppose that the set W of possible worlds is finite with cardinality $|W| \geq 3$. Then there is no relation R that satisfies all of the following six conditions:*

1. **Domain of Application.** *For every P and every Bel such that $\mathsf{R}(P, Bel)$, P is a probability measure over W and Bel is a belief set over W that is consistent and deductively closed.*

2. **Probability 1/2 Rule.** *For every probability measure P and every proposition A, if $\mathsf{R}(P, \Diamond(A))$, then $P(A) > 1/2$.*

3. **Probability 1 Rule.** *For every probability measure P and every proposition A, if $P(A) = 1$, then $\mathsf{R}(P, \Box(A))$.*

4. **Non-Skepticism.** *There exist some world $w_i \in W$ and some probability measure P such that $\mathsf{R}(P, \Diamond(\{w_i\}))$ and $P(\{w_i\}) < 1$.*

5. **Diachronic Admissibility of Accretive Belief Revision.**

6. **Diachronic Admissibility of Case Reasoning.**

Conditions 1-3 are central to Leitgeb's Thesis. Condition 1 is P1 plus P2. Condition 3 follows from the requirement in P3 that $r \leq 1$. Condition 2 (the Probability 1/2 Rule) follows from the range of the threshold r in Leitgeb's P3. Condition 4 is a minimal requirement of adequacy for any joint theory of beliefs and probabilities. Under those four conditions, we have to give up either condition 5 (Diachronic Admissibility of Accretive Belief Revision) or condition 6 (Diachronic Admissibility of Case Reasoning).

In light of the impossibility theorem just presented, something has to go. We resolutely propose to drop the AGM principle of accretive belief revision in a precisely delimited range of circumstances. In our theory of joint revision of beliefs and probabilities (Lin & Kelly 2012), the rankings that represent AGM belief revision are replaced by partial orders,[7] which satisfy case reasoning but sometimes violate accretive belief revision. Our choice is justified by the fruitful results it yields (Lin & Kelly 2012, Lin 2013). Insistence on case reasoning connects qualitative belief with dominance arguments in decision-making and policy-making. Relaxation of accretive belief revision enables us to construct a Bayesian model of Lehrer's (1965) no-false-lemma variant of Gettier counterexample to the thesis that knowledge is justified true belief. Furthermore, relaxation of accretive belief revision also enables one to extend the theory of acceptance to systematic *policies* for *conditional acceptance* that determines, for each P and E, ~~what an agent with degrees of belief P should accept given information E~~. Let B denote such a policy, so that $\mathsf{B}(P, E)$ denotes the strongest proposition an agent with credal state P ~~should~~ have after receiving the information that E. It is very natural to say that such a policy is diachronically admissible if and only if: $\mathsf{B}(P, E) = \mathsf{B}(P_E, \top)$. A conditional acceptance policy is *accretive* if and only if it satisfies $\mathsf{B}(P, E) = \mathsf{B}(P, \top) + E$ when E is logically compatible with $\mathsf{B}(P, \top)$. In Lin & Kelly (2012), we show that every diachronically admissible, accretive belief revision policy is trivial, in the sense that it is either (i) skeptical (it accepts a proposition only at probability 1), (ii) ~~gullible (almost surely fails to accept some strongest proposition)~~, or (iii) bizarrely non-monotonic with respect to probability (following a straight line toward the credal state that puts probability 1 on proposition $\{w_i\}$ makes one *drop* one's belief that $\{w_i\}$). Leitgeb's minimalist approach to acceptance theory is immune to that result, but at the expense of leaving it to the judgment

[7]Shoham (1987) invented revision based on partial orders in the context of nonmonotonic logic.

There is no unique strongest proposition that one should believe.

of the reader how belief revision should fit together across Bayesian credal states.[8] Indeed, standard AGM belief revision theory does not specify such a policy, but we respond that belief revision policies cannot be avoided if one wishes to discuss (a) the *intersubjectivity* of assertions between Bayesian agents, (b) counterfactual conditionals about what a fixed agent *would have* believed in light of E had her degrees of belief been different from what they actually are, (c) the theory of acceptance of a single agent when information is *not certain* (i.e., Jeffrey conditioning), or (d) how to write an A. I. program for acceptance for a specific individual that takes that individual's underlying credal function P as a free parameter.

Acknowledgement

The authors are indebted to Hannes Leitgeb for stimulating questions during Hanti Lin's PhD dissertation defense, which directly inspires this note.

References

Alchourrón, C.E., P. Gärdenfors, and D. Makinson (1985), "On the Logic of Theory Change: Partial Meet Contraction and Revision Functions", *The Journal of Symbolic Logic*, 50: 510-530.

Douven, I. (2002) "A New Solution to the Paradoxes of Rational Acceptability", British Journal for the Philosophy of Science 53: 391-410.

Lehrer, K. (1965) "Knowledge, Truth, and Evidence", *Analysis* 25: 168-175.

Leitgeb, H. (manuscript) "The Stability Theory of Belief".

Levi, I. (1967) *Gambling With Truth: An Essay on Induction and the Aims of Science*, New York: Harper & Row. 2nd ed. Cambridge, Mass.: The MIT Press, 1973.

Lin, H. and Kelly, K. T. (2012) "Propositional Reasoning that Tracks Probabilistic Reasoning", *Journal of Philosophical Logic*, 41(6): 957-981.

Lin, H. (2013) *PhD Dissertation: Probabilistic Reasoning that Tracks Probabilistic Reasoning.*

Shoham, Y. (1987) "A Semantical Approach to Nonmonotonic Logics", in M. Ginsberg (ed.) *Readings in Nonmonotonic Reasoning*, Los Altos, CA: Morgan Kauffman.

[8] Our late colleague Horacio Arlo-Costa responded to our result in a similar way and we responded to him in a similar way.

Appendix: Proof of the Theorem

Suppose for reductio that R satisfies all the six conditions. Since $|W| \geq 3$, let w_1, w_2, w_3 be three distinct possible worlds in W. Denote proposition $\{w_i\}$ by E_i. The probability measure that assigns 1 to E_i is denoted by V_i. Let $\underline{Bel}$ denote the strongest proposition in Bel if it exists.

Let each probability measure P with $P(E_1 \vee E_2 \vee E_3) = 1$ be identified with the triple $(P(E_1), P(E_2), P(E_3))$ of real numbers. Let:

$$P \;=\; (x, 0, 1-x)\,.$$

Prove as follows that $\neg\mathsf{R}(P, \Diamond(E_1))$ for every $P = (x, 0, 1-x)$ with $0 \leq x < 1$.

Case (i): Suppose that $P = (x, 0, 1-x)$ with $0 \leq x \leq 1/2$ (figure 2.a). It follows immediately from the probability 1/2 rule (condition 2) that $\neg\mathsf{R}(P, \Diamond(E_1))$.

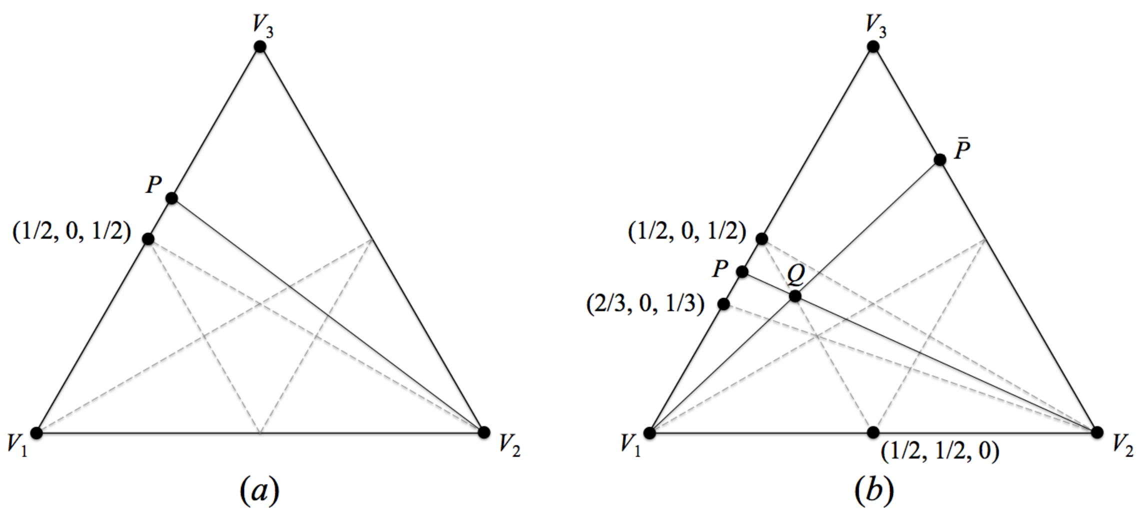

Figure 2: Case (i) in the left, case (ii) in the right

Case (ii): Suppose that $P = (x, 0, 1-x)$ with $1/2 < x \leq 2/3$ (figure 2.b). Suppose for reductio that $\mathsf{R}(P, \Diamond(E_1))$. Define Q as the unique probability measure such that $Q(E_1) = 1/2$ and $Q_{\neg E_2} = P$:

$$Q \;=\; \left(\frac{1}{2}, \frac{2x-1}{2x}, \frac{1-x}{2x}\right).$$

Namely, Q is the intersection of line $\overline{PV_2}$ and the line that passes through $(1/2, 0, 1/2)$ and $(1/2, 1/2, 0)$ (figure 2.b). We have that $\mathsf{R}(V_2, \Box(E_2))$, by the probability 1 rule (condition 3).

Then, since $Q_{E_2} = V_2$ and $Q_{\neg E_2} = P$ by construction, we have that $\mathsf{R}(Q_{E_2}, \Box(E_2))$ and that $\mathsf{R}(Q_{\neg E_2}, \Diamond(E_1))$. So, by closure under deduction (condition 1), we have that $\mathsf{R}(Q_{E_2}, \Box(E_1 \vee E_2))$ and that $\mathsf{R}(Q_{\neg E_2}, \Diamond(E_1 \vee E_2))$. Then, by the Probabilistic Norm of Case Reasoning (condition 6), we have that $\mathsf{R}(Q, \Diamond(E_1 \vee E_2))$. So there exists Bel such that $\mathsf{R}(Q, Bel)$ and $Bel(E_1 \vee E_2)$. Argue as follows that $\underline{Bel} = E_1 \vee E_2$. Since $x < 1$, it is routine to verify that both $Q(E_1)$ and $Q(E_2)$ are no more than 1/2. So, by the probability 1/2 rule (condition 2), $\underline{Bel}$ is neither E_1 nor E_2 and, hence, must be $E_1 \vee E_2$. Define $\bar{P}$ as the unique probability measure such that $\bar{P} = Q_{\neg E_1}$:

$$\bar{P} \quad = \quad \left(0, \frac{2x-1}{x}, \frac{1-x}{x}\right).$$

Namely, $\bar{P}$ is the intersection of line $\overline{V_2 V_3}$ and the line that passes through V_1 and Q. Note that $\neg E_1$ is consistent with $\underline{Bel}$. Then it follows from Diachronic Admissibility of Accretive Belief Revision (condition 5) that $\mathsf{R}(Q_{\neg E_1}, Bel{+}\neg E_1)$. Namely, $\mathsf{R}(\bar{P}, Bel')$, where $Bel' = Bel{+}\neg E_1$. Since $\underline{Bel} = E_1 \vee E_2$, we have that $Bel'(E_2)$. It follows that $\mathsf{R}(\bar{P}, \Diamond(E_2))$. But $\bar{P}(E_2) \leq 1/2$, which it is routine verify given the assumed range of x. So the probability 1/2 rule (condition 2) is violated—contradiction.

Case (iii): Suppose that $P = (x, 0, 1-x)$ with $2/3 < x < 1$. Recall how we define $\bar{P}$ from P; define the map f that sends P to $\bar{P}$ as follows:

$$\begin{aligned} f(P) &= f(x, 0, 1-x) \\ &= \left(0, \frac{2x-1}{x}, \frac{1-x}{x}\right) \\ &= \bar{P}. \end{aligned}$$

Define reflection map r with respect to vertices V_1 and V_2 as follows:

$$r(x_1, x_2, x_3) \quad = \quad (x_2, x_1, x_3).$$

Let $P_1 = P$. Whenever $P_i = (t, 0, 1-t)$ with $2/3 < t < 1$, define:

$$\begin{aligned} \bar{P}_i &= f(P_i), \\ P_{i+1} &= r^{-1} \circ f \circ r(\bar{P}_i). \end{aligned}$$

Namely, define $\bar{P}_i$ from P_i in the same way that we define $\bar{P}$ from P in case (ii), and define P_{i+1} from $\bar{P}_i$ also in the same way *except* that the roles of V_1 and V_2 are exchanged (figure 3.a). Now proceed to establish what we call the *snaking-up* lemma (cf. figure 3):

Lemma (Snaking-Up)**.** *For each $P_i = (t, 0, 1-t)$ with $2/3 < t < 1$, if $\mathsf{R}(P_i, \Diamond(E_1))$, then $\mathsf{R}(P_{i+1}, \Diamond(E_1))$.*

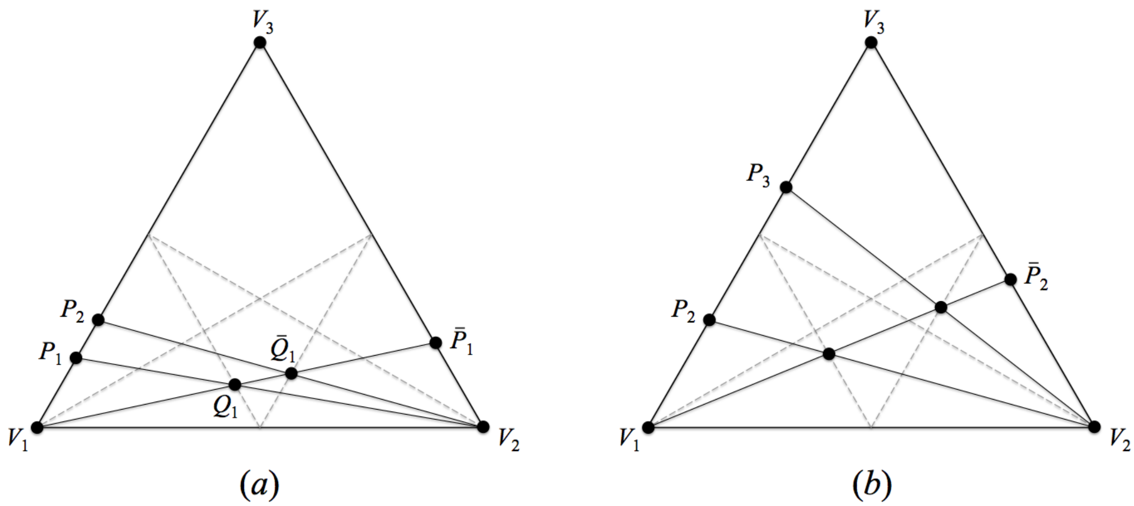

Figure 3: Case (iii)

Suppose that $\mathsf{R}(P_i, \Diamond(E_1))$, where $P_i = (t, 0, 1-t)$ with $2/3 < t < 1$. By the same argument as in case (ii) from $\mathsf{R}(P, \Diamond(E_1))$ to $\mathsf{R}(\bar{P}, \Diamond(E_2))$, we have that $\mathsf{R}(\bar{P}_i, \Diamond(E_2))$, namely that $\mathsf{R}(f(P_i), \Diamond(E_2))$. Now apply the very same argument—except with the roles of E_1 and E_2 exchanged—to the fact that $\mathsf{R}(\bar{P}_i, \Diamond(E_2))$. Then we have that $\mathsf{R}(r^{-1} \circ f \circ r(\bar{P}_i), \Diamond(E_1))$, namely that $\mathsf{R}(P_{i+1}, \Diamond(E_1))$, which completes the proof of the lemma. It is routine to verify that the function g that sends P_i to P_{i+1} can be expressed algebraically as follows:

$$g(t, 0, 1-t) \quad = \quad \left(\frac{3t-2}{2t-1}, 0, \frac{1-t}{2t-1}\right).$$

Let S be the following sequence of points on the edge $\overline{V_1V_3}$: P_1 $(= P), P_2, \ldots, P_i, P_{i+1}, \ldots$. It is routine to verify that, in sequence S, the probability of E_1 is decreasing. It is also routine to verify that, in sequence S, the amount of each decrease in the probability of E_1 is no less than a fixed number δ determined by $P_1 = (x, 0, 1-x)$, where:

$$\begin{aligned} \delta \quad &= \quad \left|\frac{3x-2}{2x-1} - x\right| \\ &= \quad \frac{(x-1)^2}{(x-\frac{1}{2})} \;>\; 0, \end{aligned}$$

because $2/3 < x < 1$. So the probability of E_1 in sequence S will eventually cease to be greater than $2/3$. Namely, there exists $n \geq 2$ such that:

$$P_n(E_1) \;\leq\; 2/3 < P_{n-1}(E_1) \;\leq\; P_1(E_1) \;=\; x.$$

An example with $n = 3$ is depicted in figure 3: the construction from P_1 to P_2 is in figure 3.a, and the construction from P_2 to P_3 is in figure 3.b. If $\mathsf{R}(P, \Diamond(E_1))$, namely if $\mathsf{R}(P_1, \Diamond(E_1)))$, then it follows from $n - 1$ applications of the snaking-up lemma that $\mathsf{R}(P_n, \Diamond(E_1))$, with $P_n(E_1) \leq 2/3$. The consequent of the conditional contradicts the results either in case (i) or in case (ii), so the antecedent must be false. That is, $\neg\mathsf{R}(P, \Diamond(E_1))$, as required.

We have established that $\neg\mathsf{R}(P, \Diamond(E_1))$ for every $P = (x, 0, 1 - x)$ with $0 \leq x < 1$. With the help of that result, argue as follows that, for every probability measure P, if $\mathsf{R}(P, \Diamond(E_1))$ then $P(E_1) = 1$. Suppose that $\mathsf{R}(P, \Diamond(E_1))$. By the probability 1/2 rule (condition 2), $P(E_1) > 1/2$, so $P_{E_1 \vee E_3}$ exists. Also, there exists Bel such that $\mathsf{R}(P, Bel)$ and $Bel(E_1)$. By condition 1, $\underline{Bel} = E_1$. So $E_1 \vee E_3$ is consistent with Bel. By Diachronic Admissibility of Accretive Belief Revision (condition 5), $\mathsf{R}(P_{E_1 \vee E_3}, Bel + (E_1 \vee E_3))$. So $\mathsf{R}(P_{E_1 \vee E_3}, \Diamond(E_1))$. Since $P_{E_1 \vee E_3} = (x, 0, 1 - x)$ for some x, it follows that $x = 1$, i.e. $P_{E_1 \vee E_3} = (1, 0, 0)$. Hence $P(E_1 \mid E_1 \vee E_3) = 1$. But the choice of w_3 is arbitrary as long as it is distinct fro w_1, so the result generalizes: $P(E_1 \mid E_1 \vee E_j) = 1$ for every possible world $w_j \in W$ distinct from w_1. It follows that $P(E_1) = 1$.

We have established that, whenever $\mathsf{R}(P, \Diamond(E_1))$, then $P(E_1) = 1$. But the choice of w_1 is arbitrary. So the result generalizes: for every possible world $w_i \in E$ and every probability measure P, if $\mathsf{R}(P, \Diamond(E_i))$ then $P(E_i) = 1$. That contradicts non-skepticism (condition 4), which completes the proof of the theorem.