Two Approaches to Belief Revision

Ted Shear[1] Branden Fitelson[2] Jonathan Weisberg[3]

[1]UC-Davis & University of Queensland
ebshear@ucdavis.edu

[2]Northeastern University
branden@fitelson.org

[3]University of Toronto
jonathan.weisberg@utoronto.ca

February 6, 2017

Today, we will be comparing two approaches to belief revision: a *Bayesian*/Lockean approach, and a *Logical*/AGM approach.

Here is an outline of the talk:

- Explain how the Bayesian/Lockean approach works,
- Explain how the Logical/AGM approach works,
- Compare and contrast these two approaches.
  - Along the way, we'll report some new results regarding the relationship between these approaches, and point out some connections to recent work (esp. Hannes's work on *stability*).
- First, some formal background and setup.

Our agents will possess *both* qualitative (yes/no) belief sets, $\mathbf{B}$, *and* numerical confidence/credence functions, $b(\cdot)$.

On the *belief* side, our agents entertain (classical, possible worlds) propositions on some finite *agenda* $\mathcal{A}$.

(1) $\mathbf{B}$ is the set of propositions in $\mathcal{A}$ believed by our agent.
  - *Note:* when $p \in \mathbf{B}$, we write $\mathsf{B}(p)$.

(2) Given a *prior* belief set $\mathbf{B}$, the *posterior* belief set $\mathbf{B}'$ is generated by revising the prior by E — *i.e.*, $\mathbf{B}' = \mathbf{B} \star E$.

On the *credence* side:

(3) $b(\cdot)$ is a classical (Kolmogorov) probability function.

(4) Given a *prior* $b(\cdot)$, the *posterior* $b'(\cdot)$ is generated *via conditionalization* by E — *i.e.*, $b'(\cdot) = b(\cdot \mid E)$.[1]

[1]Our results generalize to "minimum distance" [4] Bayesian updates satisfying (i) $b'(E) > b(E)$, (ii) $b'(E) \geq t$, and (iii) $b'(X) \geq t \Rightarrow b(E \supset X) \geq t$.

The Bayesian approach implies a (*weak*) *Lockean thesis* [5, 7].

$$\mathsf{B}(p) \quad \textit{only if} \quad b(p) \geq t.$$

It is well known [2, 9] that *strong* (*iff*) Bayesianism/Lockeanism leads to belief sets (*e.g.*, lotteries) that are *not deductively cogent*.

Cogency. An agent's belief set $\mathbf{B}$ is *cogent* iff it is (a) deductively *consistent* and (b) *closed* under logic.

We will not dwell on these well-known differences. Mainly, our talk will be about *cogent* Bayesian agents (similar to Leitgeb's).

But, first, we will briefly discuss "pure" Bayesianism. The *pure* Bayesian is a *strong Lockean* (*synchronically & diachronically*).

Pure Bayesian Revision. When one revises one's beliefs, one's posterior belief set $\mathbf{B}'$ *is Lockean* (in a *pure* sense), *i.e.*,

$$\mathbf{B}' = \mathbf{B} \circledast E := \{\, p \mid b(p \mid E) \geq t \,\}$$

To get a feel for how (pure) Bayesian revision revision works, it is helpful to note that it satisfies the following principle.

***Very* Weak Preservation.** If $E, X \in \mathbf{B}$, then $\neg X \notin \mathbf{B} \star E$.

In words: *learning something you already believe should never cause you to* ***dis****believe something you already believed.*

Strictly speaking, PBR does *not universally* satisfy VWP. But, it will — *so long as the Lockean threshold t is sufficiently large.*

Specifically, Pure Bayesian Revision is guaranteed to satisfy Very Weak Preservation, *provided that the Lockean threshold t is at least* $\phi^{-1} \approx 0.618$, where ϕ is the *Golden Ratio.*

Figure is helpful for visualizing why Pure Bayesian Revision, Very Weak Preservation, and The Golden Ratio are related in this way.

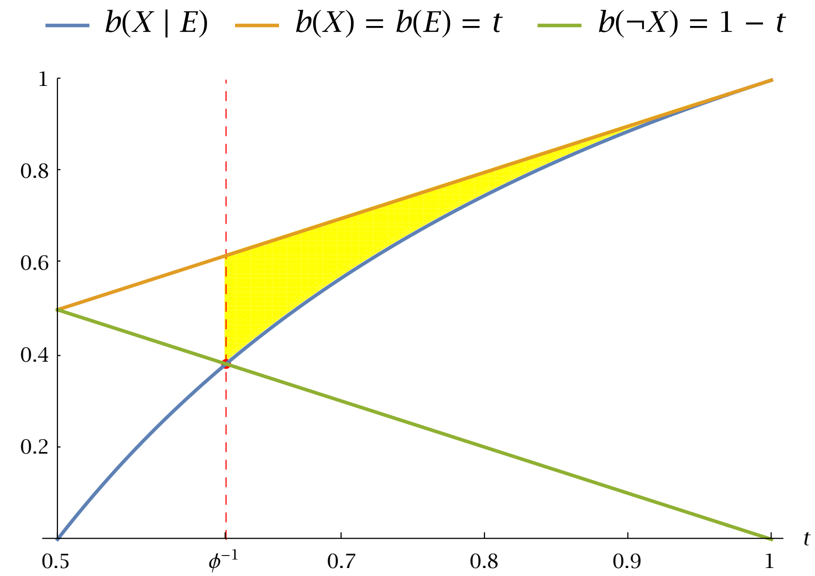


Figure: Pure Bayesian Revision, Very Weak Preservation, and ϕ

A stronger (but still pretty weak) version of Preservation has been discussed (more extensively) in the literature (*e.g.*, [13]).

Weak Preservation. If $E, X \in \mathbf{B}$, then $X \in \mathbf{B} \star E$.

In words: *learning something you already believe should never cause you to* ***not*** *believe something you already believed.*

Interestingly, Pure Bayesian Revision can *violate* Weak Preservation — *even for high Lockean thresholds* $t \geq \phi^{-1}$.

These high-credence (pure Bayesian) violations of Weak Preservation can be visualized *via* the *yellow region* in Figure 1.

However, *if* a Pure Bayesian *violates* Weak Preservation, *then* their prior belief set **B** *must not have been closed under logic.*

Proof.

Suppose (a) $E, X \in \mathbf{B}$, *but* (b) $X \notin \mathbf{B} \divideontimes E$. (b) implies $\Pr(X \mid E) < t$. Therefore, $\Pr(X \,\&\, E) < t \cdot \Pr(E) < t$. So, $X \,\&\, E \notin \mathbf{B}$. □

AGM theory is the orthodox, *logical* account of belief revision and its constraints on revisions may *seem* quite modest.

AGM can be understood as embodying a *principle of* ***conservativity*** (*aka.*, *informational economy/minimal mutilation*).

Conservativity. When an agent with a prior belief set **B** learns (exactly) E, she should revise to a posterior belief set $\mathbf{B}'$ that:

(1) *includes* E,
(2) is *cogent*, and
(3) constitutes a *minimal change*[2] to **B**.

AGM revision has a number of nice logico-mathematical properties that have made for fruitful investigation.

Specifically, it has a simple and straightforward *axiomatization.*

[2] "Minimal change" can be explicated (in *geodesic* terms) *via* a wide variety of measures of distance between prior and posterior belief sets [12, 3, 6].

Basic Gärdenfors postulates:

(*1) $\mathbf{B} * E = \mathsf{Cn}(\mathbf{B} * E)$ **Closure**

(*2) $E \in \mathbf{B} * E$ **Success**

(*3) $\mathbf{B} * E \subseteq \mathsf{Cn}(\mathbf{B} \cup \{E\})$ **Inclusion**

(*4) If E is consistent with $\mathbf{B}$, then $\mathbf{B} \star E \supseteq \mathbf{B}$ **Preservation**

(*5) If E is non-contradictory, then $\mathbf{B} * E$ is consistent **Consistency**

(*6) If $X \Leftrightarrow Y$, then $\mathbf{B} * X = \mathbf{B} * Y$ **Extensionality**

[Note: Strictly speaking, Gärdenfors's (*4) was **Vacuity** and not **Preservation**.

(*4) If E is consistent with $\mathbf{B}$, then $\mathbf{B} \star E \supseteq \mathsf{Cn}(\mathbf{B} \cup \{E\})$ **Vacuity**

But, given the other postulates, these two axioms are equivalent. And, it makes our presentation more elegant/continuous to use **Preservation** here, because it implies both of the weaker forms of Preservation above.]

(*3) $\mathbf{B} \star E \subseteq \mathsf{Cn}(\mathbf{B} \cup \{E\})$ **Inclusion**

Proposition. $\divideontimes$ satisfies **Inclusion.**

Proof: Suppose $X \in \mathbf{B} \divideontimes E$. By the definition of $\divideontimes$,

$$b(X \mid E) \geq t.$$

It is a theorem of the probability calculus that

$$\Pr(E \supset X) \geq \Pr(X \mid E).$$

Therefore,

$$b(E \supset X) \geq t.$$

Hence, $E \supset X \in \mathbf{B}$. So, by *modus ponens* (for $\supset$), we arrive at

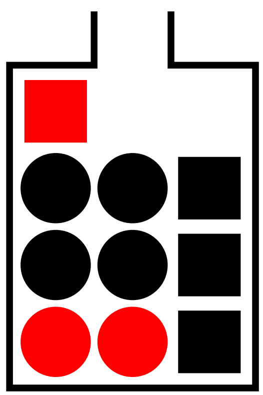
$$X \in \mathsf{Cn}(\mathbf{B} \cup \{E\}).$$

Therefore, $\mathbf{B} \star E \subseteq \mathsf{Cn}(\mathbf{B} \cup \{E\})$. □

(*4) If E is consistent with $\mathbf{B}$, then $\mathbf{B} \star E \supseteq \mathbf{B}$ **Preservation**

Proposition. *Even **Cogent** Bayesians can **violate*** **Preservation.**

Proof: Our agent will sample a single object at random from the urn on the right.



E = 'The object sampled will be red'

X = 'The object sampled will be a circle'

Assume our (Cogent) Bayesian agent has a Lockean threshold of $t = 0.85$.

Then, $E \supset X$ is *the only proposition* initially believed by our agent (see slide 16). *I.e.*,

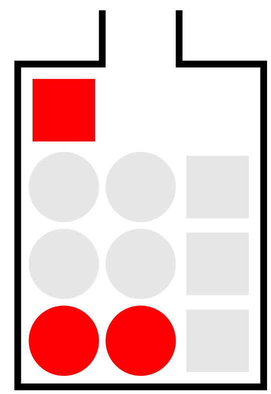
$$\mathbf{B} = \{E \supset X\}.$$

Now, *what happens if she learns E*?

(*4) If E is consistent with $\mathbf{B}$, then $\mathbf{B} \star E \supseteq \mathbf{B}$ **Preservation**

Proposition. *Even **Cogent** Bayesians can **violate*** **Preservation.**

Proof: (continued)



If she learns E (which is consistent with $\mathbf{B}$), this *rules out the black objects.* Hence,

$$b(E \supset X \mid E) = {}^{2}/_{3} < 0.85.$$

Thus, $E \supset X \notin \mathbf{B} \divideontimes E$, but $E \supset X \in \mathbf{B}$. □

Note: our Bayesian agent *is cogent at both the prior and posterior times.* To see this, note that her posterior belief set will be:

$$\mathbf{B}' = \mathbf{B} \divideontimes E = \{E, E \vee X, E \vee \neg X\}$$

We conclude with two final theorems concerning Bayesian failures of **Preservation**. The first result suggests Bayesian revision can be seen as being *more epistemically risk-averse* (in one sense) than AGM (which can be seen as *more risk-seeking*).[3]

Theorem. $\divideontimes$ violates **Preservation** (wrt $E, \mathbf{B}$) ***iff*** $\mathbf{B} \divideontimes E \subset \mathbf{B} * E$.

Proof: See Extras Slide 17

So, whenever Bayesian and AGM revision diverge, AGM is more "epistemically demanding" on the agent's beliefs — in the sense that AGM requires the agent to have *strictly more beliefs* than Bayesianism requires. In this sense, AGM may be seen to be more *epistemically risk-seeking* than Bayesianism.

Our final result brings us back to *The Golden Ratio*. . .

[3] Pettigrew [16] has independently argued (*via* the use of an epistemic *Hurwicz Criterion*) that **Cogency** implies its own variety of *risk-seeking*.

Our final theorem reveals an interesting relationship between *Cogent* Bayesian Revision, **Preservation**, and *The Golden Ratio*.

Theorem. If a *cogent* Bayesian *violates* **Preservation** (wrt $E, \mathbf{B}$), then their Lockean threshold t *must lie on the half-open interval* $[\phi^{-1}, 1)$.

Proof: See the Appendix of our paper.

Corollary. If a *cogent* Bayesian's Lockean threshold t lies *outside* the half-open interval $[\phi^{-1}, 1)$, then *they must obey* AGM.

To sum up: we have the following three central results.

(*Pure* Bayesian Revision & $t \in [\phi^{-1}, 1]$) $\Rightarrow$ V.W. Preservation

Cogent Bayesian Revision $\Rightarrow$ W. Preservation

(*Cogent* Bayesian Revision & $t \notin [\phi^{-1}, 1)$) $\Rightarrow$ Preservation/AGM

- *The Holy Grail* would be to find a *purely qualitative axiomatization* of (pure or cogent) Bayesian revision.
  - van Eijck & Renne [8] recently provided an axiomatization of (strong) *Lockeanism with* $t = 1/2$, which can be used to axiomatize *pure* Bayesianism (with a $t = 1/2$ threshold).
  - And, the $t = 1$ (extremal) case *just is* AGM revision [10, 11]. The problem of axiomatizing Bayesian revision for *intermediate* thresholds $t \in (1/2, 1)$ remains open.
- Since pure Bayesian revision may be described in "minimal distance" terms [4], this suggests a general "geodesic update" framework in which we plan to investigate *contraction* ($\div$) and other kinds revision. For instance:
  1. Let $b^\star$ be **the closest probability function** to b s.t. $b^\star(p) < t$.
  2. $\mathbf{B} \div p := \{p \mid b^\star(p) \geq t\}$

p	$b(p)$	$b(p \mid E)$	$p \in \mathbf{B}$?	$p \in \mathbf{B} \divideontimes E$?	$p \in \mathbf{B} * E$?	$p \in \mathsf{Cn}(\mathbf{B} \cup \{E\})$?
$E \wedge X$	2/10	2/3	No	**No**	**Yes**	**Yes**
$E \wedge \neg X$	1/10	1/3	No	No	No	No
$\neg E \wedge X$	4/10	0	No	No	No	No
$\neg E \wedge \neg X$	3/10	0	No	No	No	No
E	3/10	1	No	Yes	Yes	Yes
X	6/10	2/3	No	**No**	**Yes**	**Yes**
$E \equiv X$	5/10	2/3	No	**No**	**Yes**	**Yes**
$E \equiv \neg X$	5/10	1/3	No	No	No	No
$\neg E$	7/10	0	No	No	No	No
$\neg X$	4/10	1/3	No	No	No	No
$E \vee X$	7/10	1	No	Yes	Yes	Yes
$E \vee \neg X$	6/10	1	No	Yes	Yes	Yes
$\neg E \vee X$	9/10	2/3	Yes	No	No	**Yes**
$\neg E \vee \neg X$	8/10	1/3	No	No	No	No

Table: Cogent Bayesian counterexample to **Preservation**

Theorem. $\circledast$ *violates* **Vacuity**[4] (wrt $E, \mathbf{B}$) ***iff*** $\mathbf{B} \circledast E \subset \mathbf{B} * E$.

Proof: ($\Rightarrow$) Suppose $\circledast$ *violates* **Vacuity** (wrt **B** and E). Then, (a) E is consistent with **B**; and, (b) $\mathbf{B} \circledast E \not\supseteq \mathsf{Cn}(\mathbf{B} \cup \{E\})$. By (b), there exists an X such that $X \in \mathsf{Cn}(\mathbf{B} \cup \{E\})$ but $X \notin \mathbf{B} \circledast E$. It follows from (a), **Vacuity**, and **Inclusion** (for $*$) that $\mathsf{Cn}(\mathbf{B} \cup \{E\}) = \mathbf{B} * E$. Therefore, $X \in \mathbf{B} * E$ and $X \notin \mathbf{B} \circledast E$. And, by **Inclusion**, $\mathbf{B} \circledast E \subseteq \mathsf{Cn}(\mathbf{B} \cup \{E\}) = \mathbf{B} * E$. □

($\Leftarrow$) Suppose E is consistent with **B** and $\mathbf{B} \circledast E \subset \mathbf{B} * E$. Then, there exists an X such that $X \in \mathbf{B} * E$ but $X \notin \mathbf{B} \circledast E$. Because E is consistent with **B**, **Vacuity**, and **Inclusion** (for $*$) imply that $\mathbf{B} * E = \mathsf{Cn}(\mathbf{B} \cup \{E\})$. So, $X \in \mathsf{Cn}(\mathbf{B} \cup \{E\})$; but, $X \notin \mathbf{B} \circledast E$. □

[4] This proof is carried out for **Vacuity** rather than **Preservation**. But, as we mentioned above, these are equivalent, given the other AGM postulates.

[1] C. Alchourron, P. Gärdenfors, and D. Makinson. *On the logic of theory change: Partial meet contraction and revision functions*, 1985.

[2] D. Christensen, *Putting Logic in its Place*, 2007.

[3] M. Deza and E. Deza, *Encyclopedia of Distances*, 2009.

[4] P. Diaconis and S. Zabell, *Updating Subjective Probability*, 1982.

[5] K. Dorst, *Lockeans Maximize Expected Accuracy*, 2015.

[6] C. Duddy and A. Piggins, *A measure of distance between judgment sets*, 2012.

[7] K. Easwaran, *Dr. Truthlove or: How I Learned to Stop Worrying and Love Bayesian Probability*, 2015.

[8] J. van Eijck and B. Renne, *Belief as Willingness to Bet*, 2014.

[9] R. Foley, *Working Without a Net*, 1992.

[10] W. Harper, *Rational belief change, popper functions and counterfactuals*, 1975.

[11] J. Hawthorne, *A primer on rational consequence relations, popper functions, and their ranked structure*, 2013.

[12] K. Georgatos, *Geodesic Revision*, 2008.

[13] H. Leitgeb, *The review paradox: On the diachronic costs of not closing rational belief under conjunction*, 2013.

[14] Hannes Leitgeb. *Stability Theory of Belief*. Oxford University Press, 2016.

[15] I. Levi, *Gambling with Truth*, 1967.

[16] R. Pettigrew, *Accuracy and Risk: a Jamesian investigation in formal epistemology*, 2015.