Session 1

The aim of today's session is to get acquainted with belief revision theory, review some preliminary notions, and think about why it is worth getting our heads around it as PhD students in Philosophy.

In very general terms, belief revision theory is a branch of Philosophy (and Logic) that tackles the following question:

> What are the general principles that a reasoner (whether human or artificial) should follow as they acquire more and more information from the world?

Consider the following example.

Example 1 (Three Composers)**.**

You are a newbie when it comes to opera.

1. *A friend of yours who's really into this topic tells you that Verdi is Italian, while Bizet and Satie are both French, and you come to believe so (t_1).*
2. *Later on, you find out that Verdi and Bizet are compatriots (t_2). So, you are now uncertain as to whether Verdi and Bizet are both Italian or both French, but you still believe that Satie is French.*
3. *Eventually, you find out on Wikipedia that they are in fact all compatriots (t_3). As a result, you are now uncertain as to whether they are all Italian or all French.*

Question: Is the way you have changed your beliefs through t_1, t_2 and t_3 rational?

Answer: It depends!

What belief revision theory can do for us is help us understand under what conditions changing our beliefs as in Example 1 (Three Composers) is rational, and under what conditions it is not. For example, under certain theories of belief revision, the following principle is true:

> **Preservation**. If you learn P, and P is consistent with what you already believe, it is not rational to drop any belief Q as a result of learning P.

In other words, **Preservation** makes you very conservative when it comes to belief. For example, a theory that validates **Preservation** is a theory that deems you *irrational* in Example 1 (Three Composers). For, at t_2, you believe that Satie is French and, between t_2 and t_3, you learn something consistent with that belief—namely, that the three composers are all compatriots (since it is possible that they are all French). However, at t_3 you drop your belief that Satie is French, and believe only that they are either all Italian or all French.

Belief revision theory allows us to discuss such cases in a very general way and to compare theories clearly and unambiguously. In particular, what we are interested in when studying belief revision theory is establishing results like the following:

Definition 2 (Representation Result (Schema))**.**

Let $R_1, R_2, \ldots, R_n$ be a set of belief revision rules, and S some kind of structure. A representation result for $R_1, R_2, \ldots, R_n$ and structures of kind S is a proof of the following statement:

> The rules $R_1, R_2, \ldots, R_n$ are valid on a structure if, and only if, that structure is of kind S.

Why should we study these kinds of results?

1. One reason is that representation theorems demonstrate a formal **equivalence between axiomatic and semantic approaches**. Specifying correct belief revision rules ($R_1, R_2, \ldots, R_n$) and defining a mathematical structure of kind S to interpret a logical language amount to the same theoretical commitment. The rule set $R_1, R_2, \ldots, R_n$ and the structure S yield identical predictions regarding the rationality of a given belief change.
2. This presupposes, of course, that studying belief revision is itself worthwhile. There are two primary reasons why it is. First, the rules of belief revision govern rational outright belief in **the same way probability theory governs rational credence**. For those interested in the epistemic rationality of full belief, this framework is indispensable. Second, even for those less interested in epistemology, studying these results provides an opportunity to apply mathematical logic to a **domain where the formalism retains a strong intuitive grip** – namely, modeling rational agents who accept and retract beliefs upon acquiring new information.

Please find a summary of all definitions and propositions discussed in this note here: link**.**

1. Formal Preliminaries

1.1. Language and Interpretations

To do belief revision theory, we need some basic logical preliminaries. I assume everybody has done some courses or standalone study of classical propositional logic, so I won't revise that completely, but I will simply set the stage on which we are going to elaborate.

First, let us define a formal language of interest. We will concern ourselves with a language $\mathcal{L}$ so constructed.

Definition 3 (Language $\mathcal{L}$)**.**

A formal language $\mathcal{L}$ is defined as follows. Let Φ be a (countable) set of objects called "propositional variables", that is $\Phi := \{p_1, p_2, p_3, \ldots\}$. Then, we define $\mathcal{L}$ via induction as follows.

1. Base: $\Phi \subseteq \mathcal{L}$.
2. Step: For all $\phi, \psi \in \mathcal{L}$:
    1. If $\phi \in \mathcal{L}$, then $\neg\phi \in \mathcal{L}$;

2. If $\phi \in \mathcal{L}$ and $\psi \in \mathcal{L}$ then $\phi \wedge \psi \in \mathcal{L}$;
3. If $\phi \in \mathcal{L}$ and $\psi \in \mathcal{L}$ then $\phi \vee \psi \in \mathcal{L}$;
4. If $\phi \in \mathcal{L}$ and $\psi \in \mathcal{L}$ then $\phi \to \psi \in \mathcal{L}$.

3. Nothing else is in $\mathcal{L}$.

Note the greek letters $\phi, \psi, \chi, \ldots$ are *variables* for the sentences in $\mathcal{L}$, not sentences themselves. $\mathcal{L}$ contains *only* objects like $p, p_1, (p_2 \wedge p_{124}) \vee \neg p_{17}$, and so on.

Next, we need to define the semantics for our language. As it is standard, we can think of a possible world (or interpretation) simply as an assignment of truth values to our basic propositional variables. Then, we define what it means for any sentence ϕ to be true at a world using the satisfaction relation $\models$. As you know, some sentences ϕ have a special status: some are true irrespective of the interpretation you consider (i.e., tautologies) and others are false irrespective of the interpretation you consider (i.e., contradictions).

Definition 4 (Interpretations and Truth ($\models$)).

Let $\mathcal{L}$ be given. A possible world (or interpretation) w is a function $w : \Phi \to \{0, 1\}$ that assigns a truth value to each propositional variable $p \in \Phi$. Let W be the set of **all** possible worlds.

We define the satisfaction relation $w \models \phi$ (read as "ϕ is true at w") via induction on the structure of ϕ:

1. Base: For all $p \in \Phi$, $w \models p \iff w(p) = 1$.
2. Step: For all $\phi, \psi \in \mathcal{L}$:
   1. $w \models \neg\phi \iff w \not\models \phi$;
   2. $w \models \phi \wedge \psi \iff w \models \phi \text{ and } w \models \psi$.

⚠ Sentences and Propositions

As you already know, there are two ways in which we may make reference to a sentence ϕ. First, as a syntactic object in $\mathcal{L}$. This is the most obvious way, given Definition 4 (Interpretations and Truth ($\models$)). However, another way of referring to a sentence is by considering its "truth-set", i.e., the set of worlds at which the sentence is true. Let $[\![\cdot]\!]$ be a function that takes a sentence $\phi \in \mathcal{L}$ and maps it to its truth-set $[\![\phi]\!] \subseteq W$, that is:

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

Usually, when philosophers talk about **propositions** rather than sentences, they refer to what we formalize through a truth-set (in a way, the truth-set of ϕ is the "content" of ϕ). Note that propositions and sentences are not the same thing, and also they are not "equivalent", in the sense that $[\![\cdot]\!]$ "looses" some information. For instance, consider that while $\neg(\neg p_1 \wedge \neg p_2)$ and $p_1 \vee p_2$ are different objects in $\mathcal{L}$, $[\![\neg(\neg p_1 \wedge \neg p_2)]\!] = [\![p_1 \vee p_2]\!]$.

Before moving on, let me make an important remark regarding the cardinality of W. As we will see, how many worlds are in W will make a difference when doing belief revision theory.

1. Suppose first that Φ_{fin}, the set of propositional variables, is finite. That is, $|\Phi_{fin}| = n$ for some $n \in \mathbb{N}$. How many possible worlds are in W?
2. Suppose now that Φ is countable, i.e., $|\Phi| = |\mathbb{N}|$. How many possible worlds are in W?

1.2. Logical Consequence

We can already say something about belief revision theory using what we established in 1. Formal Preliminaries. First, **an agent will be represented by the propositions they believe**. That is, a rational agent will be represented by a *belief set*, which is a special kind of set $B \subseteq \mathcal{L}$ of sentences in our formal language (we will define exactly what makes it special in a moment) representing the sentences the agent takes to be true. Second, belief revision will be defined as an operator on those sets B. Suppose an agent believes all the sentences in B and that they learn ϕ. The new belief set they obtain by revising B by ϕ is denoted as:

$$B * \phi$$

Clearly, we need more information to decide what should actually be in $B * \phi$ (given the initial content of B). To do so, however, **we first need to clarify what *belief sets* are**, given that they are not just any subset of $\mathcal{L}$. This is exactly why we need the notion of logical consequence.

In general, we define logical consequence based on the formal work done in 1. Formal Preliminaries as follows.

Definition 5 (Logical Consequence ($\vdash$)).

Let $\phi, \psi \in \mathcal{L}$ and W be given. We say that ψ follows logically from ϕ (in symbols, $\phi \vdash \psi$) iff the latter is true at all the worlds w that make the former true. In symbols:

$$\phi \vdash \psi \quad :\Longleftrightarrow \quad \text{For all } w \in W : w \models \phi \implies w \models \psi$$

Note that:

1. Definition 5 (Logical Consequence ($\vdash$)) may be extended, with a slight abuse of notation, to include the consequence relation between **sets of formulas** Γ and a sentence ϕ, i.e., $\Gamma \vdash \phi$. We do this by requiring that $\Gamma \vdash \phi$ iff ϕ is true at all the worlds w which satisfy *all* the sentences $\gamma \in \Gamma$.
2. I distinguish the satisfaction ($\models$) and consequence ($\vdash$) relations symbolically. We won't need a third relation for syntactic deducibility (which is what "$\vdash$" usually denotes, strictly speaking), so we will use $\vdash$ for semantic consequence here to keep things simple.

When doing belief revision theory, we will frequently use a *consequence operator* Cn instead of a consequence relation, as it makes it easier to express certain properties in a concise and brief way. Note that Cn and $\vdash$ are inter-definable. Since we started with $\vdash$, we take it as basic and define Cn in terms of $\vdash$.

Definition 6 (Consequence Operator Cn).

Let $\vdash$ be given. We define the consequence operator Cn (of $\vdash$) as a function from sets of sentences $A \subseteq \mathcal{L}$ to sets of sentences $B \subseteq \mathcal{L}$:

$$Cn(A) := \{\phi \in \mathcal{L} : A \vdash \phi\}$$

N.b., Cn is a function from $\mathcal{P}(\mathcal{L})$ to $\mathcal{P}(\mathcal{L})$.

Now, we can finally define what a belief set is.

Definition 7 (Belief Sets)**.**

Let B be a subset of $\mathcal{L}$. B is a belief set if, and only if, the following equivalence holds:

$$B = Cn(B)$$

Let us call $\mathbf{B}$ the set of all belief sets.

In other words, belief sets are not just sets of sentences, but they are sets of sentences "closed under logical entailment". We usually explain this requirement by saying that logical closure is—in this context—the "mark of rationality". That is, a rational agent (as modeled in this simplified setting) is an agent that believes all the logical consequences of what they believe.

Before concluding this section, let me point out that logical consequence has many important properties, independently of whether we represent it via Cn or $\vdash$. In the following, I will introduce some of them. I will first define some of them in general terms (i.e., for arbitrary consequence relation $\triangleright$ and consequence operation C).

Definition 8 (Reflexivity of $\triangleright$)**.**

Let $\triangleright$ be a consequence relation on $\mathcal{L}$. $\triangleright$ is reflexive iff:

$$\text{For all } \gamma \in \Gamma: \quad \Gamma \triangleright \gamma$$

Definition 9 (Transitivity of $\triangleright$)**.**

Let $\triangleright$ be a consequence relation on $\mathcal{L}$. $\triangleright$ is transitive iff:

$$\text{If } \Gamma \triangleright \delta \text{ for all } \delta \in \Delta, \text{ and } \Delta \triangleright \phi \quad \Longrightarrow \quad \Gamma \triangleright \phi$$

Definition 10 (Monotonicity of $\triangleright$)**.**

Let $\triangleright$ be a consequence relation on $\mathcal{L}$. $\triangleright$ is monotonic iff:

$$\text{If } \Gamma \triangleright \phi \text{ and } \Gamma \subseteq \Delta \quad \Longrightarrow \quad \Delta \triangleright \phi$$

It is straightforward to prove the following:

Proposition 11.

Let $\vdash$ be the consequence relation defined in Definition 5 (Logical Consequence ($\vdash$)). $\vdash$ is reflexive, transitive, and monotonic.

Proof. See 4. Exercises. □

Note that these properties of the consequence relation correspond to analogous properties of the consequence operator Cn. I will omit the details here, but the following result is derivable from the definitions above and Proposition 11.

Proposition 12.

Let $\vdash$ be the consequence relation defined in Definition 5 (Logical Consequence ($\vdash$)), and Cn the consequence operator defined in Definition 6 (Consequence Operator Cn) based on $\vdash$. Cn satisfies the following properties:

1. ***Reflexivity**: For all $\Gamma \subseteq \mathcal{L}$, $\Gamma \subseteq Cn(\Gamma)$.*
2. ***Transitivity**: For all $\Gamma, \Delta \subseteq \mathcal{L}$, if $\Delta \subseteq Cn(\Gamma)$ then $Cn(\Delta) \subseteq Cn(\Gamma)$.*
3. ***Monotonicity**: For all $\Gamma, \Delta \subseteq \mathcal{L}$, if $\Delta \subseteq \Gamma$ then $Cn(\Delta) \subseteq Cn(\Gamma)$.*
4. ***Idempotence**: For all $\Gamma \subseteq \mathcal{L}$, $Cn(Cn(\Gamma)) = Cn(\Gamma)$.*

Proof. It is possible to prove this result from Proposition 11 and Definition 6 (Consequence Operator Cn). See also 4. Exercises. □

The classical consequence relation $\vdash$ has other important properties

Proposition 13.

Let $\vdash$ be the consequence relation defined in Definition 5 (Logical Consequence ($\vdash$)). The following properties hold for $\vdash$.

1. ***Compactness**: $\Gamma \vdash \phi \implies \exists \Gamma_{fin} \subseteq \Gamma \text{ such that } \Gamma_{fin} \vdash \phi$, where Γ_{fin} is a finite subset of Γ.*
2. ***Deduction Theorem**: If $\Gamma \cup \{\phi\} \vdash \psi$, then $\Gamma \vdash \phi \to \psi$.*
3. ***Disjunction in the Premises**: If $\Gamma \cup \{\phi_1\} \vdash \psi$ and $\Gamma \cup \{\phi_2\} \vdash \psi$, then $\Gamma \cup \{\phi_1 \vee \phi_2\} \vdash \psi$.*

Proof. See Compactness. □

The proof of compactness is mathematically a bit more involved (I have prepared a supplementary note on this that we can discuss separately if you are interested!).

If you want to dive deeper into the properties of classical consequence, or if you are looking for a beautifully written explanation of consequence relations *in general* (abstracting away from our

specific semantic definition in Definition 5 (Logical Consequence ($\vdash$)), I highly recommend David Makinson's *Bridges from Classical to Nonmonotonic Logic* (2005).

Now we are ready to introduce belief revision theory.

2. Belief Revision Theory

The best way to get a feel for this topic is by diving into the most famous framework for belief change: **AGM theory**. Named after the three scholars who introduced it in 1985—Carlos Alchourrón, Peter Gärdenfors, and David Makinson—it is essentially the gold standard in the field!

I'll start by **introducing AGM axiomatically**. This simply means we'll look at the basic rules (or postulates) that describe how an ideally rational agent—let's call them an “AGM-rational” agent—ought to update their beliefs. After that, we'll explore **the semantic structures we use to actually model these agents**. Our main goal here is to build up to AGM's famous completeness result (see Definition 2 (Representation Result (Schema))).

> ⓘ **A Quick Heads-Up: Mathematics vs. Philosophy**
>
> As we go through this, it's really important to separate the hard math from the philosophical interpretations. For instance, proving that the AGM postulates correspond to a specific mathematical structure is an uncontroversial mathematical theorem. However, claiming that “AGM-rationality” is rationality simpliciter is a philosophical claim—and a highly debated one!

To formalize the AGM framework, let $B \in \mathbf{B}$ be any belief set, $\phi \in \mathcal{L}$ be any sentence, and $* : \mathbf{B} \times \mathcal{L} \to \mathbf{B}$ be a candidate revision function. In plain English, $*$ is just a function that takes your initial belief set B (from the collection of all possible belief sets, $\mathbf{B}$) and a new piece of information ϕ (a sentence from our language $\mathcal{L}$), and gives you back a brand-new set: $B * \phi$. The set $B * \phi$ is the **revision of B by ϕ** iff $*$ validates the postulates before.

Before detailing these postulates, however, it is helpful to understand the **extra-logical motivations** underpinning them:

1. **Consistency and Closure:** We require the revised belief set, $B * \phi$ (the result of revising B by ϕ), to be logically closed, consistent (unless ϕ is logically contradictory), and to successfully incorporate the new information ϕ.
2. **Informational Economy:** The new belief set $B * \phi$ should differ from the original set B as little as possible. Revision must be conservative, retaining as much prior information as logical consistency allows.

These “meta-principles” – particularly the goal of informational economy – drive the specific characterization of the revision operator $*$. Again: these meta-principles **are not true as a matter of logic, but are decided in advance on extra-logical grounds**.

The function $*$ is a basic AGM belief revision function if and only if it satisfies the following postulates. Let us start with **Closure**:

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

This has a very straightforward meaning: when you revise your beliefs, you shouldn't end up with a fragmented, messy pile of isolated sentences. You should always arrive at a new, stable epistemic state that is logically closed.

(You might be thinking: "Wait, isn't this redundant if we already defined $$ as a function that outputs a belief set?" Mathematically, yes! But we state it explicitly as an axiom to ensure this property of revision independently of the definition of $*$ as some kind of function.)

Also, keep in mind that **the AGM axioms do not single out just *one* unique way to revise beliefs**. Knowing the rules of logic alone isn't enough to tell us exactly what ends up inside $B * \phi$ for a specific agent. The postulates merely set the boundary lines for what counts as a "rational" change. As we clarified earlier (see Example 1 (Three Composers)), to figure out the exact output of a revision, we have to bring in extra-logical information—like how strongly the agent values, or "entrenches," their specific beliefs!

Before moving to the next postulate, consider the following mathematical consequence of **Closure**.

Closure Implies that Every Belief Set Contains all Tautologies ›

It is a trivial mathematical truth that $\emptyset \subseteq B$, where B is a belief set. However, this has an important consequence given the fact that classical logical consequence is monotonic. For it follows by monotonicity that, since $\emptyset \subseteq B$, then $Cn(\emptyset) \subseteq Cn(B)$. Since B is a belief set, we know $Cn(B) = B$, which means $Cn(\emptyset) \subseteq B$. But what exactly is $Cn(\emptyset)$?

$Cn(\emptyset)$ is the set of all propositional tautologies. This can be easily seen by unpacking the meaning of Cn as applied to $\emptyset$:

$$Cn(\emptyset) = \{\phi \in \mathcal{L} : \emptyset \vdash \phi\}$$

In turn, $\emptyset \vdash \phi$ is true iff for all possible worlds $w \in W$: if w satisfies all sentences in $\emptyset$, then $w \models \phi$.

The condition "w satisfies all sentences in $\emptyset$" can be logically spelled out as follows: for all $\psi \in \mathcal{L}$, if $\psi \in \emptyset$, then $w \models \psi$. Since $\psi \notin \emptyset$ for any choice of $\psi \in \mathcal{L}$ by the very definition of the empty set, the antecedent "$\psi \in \emptyset$" is always false. This implies that the conditional "w satisfies all sentences in $\emptyset$" is vacuously true for every possible world $w \in W$. Therefore, the statement $\emptyset \vdash \phi$ is true if, and only if, $w \models \phi$ for all $w \in W$. That is, ϕ is a propositional tautology.

Therefore, any belief set B logically contains all the propositional tautologies.

Also, note that, given **Closure** and the properties of $\vdash$ (and Cn), saying $B \vdash \phi$ (or $\phi \in Cn(B)$) is exactly the same as saying $\phi \in B$.

The second postulate is called **Success**:

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

This postulate has a rather simple meaning: revising by ϕ should result in believing ϕ. The third postulate is called **Consistency**:

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

This postulate has a rather straightforward meaning, but it forces us to deeply understand the meaning of the objects and notation we have introduced. **There are three "things" to clarify here**: the meaning of the antecedent, that of the consequent, and the reason why this postulate is in "If-Then" form.

1. $\phi \nvdash \bot$ (where $\bot$ is just any contradiction you like, e.g., $p_{154} \wedge \neg p_{154} \in \mathcal{L}$) means that ϕ **is *not* a contradiction**. Here is why. Recall that, in general, $\phi \nvdash \psi$ means that there exists a world w such that $w \models \phi$ but $w \nvDash \psi$. So, $\phi \nvdash \bot$ means that there exists *some* world w such that $w \models \phi$ and $w \nvDash \bot$. Since for any world w we have $w \nvDash \bot$, $\phi \nvdash \bot$ **simply means that ϕ is non-contradictory or satisfiable**. (Note: ϕ might be either a tautology or a contingent proposition.)
2. $Cn(B * \phi) \neq \mathcal{L}$ is just an equivalent way of saying that $B * \phi$ **contains no contradictions**. Let me explain this by showing that $Cn(B * \phi) = \mathcal{L}$ if, and only if, $B * \phi$ contains a contradiction.
   - Suppose $Cn(B * \phi) = \mathcal{L}$. Therefore, $\bot \in Cn(B * \phi)$, i.e., $B * \phi \vdash \bot$. Note that since $B * \phi$ is a belief set, it is closed under logical consequence by **Closure**. Therefore, if $B * \phi \vdash \bot$, it must be that $\bot \in B * \phi$.
   - Suppose now that $\bot \in B * \phi$. In classical logic, $\bot \vdash \psi$ for any $\psi \in \mathcal{L}$ (the principle of explosion). So, since classical consequence is monotonic and $\{\bot\} \subseteq B * \phi$, it follows that $Cn(\{\bot\}) \subseteq Cn(B * \phi)$. Since $Cn(\{\bot\}) = \mathcal{L}$, we get $\mathcal{L} \subseteq Cn(B * \phi)$. Ultimately, recall that $Cn(B * \phi)$ is a subset of our language $\mathcal{L}$, hence $Cn(B * \phi) \subseteq \mathcal{L}$ by definition, which implies that $Cn(B * \phi) = \mathcal{L}$.
3. Let me explain now the last bit of information contained in **Consistency**. Why require that $Cn(B * \phi) \neq \mathcal{L}$ only if $\phi \nvdash \bot$, and not in general? **The reason is that requiring it in general would conflict with Success.** For suppose that the agent revises their belief set by an outright contradiction, $\bot$. By **Success**, we have that $\bot \in B * \bot$, which (as we just saw in point 2) implies that $Cn(B * \bot) = \mathcal{L}$.

Escaping an Inconsistent Belief Set

The **Consistency** postulate has an important consequence. Suppose that B is an inconsistent belief set, i.e., $B = \mathcal{L}$. Suppose that ϕ is a non-contradictory sentence. By **Consistency**, it follows that $B * \phi \neq \mathcal{L}$. That is, revising an inconsistent set of beliefs by a non-contradictory sentence restores consistency.

(2) Why Finding the Right Belief Revision Operator Is So Hard

Closure, **Success**, and **Consistency** are rather reasonable constraints on $*$, aren't they? However, accepting them already forces us to reject simple, naive revision operators. Suppose we want to revise a belief set B by some new information ϕ, and currently B contains $\neg\phi$. You might think we can define a simple revision operator $*_{naive}$ that just removes the direct contradiction $\neg\phi$, adds the new information ϕ, and then logically closes the result:

$$B *_{naive} \phi := Cn((B \setminus \{\neg\phi\}) \cup \{\phi\})$$

But this fails precisely because beliefs are logically entangled, which inevitably leads to a violation of **Consistency**. Suppose B contains the belief p (“It is raining”), the belief $p \to \neg q$ (“If it is raining, the picnic is cancelled”), and consequently the belief $\neg q$ (“The picnic is cancelled”). Let's focus on these core beliefs: $\{p, p \to \neg q, \neg q\} \subseteq B$.

Suppose you learn q (“The picnic is on!”) and revise B using $*_{naive}$. First, the operator removes $\neg q$ and adds q, leaving us with the intermediate set containing $\{p, p \to \neg q, q\}$. However, the definition of $*_{naive}$ then requires us to apply Cn to satisfy **Closure**. Since p and $p \to \neg q$ are still in the set and they logically entail $\neg q$, applying Cn brings $\neg q$ right back! Because the closed set now logically contains both q and $\neg q$, it logically explodes into the absurd belief set $\mathcal{L}$ (i.e., $B *_{naive} q = \mathcal{L}$).

Notice that the new information q is perfectly non-contradictory on its own ($q \nvdash \bot$). Yet, our revised belief set became $\mathcal{L}$, which is a direct violation of **Consistency**.

This shows that belief revision cannot just be simple set addition and subtraction of elements from B. When you add new information that contradicts the old, you don't just have to remove the direct contradiction; you have to track down and modify the underlying beliefs that logically entail it.

The fourth postulate is called **Inclusion**:

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

and it sets an “upper bound” to the effects of revision. To fully grasp this, let me clarify what $Cn(B \cup \{\phi\})$ is. In the belief revision literature, the operation that takes a set B and a sentence ϕ and returns $Cn(B \cup \{\phi\})$ is called **expansion** (or sometimes *augmentation*). It is usually denoted by $+$, meaning $B + \phi := Cn(B \cup \{\phi\})$. It has a very straightforward meaning: you simply add the new information to your set of beliefs and take the logical closure of the result. So, **Inclusion** simply says that, at most, belief revision may amount to expansion.

The fifth postulate is called **Vacuity**:

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

This postulate requires more discussion, for its meaning and its role are not completely obvious. As we will see, **Vacuity** is motivated by the informational economy idea briefly mentioned above.

1. First, notice that the right-hand side of the equation is exactly the expansion operation $+$ we just defined above. So, **Vacuity** tells us that revision $*$ is strictly identical with expansion $+$ in certain cases.
2. Let us look closely at the condition $B \nvdash \neg\phi$. This condition tells us **when** revision and expansion deliver the same result.
    1. First of all, let's intuitively clarify its meaning. Informally, $B \nvdash \neg\phi$ means two things.
        1. First, it means that B is consistent. Recall that $B = Cn(B)$, so if B were inconsistent, then $Cn(B) = \mathcal{L}$, which means $B \vdash \psi$ for *all* $\psi \in \mathcal{L}$ (including $\neg\phi$). So, if $B \nvdash \neg\phi$, it must be consistent.
        2. Second, $B \nvdash \neg\phi$ means that B is “compatible” with ϕ. To see this, recall that $\Gamma \vdash \psi$ is a universal statement, saying that *every* world w satisfying Γ satisfies ψ as well. Its

negation, then, is an existential statement, saying that there exists a world where all $\gamma \in \Gamma$ are true but ψ is false. Thus, $B \nvdash \neg\phi$ **means that there exists at least a world** w **where all** $b \in B$ **are true but** $\neg\phi$ **is false, i.e.** ϕ **is true**. The fact that there exists a world where both B and ϕ are true means that the two are **compatible**.

2. Now that we understand what the condition "$B \nvdash \neg\phi$" says: why does AGM say that revision and expansion coincide *if* $B \nvdash \neg\phi$ is the case? That is, why not define $*$ in such a way that it just is $+$ for *every case*? After all, $+$ is a very simple operation. **The problem is that, precisely when** $B \vdash \neg\phi$**, claiming that** $* = +$ **has disastrous consequences**. Suppose that $B \vdash \neg\phi$, i.e., $\neg\phi \in Cn(B)$. Since $B \subseteq B \cup \{\phi\}$ obviously holds, it follows by the monotonicity of Cn that $Cn(B) \subseteq Cn(B \cup \{\phi\})$. So, $\neg\phi \in Cn(B \cup \{\phi\})$, and we also have $\phi \in Cn(B \cup \{\phi\})$ by the reflexivity of Cn. Therefore, their conjunction $\phi \wedge \neg\phi \in Cn(B \cup \{\phi\})$, which means $Cn(B \cup \{\phi\}) = \mathcal{L}$. If we forced $* = +$ across the board, it would follow by **Closure** that $B * \phi = \mathcal{L}$ whenever the new information contradicts our prior beliefs. The problem, however, is that **if the new information** ϕ **is not contradictory in itself (**$\phi \nvdash \bot$**), this is a direct violation of the Consistency postulate!** To avoid this problem, we do not equate $*$ with $+$ across the board, but only in a specific, safe case: when $B \nvdash \neg\phi$.

The sixth postulate is called **Congruence**:

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

and it simply says that, if ϕ and ψ are logically equivalent, then revising by ϕ is exactly the same as revising by ψ. That is, **what you end up believing as a result of revision depends on the *content* of a proposition** (world-theoretically, its truth-set) and not on its syntactic presentation. (This postulate is sometimes called Dalal's Principle of Irrelevance of Syntax). Take a moment to convince yourself that ϕ and ψ are logically equivalent in the semantic sense (i.e., $[\![\phi]\!] = [\![\psi]\!]$) if, and only if, $Cn(\{\phi\}) = Cn(\{\psi\})$.

This concludes our discussion of the main AGM postulates. The following two rules, usually called the supplementary postulates, are slightly more cumbersome. However, they are necessary for proving the representation theorem, as they will significantly constrain the kind of structure S we will consider.

2.1. Supplementary Postulates

The seventh postulate is called **Superexpansion**:

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

This postulate is similar to **Inclusion** in 2. Belief Revision Theory above. It says that revising B by a conjunction may result in, at most, revising by one conjunct and then expanding by the other. Alternatively, **you are not permitted to believe more by revising by** $\phi \wedge \psi$ **than you would if you simply revised by** ϕ **and then expanded by** ψ**.**

The eighth postulate is called **Subexpansion**:

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

This postulate says that **when** $B * \phi$ **is compatible with** ψ**, expanding by** ψ **on top of revising by** ϕ **may result in having at most the same beliefs as in the case where you simply revise by the conjunction** $\phi \wedge \psi$**.** It is very important to appreciate a consequence of these two

postulates taken together: if the result of revising by one conjunct, i.e., $B * \phi$, is compatible with the other conjunct ψ, then

$$B * (\phi \wedge \psi) = (B * \phi) + \psi$$

That is, revising by the conjunction of the two is exactly the same as revising by the first and then expanding by the second. Let me clarify two points here:

1. The equality above is not affected by the order of the conjuncts, i.e., by whether we revise by $\phi \wedge \psi$ or $\psi \wedge \phi$. The reason is that $B * (\phi \wedge \psi) = B * (\psi \wedge \phi)$ as a result of **Congruence**, since $Cn(\phi \wedge \psi) = Cn(\psi \wedge \phi)$.
2. The requirement that $B * \phi$ (or $B * \psi$) is compatible with ψ (or ϕ) is crucial. For if $B * \phi \vdash \neg\psi$, we would have that $B * (\phi \wedge \psi) \neq \mathcal{L}$ as a result of **Consistency**, but $(B * \phi) + \psi = \mathcal{L}$. Here is why:
    1. $(B * \phi) + \psi = Cn((B * \phi) \cup \{\psi\})$.
    2. Obviously, $B * \phi \subseteq (B * \phi) \cup \{\psi\}$.
    3. By the monotonicity of Cn, $Cn(B * \phi) \subseteq Cn((B * \phi) \cup \{\psi\})$.
    4. $\neg\psi \in Cn(B * \phi)$ *ex hypothesi*, while $\psi \in Cn((B * \phi) \cup \{\psi\})$ by the reflexivity of Cn.
    5. Therefore, $\psi \wedge \neg\psi \in Cn((B * \phi) \cup \{\psi\}) = (B * \phi) + \psi = \mathcal{L}$.

Why We Need the Supplementary Postulates >

You might wonder why we bother with these two extra rules. The answer lies in the Completeness Result we mentioned earlier.

The first six postulates only guarantee a “weak” kind of revision (called *partial meet revision*). As we will see, the mathematical structures that allow us to encode these belief revision rules are ordered structures. As we will see, by adding **Superexpansion** and **Subexpansion** we enforce a strict, transitive order on our beliefs. That is, **adding the supplementary postulates allows us to prove that AGM-rational revision corresponds exactly to elegant semantic structures, such as a system of concentric spheres** (where we fall back to the “closest” possible worlds when revising) or an epistemic entrenchment ordering (where we always sacrifice our least valuable beliefs first).

2.2. Derivative Rules of AGM

The basic AGM postulates entail several intuitive derivative rules. Here are some of the most important ones that follow strictly from the first six postulates:

$$\text{If } B \nvdash \neg\phi, \text{ then } B \subseteq B * \phi \qquad \text{(Preservation)}$$

$$\text{If } \psi \in B * \phi_1 \text{ and } \psi \in B * \phi_2, \text{ then } \psi \in B * (\phi_1 \vee \phi_2) \qquad \text{(Or)}$$

$$\text{If } \psi \notin B * \phi \text{ and } \psi \notin B * \neg\phi, \text{ then } \psi \notin B \qquad \text{(Negation Rationality)}$$

We can also derive a very famous rule concerning conditional beliefs and material implication:

$$\text{If } \psi \in B * \phi, \text{ then } \phi \to \psi \in B \qquad \text{(Frontloading)}$$

Once we adopt **Superexpansion** and **Subexpansion**, we unlock a new set of derivative rules that govern how revision behaves when we build up more complex pieces of new information, particularly conjunctions and disjunctions.

The following rules dictate the logic of sequential and conjunctive learning:

$$\text{If } \chi \in B * \phi \text{ and } \psi \in B * \phi, \text{ then } \chi \in B * (\phi \wedge \psi) \quad (\text{Cautious Monotony})$$

$$\text{If } \chi \in B * (\phi \wedge \psi) \text{ and } \psi \in B * \phi, \text{ then } \chi \in B * \phi \quad (\text{Cut})$$

$$\text{If } \chi \in B * \phi \text{ and } \neg\psi \notin B * \phi, \text{ then } \chi \in B * (\phi \wedge \psi) \quad (\text{Rational Monotony})$$

(*Note*: **Rational Monotony** is a sort of "generalized" version of **Preservation**. To convince your self, note that **Rational Monotony** is equivalent to **Preservation** when $\phi = \top$.)

We also gain derivative rules that dictate how revision handles disjunctions (situations where we learn that at least one of two things is true, but we don't know which):

$$\text{If } \psi \in B * (\phi_1 \vee \phi_2), \text{ then } \psi \in B * \phi_1 \text{ or } \psi \in B * \phi_2 \quad (\text{Disjunction Rationalit}$$

Finally, there is an important rule, called **Disjunctive Factoring**, which is logically equivalent to the combination of both supplementary postulates (given the six basic postulates):

$$B * (\phi_1 \vee \phi_2) \text{ is equal to } B * \phi_1, \text{ or } B * \phi_2, \text{ or } B * \phi_1 \cap B * \phi_2 \quad (\text{Disjunctive Fa}$$

3. Exercises

Solutions are provided here

3.1. Exercises: Formal Preliminaries

1. Prove Proposition 11 (i.e., that classical semantic consequence $\vdash$ is reflexive, transitive, and monotonic).
2. Prove Proposition 12 (i.e., that the classical consequence operator Cn satisfies reflexivity, transitivity, monotonicity, and idempotence).

3.2. Exercises: Consequence Relations and Consequence Operators

3. The properties of consequence relations $\triangleright$ and consequence operators C are not all logically independent. Explore their logical interactions by proving the following claims:

1. **Reflexivity + Transitivity $\implies$ Monotonicity:** Prove that if a consequence relation $\triangleright$ (or operator C) is reflexive and transitive, it must necessarily be monotonic.
2. **Monotonicity + Idempotence $\implies$ Transitivity:** For a consequence operator C, prove that if C is monotonic and idempotent, then C is transitive.
3. **Reflexivity + Transitivity $\implies$ Idempotence:** For a consequence operator C, prove that if C is reflexive and transitive, then C is idempotent.
4. **Reflexivity + Monotonicity $\not\implies$ Transitivity:** Provide a counterexample (e.g., a restricted consequence operator on $\mathcal{L}$) that is reflexive and monotonic, but fails to be transitive.

(*N.b.*, 4 implies that **Reflexivity + Monotonicity \cancel{ \implies } Idempotence**.)

4. Cumulative Transitivity and Non-Monotonicity. As established in Exercise 3.1, standard transitivity is sufficiently strong that, when paired with reflexivity, it entails monotonicity. Consequently, it is theoretically useful to identify a weaker formulation of transitivity that does not entail monotonicity in the presence of reflexivity. This allows for the formal study of non-monotonic consequence relations – such as those modeling default reasoning, where acquiring new information may lead to the retraction of prior conclusions – without reducing the consequence operator to mere reflexivity. Consider the property of Cumulative Transitivity.

Definition 14 (Cumulative Transitivity (Cut))**.**

A consequence relation $\triangleright$ is cumulatively transitive iff:

$$\text{If} \quad \Gamma \triangleright \delta \text{ for all } \delta \in \Delta \quad \text{and} \quad \Gamma \cup \Delta \triangleright \phi, \quad \text{then} \quad \Gamma \triangleright \phi$$

Equivalently, for a consequence operator C, cumulative transitivity is defined as:

$$\text{For all } \Gamma, \Delta \subseteq \mathcal{L}: \quad \text{if} \quad \Gamma \subseteq \Delta \subseteq C(\Gamma), \quad \text{then} \quad C(\Delta) \subseteq C(\Gamma)$$

Explore the limits of this property by proving the following:

1. **Reflexivity + Cut + Idempotence $\not\Longrightarrow$ Monotonicity:** Provide a counterexample demonstrating that a consequence relation $\triangleright$ (or operator C) can be reflexive, cumulatively transitive, and idempotent, yet fail to be monotonic.
2. **Reflexivity + Monotonicity $\Longrightarrow$ (Transitivity $\Longleftrightarrow$ Cut):** Prove this logical equivalence.
3. **Reflexivity + Monotonicity $\Longrightarrow$ (Idempotence $\Longleftrightarrow$ Cut):** Prove this logical equivalence.

3.3. Exercises: Derivative Belief Revision Postulates

5. Derivative Rules for Revision. Prove that the derivative rules listed in 2.2. Derivative Rules of AGM follow from the 8 postulates listed in 2. Belief Revision Theory.

6. Derivative Rules: Rational Monotony. Prove that if we replace ϕ with a tautology $\top$ in Rational Monotony (see 2.2. Derivative Rules of AGM), it becomes logically equivalent to the Preservation rule.