Session 2

In our last session, I introduced the standard AGM theory of belief revision from an axiomatic perspective. We saw how an agent's belief revision in light of new information can be represented through set-theoretic operations. Specifically, I showed you that:

1. Our ideal agent's epistemic state can be represented by a **belief set**, i.e., a set $B \subseteq \mathcal{L}$ (where $\mathcal{L}$ is our formal language) that is **closed under logical consequence**.
2. The new information the agent receives is represented by a sentence $\phi \in \mathcal{L}$.
3. The agent's act of revising their original beliefs B by ϕ simply amounts to "dropping" B and "adopting" a new belief set, $B * \phi$.

But this is just the beginning of the story. I also emphasized that *whatever the actual mechanics of revision are*, on the standard (AGM) approach, the revision process must obey certain rules. The revision operator $*$, which maps a belief set B and a sentence ϕ to a new belief set $B * \phi$, must satisfy the following six foundational postulates:

$$B * \phi \text{ is a belief set} \qquad \text{(Closure)}$$

$$\phi \in B * \phi \qquad \text{(Success)}$$

$$\text{If } \phi \nvdash \bot, \text{then } Cn(B * \phi) \neq \mathcal{L} \qquad \text{(Consistency)}$$

$$B * \phi \subseteq Cn(B \cup \{\phi\}) \qquad \text{(Inclusion)}$$

$$\text{If } B \nvdash \neg\phi, \text{then } B * \phi = Cn(B \cup \{\phi\}) \qquad \text{(Vacuity)}$$

$$\text{If } Cn(\{\phi\}) = Cn(\{\psi\}), \text{then } B * \phi = B * \psi \qquad \text{(Congruence)}$$

Furthermore, to properly handle composite information, we want $*$ to validate two supplementary postulates:

$$B * (\phi \wedge \psi) \subseteq (B * \phi) + \psi \qquad \text{(Superexpansion)}$$

$$\text{If } B * \phi \nvdash \neg\psi, \text{then } (B * \phi) + \psi \subseteq B * (\phi \wedge \psi) \qquad \text{(Subexpansion)}$$

Until now, our study of belief revision has been highly abstract. We have essentially treated the revision operator $*$ as a **black box**: we defined the class of available inputs (belief sets and sentences), the kind of outputs it returns (new belief sets), and the higher-order rules it must respect. However, we have not provided a concrete **construction** for this operator—a method that, given a specific belief set B and a specific sentence ϕ, actually allows us to compute $B * \phi$.

Today, we will open that black box and study one such construction.

The Value of the Axiomatic Approach

The fact that we treated $*$ as a black box should not make you think our previous endeavors were useless! In fact, characterizing $*$ purely in terms of the rules it must respect has already given us a massive theoretical advantage:

1. **Deriving New Rules:** It allowed us to prove a host of **derivative rules** that $*$ must also respect.
2. **Generality:** Working at the abstract level of postulates means that if we prove a theorem based *only* on those axioms (e.g., the basic six postulates), we have ipso facto proved that **any**

suitable (!) construction for $*$ must validate that consequence. We secure results that are completely independent of the underlying mechanics. Conversely, if we start with a specific construction for $*$ and prove a property, we cannot guarantee that property holds for *other* suitable (!) constructions, even if both validate the basic postulates.

Before moving on, make sure that you remember the notions we are keeping in the background, that is:

1. our boolean language $\mathcal{L}$ (definition);
2. the notion of interpretation (or valuation, or possible world) (definition);
3. logical consequence relation (definition) and operator (definition);
4. belief sets (definition); and
5. the properties of classical consequence relations and operator.

1. Constructing the Belief Revision Operator: Introduction

There are many ways in which we can construct a concrete belief revision operator. We can generally group them into two main categories: **syntactic** and **semantic** approaches.

Syntactic Approaches. In the syntactic approach, the belief revision operator $*$ is obtained as a result of some manipulation of sentences, or sets of sentences, within our belief set B. There are two main methods within this category:

1. **"Selection Function" approach**: On this approach, when we want to revise B by ϕ, we perform two operations . First, we perform a *contraction*: we take B and single out all the maximal subsets of B that **do not** entail $\neg\phi$ (i.e., we single out the largest possible subsets of B that are logically compatible with the new information ϕ). Because there might be several such maximal subsets, we use a "selection function" to pick the "best" ones—or we take their intersection—to form a contracted, consistent base. Second, we perform an *expansion*: we simply add the new sentence ϕ to this contracted base and close it under logical consequence to obtain $B * \phi$.
2. **Epistemic Entrenchment**: This approach represents the agent's doxastic state by pairing the belief set B with an *entrenchment ordering* over the sentences in the language . This ordering reflects how firmly the agent holds specific beliefs. When the agent receives new information ϕ that contradicts the current belief set, the entrenchment ordering determines which beliefs must be surrendered first . The least entrenched beliefs are discarded to restore consistency, while the more entrenched ones are preserved . The revised belief set $B * \phi$ is constructed by keeping all old beliefs that are strictly more entrenched than $\neg\phi$, and then expanding by ϕ .

Semantic Approaches. In the semantic approach, the revision operator is constructed not by directly manipulating sentences, but by evaluating the **possible worlds** (or models) that make those sentences true.

1. **Implausibility Orderings and Ranking Functions**: The space of possible worlds is equipped with an implausibility ordering, where worlds are ranked according to their implausibility. The center consists of the least implausible worlds, representing the current belief set . When revising by ϕ, the new belief set is determined by the theory of the *least implausible* possible worlds in which ϕ is true .

Today, we will discuss a “semantic” construction.

2. Plausibility Orderings

Recall that when we introduced our propositional language $\mathcal{L}$ and its logical consequence relation $\vdash$, we defined a set, W, of all possible valuations for $\mathcal{L}$. This set contains all possible functions $w : \Phi \to \{0, 1\}$—

that is, all possible assignments of truth values to the atomic propositions in $\mathcal{L}$, which in turn recursively determine the truth value of any complex sentence in $\mathcal{L}$.

Conceptually, these abstract objects can be understood as *possible worlds*. A valuation function w describes a maximally specific way things could be (given the expressive limits of $\mathcal{L}$) because it takes a stance on every "fact" expressible in the language. To emphasize this philosophical interpretation, I will refer to W as the set of possible worlds from now on.1

As anticipated in Session 1, §1.1, every sentence $\phi \in \mathcal{L}$ corresponds to a specific subset of W called its **truth-set**—the set of all worlds where ϕ is true. Philosophically, this set is often identified with the *proposition* expressed by ϕ, representing its "content." Regardless of the philosophical backstory, this extensional notion is mathematically crucial for the constructions we will explore.

Definition 1 (Truth-Set).

Let $\phi \in \mathcal{L}$. The truth-set function $[\cdot] : \mathcal{L} \to \mathcal{P}(W)$ maps a sentence to the set of its models:

$$[\phi] := \{w \in W : w \models \phi\}$$

We can extend the domain of $[\cdot]$ to include *sets* of sentences. Let $\Gamma \subseteq \mathcal{L}$ be a set of sentences. Its truth-set is the intersection of the truth-sets of all its members:

$$[\Gamma] := \bigcap_{\phi \in \Gamma} [\phi] = \{w \in W : w \models \phi \text{ for all } \phi \in \Gamma\}$$

With a slight **abuse of notation**, I have extended the function $[\cdot]$ so that it takes sets of sentences as arguments, not just individual sentences. Intuitively, this intersection yields the exact set of worlds where every sentence in Γ is simultaneously true. This semantic mapping gives rise to the following **fundamental properties** that are crucial for the proofs ahead.

Lemma 2.

Let $\phi, \psi \in \mathcal{L}$ and let $\Gamma, \Delta \in \mathcal{P}(\mathcal{L})$. The semantic sets behave according to the following logical rules:

1. $\phi \vdash \psi \iff [\phi] \subseteq [\psi]$
2. $\Gamma \subseteq \Delta \implies [\Delta] \subseteq [\Gamma]$
3. $\Gamma \subseteq \Delta \iff [\Delta] \subseteq [\Gamma] \quad (\text{if } \Delta \text{ is a belief set})$
4. $[Cn(\Gamma \cup \Delta)] = [\Gamma] \cap [\Delta]$

Proof. See the solutions to the exercises. □

Now, we are ready to see exactly **how to construct the revision operator $*$ using possible worlds**. Let's start with the informal intuition before diving into the formal mathematical construction.

In real-world scenarios, we rarely restrict our beliefs solely to what is strictly entailed by our evidence. We regularly make **inferences that go "beyond" the evidence**—accepting the risk that our beliefs might be false despite the evidence we possess. However, when we reason *inductively* like this, we do not just believe anything. That would be irrational. Instead, we constrain these "inductive leaps": **we believe those things that, given our evidence, are as plausible as they can possibly be**.

This informal but highly intuitive idea validates the following powerful claim:

Claim 3.

Given your evidence E, it is rational to believe any proposition that is true across all the most plausible possible worlds, given E.

Clearly, to have a fully-fledged theory of rational inductive belief, we would eventually need to define exactly what "plausibility" amounts to. However, the beauty of the representation result we are about to prove is that **we can discover profound results without providing a concrete characterization of plausibility just yet**. Specifically, the theorem will show that if rational belief (and belief revision) behaves according to Claim 3, then imposing certain structural properties on this abstract "plausibility" ordering will generate a revision operator $*$ that perfectly satisfies the standard AGM postulates.

Just to give you an idea of what plausibility *could* mean in practice, it might be interpreted as:

1. **Probability:** Under this view, Claim 3 states that it is rational to believe whatever is most probable given E.
2. **Normality (in Martin Smith's sense):** Under this view, Claim 3 states that you should believe whatever is true in the set of worlds that demand the fewest explanations (i.e., the most "normal" worlds).
3. **Modality/Similarity (in Duncan Pritchard and Jaakko Hirvelä's sense):** Under this view, Claim 3 states that you should believe whatever is true in the possible worlds that are most similar to the actual world.2

To generate a belief revision operator $*$ that "corresponds" to the AGM postulates (in a formal sense we will clarify later), we must assume that the possible worlds in W can be ordered by a relation $\preceq$, and that this relation satisfies specific structural properties.

It is crucial to emphasize a conceptual point here: we treat the plausibility ordering $\preceq$ mathematically as a *relative* ordering rather than an absolute one. **The relative plausibility of two worlds w_1, w_2 strictly depends on the agent's current background belief set B.** Formally, this means:

1. We assume that every belief set B has an associated ordering $\preceq_B$ satisfying certain structural properties.
2. We **do not** assume that $\preceq_B = \preceq_{B'}$ for distinct belief sets $B \neq B'$, nor do we enforce that they must be different.

By remaining agnostic on this, the results we prove will be entirely independent of this debate. Our theorems hold perfectly whether you interpret plausibility as an absolute, objective hierarchy or as an evidence-relative relation.

(Note: The idea that plausibility shifts based on current evidence is highly intuitive. Consider probability theory: an agent can define entirely different probability distributions over the exact same space of possible worlds simply by conditioning on different prior evidence.)

Because the ordering is tied to a specific belief set B, we will write $\preceq_B$ when the context requires strict precision. Otherwise, to keep the notation clean, I will simply write $\preceq$.

The following structural properties are required to characterize the AGM belief revision operator $*$ in terms of possible world plausibility. In words, we assume that plausibility has the following properties:

1. All worlds are comparable in terms of plausibility.
2. Plausibility is transitive.
3. All and only the models of our current belief set B are the most plausible worlds.
4. There is always at least one "most plausible" world among the models of any consistent sentence.

Let us flesh out these properties in a more precise way. Let B be a belief set and $\preceq_B$ be its associated plausibility ordering (where $w_1 \preceq_B w_2$ is read as "w_1 is at least as plausible as w_2", and we write $w_1 \prec_B w_2$ for strict plausibility):

1. **Connectedness**: For any worlds $w_1, w_2 \in W$: Either $w_1 \preceq_B w_2$ or $w_2 \preceq_B w_1$.
2. **Transitivity**: For any worlds $w_1, w_2, w_3 \in W$: If $w_1 \preceq_B w_2$ and $w_2 \preceq_B w_3$, then $w_1 \preceq_B w_3$.
3. **Centeredness**:
   1. If $w_1, w_2 \in [B]$, then $w_1 \preceq_B w_2$.
   2. If $w_1 \in [B]$ and $w_2 \notin [B]$, then $w_1 \prec_B w_2$.
4. **Limit Assumption**: For any sentence $\phi \in \mathcal{L}$: If $[\phi] \neq \emptyset$, then there exists at least one world $w_1 \in [\phi]$ such that $w_1 \preceq_B w_2$ for all $w_2 \in [\phi]$.

If $\preceq_B$ satisfies (1)-(2), then $\preceq_B$ is a total preorder.

Now, before stating and proving our representation theorem, we must clarify how we build $B * \phi$ given our plausibility ordering $\preceq$. The intuitive idea is that when we revise B by ϕ, the set of sentences $B * \phi$ (i.e., the set of all the ideal agent's beliefs) will be the set of all and only the sentences ψ that are true throughout the most plausible worlds w in $[\phi]$ given $\preceq_B$. That is, we want our new belief set to accept the sentence ϕ, but also to do so in the most parsimonious way possible: we take as the “world basis” for $B * \phi$ all and only those ϕ-worlds that are most plausible according to our initial belief set B.

Define:

$$\min_B([\phi]) := \{w \in [\phi] : \text{For all } w' \in [\phi], w \preceq_B w'\}$$

That is, $\min_B([\phi])$ selects the set of most plausible worlds w that satisfy ϕ.

Next, define the theory function $T : \mathcal{P}(W) \to \mathcal{P}(\mathcal{L})$ such that, for any $V \subseteq W$:

$$T(V) := \{\phi \in \mathcal{L} : V \subseteq [\phi]\}$$

In other words, T takes a world-base V and returns the set of sentences $\phi \in \mathcal{L}$ that are true in all the worlds in V. There are three important properties connecting our semantic sets and our theory function to be aware of:

Lemma 4.

Let $V, V_1, V_2 \subseteq W$ be any sets of worlds, and $\Gamma, \Delta \subseteq \mathcal{L}$ be any sets of sentences.

1. *$T(V)$ is a belief set, i.e. $T(V) = Cn(T(V))$.*
2. *$T([\Gamma]) = Cn(\Gamma)$. Consequently, $T([\Gamma]) = \Gamma \iff \Gamma$ is a belief set.*
3. *$V \subseteq [T(V)]$. Furthermore, the equality $[T(V)] = V$ holds for all $V \subseteq W$ if, and only if, Φ is finite.*
4. *If $V_1 \subseteq V_2$, then $T(V_2) \subseteq T(V_1)$.*

Proof. See the solutions to the exercises. ◻

This third property is conceptually crucial. It shows that while we can map sentences to worlds and back without losing anything (provided the sentences are logically closed), moving from arbitrary sets of worlds to syntax and back generally results in a loss of information—unless our language is strictly constrained to be finite.

Finally, we define our revision operator $*$. Assuming that for every belief set $B \in \mathbf{B}$ there exists an associated plausibility ordering $\preceq_B$, we can define the global revision function $* : \mathbf{B} \times \mathcal{L} \to \mathbf{B}$ as follows:

$$B * \phi := T(\min_B([\phi]))$$

Informally, the revised belief set $B * \phi$ is exactly the set of sentences ψ that are true throughout $\min_B([\phi])$—that is, the set of sentences true in the most B-plausible worlds that satisfy ϕ.

Before moving to the formal statement of the theorem and the first part of its proof, let us look at a concrete example to see exactly how this semantic revision operator works in practice.

2.1. Example

Imagine a biology professor who is currently waiting to hear whether their new research project will be approved. To model the epistemic state they are in, we can construct a space of possible worlds W using just four propositional variables:

- p: The research project is approved.
- f: The project's funding is secured.
- h: Hiring for the project is permitted.
- l: Lab space is allocated.

Our agent firmly believes two conditional "laws" about how the university operates. The first is a university-wide rule, while the second is a departmental rule. **Consequently, the agent holds the first law to be significantly "more important" than the second**.

1. **Law 1 (University Policy):** If a project is approved, funding must be secured.

$$L_1 := p \to f$$

2. **Law 2 (Department Practice):** If funding is secured, hiring is permitted and lab space is allocated.

$$L_2 := f \to (h \land l)$$

Note: These "laws" are not necessary, logical truths—you can easily point to logically possible worlds where they fail. Instead, they should be understood as *ceteris paribus* laws, or extremely robust generalizations the professor has come to trust after years of administrative experience.

Now, let us assume the agent's initial belief set B consists of these two laws, plus the belief that the project is approved ($\neg p$).

$$\begin{aligned} B &= Cn(\{L_1, L_2, \neg p\}) \\ &= Cn(\{\neg p \land (f \to (h \land l))\}) \end{aligned}$$

By the **Centeredness** property, the most plausible worlds in W are exactly those inside the truth-set $[B]$. These are the worlds where no project is approved and the laws hold. They sit at the very bottom of the ordering $\preceq_B$.

Now, imagine that our biology professor receives an email from the dean with new information ϕ: the new project has just been approved, **but no new hiring is permitted**.

$$\phi := p \land \neg h$$

Before we dive into the formal details, let us **informally summarize what should happen**. The agent started out believing they did not get the project approved ($\neg p$) and that two laws hold, with L_1 being more important than L_2. Upon learning from the dean that p and $\neg h$ are true, they must revise their beliefs in a conservative way: they must accept p and $\neg h$, and to make room for this, they will retain the strong university rule L_1 while dropping the weaker departmental rule L_2. In other words, given the dean's email, the professor expects to get the funding (f) and simply accepts that the standard department practice (L_2) has been overridden (for they won't be permitted to hire new people for the moment).

Let us now see how the formal possible-worlds construction predicts this intuitive result.

Clearly, ϕ is false in $[B]$. Therefore, the most plausible worlds where ϕ is true cannot be worlds inside $[B]$, for every world in $[B]$ satisfies $\neg p$, and hence $\neg\phi$. As a result, our proposed construction requires us to step outside the set of most plausible worlds (i.e., the B-worlds). We must move away from the center of the plausibility ordering and inspect the truth-set $[\phi]$.

First, let us graphically represent the space of possible worlds, highlighting which worlds are in $[B]$ and which are in $[\phi]$:

	p	f	h	l	$\in [B]$	$\in [\phi]$
w_1	1	1	1	1		
w_2	1	1	1	0		
w_3	1	1	0	1		✓
w_4	1	1	0	0		✓
w_5	1	0	1	1		
w_6	1	0	1	0		
w_7	1	0	0	1		✓
w_8	1	0	0	0		✓
w_9	0	1	1	1	✓	
w_{10}	0	1	1	0		
w_{11}	0	1	0	1		
w_{12}	0	1	0	0		
w_{13}	0	0	1	1	✓	
w_{14}	0	0	1	0	✓	
w_{15}	0	0	0	1	✓	
w_{16}	0	0	0	0	✓	

Notice that in *all* four worlds in $[\phi] = \{w_3, w_4, w_7, w_8\}$, at least one of the laws L_1 or L_2 (which the agent initially believed) is falsified:

- In w_3 and w_4, L_2 is flouted, because the funding is secured (f) but hiring is *not* permitted ($\neg h$). However, L_1 is still true.
- In w_7 and w_8, L_1 is flouted, because the research project is approved (p) but the funding is *not* secured ($\neg f$).

How does the plausibility ordering $\preceq_B$ rank these four worlds? We **want this ordering to reflect the agent's expectations**. Because L_1 is a university-wide mandate, while L_2 is merely a departmental rule, worlds where the university rule L_1 holds (w_3 and w_4) are *more plausible* than worlds where L_1 fails (w_7 and w_8).

To visualize this, we can map out the relevant tiers of the plausibility ordering. Below, worlds that are equally plausible are piled up vertically. The worlds in $[B]$ sit at the left, followed by the increasingly less implausible tiers (using $[\ldots]$ for intermediate worlds that are not relevant to our current discussion):

$$\begin{matrix} w_9 \\ w_{13} \\ w_{14} \\ w_{15} \\ w_{16} \end{matrix} \prec_B [\ldots] \preceq_B \begin{matrix} w_3 \\ w_4 \end{matrix} \prec_B [\ldots] \preceq_B \begin{matrix} w_7 \\ w_8 \end{matrix} \preceq_B [\ldots]$$

Legend

■ Worlds in $[B]$

■ Worlds in $[\phi]$

Because w_3 and w_4 sit strictly to the left of w_7 and w_8, we know that $w_3 \preceq_B w_7$, $w_4 \preceq_B w_7$, and so on. Now, to compute the new belief set $B * \phi$, we apply our formula:

$$B * \phi = T(\min_B([\phi]))$$

1. First, we identify $\min_B([\phi])$. Based on our ordering, the most plausible worlds *in* $[\phi]$ are those that preserve the stronger law L_1. Thus,

$$\min_B([\phi]) = \{w_3, w_4\}$$

2. Second, we extract the theory of this set:

$$T(\{w_3, w_4\}) = Cn(\{p \wedge f \wedge \neg h\})$$

Therefore, **the revised belief set $B * \phi$ encodes the following epistemic attitudes**:

1. The agent believes the project is approved (p) and that hiring is *not* permitted ($\neg h$). This is the information ϕ they received.
2. The agent believes that the funding is secured (f). This is a consequence of L_1.
3. The agent suspends judgment on whether the lab space is allocated (l is true at w_3 and false at w_4, meaning neither l nor $\neg l$ is true across $\min_B([\phi]) = \{w_3, w_4\}$).

Note that **law L_2 has been surrendered to accommodate the new information** (i.e., the agent now believes $\neg L_2$).

Finally, notice a crucial philosophical feature of this mathematical construction: **the agent is permitted to rationally believe *more* than what the new evidence strictly entails**. The email (ϕ) only entails $p \wedge \neg h$. It says absolutely nothing about funding (f). Yet, the agent still ends up believing f. Why? Because f is true in all the *most plausible* ϕ-worlds (w_3 and w_4).

3. The Representation Theorem

In the previous section, we saw how to construct a candidate revision operator $*$ using possible worlds and plausibility orderings $\preceq_B$. We also worked through examples to see exactly how this semantic machinery behaves in practice.

Now, it is time to step back up to the theoretical level. The main result we are going to prove is that the plausibility-based construction perfectly captures the rational constraints we previously defined using axioms. The Representation Theorem formally proves this exact equivalence.

Theorem 5 (Representation Theorem for Belief Revision).

Let $\mathcal{L}$ be the propositional language defined previously and W be the corresponding space of possible worlds. Let $B \subseteq \mathcal{L}$ be a belief set.

(1) *If $\preceq_B$ is a plausibility ordering on W that satisfies Connectedness, Transitivity, Centeredness, and the Limit Assumption, then the revision operator $*$ defined by:*

$$B * \phi := T(\min_B([\phi]))$$

satisfies all basic and supplementary AGM postulates.

(2) *Assume Φ is finite. If $*$ is a revision operator on B that satisfies all eight AGM postulates, then there exists a plausibility ordering $\preceq_B$ on W satisfying Connectedness, Transitivity, Centeredness, and the Limit Assumption, such that for every formula $\phi \in \mathcal{L}$:*

$$B * \phi = T(\min_B([\phi]))$$

4. Part One: "Soundness"

Let us first prove the first half of the Representation Theorem Theorem 5 (Representation Theorem for Belief Revision). This result establishes that our semantic construction $*$ qualifies as a belief revision operator, for it successfully validates all the AGM postulates we discussed in Session 1.

Lemma 6 (Soundness)**.**

Let $\mathcal{L}$ be the propositional language defined previously and W be the corresponding space of possible worlds. Let $B \subseteq \mathcal{L}$ be a belief set. If $\preceq_B$ is a plausibility ordering on W that satisfies Connectedness, Transitivity, Centeredness, and the Limit Assumption, then the revision operator $*$ defined by:

$$B * \phi := T(\min_B([\phi]))$$

satisfies all eight basic and supplementary AGM postulates.

Proof. Assume that $B \subseteq \mathcal{L}$ is a belief set and $\preceq_B$ (or simply $\preceq$) is a plausibility ordering on W satisfying Connectedness, Transitivity, Centeredness, and the Limit Assumption. We must show that the revision operator defined by

$$B * \phi := T(\min_B([\phi]))$$

satisfies the AGM postulates.

1. Closure. We must show that $B * \phi$ is a logically closed belief set, i.e., $B * \phi = Cn(B * \phi)$. By substituting our definition of $*$, this amounts to showing:

$$T(\min_B([\phi])) = Cn(T(\min_B([\phi])))$$

Recall from our earlier results (see Lemma 4) that for *any* arbitrary set of worlds $V \subseteq W$, its theory $T(V)$ is automatically a logically closed belief set. Since $\min_B([\phi])$ is simply a set of worlds, its theory is logically closed.

2. Success. We must show that $\phi \in B * \phi$. By the definition of the theory function:

$$T(\min_B([\phi])) = \{\psi \in \mathcal{L} : \min_B([\phi]) \subseteq [\psi]\}$$

By definition, $\min_B([\phi])$ selects minimal worlds *within* $[\phi]$. Thus, $\min_B([\phi]) \subseteq [\phi]$ is true. Therefore, $\phi \in T(\min_B([\phi]))$.

3. Inclusion. We must show that $B * \phi \subseteq Cn(B \cup \{\phi\})$. Let $\psi \in B * \phi$, which means $\psi \in T(\min_B([\phi]))$, i.e., $\min_B([\phi]) \subseteq [\psi]$. We need to prove that $\psi \in Cn(B \cup \{\phi\})$.
First, recall the semantic equivalence for logical consequence:

$$B \cup \{\phi\} \vdash \psi \iff [B] \cap [\phi] \subseteq [\psi]$$

So, it suffices to prove that $[B] \cap [\phi] \subseteq [\psi]$. We evaluate this by looking at two cases.

1. **Case One:** Suppose $[B] \cap [\phi] = \emptyset$ (i.e., $B \vdash \neg\phi$). The empty set is trivially a subset of any set, so $[B] \cap [\phi] \subseteq [\psi]$ automatically holds.
2. **Case Two:** Suppose $[B] \cap [\phi] \neq \emptyset$. We will prove that, in this case, $[B] \cap [\phi] = \min_B([\phi])$.
   - ($\subseteq$) Let $w \in [B] \cap [\phi]$. Because $w \in [B]$, Centeredness dictates that $w \preceq w'$ for all $w' \in W$. Thus, $w \preceq w'$ for all $w' \in [\phi]$. Since $w \in [\phi]$ and is at least as plausible as any other world in $[\phi]$, we have $w \in \min_B([\phi])$.
   - ($\supseteq$) Let $w \in \min_B([\phi])$. Clearly, $w \in [\phi]$. Is $w \in [B]$? Suppose for contradiction that $w \notin [B]$. Since we are in Case Two, we know there is at least one world $w^* \in [B] \cap [\phi]$. By Centeredness, any model of B is strictly more plausible than any non-model, meaning $w^* \prec w$. But since $w^* \in [\phi]$, this strictly contradicts the assumption that w is minimal in $[\phi]$. Hence, w must be in $[B]$, meaning $w \in [B] \cap [\phi]$.
   - So, we have established $\min_B([\phi]) = [B] \cap [\phi]$. Recall that, *ex hypothesi*, $\psi \in B * \phi$, meaning that $\min_B([\phi]) \subseteq [\psi]$. So, it follows immediately that $[B] \cap [\phi] \subseteq [\psi]$.

In both cases, we established $[B] \cap [\phi] \subseteq [\psi]$, which means $B \cup \{\phi\} \vdash \psi$. Since ψ is an arbitrary formula in $B * \phi$, we conclude $B * \phi \subseteq Cn(B \cup \{\phi\})$.

4. Vacuity. Suppose that $B \nvdash \neg\phi$, meaning that $[B] \nsubseteq [\neg\phi] = W \setminus [\phi]$. In other words, there exists at least one world $w \in [B]$ which is also in $[\phi]$, i.e. $[B] \cap [\phi] \neq \emptyset$.

As I have shown in **Case Two** of step **(3)**, if $[B] \cap [\phi] \neq \emptyset$, it follows that

$$\min_B([\phi]) = [B] \cap [\phi]$$

Hence, it follows that

$$T(\min_b([\phi])) = T([B] \cap [\phi])$$

By definition, $T(\min_B([\phi])) = B * \phi$. Let me show that $T([B] \cap [\phi]) = Cn(B \cup \{\phi\})$.

- ($\subseteq$) Suppose $\psi \in T([B] \cap [\phi])$. By definition, $[B] \cap [\phi] \subseteq [\psi]$. Again by definition, $B \cup \{\phi\} \vdash \psi$. Hence, $\psi \in Cn(B \cup \{\phi\})$.
- ($\supseteq$) Suppose $\psi \in Cn(B \cup \{\phi\})$. By definition, $B \cup \{\phi\} \vdash \psi$. Again by definition, $[B] \cap [\phi] \subseteq [\psi]$. Hence, $\psi \in T([B] \cap [\phi])$.

Therefore, $T([B] \cap [\phi]) = Cn(B \cup \{\phi\})$. Substituting in the equivalence above, we get

$$B * \phi = Cn(B \cup \{\phi\})$$

5. Consistency. Suppose $\phi \nvdash \bot$. Note that $\phi \nvdash \bot$ is equivalent to the claim that $[\phi] \nsubseteq [\bot] = \emptyset$, which in turn is strictly equivalent to $[\phi] \neq \emptyset$.

We need to prove that $Cn(B * \phi) \neq \mathcal{L}$. Since it is trivially true that $Cn(B * \phi) \subseteq \mathcal{L}$, we must prove that $Cn(B * \phi) \subset \mathcal{L}$. This means there exists some $\psi \in \mathcal{L}$ such that $\psi \notin Cn(B * \phi)$. Because $B * \phi$ is a logically closed belief set, $Cn(B * \phi) = B * \phi$. Thus, we must find some $\psi \in \mathcal{L}$ such that $\psi \notin B * \phi$. Semantically, this requires finding a ψ such that $\min_B([\phi]) \nsubseteq [\psi]$.

Recall that $Cn(B * \phi) = \mathcal{L}$ means $B * \phi$ is inconsistent. The proof is complete if we can show that $B * \phi$ does not entail a blatant contradiction. Consider the contradiction $p \wedge \neg p \in \mathcal{L}$ (for some atomic $p \in \Phi$).

Obviously, $[p \wedge \neg p] = \emptyset$. So, we must prove:

$$\min_B([\phi]) \nsubseteq [p \wedge \neg p] = \emptyset$$

Since the only set that is a subset of the empty set is the empty set itself, this inclusion fails if and only if $\min_B([\phi]) \neq \emptyset$.

To prove that this set is non-empty, we will proceed by *reductio ad absurdum*. Suppose that $\min_B([\phi]) = \emptyset$.

By definition, a world is in the minimal set iff there is no strictly more plausible world in the truth-set. If the minimal set is empty, it means that *no* world in $[\phi]$ is minimal. Consequently, for every world $w \in [\phi]$, there must exist a distinct world $w' \in [\phi]$ such that $w' \prec_B w$.

This implies that any $w \in [\phi]$ is merely a link in an infinite, strictly descending chain of ever-more-plausible worlds:

$$\cdots \prec_B w_5 \prec_B w_4 \prec_B w_3 \prec_B w_2 \prec_B w_1 \prec_B w$$

However, the Limit Assumption explicitly forbids the existence of such chains without a lower bound. Recall that the Limit Assumption reads as follows:

$$\text{For any sentence } \phi \in \mathcal{L}\text{: If } [\phi] \neq \emptyset \text{, there exists } w^* \in [\phi] \text{ such that } w^* \preceq_B x \text{ for all } x \in [\phi]$$

Since we know from our initial premise that $[\phi] \neq \emptyset$, the Limit Assumption guarantees the existence of this world w^*.

Since $w^* \in [\phi]$ and we assumed that for *every* world in $[\phi]$ there is a strictly more plausible one, there must exist some world $w^\dagger \in [\phi]$ such that $w^\dagger \prec_B w^*$. Recall that $w^\dagger \prec_B w^*$ means $w^\dagger \preceq_B w^*$ and $w^* \npreceq_B w^\dagger$. But the Limit Assumption stated that $w^* \preceq_B x$ for *all* $x \in [\phi]$, which strictly entails $w^* \preceq_B w^\dagger$.

We have derived both $w^* \npreceq_B w^\dagger$ and $w^* \preceq_B w^\dagger$: contradiction.

Therefore, we must reject the *reductio* assumption. The set $\min_B([\phi])$ is non-empty. Because it is non-empty, it implies that $\min_B([\phi]) \nsubseteq [p \wedge \neg p] = \emptyset$, meaning that $p \wedge \neg p \notin B * \phi$. Because there is at least one formula in $\mathcal{L}$ not entailed by $B * \phi$, we conclude that $Cn(B * \phi) \neq \mathcal{L}$.

6. Congruence. Suppose that $Cn(\{\phi\}) = Cn(\{\psi\})$. Therefore, we have $\phi \vdash \psi$ and $\psi \vdash \phi$, meaning that

$$[\phi] = [\psi]$$

Hence, since $\min_B(\cdot)$ is a function, we have that

$$\begin{aligned} B * \phi &= T(\min_B([\phi])) \\ &= T(\min_B([\psi])) \\ &= B * \psi \end{aligned}$$

7. Superexpansion. We must prove that $B * (\phi \wedge \psi) \subseteq (B * \phi) + \psi$.

Let $\chi \in B * (\phi \wedge \psi)$. By definition, this assumption means $\chi \in T(\min_B([\phi \wedge \psi]))$, i.e.:

$$\min_B([\phi] \cap [\psi]) \subseteq [\chi]$$

Boolean and Set-Theoretic Operations >

Although we did not prove this together, it is straightforward to verify that, by means of the truth-set function $[\cdot]$, there exists a "correspondence" between Boolean operators (i.e., the connectives of our language $\mathcal{L}$, $\wedge, \vee, \neg, \rightarrow, \leftrightarrow$) and set-theoretic operators (i.e., $\cup, \cap, ^c$). (N.b., recall that for any universe set U and set $A \subseteq U$, $A^c = U \setminus A$.)

$$[\neg\phi] = [\phi]^c = W \setminus [\phi]$$
$$[\phi \wedge \psi] = [\phi] \cap [\psi]$$
$$[\phi \vee \psi] = [\phi] \cup [\psi]$$
$$[\phi \rightarrow \psi] = [\phi]^c \cup [\psi]$$
$$[\phi \leftrightarrow \psi] = ([\phi] \cap [\psi]) \cup ([\phi]^c \cap [\psi]^c)$$

In the following, I will make use of this correspondence.

We need to prove that $\chi \in (B * \phi) + \psi$. By definition, $(B * \phi) + \psi = Cn((B * \phi) \cup \{\psi\})$. The problem here is that we cannot easily use the semantic definition of $*$ or the postulates we already proved directly on $Cn((B * \phi) \cup \{\psi\})$. We need to find a way to "remove" $\{\psi\}$. Luckily, there exists a very straightforward way of doing so, which is the **Deduction Theorem** for classical logic – see Session 1 – which states that, for any $\Gamma \subseteq \mathcal{L}$ and $x, y \in \mathcal{L}$:

$$y \in Cn(\Gamma \cup \{x\}) \iff x \rightarrow y \in Cn(\Gamma)$$

As stated above, we need to prove $\chi \in (B * \phi) + \psi = Cn((B * \phi) \cup \{\psi\})$. So, by the Deduction Theorem, we need to prove:

$$\begin{aligned} \chi \in Cn((B * \phi) \cup \{\psi\}) &\iff \psi \rightarrow \chi \in Cn(B * \phi) \\ &\iff \psi \rightarrow \chi \in B * \phi \qquad \text{for } B * \phi \text{ is a belief set} \end{aligned}$$

So, I will prove that $\psi \rightarrow \chi$ is in $B * \phi$. That is, we need to prove that $\psi \rightarrow \chi \in T(\min_B([\phi]))$. Semantically, this means we must prove:

$$\min_B([\phi]) \subseteq [\psi \rightarrow \chi]$$

How can we further simplify the statement we need to prove, i.e., $\min_B([\phi]) \subseteq [\psi \rightarrow \chi]$? The easiest approach is to simplify the truth-set of the implication $\psi \rightarrow \chi$ by considering the conditions under which a world w satisfies $\psi \rightarrow \chi$.

Recall the truth table for "$\rightarrow$". $\psi \rightarrow \chi$ is true iff either ψ is false or χ is true. So, $[\psi \rightarrow \chi] = (W \setminus [\psi]) \cup [\chi]$, which we can write as $[\psi]^c \cup [\chi]$. You can easily verify that a world w is in $[\psi]^c \cup [\chi]$ if, and only if, $w \models \psi \rightarrow \chi$.

So, we need to prove:

$$\min_B([\phi]) \subseteq [\psi]^c \cup [\chi]$$

Let w be an arbitrary world in $\min_B([\phi])$. We want to show that $w \in [\psi]^c \cup [\chi]$. We have two cases:

1. **Case One:** $w \notin [\psi]$. Hence, $w \in [\psi]^c$. Therefore, $w \in [\psi]^c \cup [\chi]$ trivially holds.
2. **Case Two:** $w \in [\psi]$. In this case, $w \in \min_B([\phi]) \cap [\psi] \subseteq [\phi] \cap [\psi]$. I will show that w must actually be minimal in the intersection, i.e., $w \in \min_B([\phi] \cap [\psi])$, which together with our assumption implies that $w \in [\chi]$. Suppose the opposite is true, i.e. $w \notin \min_B([\phi] \cap [\psi])$. So, there exists some world $w' \in [\phi] \cap [\psi]$ such that $w' \prec_B w$. Since $w, w' \in [\phi]$, the fact that $w' \prec_B w$ contradicts our starting assumption that $w \in \min_B([\phi])$. Thus, no such w' exists, and $w \in \min_B([\phi] \cap [\psi])$. Because we know from our initial premise that $\min_B([\phi] \cap [\psi]) \subseteq [\chi]$, it follows that $w \in [\chi]$. Therefore, $w \in [\psi]^c \cup [\chi]$.

Either way, $w \in [\psi]^c \cup [\chi]$. Since w is an arbitrary world in $\min_B([\phi])$, we have established $\min_B([\phi]) \subseteq [\psi \rightarrow \chi]$. Hence, $(\psi \rightarrow \chi) \in T(\min_B([\phi])) = B * \phi$. This gives us $\chi \in (B * \phi) + \psi$. Since χ is an arbitrary formula from $B * (\phi \wedge \psi)$, we conclude $B * (\phi \wedge \psi) \subseteq (B * \phi) + \psi$. □

8. Subexpansion. Suppose that $\neg\psi \notin B * \phi$. We must show that $(B * \phi) + \psi \subseteq B * (\phi \wedge \psi)$. First of all, let me clarify what exactly we need to prove here. As we did in the proof of **Superexpansion**, we need to “unpack” the statement "$(B * \phi) + \psi \subseteq B * (\phi \wedge \psi)$" into something easier to prove by applying the definitions.

$$\begin{aligned}(B*\phi)+\psi\subseteq B*(\phi\wedge\psi) &\iff Cn(T(\min_B([\phi]))\cup\{\psi\})\subseteq T(\min_B([\phi\wedge\psi]))\\ &\iff \text{For all } \chi: \chi\in Cn(T(\min_B([\phi]))\cup\{\psi\}) \implies \chi\in T(\min_B([\phi\wedge\psi]))\\ &\iff \text{For all } \chi: \psi\to\chi\in Cn(T(\min_B([\phi]))) \implies \chi\in T(\min_B([\phi\wedge\psi]))\\ &\iff \text{For all } \chi: \psi\to\chi\in T(\min_B([\phi])) \implies \chi\in T(\min_B([\phi\wedge\psi]))\\ &\iff \text{For all } \chi: \min_B([\phi])\subseteq[\psi\to\chi] \implies \min_B([\phi\wedge\psi])\subseteq[\chi]\\ &\iff \text{For all } \chi: \min_B([\phi])\subseteq[\psi]^c\cup[\chi] \implies \min_B([\phi\wedge\psi])\subseteq[\chi]\end{aligned}$$

Therefore, we need to prove that, for any sentence χ, if $\min_B([\phi])\subseteq[\psi]^c\cup[\chi]$, then $\min_B([\phi\wedge\psi])\subseteq[\chi]$. So, let us suppose that $\min_B([\phi])\subseteq[\psi]^c\cup[\chi]$, and let us go back to the starting premise we made at the beginning of this section. That premise may be unpacked as follows:

$$\begin{aligned}\neg\psi\notin B*\phi &\iff \neg\psi\notin T(\min_B([\phi]))\\ &\iff \min_B([\phi])\not\subseteq[\psi]^c\\ &\iff \min_B([\phi])\cap[\psi]\neq\emptyset\end{aligned}$$

We need to prove that $\min_B([\phi\wedge\psi])\subseteq[\chi]$, or alternatively that $\min_B([\phi]\cap[\psi])\subseteq[\chi]$. First of all, note that the following statement is true given our supposition that $\min_B([\phi])\subseteq[\psi]^c\cup[\chi]$:

$$\min_B([\phi])\cap[\psi]\subseteq[\chi] \tag{I}$$

For if x belongs to the left-hand side, then in particular $x\in\min_B([\phi])$, and by our supposition $x\in[\psi]^c\cup[\chi]$. But since x is in the left-hand side, it must be that $x\in[\psi]$, meaning $x\notin[\psi]^c$. Because $x\in[\psi]^c\cup[\chi]$ and $x\notin[\psi]^c$, it strictly follows that $x\in[\chi]$, thereby establishing the subset inclusion above.

So, to prove $\min_B([\phi\wedge\psi])\subseteq[\chi]$, I will first prove the following identity:

$$\min_B([\phi\wedge\psi])=\min_B([\phi])\cap[\psi] \tag{II}$$

From (I) and (II) it follows immediately that our conclusion holds, i.e., that $\min_B([\phi\wedge\psi])\subseteq[\chi]$. So, let us establish (II).

($\supseteq$) Suppose for *reductio* that $w\in\min_B([\phi])\cap[\psi]$ but $w\notin\min_B([\phi\wedge\psi])$. Because $w\in[\phi]$ and $w\in[\psi]$, w satisfies $\phi\wedge\psi$. If it is not minimal in $[\phi\wedge\psi]$, there must exist some $w'\in[\phi\wedge\psi]$ such that $w'\prec_B w$. However, since $w'\in[\phi\wedge\psi]$, we know $w'\in[\phi]$. The fact that $w'\in[\phi]$ and $w'\prec_B w$ strictly contradicts our starting premise that $w\in\min_B([\phi])$.

Hence, no such w' exists, and $w\in\min_B([\phi\wedge\psi])$, which implies that:

$$\min_B([\phi])\cap[\psi]\subseteq\min_B([\phi\wedge\psi])$$

($\subseteq$) Let us prove that the inclusion also holds the other way around. It is crucial to remember here that $\min_B([\phi])\cap[\psi]\neq\emptyset$ (our initial assumption). Let us call $w^\dagger$ a world in this intersection. Now, suppose for *reductio* that $w\in\min_B([\phi\wedge\psi])$ but $w\notin\min_B([\phi])\cap[\psi]$. Since $w\in[\phi\wedge\psi]$, we know $w\in[\psi]$. Therefore, the only way w is excluded from the intersection is if $w\notin\min_B([\phi])$. If $w\notin\min_B([\phi])$, there must exist some $w^*\in[\phi]$ such that $w^*\prec_B w$. Now consider our three worlds:

1. Because $w\in\min_B([\phi\wedge\psi])$ and $w^\dagger\in[\phi\wedge\psi]$, it must be that $w\preceq_B w^\dagger$.
2. Because $w^\dagger\in\min_B([\phi])$ and $w^*\in[\phi]$, it must be that $w^\dagger\preceq_B w^*$.
3. By our assumption above, $w^*\prec_B w$.

Stringing these together yields $w\preceq_B w^\dagger\preceq_B w^*\prec_B w$, which implies $w\prec_B w$. This is a contradiction, since $w\prec_B w$ means $w\preceq_B w$ and $w\not\preceq_B w$.

Thus, we reject the *reductio* assumption, meaning $w\in\min_B([\phi])\cap[\psi]$, which implies that $\min_B([\phi\wedge\psi])\subseteq\min_B([\phi])\cap[\psi]$.

Now that we have established $\min_B([\phi \wedge \psi]) = \min_B([\phi]) \cap [\psi]$, we can easily prove our desired claim. By (II), $\min_B([\phi \wedge \psi]) = \min_B([\phi]) \cap [\psi]$ and by (I), $\min_B([\phi]) \cap [\psi] \subseteq [\chi]$. So, we conclude that:

$$\min_B([\phi \wedge \psi]) \subseteq [\chi]$$

Recall that χ is just an arbitrary formula in $\mathcal{L}$. So, we have proved that:

$$\text{For all } \chi : \min_B([\phi]) \subseteq [\psi]^c \cup [\chi] \implies \min_B([\phi \wedge \psi]) \subseteq [\chi]$$

which is logically equivalent to Subexpansion, as shown at the beginning of the sub-proof, and hence we have proved Subexpansion.

Conclusion. Therefore, we have proved that, for any belief set B associated with an order $\preceq_B$ on W satisfying the properties listed above, and any proposition ϕ, if we define $B * \phi$ as $T(\min_B([\phi]))$, all the eight postulates of AGM belief revision are satisfied. $\square$

5. Exercisess

1. Proofs. Prove all the lemmas and propositions we did not prove together. See this file for the solutions.