A matter of trust: Dynamic attitudes in epistemic logic
Rodenhauser, L.B.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 1.
Dynamic Attitudes

This chapter introduces the setting that we will work with throughout this dissertation. Plausibility orders (§1.1) are representations of the epistemic state of an agent; propositional attitudes (§1.2) capture static properties of such orders; upgrades (§1.3) are relational transformers, corresponding to types of changes that may be applied to plausibility orders. Uptake operators (§1.4) are families of upgrades indexed by propositions. Finally, the class of dynamic attitudes (§1.5) is given by a subclass of the class of all uptake operators, subject to a number of additional requirements.

Our setting is purely semantic and language-independent, an important fact that we emphasize in §1.6, focusing on a key constraint imposed on dynamic attitudes: idempotence.

Having introduced our formal machinery, we begin the investigation of our setting in §1.7 and §1.8: fixed points of dynamic attitudes (§1.7) represent the propositional attitudes which are realized by applying particular types of transformations. The notion of subsumption (§1.8) provides a natural way to compare the strength of given dynamic attitudes, which ties in naturally with the familiar entailment order on propositional attitudes.

1.1. Plausibility Orders

A set of possible worlds is a non-empty set. Given a set of possible worlds $W$, a proposition is a subset of $W$. Intuitively speaking, a proposition $P$ is “a representation of the world as being a certain way” (Stalnaker 1978), which is to say that $P$ is satisfied in the worlds that are “that way” (i.e., as represented by $P$), and not satisfied in the worlds that are not “that way”. Here, the worlds that are that way are simply the members of the set $P$, and the worlds that are not that way are the worlds that are not members of $P$. So, formally, a world $w \in W$ satisfies $P$ iff $w \in P$, and a world $v \in W$ does not satisfy $P$ iff $v \notin P$. 

Chapter 1. Dynamic Attitudes

We use the usual notation for set-theoretic operations on propositions, in particular:

\[ \neg P := \{ w \in W \mid w \not\in P \}, \quad P \cap Q := \{ w \in W \mid w \in P \text{ and } w \in Q \} \]
\[ P \cup Q := \{ w \in W \mid w \in P \text{ or } w \in Q \}, \quad P \Rightarrow Q := \{ w \in W \mid \text{if } w \in P, \text{ then } w \in Q \}. \]

In the remainder of the dissertation, unless specifically noted otherwise, we assume a fixed, but arbitrary set of possible worlds \( W \) as given.

1.1.1. Plausibility Orders. A plausibility order \( S \) (on \( W \)) is a pair

\[ S := (S, \leq_S), \]

where \( S \subseteq W \) is a finite set of possible worlds (called the domain of \( S \)), and \( \leq_S \subseteq S \times S \) is a total preorder on \( S \), i.e., a transitive and connected (and thus reflexive) relation.\(^1\)

A plausibility order represents the epistemic state of an (implicit) agent. While the set of possible worlds \( W \) comprises the totality of all possibilities that are consistent with some basic (unchangeable, context-independent and time-independent) implicit information about the world, over time, the agent will gain more information about the (real state of the) world, information that allows her to cut down the set of possibilities from the initial set \( W \) to a proper subset thereof. The latter set, represented by the domain \( S \) of a plausibility order, embodies what we call the agent’s hard information, assumed to be absolutely certain, irrevocable and truthful. Going further, the agent may also possess soft information, that is not absolutely certain, and subject to revision if the need arises. This information only allows her to hierarchize the possibilities consistent with her hard information according to their subjective “plausibility”, but not to actually discard any of them. This relative hierarchy is given by the relation \( \leq_S \).\(^2\)

Here, our assumption is that \textit{the smaller the better}, as in, for example, a ranking of soccer teams: the teams who have smaller numbers in the ranking have

---

1A binary relation \( R \) on a given set \( S \) is transitive if for any \( w, v, u \in S \): if \( (w, v), (v, u) \in R \), then also \( (w, u) \in R \); connected if for any \( w, v \in S \), either \( (w, v) \in R \) or \( (v, w) \in R \); and reflexive if for any \( w \in S \): \( (w, w) \in R \).

2For the distinction between hard and soft information, cf. [van Benthem (2007)], who illustrates the difference using the example of a card game: the total number of cards in the deck and the cards I hold myself would typically be taken to be hard information. On the other hand, there is also soft information: “I see you smile. This makes it more likely that you hold a trump card, but it does not rule out that you have not got one” (ibid.). My seeing you smile might lead me to believe that you hold a trump card, but this belief is open to revision as further evidence becomes available.
1.1. Plausibility Orders

performed better according to the ranking than those with higher numbers (for example, the team in second place is better than the team in third place, etc.); the difference between a plausibility order and your typical ranking of soccer teams is that several worlds may share the same rank. So the fact that \( w \leq_S v \) indicates that the world \( w \) is at least as plausible as the world \( v \) (from the perspective of our agent). Taken together, the structure \( S \) represents the epistemic state of an (implicit) agent, at a given point in time and considered in isolation from other agents.

We shall write \( w \approx_S v \) to indicate that \( w \) and \( v \) are equiplausible, i.e., \( w \approx_S v \) iff both \( w \leq_S v \) and \( v \leq_S w \).

We also notice that, in case \( W \) is finite, there is a special plausibility order \( W = (W, W \times W) \), representing the initial state of ignorance, in which the agent has not been able to exclude any world from consideration, and also has not been able to impose any hierarchy on the worlds she considers possible. Another special order is given by \( \emptyset = (\emptyset, \emptyset) \): this order represents the absurd state, in which the agent has excluded all possible worlds from consideration.

Given a plausibility order \( S \), we often use the infix notation for the pre-order \( \leq_S \), writing, for example, \( "w \leq_S v" \) rather than \( "(w, v) \in \leq_S" \). If \( S \) is clear from the context, we often drop the subscript, writing \( "\leq" \) rather than \( "\leq_S" \). Also, we sometimes write \( "(w, v) \in S" \) (rather than \( "(w, v) \in \leq_S" \)); and \( "w \in S" \) (rather than \( "w \in \leq_S" \)).

An example of a plausibility order is provided in Figure 2. According to the diagram, the hard information of the agent is currently given by the proposition that the actual world is among the worlds in \( \{w_1, \ldots, w_5\} \). We adopt the convention that worlds that are higher on the page are more plausible than those that are lower (again, similarly as one would write a ranking of soccer teams: starting with the best teams higher up in the list). So in our example, \( w_1 \) is more plausible than \( w_2 \), \( w_2 \) is more plausible than \( w_4 \), etc. On the other hand (unlike in soccer team rankings), two worlds may be equiplausible, and this is represented by drawing them on the same level: so \( w_2 \) and \( w_3 \), for example, are equiplausible: \( w_2 \approx w_3 \) (which means, by definition of equiplausibility, that both \( w_2 \leq w_3 \) and \( w_3 \leq w_2 \)).

1.1.2. Best Worlds. For any proposition \( P \), the best \( P \)-worlds (or most plausible \( P \)-worlds) in a plausibility order \( S \), denoted with \( \text{best}_S P \), are the \( \leq_S \)-minimal elements of \( P \), given by the proposition

\[
\text{best}_S P := \{ w \in P \cap S \mid \forall v \in P \cap S : w \leq_S v \}.
\]

So the best \( P \)-worlds in \( S \) are those \( P \)-worlds in \( S \) that are at least as good as any \( P \)-world in \( S \) according to the hierarchy given by \( \leq_S \).
Figure 2. A plausibility order with domain \{w_1, \ldots, w_5\}. By our drawing convention, worlds that are higher up on the page are more plausible, and worlds on the same level are equiplausible. So, for example, the world \(w_1\) is more plausible than all other worlds.

The best worlds (or most plausible worlds) in \(S\) are given by \(\text{best}_S := \text{best}_S S\). Note that \(\text{best}_S = \emptyset\) iff \(S = \emptyset\). In the example given by Figure 2, \(\text{best}_S = \{w_1\}\).

1.1.3. Union, Intersection, Conditionalization. For convenience, we lift the set-theoretic operations of intersection and union to plausibility orders. Given plausibility orders \(S_1\) and \(S_2\), the intersection \(S_1 \cap S_2\) of \(S_1\) and \(S_2\), and the union \(S_1 \cup S_2\) of \(S_1\) and \(S_2\) are, respectively, given by the ordered pairs

\[
S_1 \cap S_2 := (S_1 \cap S, \leq_{S_1} \cap \leq_{S_2}), \quad S_1 \cup S_2 := (S_1 \cup S, \leq_{S_1} \cup \leq_{S_2}).
\]

Notice that neither \(S_1 \cup S_2\) nor \(S_1 \cap S_2\) is in general guaranteed to be a plausibility order, as we may encounter failures of both transitivity and connectedness. In using these notations in what follows, we will take care to apply them only when given plausibility orders \(S_1, S_2\) such that \(S_1 \cup S_2, S_1 \cap S_2\) are plausibility orders.

A more interesting operation on a given plausibility order is what we call “conditionalization”. If \(S\) is a plausibility order, and \(P\) a proposition, we denote with \(S|_P\) the conditionalization of \(S\) on \(P\), given by

\[
S|_P := (S \cap P, \{(w,v) \in S \mid w,v \in P\}).
\]

So conditionalizing a plausibility order \(S\) on a proposition \(P\) amounts to simply “throwing away” all the non-\(P\)-worlds in \(S\), and restricting the relation...
\( \leq_S \) accordingly. We may interpret this as corresponding to an event in which the agent receives hard information that \( P \), allowing her to exclude all non-\( P \)-worlds from consideration.

1.1.4. Systems of Spheres. A system of spheres (sometimes also called an onion) is a finite nested set of finite non-empty propositions, that is, a set of sets
\[
O = \{O_1, \ldots, O_n\},
\]
where \( n \in \omega \), \( O_k \subseteq W \) for \( 1 \leq k \leq n \), and \( O_1 \subseteq \ldots \subseteq O_n \). The propositions \( O_k \) \((1 \leq k \leq n)\) are called spheres in \( O \). As with plausibility orders, we allow systems of spheres to be empty.\(^3\) A picture of a system of spheres is provided in Figure 3. In such a drawing, the region covered by each of the circles corresponds to one of the spheres.

For our purposes, systems of spheres will prove useful as equivalent representations of plausibility orders that sometimes allow for a more compact representation in diagrams. The connection between plausibility orders and systems of spheres is as follows. Every plausibility order \( S \) comes equipped with a system of spheres. Given a world \( w \in S \), let \( w^{\text{up}} \) denote
\[
w^{\text{up}} := \{ x \in S \mid x \leq w \}.
\]
The proposition \( w^{\text{up}} \) collects all worlds that are at least as plausible as \( w \). Now the set of propositions
\[
\text{sph}(S) := \{ w^{\text{up}} \mid w \in S \}
\]
is a system of spheres, called the system of spheres for \( S \). Notice that the innermost sphere of \( \text{sph}(S) \) is given by \( \text{best}_S \).

It is easy to see that the function \( \text{sph} \) is actually a bijection, so plausibility orders and systems of spheres are in one-to-one correspondence: they are just notational variants. We denote the inverse of \( \text{sph} \) with \( \text{ord} \), that is, for any system of spheres \( O \), \( \text{ord}(O) \) is the unique plausibility order \( S \) such that \( \text{sph}(S) = O \).

Given a system of spheres \( O = \{O_1, \ldots, O_n\} \), we have
\[
w \leq_{\text{ord}(O)} v \iff w, v \in O_n \quad \text{and} \quad \forall O \in O : v \in O \Rightarrow w \in O.
\]
The one-to-one correspondence between plausibility orders and systems of spheres allows us to define a plausibility order \( S \) by fixing a system of spheres \( O \): when doing this, it is always understood that \( S := \text{ord}(O) \). To simplify notation, we sometimes write \( P \in S \) to indicate that \( P \in \text{sph}(S) \).

\(^3\)Systems of spheres were introduced by Lewis (1973) in the context of his work on counterfactuals. Lewis also noted that a system of spheres in his sense is equivalent to an ordering on possible worlds. The significance of this kind of structure for belief revision theory was realized by Grove (1988).
1.1.5. Spohn Ordinals. As suggested by our notation, one can assign a natural number to each world \( w \) in a given system of spheres \( \mathcal{O} \), called the “Spohn ordinal” of \( w \), essentially given by the “rank” of the smallest sphere containing \( w \).

Formally, given a system of spheres \( \mathcal{O} = \{O_1, \ldots, O_n\} \), the Spohn ordinal \( \kappa_{\mathcal{O}}(w) \) of a world \( w \in O_n \) is the least natural number \( k \) such that \( w \in O_k \), i.e.,

\[
\kappa_{\mathcal{O}}(w) := \min\{k \in \omega \mid w \in O_k\}.
\]

It also makes sense to talk about the Spohn ordinal \( \kappa_{\mathcal{O}}(P) \) of a proposition in a given system of spheres: essentially, \( \kappa_{\mathcal{O}}(P) \) is the smallest number \( n \) such that we can find \( P \)-worlds with Spohn ordinal \( n \) in \( \mathcal{O} \). If there are no \( P \)-worlds in \( \mathcal{O} \), then we put \( \kappa_{\mathcal{O}}(P) = 0 \). Formally:

\[
\kappa_{\mathcal{O}}(P) := \begin{cases} 
\min\{\kappa_{\mathcal{O}}(w) \mid w \in P \cap S\} & P \cap S \neq \emptyset, \\
0 & \text{otherwise.}
\end{cases}
\]

While we can understand Spohn ordinals of possible worlds in given orders to give a numerical measure of the plausibility of a given world (with the number 1 corresponding to maximal plausibility), Spohn ordinals of propositions in given order capture a numerical measure of the plausibility of a proposition.
Given the one-to-one correspondence between systems of spheres and plausibility orders noted above, it makes sense to directly speak of Spohn ordinals of worlds in plausibility orders, putting $\kappa_S(w) := \kappa_{\text{sph}(S)}(w)$. Analogously, we put $\kappa_S(P) := \kappa_{\text{sph}(S)}(P)$.

Notice that a plausibility order $S$ may be uniquely specified by determining its domain $S$ and the Spohn ordinal $\kappa_S(w)$ of each world $w \in S$. When proceeding in this way, it is understood that $w \leq_S v$ iff $\kappa_S(w) \leq \kappa_S(v)$.

1.2. Propositional Attitudes

We routinely ascribe doxastic propositional attitudes to agents, by saying, for example, that an agent believes, knows, or doubts certain things. Such ascriptions tell us something about the opinion of the agent in question about those things. Leaving the agent entertaining the attitude implicit, we can model propositional attitudes as families of doxastic propositions, in the following manner.

1.2.1. Doxastic Propositions. A doxastic proposition $P$ (on $W$) is a function

$$S \mapsto P(S)$$

that assigns a proposition $P(S) \subseteq S$ to each plausibility order $S$ (on $W$). We call $P(S)$ the proposition denoted by $P$ in $S$.

So a doxastic proposition $P$ gives us a proposition $P$ for each plausibility order $S$. In this way, doxastic propositions can be seen as “intensionalized propositions”, since just which proposition $P$ gives us may depend on $S$, and thus on the epistemic state of our agent.

Our main use for the concept of a doxastic proposition is in stating the next definition.

1.2.2. Propositional Attitudes. A (doxastic) propositional attitude is an indexed family of doxastic propositions

$$A := \{AP\}_{P \subseteq W},$$

with the doxastic propositions $AP$ indexed by arbitrary propositions $P \subseteq W$ and satisfying

$$AP(S) = A(P \cap S)(S). \quad \text{(Conservativity)}$$

The conservativity constraint imposed on propositional attitudes derives from the intuition that evaluation in a given structure should not depend on objects
that are not part of that structure, a rather minimal well-behavedness assumption. Note that conservativity will be automatically satisfied if a modal-logical language is used in the usual way, since the denotation of a sentence in a given structure is customarily given by a subset of the domain of that structure.\textsuperscript{4}

A propositional attitude can equivalently be seen as a function

$$P \rightarrow^A AP$$

taking as input a proposition $P$, and returning as output a doxastic proposition $AP$, i.e., another function, from plausibility orders to propositions.

The intuitive way to “read” a given propositional attitude $A$ is as follows: given an order $S$, and a proposition $P$, the proposition $AP(S)$ collects the worlds in $S$ in which the agent has the propositional attitude $A$ towards $P$.

1.2.3. Notation. Given a propositional attitude $A$, a proposition $P$, and a plausibility order $S$, we put

$$A_S P := AP(S).$$

We also frequently write $S, w \models AP$ to mean that $w \in A_S P$, using the notation familiar from modal logic; and we write $S \models AP$ to mean that for any $w \in S$: $S, w \models AP$.

1.2.4. Irrevocable Knowledge and Simple Belief. Using the concept of a propositional attitude, we can give more formal content to the notions of “hard information” and “soft information” we have introduced above. As we have said, an agent gains hard information by excluding possible worlds from consideration. One can think of hard information as what the agent (irrevocably) knows. On the other hand, soft information allows the agent to hierarchize the possible worlds that she still considers as candidates for being the actual world. Of special interest are the worlds she considers most plausible, and one can think of the most plausible worlds as what the agent (simply) believes in view of her soft information.

Here, the qualifiers “irrevocably” and “simply” are used to distinguish these concepts of knowledge and belief from other, similar notions that we

\textsuperscript{4}Our terminology is borrowed from the theory of generalized quantifiers: in this literature, a type $< 1, 1 >$ quantifier $Q$ is called conservative if $Q(A, B) = Q(A, A \cap B)$, i.e., given sets $A$ and $B$, the denotation of $Q$ only depends on the $A$-part of $B$. In the present context, the domain $S$ plays roughly the role of $A$, while the propositional argument $P$ plays the role of $B$, and the content of our requirement is that the proposition determined by the attitude should only depend on the $S$-part of $P$, i.e., $S \cap P$. 
will make use of. Let us now define irrecovable knowledge and simple belief formally, using the concept of a propositional attitude:

— (Irrevocable) knowledge, denoted by $K$, is the propositional attitude defined by

$$K_S P := \{w \in S \mid S \subseteq P\},$$

i.e., the agent (irrevocably) knows $P$ iff $P$ is satisfied in all worlds in $S$.

— (Simple) belief, denoted by $B$, is the propositional attitude defined by

$$B_S P := \{w \in S \mid \text{best } S \subseteq P\},$$

i.e., the agent (simply) believes that $P$ iff $P$ is satisfied in all the most plausible worlds in $S$.

Irrevocable knowledge is the strongest reasonable formalization of the (pre-formal) concept of knowledge that our setting gives rise to; simple belief, on the other hand, is the weakest reasonable formalization of the (pre-formal) concept of belief that our setting gives rise to.

1.2.5. Introspectiveness. Irrevocable knowledge and simple belief are both examples of introspective propositional attitudes.

We call a propositional attitude $A$ introspective if, for any proposition $P$, and for any plausibility order $S$:

$$A_S P \in \{S, \emptyset\}.$$ 

In other words, introspective attitudes are global properties of a given plausibility order: either the agent has the attitude $A$ towards $P$ in all worlds in $S$, or in no world in $S$.

Notice that $K$ as well as $B$ are indeed introspective propositional attitudes. Whether the agent (irrevocably) knows that $P$ in a given order $S$ does not depend on any particular world in $S$; it merely depends on the question whether all worlds in $S$ are contained in $P$. Analogously: whether the agent (simply) believes that $P$ in $S$ does not depend on any particular world in $S$; it merely depends on the question whether all world in best $S$ are contained in $P$.

To motivate why we call this property “introspectiveness”, let us introduce a number of operations on propositional attitudes that will be useful in the remainder of this dissertation.
1.2.6. Operations on Propositional Attitudes. Given propositional attitudes $A$ and $A'$, one easily defines new ones by means of standard operations. In particular,

- the opposite $A^\sim$ of $A$ is given by $A^\sim_S P := A_S(\neg P)$;
- the complement $\neg A$ of $A$ is given by $(\neg A)_S P := S \setminus A_S P$;
- the dual $A^\sim$ of $A$ is given by $A^\sim_S P := \neg A_S P$;
- the composition $A A'$ of $A$ and $A'$ is given by $(A A')_S P := A_S S P \setminus A'_S P$;
- the conjunction $A \land A'$ of $A$ and $A'$ is given by $(A \land A')_S P := A_S P \cap A'_S P$;
- the disjunction $A \lor A'$ of $A$ and $A'$ is given by $(A \lor A')_S P := A_S P \cup A'_S P$;
- the material implication $A \rightarrow A'$ of $A$ and $A'$ is given by $(A \rightarrow A')_S P = (\neg A \lor (A \land A'))_S P$.

We use the notation familiar from modal logic, writing $S \models AP \land A'P$ iff $S \models (A \land A')P$, or $S \models AP \rightarrow A'P$ iff $S \models (A \rightarrow A')P$ etc.

Notice now that a propositional attitude $A$ is introspective (cf. §1.2.5 above) iff for any plausibility order $S$ and proposition $P$,

$$S \models AP \rightarrow KAP$$

and

$$S \models \neg AP \rightarrow K\neg AP$$

are both satisfied, that is: if the agent has the attitude $A$ towards $P$ in $S$, then the agent knows that she has the attitude $A$ towards $P$ ("positive introspection"), and if the agent does not have the attitude $A$ towards $P$ in $S$, then the agent knows that she does not have the attitude $A$ towards $P$ ("negative introspection"). And this is just how introspectiveness (understood as the conjunction of positive and negative introspection) is usually defined.

An introspective propositional attitude $A$ may be defined by specifying, for each proposition $P$, the set of plausibility orders $S$ such that $S \models AP$. This is sufficient since, by definition, $S \models AP$ iff $A_S P = S$, and by the fact that $A$ is introspective, $S \not\models AP$ iff $A_S P = \emptyset$. Hence, for an introspective propositional attitude $A$, a specification of the plausibility orders $S$ such that $S \models AP$ fixes, for each plausibility order $S$ and world $w \in S$, whether $S, w \models AP$.

1.2.7. Conditional Belief. For any proposition $Q$, belief conditional on $Q$, denoted by $B_Q$, is the introspective propositional attitude defined, for each proposition $P$, by

$$S \models B_Q P \iff \text{best}_S Q \subseteq P,$$

i.e., $P$ is believed conditional on $Q$ iff $P$ is satisfied in all the most plausible $Q$-worlds in $S$. 

Conditional belief may equivalently be defined in terms of the notion of conditionalization defined above (cf. §1.1.3), noticing that
\[ S \vDash B^Q P \iff S|_Q \vDash BP. \]
So \( P \) is believed conditional on \( Q \) iff \( P \) is believed in \( S|_Q \), the conditionalization of \( S \) on \( Q \). This clearly brings out the dynamic flavour of conditional belief: conditional beliefs are beliefs that are held conditional on various pieces of hard information. That is, we can interpret the fact that \( S \vDash B^Q P \) as saying that given the hard information that \( Q \) is the case, the agent would believe that \( P \) is the case.

Following up on this idea, conditional beliefs provide a way of gauging the “stability” of given propositional attitudes under the influx of hard information. As an example, consider simple belief and irrecovable knowledge, the two propositional attitudes we have defined in §1.2.4. Here, we observe:

**Proposition 1** (Baltag and Smets (2008)). Let \( S \) be a plausibility order, and let \( P \) be a proposition. Suppose that \( S \vDash BP \). Then

1. \( S \vDash BP \iff S \vDash B^{S} P. \)
2. \( S \vDash KP \iff S \vDash B^Q P \) for any \( Q \subseteq W. \)

As it turns out, then, simple beliefs are only guaranteed to be stable when receiving hard information that the agent already has. In this sense, simple beliefs are easily defeated. Irrecovable knowledge, on the other hand, are those beliefs that are stable under receiving any new hard information. In this sense, irrecovable knowledge is indefeasible.

**1.2.8. Further Examples.** We now introduce a number of further important examples of propositional attitudes that will be relevant in the remainder of this dissertation.

1. **Strong belief**, denoted by \( Sb \), is the introspective propositional attitude defined by

   \[ S \vDash Sb P \iff \text{best}_S \subseteq P \text{ and } \forall x \in P \forall y \notin P : x < y, \]

   i.e., \( P \) is strongly believed iff \( P \) is believed, and moreover, all \( P \)-worlds are strictly more plausible than all non-\( P \)-worlds.

---

6Strong belief is considered by Stalnaker (1996)—who calls the notion “robust belief”—, and by Battigalli and Siniscalchi (2002).
— **Refined belief**, denoted by $Rb$, is the propositional attitude defined by

$$S \models RbP \text{ iff } \text{best} S \subseteq P \text{ and } \forall x \in P, y \notin P : x < y \text{ or } y < x,$$

i.e., $P$ is refined believed iff $P$ is believed, and moreover, there are no ties between $P$-worlds and non-$P$-worlds in the ranking given by $\leq^7$.

— **Belief to degree $n$**, denoted by $B^n$, is the introspective propositional attitude defined, for any natural number $n \geq 1$, by

$$S \models B^nP \text{ iff } \forall w \in S : \kappa_S(w) \leq n \Rightarrow w \in P.$$

Notice that $B^1$ is just $B$, i.e., $B^1P$ is satisfied in a given plausibility order $S$ iff the most plausible worlds in $S$ are $P$-worlds. More generally, we have $S \models B^nP$ iff all worlds with Spohn ordinal from 1 to $n$ are $P$-worlds.$^8$

— **Defeasible knowledge**, denoted by $\Box$, is the propositional attitude defined by

$$S, w \models \Box P \text{ iff } \forall v \in S : v \leq w \Rightarrow v \in P,$$

i.e., $P$ is defeasibly known at $w$ iff $P$ is satisfied in all worlds that are at least as plausible as $w$. This formalizes a notion of knowledge that is weaker than irrevocable knowledge, but still **veridical** in the sense that from that fact that $S, w \models \Box P$ it follows that $w \in P$.\(^9\)

Observe that defeasible knowledge is not introspective (in the sense of the definition given in §1.2.5): different propositions may be defeasibly known at distinct worlds in a given plausibility order.

All propositional attitudes in the above list allow for natural characterizations in terms of conditional belief. The first and the fourth item of the next proposition are due to Baltag and Smets (2008).

**Proposition 2.** *Let $S$ be a plausibility order, and let $P$ be a proposition. Suppose that $S \models BP$. Then*

$^7$To the best of my knowledge, refined belief has not been considered in the previous literature. However, as we shall see later on, refined belief is closely connected to a dynamic attitude called ***moderate trust*** (defined in §2.5.1), which appears for the first time (under the name of “restrained revision”) in the work of Booth and Meyer (2006).


$^9$Defeasible knowledge was defined by Stalnaker (2006) in his formalization of Lehrer’s defeasibility theory of knowledge (Lehrer 1990). Stalnaker defined defeasible knowledge in terms of conditional belief (i.e., as in item (4.) of Proposition 2 below); that defeasible knowledge is simply the Kripke modality for the converse $\geq$ of the plausibility order $\leq$ was discovered by Baltag and Smets (2008), who initially called the notion “safe belief.”
1.2. Propositional Attitudes

1. $S \vdash SbP$ iff $S \vdash B^QP$ for any $Q$ such that $Q \cap P \neq \emptyset$.

2. $S \vdash RbP$ iff $S \vdash B^QP$ for any $Q$ such that best $Q \cap P \neq \emptyset$.

3. $S \vdash B^nP$ iff $S \vdash B^QP$ for any $Q$ such that $Q = S \setminus \{w \in S \mid \kappa_S(w) \leq k\}$ for some natural number $k < n$.

4. $S, w \vdash \Box P$ iff $S, w \vdash B^QP$ for any $Q$ such that $w \in Q$.

**Proof.** We prove the second and third item. Start with (2.). From left to right, suppose that $S \vdash RbP$. Take any $Q$ such that best $Q \cap P \neq \emptyset$. Then $S \vdash B^QP$ iff $best \in P$. Take any $w \in best$ and suppose $w \notin P$. Since best $Q \cap P \neq \emptyset$, there exists $v$ such that $w \approx v, v \notin P$. Thus, $S \notin RbP$, contradiction. So $S \vdash B^QP$. This finishes the left to right direction.

From right to left, suppose that $S \vdash BP$ and for any $Q$ such that best $Q \cap P \neq \emptyset$: $S \vdash B^QP$. If $S = \emptyset$, our claim holds. If $S \neq \emptyset$, then best $S \neq \emptyset$, so $P \cap S \neq \emptyset$. Take any $w \in P \cap S$. Suppose there exists $v \in \neg P \cap S$ such that $w \approx v$. Then best $\{w, v\} \cap P \neq \emptyset$ and $S \neq B^{\{w, v\}}P$. This is a contradiction. So for any $v \in \neg P \cap S$: $w < v$ or $v < w$. It follows that $S \vdash RP$. This finishes the right to left direction, and the proof for the second item.

We continue with the proof for the third item. From left to right, suppose that $S \vdash B^nP$ for some $n \geq 1$. Take any $Q$ such that $Q = S \setminus \{w \in S \mid \kappa_S(w) \leq k\}$ for some natural number $k < n$. Suppose towards a contradiction that $S \neq B^QP$, i.e., it is not the case that best $Q \subseteq P$, and thus there exists a world $v \in $ best $Q$ such that $v \notin P$. Hence, by definition of $Q$ there exists $v$ such that $\kappa_S(v) \leq n$ and $v \notin P$. It follows that $S \neq B^nP$. This is a contradiction. Thus $S \vdash B^QP$ after all, which completes the left to right direction.

From right to left, we argue by contraposition. Suppose that $S \neq B^nP$. Then there exists some world $w \in S$ such that $\kappa_S(w) \leq n$ and $w \notin P$. Hence best $Q \notin P$, where $Q = (S \setminus \{w \in S \mid \kappa_S(w) \leq n - 1\})$. So $S \neq B^QP$. Hence it is not the case that $S \vdash B^QP$ for any $Q$ such that $Q = S \setminus \{w \in S \mid \kappa_S(w) \leq k\}$ for some natural number $k < n$. This finishes the right to left direction, and the proof for the third item.

Let me comment on the individual items of the previous proposition in turn. Throughout, we assume as given a plausibility order $S$ and a proposition $P$ such that $S \vdash BP$. Item (1.) says that strong beliefs are particularly robust: $P$ is strongly believed in $S$ iff it is simply believed conditional on any $Q$ consistent with $P$. Item (2.) says that refined beliefs are robust in a weaker sense. We can say that a proposition $P$ is “not implausible given $Q$” iff $P \cap \text{best} \neq \emptyset$. Then, according to item (2.), “refined belief in $P$” is the same as “belief in $P$ conditional on any $Q$ such that $P$ is not implausible given $Q$". 

$\neg$
Item (3.) says that $P$ is believed to degree $n$ in $S$ iff conditionalizing on the set of worlds with Spohn ordinal less than or equal to $k$ yields an order in which $P$ is believed, for any $k$ smaller than $n$. Item (4.), finally, says that $P$ is defeasibly known at a world $w$ in $S$ iff the belief in $P$ can be defeated only by receiving hard information that is actually false at $w$, i.e., $P$ is defeasibly known at world $w$ iff $P$ is believed conditional on any proposition $Q$ satisfied at $w$.

1.2.9. Triviality and Absurdity. Two further special examples of introspective propositional attitudes are given by triviality $\top$ and absurdity $\bot$, defined by

\[ S \models \top P \text{ iff } P \subseteq W, \]
\[ S \models \bot P \text{ iff } S = \emptyset. \]

Observe that $\top_S P = Kw$ and $\bot_S P = K\emptyset$ for any $S$ and $P$. Triviality is the propositional attitude the agent has in any possible world of any order; absurdity is the propositional attitude the agent has in no possible world of any order.

1.2.10. Duals. Recall from above (§1.2.6) that the dual of a propositional attitude $A$ is given by $A_S^\sim := \neg A_S^\sim$. By way of illustration, we give explicit clauses for four examples of propositional attitudes introduced so far below:

- The dual $K^\sim$ of irrevocable knowledge $K$ is given by
  \[ S \models K^\sim P \text{ iff } P \cap S \neq \emptyset. \]

- The dual $Sb^\sim$ of strong belief $Sb$ is given by
  \[ S \models Sb^\sim P \text{ iff } \exists w, v \in S : w \in P, v \notin P \text{ and } v \leq w. \]

- The dual $B^\sim$ of simple belief is given by
  \[ S \models B^\sim P \text{ iff } \text{best}_S \cap P \neq \emptyset. \]

- The dual $\Diamond := \Box^\sim$ of defeasible knowledge $\Box$ is given by
  \[ S, w \models \Diamond P \text{ iff } \exists v : v \leq w \text{ and } v \in P. \]

$K^\sim$ expresses the possibility of $P$ (i.e., it is not the case that the agent has hard information that not $P$); $Sb^\sim$ expresses the remote plausibility of $P$ (i.e., it is not the case that all non-$P$-worlds are strictly more plausible than all $P$-worlds); $B^\sim$ expresses the plausibility of $P$ (i.e., it is not the case that all $\leq$-minimal elements are non-$P$-worlds); finally, $\Diamond$ expresses a “defeasible possibility”, that is, $S, w \models \Diamond P$ iff there is a world $v$ that is at least as plausible as $w$ such that $v$ satisfies $P$. 
1.3. Doxastic Upgrades

As time passes, an agent’s epistemic state may change in various ways. While plausibility orders do not track time explicitly, we can model changes over time in a step-wise manner, by changing a given order and moving to another one. For this purpose, we use the notion of an upgrade, i.e., a specific type of relational transformer, that takes given plausibility orders \( S \) to new ones, by (possibly) restricting the domain \( S \), and (possibly) reordering worlds in the relation \( \leq \).

1.3.1. Doxastic Upgrades. A (doxastic) upgrade \( u \) on \( W \) is a function

\[
S \xrightarrow{u} S^u
\]

that takes a given plausibility order \( S = (S, \leq_S) \) (on \( W \)) to a plausibility order \( S^u = (S^u, \leq_{S^u}) \) (on \( W \)), satisfying \( S^u \subseteq S \).

The requirement that \( S^u \subseteq S \) means that upgrades either grow the hard information of an agent (by eliminating worlds) or leave it the same (by not eliminating worlds). This embodies our understanding that hard information is really hard: it does not get lost as the agent’s information state changes (hence the “up” in “upgrade”). But crucially, upgrades may reorder the plausibility hierarchy (given by the relation \( \leq_S \)) that represents the agent’s soft information (hence the “grade” in upgrade).

1.3.2. Executability. Given a plausibility order \( S \) and a world \( w \in S \), an upgrade \( u \) is executable in \( w \) iff \( w \in S^u \). An upgrade is executable in \( S \) if it is executable in some \( w \in S \), i.e., iff \( S^u \neq \emptyset \).

1.3.3. Examples. Three examples of upgrades have received particular attention in the recent literature: \(!P\), \( \uparrow P \) and \( \hat{\uparrow} P \). We add two obviously interesting “edge cases”: \( \emptyset \) and \( id \). So we have five first examples, defined as follows. For each proposition \( P \),

— the update \( !P \) maps each plausibility order \( S \) to the conditionalization of the order on \( P \), i.e., we simply put \( S^{!P} := S|_P \): all non-\( P \)-worlds are deleted, while the new order on the remaining worlds is inherited from \( S \)\(^{11}\)

\(^{10}\)Cf. in particular the papers by van Benthem (2007) and Baltag and Smets (2008), and the monograph by van Benthem (2011).

\(^{11}\)This notion seems to be folklore. It is implicit in Stalnaker (1978) and appears frequently in the dynamic semantics literature (cf. van Eijck and Visser (2008) for an overview). The present notion of update may also be seen as the natural qualitative analogue of Bayesian update.
Chapter 1. Dynamic Attitudes

---

Figure 4. Lexicographic Upgrade \( \uparrow P \) (4.1) and Minimal Upgrade \( \uparrow \neg P \) (4.2). Applying \( \uparrow P \) amounts to putting all \( P \)-worlds on top of all non-\( P \)-worlds, i.e., after the upgrade, all \( P \)-worlds are closer to the center than all non-\( P \)-worlds (while the order within the two zones remains unchanged). Applying \( \uparrow \neg P \) amounts to putting only the best \( P \)-worlds on top of all the non-\( P \)-worlds, leaving everything else unchanged.

1. **Lexicographic upgrade** \( \uparrow P \) makes all \( P \)-worlds strictly better than all non-\( P \)-worlds. No worlds are deleted, and in-between the two “zones” \( P \) and \( \neg P \), the order is not affected by the upgrade (see Figure 4.1 for illustration).

2. **Minimal upgrade** \( \uparrow \neg P \) makes the best \( P \)-worlds the best worlds overall. No worlds are deleted, and the relation among all worlds that are not in \( \text{best}_S P \) remains the same (see Figure 4.2 for illustration).

3. **Null upgrade** \( \varnothing \) maps every plausibility order to the empty plausibility order.

4. **Trivial upgrade** \( \text{id} \) maps every plausibility order to itself.

---

1.3.4. Composition. Upgrades are functions, so they can be composed to obtain new functions. Given arbitrary upgrades \( u \) and \( u' \), the composition \( u \cdot u' \) of \( u \) and \( u' \) is given by \( S^{u \cdot u'} := (S^u)u' \).

**Proposition 3.** For any upgrade \( u \):

1. \( u \cdot \text{id} = \text{id} \cdot u = u \).

2. \( u \cdot \varnothing = \varnothing \cdot u = \varnothing \).

---

\[ ^{12}\text{Cf. Nayak (1994).} \]
\[ ^{13}\text{Cf. Boutilier (1996).} \]
1.4. Uptake Operators

Proof. The first item is obvious. The second item follows from the fact that \( \emptyset^u = \emptyset \) by definition of upgrades.

1.3.5. Subsumption. Let \( S \) be a plausibility order, and let \( u \) be an upgrade. We say that \( S \) subsumes \( u \) (or: \( S \) is in the fixed point of \( u \)) iff \( \mathcal{S}^u = S \).

This captures that the information carried by the upgrade \( u \) is “already present” in the order \( S \): actually applying the upgrade is thus redundant.

This naturally gives rise to a notion of relative subsumption of upgrades: an upgrade \( u \) is said to subsume an upgrade \( u' \) if applying \( u \) generally makes applying \( u' \) redundant.

Formally, let \( u, u' \) be upgrades. We say that \( u \) subsumes \( u' \) iff applying \( u \) generally yields an order in the fixed point of \( u' \); that is, for any plausibility order \( S \):

\[
(\mathcal{S}^u)u' = \mathcal{S}^u.
\]

We write \( u \rightarrow u' \) if \( u \) subsumes \( u' \).

Two examples: one may easily check that \( \uparrow P \) subsumes \( \uparrow P \), which in turn subsumes \( \uparrow P \), for any proposition \( P \).

The concept of subsumption is familiar from the dynamic semantics literature. There, the analogue of subsumption of an upgrade by an order is often called support, or acceptance, and the analogue of subsumption of an upgrade by another is called, simply, dynamic entailment.\footnote{Cf. Veltman (1996). The notion of (“relative”) subsumption discussed here corresponds to Veltman’s entailment relation “validity_2”: “an argument is valid_2 iff updating any information state \( \sigma \) with the premises \( \psi_1, \ldots, \psi_n \) in that order, yields an information state in which the conclusion \( \varphi \) is accepted.” (ibid.) In the simplest case we have just a single premise, and then, \( \varphi \) dynamically entails \( \psi \) (in the sense of validity_2) if updating with \( \varphi \) yields a state in which \( \psi \) is accepted. “Acceptance”, in turn, is defined in terms of fixed points: “every now and then it may happen that \( \sigma[\varphi] = \sigma \). If so, the information conveyed by \( \varphi \) is already subsumed by \( \sigma \). In such a case we (…) say that \( \varphi \) is accepted in \( \sigma \).” (ibid.) So Veltman’s notion of acceptance corresponds to our notion of a state subsuming an upgrade.}

1.4. Uptake Operators

Our examples of upgrades naturally suggest to isolate the propositional argument already implicit in the definition of concrete upgrades like \( !P \) or \( \uparrow P \). This leads to the notion of an “uptake operator”.

1.4.1. Uptake Operators. An \( \text{(uptake) operator} \ \tau \ \text{(on} \ W) \) is a family of upgrades

\[
\{ \tau P \}_{P \subseteq W},
\]
indexed by arbitrary propositions $P \subseteq W$. Equivalently, uptake operators can be defined as functions from propositions to upgrades, i.e., $\tau$ can be seen as a function

$$P \xrightarrow{\tau} \tau P$$

associating each proposition $P$ with an upgrade $\tau P$.

Uptake operators represent modes of processing informational inputs given by propositions. Some but, as we will argue below, not all uptake operators can be seen as describing strategies for belief change. Just which uptake operators do correspond to such strategies will be the topic of §1.5.

An uptake operator $\tau$ which describes a strategy for belief change, does so in the following way, viewed from the perspective of the agent:

"Whenever I receive the information that $P$ from a $\tau$-source, I will change my belief state from my current plausibility order $S$ to $S^{\tau P}$.”

So an uptake operator $\tau$, insofar as it corresponds to a dynamic attitude, captures an assessment of reliability by means of exhaustively describing behavioural dispositions: $\tau$ encodes, for each possible epistemic state, represented by a plausibility order $S$, and informational input, represented by a proposition $P$, how the agent reacts when receiving the input $P$ in state $S$: by proceeding to a new epistemic state, represented by the plausibility order $S^{\tau P}$.

1.4.2. Infallible, Strong and Minimal Trust. Our three examples of standard upgrades readily give rise to examples of uptake operators.

— **Infallible trust** $!$ is the operator that maps each proposition $P$ to the update $!P$. The operator $!$ captures that the source is known to be infallible.

— **Strong trust** $\gg$ is the operator that maps each proposition $P$ to the lexicographic upgrade $\gg P$. The operator $\gg$ captures that the source of information is strongly believed (but not known) to be trustworthy.

— **Minimal trust** $\dagger$ is the operator that maps each proposition $P$ to the minimal upgrade $\dagger P$. The operator $\dagger$ captures that the source of information is simply believed (but not known or strongly believed) to be trustworthy.

These operations formalize three distinct levels of trust. It is natural to relate them to Spohn’s tiger example discussed in the introduction of this dissertation. Recall our set of scenarios:

— **I read a somewhat sensationalist coverage in the yellow press claiming that there are tigers in the Amazon jungle.**
1.4. Uptake Operators

— I read a serious article in a serious newspaper claiming this.
— I read the Brazilian government officially announcing that tigers have been discovered in the Amazon area.
— I see a documentary on TV claiming to show tigers in the Amazon jungle.
— I read an article in Nature by a famous zoologist reporting of tigers there.
— I travel to the Amazon jungle, and see the tigers.

Infallible trust is bestowed upon a source that is considered to be an absolute, unquestionable authority concerning the truth of the information received. Infallible trust might thus be applicable to “seeing the tigers with one’s own eyes”, or perhaps to “reading about the tigers in Nature.” But the latter case could also be treated as a case of strong trust, which captures a strong, but still defeasible form of trust in a source—upon reading about the tigers in Nature, one might still want to travel there and check whether the information obtained is correct. A documentary on TV, or an official government announcement on the other hand, might be taken to correspond to minimal trust, belief-inducing, but at the same time, easily defeasible.

1.4.3. Neutrality and Isolation. The trivial upgrade and the null upgrade introduced earlier give rise to two special uptake operators, called “neutrality” and “isolation”.

— Neutrality id is the operator given by $S^{idP} := S$.
— Isolation $\emptyset$ is the operator given by $S^{\emptyset P} := \emptyset$.

The uptake operator id formalizes the concept of “ignoring a source”: receiving the information that $P$ from an id-source is completely inconsequential. Let us consider a “real life” example: Sometimes, information received from a source is best ignored.

Bart to Jessica: I really am the man of your dreams.

Jessica dismisses what Bart says. From a certain perspective, it is as if she had not received any information at all; of course, in some way, she will record that Bart uttered that sentence; possibly, that might affect her perception of Bart; etc—but we disregard such aspects here, focusing just on the question how Jessica’s opinion about the proposition in question is affected by the information obtained from Bart. And the plausible answer is: not at all. More generally, input from a source will be ignored if the source is considered
to be irrelevant; in the narrow sense described, the information received will then not register at all in the epistemic state of the hearer.

The operator $\emptyset$, on the other hand, captures, as the name suggests, isolation from a source: no information received from a $\emptyset$-source will ever give rise to an executable upgrade. This is encoded in the fact that isolation encodes an irreversible crash of an agent’s belief system due to an arbitrary piece of information: for any order $S$, and proposition $P$, $S^{\emptyset P} = \emptyset$, and for any proposition $Q$ and operator $\tau$, $\emptyset^{\tau Q} = \emptyset$: no information received from another source will be able to repair the damage done by upgrading with information from a $\emptyset$-source. This dynamic attitude, $\emptyset$, is useful if we want to model that the communication channel from a particular source to our agent is blocked: the agent cannot, actually, receive any information from that source.

1.4.4. Creating Propositional Attitudes. Given an uptake operator $\tau$, it is natural to ask what propositional attitudes are induced by upgrades $\tau P$ applied to arbitrary plausibility orders.

For an uptake operator $\tau$ and a propositional attitude $A$, we say that $\tau$ creates $A$ iff $S^{\tau P} = A P$, for any order $S$ and proposition $P$.

**PROPOSITION 4.**

— Infallible trust $\dagger$ creates knowledge $K$.

— Strong trust $\uparrow$ creates the disjunction of opposite knowledge and strong belief $K \neg \vee Sb$.

— Minimal trust $\uparrow$ creates the disjunction of opposite knowledge and belief $K \neg \vee B$.

— Isolation creates absurdity.

— Neutrality creates triviality.

**PROOF.** We do the first two items for illustration. For the first item, let $S$ be a plausibility order, and $P$ a proposition. Then $S^{\tau P} = S \cap P \subseteq P$, hence $S^{\tau P} = KP$. We conclude that infallible trust creates knowledge. For the second item, let, again, $S$ be a plausibility order, and $P$ a proposition. If $P \cap S = \emptyset$, then $P \cap S^{\tau P} = \emptyset$ since $S^{\tau P} \subseteq S$, hence $S^{\tau P} = K \neg P$. If, on the other hand, $P \cap S \neq \emptyset$, then by definition of strong trust, $S^{\tau P} = SbP$. Hence strong trust creates the disjunction of opposite knowledge and strong belief.

The connection between strong trust and minimal trust on the one hand, and strong belief and simple belief on the other hand is thus, in view of the previous proposition, not as straightforward as one might have expected: in general, strong trust does not create strong belief, but rather the disjunction of opposite knowledge and strong belief, and similarly, minimal trust does not create simple belief, but rather the disjunction of opposite knowledge and simple belief. We follow up on these observations in §2 and §1.7.
1.5. Dynamic Attitudes

We have seen a number of examples of uptake operators that intuitively correspond to different forms of trust. The question is now the following: which uptake operators can reasonably be taken to represent dynamic attitudes, understood in a pre-formal sense, as agent-internal assessments of the reliability of a source?

1.5.1. Framework- vs. Theory-Level Constraints. In the Belief Revision literature, many attempts of formulating theories of what counts as “rational revision” exist. The AGM postulates and the postulates on transformations of plausibility orders aiming at characterizing “rational iterated revision” suggested by Darwiche and Pearl are examples. But the above query is of a different kind. The framework should allow more variety than the theories (for example, theories of rational revision) one may develop in it. The following quote pertains to a different problem domain—the logical study of time—but clearly brings out the difference:

In the logical study of Time, attention is often restricted to the choice of specific axioms for the temporal precedence order matching certain desired validities in the tense logic. But, there [also] (...) exist preliminary global intuitions, such as ‘anisotropy’ or ‘homogeneity’, constituting the texture of our idea of Time, constraining rather than generating specific relational conditions.

Our investigation of the concept of a dynamic attitude begins with an attempt to identify a number of “global intuitions” in Van Benthem’s sense, i.e., framework-level constraints, embodying the “texture” of our idea of a dynamic attitude, rather than “specific axioms”, i.e., theory-level requirements. Such “specific axioms” have historically played the predominant role in the belief revision literature, as the AGM tradition basically started with a list of such constraints, known today as “the AGM postulates.” For our purposes, however, it seems methodologically helpful to avoid, as best as we can, encoding substantive requirements at the framework-level.

Another difference of our approach to the AGM tradition is its broader focus. The latter theory was developed with an interest in modeling revision with information obtained from a reliable source. Here, we are not only interested in the acceptance of new information (based on trust), but also in

---

17 Of course, what exactly counts as a framework-level rather than theory-level constraint may itself be a matter of debate.
its rejection (based on distrust), as well as in various intermediate levels, and ways of “mixing” trust and distrust. This provides another reason why the constraints we are looking for need to be found at a higher level of abstraction.

In the remainder of this section, we propose four such framework-level constraints: idempotence, dynamic conservativity, informativity and closure under opposites. We specify and motivate them, and use them to define our notion of a dynamic attitude.

1.5.2. Idempotence. An operator \( \tau \) is idempotent iff for any \( P \subseteq W : \tau P = \tau P \), where we recall that \( \tau P = \tau P \) iff for any plausibility order \( S: (S^{\tau P})^{\tau P} = S^{\tau P} \) (cf. §1.3.5).

As a constraint on dynamic attitudes, the condition is motivated by the consideration that receiving the same information (individuated semantically, as a proposition) from the same source one has just received it from should be redundant. Processing the same information twice is unnecessary. Another way to motivate this is our assumption that dynamic attitudes describe, comprehensively, in totality, how an agent processes information coming from a particular source. We can think of each particular information processing event as moving to a new stable state based on taking the input into account. Then our constraint says that a dynamic attitude \( \tau \) captures this stable state, i.e., \( S^{\tau P} \), for each proposition \( P \) and plausibility order \( S \). In other words, \( \tau \) can be taken to represent the target of revision predicated on a particular assessment of reliability.

1.5.3. Dynamic Conservativity. An operator \( \tau \) satisfies dynamic conservativity if for any plausibility order \( S \) and proposition \( P, S^\tau P = S^{\tau (P \cap S)} \).

This requirement parallels the conservativity constraint imposed on propositional attitudes, reflecting the aforementioned fundamental logical principle that properties of a structure do not (cannot, should not) depend on objects that are not part of the structure. In our domain of dynamic attitudes, this corresponds to the idea that worlds that the agent has already irrecoverably excluded from consideration (in the sense captured by irrecoverable knowledge: \( w \) is irrecoverably excluded in \( S \) if \( S \models K \neg \{w\} \), i.e., if \( w \notin S \)) should not affect how she processes an informational input. Hence the effect of applying an upgrade to an order \( S \) should not depend on (all of) \( P \), but rather only on the \( S \)-part of \( P \).

1.5.4. Informativity. An operator \( \tau \) satisfies informativity iff for any order \( S: S^{\tau \emptyset} \in \{S, \emptyset\} \).
1.5. Dynamic Attitudes

The rationale for this requirement is the intuition that absurd information is not useable: receiving the absurd proposition \( \varnothing \) from a source, the agent needs to ignore the information (this corresponds to \( S^{\varnothing} = id \)), or else the agent will acquire inconsistent beliefs (this corresponds to \( S^{\varnothing} = \varnothing \)). Put conversely, if a given input \( P \) does provide genuine information to an agent (i.e., \( S^{\tau P} \notin \{S, \varnothing\} \)), then \( P \neq \varnothing \).

1.5.5. Closure under Opposites. A class of uptake operators \( \Theta \) is closed under opposites if whenever \( \tau \in \Theta \), then also \( \tau^\circ \in \Theta \), where \( \tau^\circ \) (the opposite of \( \tau \)) is defined by \( \tau^\circ P = \tau(\neg P) \).

We will require that the class of dynamic attitudes is closed under opposites. The intuition behind this requirement is that corresponding to any particular assessment of reliability (of a source), one should be able to identify a corresponding “assessment of unreliability”, which is essentially given by the instruction to “upgrade to the contrary.” In effect, this generalizes the idea underlying Smullyan’s liar puzzles, where we are effortlessly able to think of sources as “opposite truthtellers”, to be treated just in the way formalized by the opposite operation, and based on our understanding of what it means to treat someone as a “truthteller”.

1.5.6. Dynamic Attitudes. The preceding considerations lead to the following definition which introduces the central concept of this dissertation.

The class of dynamic attitudes (on \( W \)) is the largest class \( \Delta \) of uptake operators \( \tau \) (on \( W \)) satisfying, for any plausibility order \( S \) (on \( W \)) and proposition \( P \subseteq W \):

- idempotence: \( \tau P = \tau P \),
- dynamic conservativity: \( S^{\tau P} = S^{(P \cap S)} \),
- informativity: \( S^{\tau \varnothing} \in \{S, \varnothing\} \),
- closure under opposites: \( \tau^\circ \in \Delta \).

Observe that our examples from the previous section satisfy these properties, i.e., if \( \tau \in \{!, \uparrow, \downarrow, id, \varnothing\} \), then \( \tau \in \Delta \). We will see many more examples of dynamic attitudes starting in the next chapter.

**Proposition 5.** Any dynamic attitude \( \tau \) satisfies the following, for all plausibility orders \( S \) and propositions \( P \):

1. \( S^{\tau W} \in \{S, \varnothing\} \).
2. If \( P \cap S \in \{S, \varnothing\} \), then \( S^{\tau P} \in \{S, \varnothing\} \).

\(^{18}\)Observe that the opposite operation is involutive, i.e., \((\tau^\circ)^\circ = \tau\).
Chapter 1. Dynamic Attitudes

**PROOF.** For the first item, use informativity and closure under opposites. For the second item, use the first item, informativity and dynamic conservativity. 

1.5.7. Strong Informativity. An uptake operator $\tau$ satisfies strong informativity iff $P \in \{W, \varnothing\}$ implies $S^{\tau P} \in \{S, \varnothing\}$ for any order $S$ and $P \in W$.

Strong informativity expresses that the absurd proposition $\varnothing$ as well as the trivial proposition $W$ cannot provide useable information for an agent. By its definition and the previous proposition, any dynamic attitude satisfies strong informativity.

In fact, the class of dynamic attitudes $\Delta$ can equivalently be defined as the class of uptake operators $\Delta'$ satisfying idempotence, dynamic conservativity and strong informativity.

**Lemma 6.** $\Delta = \Delta'$.

**Proof.** For the “$\subseteq$” direction, any operator $\tau \in \Delta$ satisfies idempotence and dynamic conservativity by its definition, and strong informativeness by its definition together with the first item of Proposition 5. So $\tau \in \Delta'$.

For the “$\supseteq$” direction, let $\tau \in \Delta'$. By definition, $\tau$ satisfies idempotence, dynamic conservativity and informativity. It remains to show that $\tau^{-} \in \Delta'$, which implies that $\tau \in \Delta$. To establish the latter claim, first observe that since $\tau^{-} P = \tau(\neg P)$ for any $P$, $\tau^{-}$ is idempotent by idempotence of $\tau$. Next, we have to show that $\tau^{-}$ satisfies dynamic conservativity. For this, observe that $S^{\tau^{-} P} = S^{\tau(\neg P)}$ and $S^{\tau^{-} (P \cap S)} = S^{\tau(\neg (P \cap S))}$ by definition. Since $\tau$ satisfies dynamic conservativity, $S^{\tau(\neg (P \cap S))} = S^{\tau^{-} (P \cap S) \cap S}$. But $\neg (P \cap S) \cap S = \neg P$, so $S^{\tau^{-} P} = S^{\tau^{-} (P \cap S) \cap S = S^{\tau(\neg P)}}$. Hence $S^{\tau^{-} P} = S^{\tau^{-} (P \cap S)} = S^{\tau^{-} P}$, so $\tau^{-}$ satisfies dynamic conservativity. Finally, strong informativity is obviously satisfied by $\tau^{-}$ whenever strong informativity is satisfied by $\tau$. So $\tau^{-} \in \Delta'$, thus $\tau \in \Delta$, and the proof is complete.

1.6. Idempotence and Moorean Phenomena

In this section, we discuss how the idempotence condition we have imposed on dynamic attitudes can be squared with the failures of idempotence that have received a lot of attention in the literature, both in dynamic epistemic logic and in dynamic semantics. Roughly, the message of this section is that idempotence in the current setting, and idempotence in the setting of dynamic epistemic logic do not amount to the same thing, because the former notion is formulated in terms of propositions, i.e., semantic objects, while the latter notion is formulated w.r.t. a logical language.
1.6. Idempotence and Moorean Phenomena

1.6.1. Moore Sentences. Suppose that an infallibly trusted source tells you:

\[ \text{It is raining even though you don’t know it.} \]

In a situation where your hard information excludes neither that it is raining nor the opposite, that it is not raining, this may very well be useful information. Such a sentence is called a Moore sentence. It has the peculiar property that learning it makes it become false.\(^{19}\) We may formalize our sentence in the syntax of epistemic logic (formally defined right below) as

\[ p \land \neg Kp. \]

Suppose that the set of worlds where \( p \) is true is given by the proposition \( P \). By our assumption, your plausibility order \( S \) is such that neither \( P \cap S \) nor \( \neg P \cap S \) are empty.

Upon learning the above sentence, you delete all non-\( P \)-worlds from \( S \) (remember our assumption: the source is infallibly trusted). But now suppose that the source tells you the same sentence again. This time, you delete all remaining worlds (since, by now, you know that \( P \)), ending up with the empty plausibility order. Clearly, in this case, processing the same input twice is not the same as processing it once.

Is this at odds with the setting we have introduced so far? It’s not. To make very clear that this is a non-problem, let us give the syntax and semantics for a standard logical language, and discuss the issue with more precision. Outside of the current discussion, we will not use this language for quite some time (we return to it only in §5).

1.6.2. The Epistemic Language and Its Semantics. Fix a non-empty set \( \Phi \), called the set of atomic sentences. The elements of \( \Phi \) are understood as denoting basic facts that may or may not hold in a given possible world. A valuation \([\cdot]\) is a map

\[ p \mapsto P \]

is a map that assigns a proposition \( P \subseteq W \) to every atomic sentence \( p \in \Phi \). A (single-agent) plausibility model is a pair \( M = (S, [\cdot]) \), where \( S \) is a plausibility order, and \([\cdot]\) is a valuation.

The language \( \mathcal{L} \) (called the epistemic-doxastic language) is given by the following grammar (\( p \in \Phi \)):

\[
\varphi ::= p \mid \neg \varphi \mid (\varphi \land \varphi) \mid \Box \varphi \mid K\varphi
\]

\(^{19}\text{Cf. van Ditmarsch and Kooi (2006), Holliday and Icard (2010) for more background on and technical explorations of “Moorean phenomena” in epistemic logic.}\)
Read $K\varphi$ as the agent infallibly (or: indefeasibly) knows that $\varphi$; read $\Box \varphi$ as the agent defeasibly knows that $\varphi$.

We interpret the language $\mathcal{L}$ in the usual manner, by providing, for each plausibility model $\mathcal{M} = (S, [\_])$ a map $[\_]_\mathcal{M}$ that assigns a proposition $[\varphi]_\mathcal{M} \subseteq S$ to each sentence $\varphi \in \mathcal{L}$, the proposition comprising the worlds where $\varphi$ is satisfied in $\mathcal{M}$ (equivalently: those worlds $w$ such that $\varphi$ is true at $w$ in $\mathcal{M}$).

$[\_]_\mathcal{M}$ is defined, for each $\mathcal{M}$, by induction on the construction of $\varphi$. Let $\mathcal{M} = (S, [\_])$ be a plausibility model.

- $[p]_\mathcal{M} := [p] \cap S$,
- $[\neg \varphi]_\mathcal{M} := S \setminus [\varphi]_\mathcal{M}$,
- $[\varphi \land \psi]_\mathcal{M} := [\varphi]_\mathcal{M} \cap [\psi]_\mathcal{M}$,
- $[\Box \varphi]_\mathcal{M} := \Box S [\varphi]_\mathcal{M}$,
- $[K \varphi]_\mathcal{M} := KS[\varphi]_\mathcal{M}$.

### 1.6.3. Intensional Dynamic Attitudes

We now observe that, given the above semantics, any dynamic attitude $\tau$ can be lifted to an intensional dynamic attitude, which we again denote with $\tau$, and which is given by the set of upgrades

$$\{\tau \varphi\}_{\varphi \in \mathcal{L}},$$

where for each $\varphi \in \mathcal{L}$, $\tau \varphi$ is determined by putting, for every plausibility model $\mathcal{M} = (S, [\_])$:

$$\mathcal{M}^{\tau \varphi} := \mathcal{M}^{[\varphi]}_\mathcal{M},$$

where we use the notation $\mathcal{M}^{\tau P} := (S^{\tau P}, [\_])$.

The crucial observation is now that it is not in general the case that

$$(\mathcal{M}^{\tau \varphi})^{\tau \varphi} = \mathcal{M}^{\tau \varphi}.$$

Consider the case we started with: put $\tau = !$, consider the sentence $p \land \neg Kp$, and pick a plausibility model $\mathcal{M}$ such that neither $[p]_\mathcal{M}$ nor $[\neg p]_\mathcal{M}$ are empty. Then, as we have seen,

$$(\mathcal{M}^!{(p \land \neg Kp)})^!{(p \land \neg Kp)} \neq \mathcal{M}^!{(p \land \neg Kp)}.$$

What is still the case in general, however, is that for every plausibility model $\mathcal{M} = (S, [\_])$:

$$(S^{\tau [\varphi]}_\mathcal{M})^{\tau [\varphi]}_\mathcal{M} = S^{\tau [\varphi]}_\mathcal{M}.$$
This resolves the above “puzzle”: our requirement that dynamic attitudes be idempotent is not at odds with the phenomenon of Moore sentences. While an intensional dynamic attitude \( \{ \tau \varphi \}_{\varphi \in \mathcal{L}} \), which applies to sentences, need not be idempotent, “ordinary” dynamic attitudes \( \{ \tau P \}_{P \in \mathcal{W}} \) are (by definition) idempotent. Processing the same sentence received from the same source may fail to yield the same result as processing the sentence only once. But processing the same proposition received from the same source amounts to the same as processing it once.

1.7. Fixed Points of Dynamic Attitudes

As pointed out in the introduction, one of our main interests is to study the connection between propositional and dynamic attitudes. This section and the following one, establish a concrete link that forms the basis for much of the work in the dissertation. In this section, we demonstrate that introspective propositional attitudes can be seen as fixed points of dynamic ones; in the next section, we introduce a subsumption order on dynamic attitudes that turns out to match the usual entailment order on the corresponding fixed points.

1.7.1. Fixed Points. Given a dynamic attitude \( \tau \), the fixed point \( \bar{\tau} \) of \( \tau \) is the introspective propositional attitude \( \bar{\tau} = \{ \tau P \}_{P \in \mathcal{W}} \) defined by

\[
S \models \tau P \iff S^\tau P = S.
\]

Given a propositional attitude \( A \) and a dynamic attitude \( \tau \), if \( \bar{\tau} = A \), then we often say that \( \tau \) realizes \( A \).

Notice that fixed points are unique: every dynamic attitude has exactly one fixed point. To see this, suppose that \( A_1 \) and \( A_2 \) are both fixed points of \( \tau \). Let \( S \) be a plausibility order, and \( P \) a proposition. Then \( S \models A_1 P \) iff (since \( A_1 \) is the fixed point of \( \tau \)) \( S^\tau P = S \) iff (since \( A_2 \) is the fixed point of \( \tau \)) \( S \models A_2 P \). Since \( S \) and \( P \) were arbitrarily chosen, we conclude that \( A_1 = A_2 \).

Fixed points of dynamic attitudes \( \tau \) capture the redundancy of specific upgrades given by \( \tau \) by means of the propositional attitude \( \bar{\tau} \). According to our definition, \( S \models \tau P \) iff \( S \) is in the fixed point of the upgrade \( \tau P \) (hence our choice of terminology). What this means intuitively is that \( S \models \bar{\tau} P \) iff the agent already has the information she would obtain if a \( \tau \)-source provided her with the input \( P \). Actually receiving \( P \) from such a \( \tau \)-source is thus redundant; the current plausibility order \( S \) “subsumes” the input \( \tau P \).

---

\[60\]Observe that the fixed point \( \bar{\tau} \) of a dynamic attitude \( \tau \) is indeed a propositional attitude, since the dynamic conservativity of \( \tau \) ensures the conservativity of \( \bar{\tau} \).
1.7.2. Creating and Stopping. Given a dynamic attitude \( \tau \) and a propositional attitude \( A \), the first is the fixed point of the second iff two things come together. Stating them separately is conceptually helpful. We say that \( \tau \) creates \( A \) iff \( S^{\tau P} = AP \) (for any \( S \) and \( P \)); and we say that \( A \) stops \( \tau \) iff whenever \( S = AP \), then \( S^{\tau P} = S \) (for any \( S \) and \( P \)).

**Proposition 7.** The fixed point of \( \tau \) is \( A \) iff \( \tau \) creates \( A \) and \( A \) stops \( \tau \).

**Proof.** From left to right, suppose that \( \overline{\tau} = A \). Then obviously \( A \) stops \( \tau \). Further, take any order \( S \) and proposition \( P \). Then \( S^{\tau P} = \overline{\tau} P \) by idempotence of dynamic attitudes. By the assumption, \( S^{\tau P} = AP \). So \( \tau \) creates \( A \). From right to left, suppose that \( \tau \) creates \( A \) and \( A \) stops \( \tau \). We need to show that \( \overline{\tau} = A \). One half of this follows from the fact that \( A \) stops \( \tau \). For the other half, suppose that \( S^{\tau P} = S \). Since \( \tau \) creates \( A \), \( S^{\tau P} = AP \), i.e., \( S = AP \), the desired result. So \( \overline{\tau} = A \), and the proof is complete.

So the fact that \( A \) is the fixed point of \( \tau \) indicates that (1) \( \tau \) creates \( A \) and (2) \( \tau \) leaves the order unchanged once \( A \) has been reached. In this sense, \( \tau \) “dynamically realizes” \( \overline{\tau} \).

Two examples illustrate how creating and stopping may come apart. First, consider infallible trust \( ! \) and belief \( B \). On the one hand, \( ! \) creates \( B \). On the other hand, from the fact that \( S = BP \), it does not follow that \( S^{! P} = S \) (counterexample: assume that there are non-\( P \)-worlds in \( S \)). In our terminology: \( B \) does not stop \( ! \). So the fixed point of \( ! \) is not \( B \). Second, consider absurdity \( \bot \) and minimal trust \( \uparrow \). On the one hand, \( \bot \) stops \( \uparrow \); but, obviously, \( \uparrow \) does not create \( \bot \). So the fixed point of \( \uparrow \) is not \( \bot \).

1.7.3. Definability of Propositional Attitudes. Here are five first examples of fixed points:

**Proposition 8.**

1. \( \overline{!} = K \) *(the fixed point of infallible trust is knowledge).*

2. \( \overline{\downarrow} = Sb \lor \neg K \) *(the fixed point of strong trust is the disjunction of strong belief and the opposite of knowledge).*

3. \( \overline{\uparrow} = B \lor \neg K \) *(the fixed point of minimal trust is the disjunction of simple belief and the opposite of knowledge).*

4. \( \overline{\uparrow} = \bot \) *(the fixed point of isolation is inconsistency).*

5. \( \overline{id} = \top \) *(the fixed point of neutrality is triviality).*
Notice that, perhaps unexpectedly, strong trust $\uparrow$ and strong belief $Sb$ on the one hand, and minimal trust $\uparrow$ and simple belief $B$ on the other hand, do not quite match. Take $\uparrow$. The reason for the mismatch is that a plausibility order $S$ containing no $P$-worlds does not satisfy $SbP$. However, $S\uparrow P = S$ by definition of $\uparrow$. So the fixed point of strong trust is not strong belief. Analogous remarks apply to $\hat{\uparrow}$ and $B$. We will see in §2.1 how to define dynamic attitudes whose fixed points are strong belief and simple belief, respectively.

1.7.4. Characterizing Introspectiveness. Fixed points of dynamic attitudes are, by their definition, introspective propositional attitudes. In fact, the class of introspective propositional attitudes can be characterized in terms of dynamic attitudes.

For use in the proof of the next theorem, we define, for any propositional attitude $A$, the dynamic attitude test for $A$, denoted by $?A$, given by

$$S?AP := \begin{cases} S & S \vdash AP, \\ \emptyset & \text{otherwise,} \end{cases}$$

We now observe the following:

**Theorem 9.** Let $A$ be a propositional attitude. The following are equivalent:

1. $A$ is introspective.
2. There exists a dynamic attitude $\tau$ such that $\tau = A$.

**Proof.** For the direction from (1.) to (2.), suppose that $A$ is an introspective propositional attitude. Consider the dynamic attitude $?A$, as defined ahead of this proposition. The fixed point of $?A$ is, obviously, $A$, which finishes one direction. Since the other direction is trivial, we are done.

Think of introspective propositional attitudes as possible targets of belief change. The result ensures that for each such target, there is a strategy, given by a dynamic attitude, that realizes that target. Of course, there may be more reasonable strategies than the ones actually chosen in the above proof. We further comment on this aspect below, in §1.8.

1.7.5. More on Tests. The dynamic attitude $?A$ we have defined for each propositional attitude $A$ in the previous paragraph tests, for each order $S$ and proposition $P$, whether the agent has the propositional attitude $A$ towards $P$. If the test succeeds, the order remains unchanged; if the test fails, the agent ends up in the absurd epistemic state given by the empty plausibility order $\emptyset$. Intuitively, such tests correspond to dynamic acts of introspection of the agent.
We can also use tests to lift operations on propositional attitudes to the dynamic level. Two examples that will be used in the proof of Proposition 12 in § 1.8 below: (1) given dynamic attitudes $\sigma$ and $\tau$, the disjunction test $\triangledown(\sigma \lor \tau)$ is the dynamic attitude defined by

$$
S^{\triangledown(\sigma \lor \tau)} := \begin{cases} 
S & S = \overline{\sigma} \lor \overline{\tau} P \\
\emptyset & \text{otherwise}
\end{cases}
$$

The fixed point of the disjunction test $\triangledown(\sigma \lor \tau)$ is the disjunction of the fixed points of $\sigma$ and $\tau$. (2) given dynamic attitudes $\sigma$ and $\tau$, the conjunction test $\triangledown(\sigma \land \tau)$ is the dynamic attitude defined by

$$
S^{\triangledown(\sigma \land \tau)} := \begin{cases} 
S & S = \overline{\sigma} \land \overline{\tau} P \\
\emptyset & \text{otherwise}
\end{cases}
$$

The fixed point of the conjunction test $\triangledown(\sigma \land \tau)$ is the conjunction of the fixed points of $\sigma$ and $\tau$.

1.8. Subsumption

1.8.1. Entailment. Propositional attitudes allow for qualitative comparisons along the natural entailment relation. For example, knowledge $K$ implies belief $B$, since whenever $S \models KP$, also $S \models BP$. More generally, the entailment order on (introspective) propositional attitudes is defined by

$$A \leq A' \iff \forall S \forall P: S \models AP \rightarrow A'P,$$

for any propositional attitudes $A$ and $A'$. If $A \leq A'$, then we say that $A$ entails $A'$.

More generally, given an $n$-ary operation $o$ that assigns an introspective attitude $o(A_1, \ldots, A_n)$ to given introspective attitudes $A_1, \ldots, A_n$, define the $n$-ary operation $\triangledown o$ that assigns to given dynamic attitudes $\tau_1, \ldots, \tau_n$ the test $\triangledown o(\tau_1, \ldots, \tau_n)$, defined by

$$S^{\triangledown o(\tau_1, \ldots, \tau_n)} := \begin{cases} 
S & S = o(\tau_1, \ldots, \tau_n) P \\
\emptyset & \text{otherwise}
\end{cases}
$$

Then, one may observe that $\triangledown o(\tau_1, \ldots, \tau_n) = o(\overline{\tau_1}, \ldots, \overline{\tau_n})$. Figure 5 shows this in a diagram. Arrows going down indicate applications of the fixed point operation $\tau \mapsto \tau$, while arrows going from left to right indicate applications of $\triangledown o$ (at the top), respectively $o$ (at the bottom).
1.8. Subsumption

\[ \tau_1, \ldots, \tau_n \Downarrow \? o(\tau_1, \ldots, \tau_n) \]

\[ \tau_1, \ldots, \tau_n \Downarrow o(\tau_1, \ldots, \tau_n) \]

**Figure 5.** In this diagram, downward-point arrows represent applications of the fixed point operator. Taking the fixed point of a bunch of dynamic attitudes \( \tau_1, \ldots, \tau_n \) and applying the operation \( o \) amounts to the same thing as applying the operation \( ? o \) to \( \tau_1, \ldots, \tau_n \) (as defined in the main text), and only then taking the fixed point of the resulting dynamic attitude \( ? o(\tau_1, \ldots, \tau_n) \).

1.8.2. **Subsumption.** Here, we consider the question how an analogous notion for dynamic attitudes should be defined. Our answer is that a strength order on dynamic attitudes naturally arises from the notion of upgrade subsumption defined in §3. Recall that, given upgrades \( u \) and \( u' \), \( u \) subsumes \( u' \) (notation: \( u = u' \) iff \( (S^u)^{u'} = S^u \) for any plausibility order \( S \)).

We now define the subsumption order \( \leq \) on dynamic attitudes by putting

\[ \sigma \leq \tau \iff \forall P: \sigma P \models \tau P \]

for any dynamic attitudes \( \sigma \) and \( \tau \). If \( \sigma \leq \tau \), then we say that \( \sigma \) *subsumes* \( \tau \). We write \( \sigma < \tau \) iff \( \sigma \leq \tau \) and not \( \tau \leq \sigma \); and we write \( \sigma \approx \tau \) iff both \( \sigma \leq \tau \) and \( \tau \leq \sigma \).

1.8.3. **Examples.** Here are two examples of subsumption relations among attitudes:

**Proposition 10.**

- For any dynamic attitude \( \tau \): if \( \tau \neq id \), then \( \tau < id \), and if \( \tau \neq \varnothing \), then \( \varnothing < id \).
- \( ! \Downarrow \Downarrow \Downarrow \).

We will see more uses of the subsumption order in §2.5.

1.8.4. **Subsumption as Inclusion of Fixed Points.** Equivalently, the subsumption order on dynamic attitudes can be defined in terms of fixed points of dynamic attitudes, since an attitude \( \sigma \) subsumes an attitude \( \tau \) iff the fixed point of \( \sigma \) entails the fixed point of \( \tau \). The next theorem justifies this remark.

**Theorem 11.** \( \sigma \leq \tau \) iff \( \sigma \Downarrow \tau \).
Chapter 1. Dynamic Attitudes

Proof. From (1.) to (2.), suppose that $\sigma \leq \tau$, i.e., for all $S$, for all $P$: $(S^{\sigma P})^{\tau P} = S^{\sigma P}$. Choose a plausibility order $S$ and a proposition $P$, and suppose that $S \models \sigma P$. This means that $S^{\sigma P} = S$. By the assumption, $(S^{\sigma P})^{\tau P} = S^{\sigma P}$, so $S^{\tau P} = S$, thus $S \models \tau P$, which concludes the left to right direction.

Conversely, suppose that $\sigma \leq \tau$, i.e., for all $S$, for all $P$: if $S \models \sigma P$, then $S \models \tau P$. Choose a plausibility order $S$ and a proposition $P$. By idempotence of attitudes, $S^{\sigma P} = (S^{\sigma P})^{\sigma P}$. So $S^{\sigma P} \models \sigma P$. By the initial assumption, $S^{\sigma P} \models \tau P$, so $(S^{\sigma P})^{\tau P} = S^{\sigma P}$. We have thus established that $\sigma P \cdot \tau P = \sigma P$, which concludes the right to left direction, and the proof.

So subsumption expresses inclusion of the corresponding fixed points and thus inclusion of un informativeness. That $\sigma \leq \tau$ means that whenever applying $\sigma$ is redundant, then so is applying $\tau$, i.e., whenever a plausibility order is a fixed point of the upgrade $\sigma P$, then it is also a fixed point of the upgrade $\tau P$.

1.8.5. Properties of the Subsumption Order. We observe a number of basic properties of our two strength orders. Let us first fix some standard terminology. As usual, a preorder is a reflexive and transitive binary relation; a partial order is an antisymmetric preorder\footnote{A binary relation $R$ is antisymmetric iff for any $x$ and $y$ in the domain of $R$: if $xRy$ and $yRx$, then $x = y.$}. Now let $O = (O, \leq)$ be a preorder, and let $K \subseteq O$. Then $x \in O$ is a lower bound for $K$ iff for all $y \in K$: $x \leq y$; and $x$ is a greatest lower bound for $K$ iff $x$ is a lower bound for $K$ and for all lower bounds $y$ for $K$: $y \leq x$. Similarly, $x$ is an upper bound for $K$ iff for all $y \in K$: $y \leq x$; and $x$ is a least upper bound for $K$ in $O$ iff $x$ is an upper bound for $K$ and for all upper bounds $y$ for $K$: $x \leq y$. Observe that greatest lower bounds and least upper bounds are unique if $O$ is a partial order, but not necessarily so if $O$ is a preorder. Finally, a lattice is a partial order $O$ such that for any $x, y \in O$: $\{x, y\}$ has a least upper bound and a greatest lower bound.

Using this terminology, the entailment and the subsumption order introduced above are distinguished primarily by the fact that the latter fails to be antisymmetric:

Proposition 12.

1. The entailment order on introspective propositional attitudes is a lattice.

2. The subsumption order on dynamic attitudes is a preorder $O = (O, \leq)$ such that for any $x, y \in O$: $\{x, y\}$ has a (not necessarily unique) least upper bound and a (not necessarily unique) greatest lower bound.
3. **The subsumption order on dynamic attitudes is not a partial order, hence not a lattice.**

**Proof.**

1. The entailment order on introspective propositional attitudes is obviously a partial order. Furthermore, given two attitudes \( A \) and \( A' \), the (unique) least upper bound for \( A \) and \( A' \) is given by \( A \land A' \), while the (unique) greatest lower bound for \( A \) and \( A' \) is given by \( A \lor A' \). So the entailment order is a lattice.

2. The subsumption order on dynamic attitudes is obviously a preorder. Now let \( \tau \) and \( \tau' \) be dynamic attitudes. Observe that the fixed point of the conjunction test \( ?(\tau \land \tau') \) is \( \overline{\tau} \land \overline{\tau'} \). But \( \overline{\tau} \land \overline{\tau'} \) is the (unique) least upper bound for \( \overline{\tau} \) and \( \overline{\tau'} \). It follows using Proposition 11 that \( ?(\tau \land \tau') \) is a (not necessarily unique) least upper bound for \( \tau \) and \( \tau' \). For greatest lower bounds, we argue analogously using the disjunction test \( ?(\tau \lor \tau') \). Our claim follows.

3. We exhibit a counterexample, showing that the subsumption order is not antisymmetric. The fixed point of \( ?K \) is the same as the fixed point of \( ! \), i.e., irrevocable knowledge \( K \). By Proposition 11, \( ?K \approx ! \). However, obviously not \( ?K = ! \). So the subsumption order is not antisymmetric, and thus not a partial order, hence not a lattice either.

The main point of the previous result is that the map sending dynamic attitudes to their fixed points is not one-to-one; given a propositional attitude \( A \), it is not usually possible to identify a unique attitude \( \tau \) such that \( \tau = A \). In other words: there are more dynamic attitudes than introspective propositional ones, and as a result the subsumption order fails to be a lattice by Proposition 11.

1.8.6. **Question.** If we understand introspective propositional attitudes as targets for belief change (cf. the previous section), the preceding observation raises an important question. Having chosen a suitable such target, how would an agent choose an appropriate dynamic attitude that realizes it? Is the choice completely arbitrary? Or, to put it differently: given distinct dynamic attitudes \( \tau \) and \( \tau' \) such that \( \overline{\tau} = \overline{\tau'} = A \), can a case be made that one of them is, in some important respect, a better choice for realizing \( A \)? We will study one answer to this question in Chapter 3 where we use the criterion of minimal

\[33\] There are exceptions: the only dynamic attitude whose fixed point is absurdity \( \bot \) is isolation \( \emptyset \).
change that has traditionally been important in belief revision theory to select attitudes that are “more optimal” for a given fixed point over ones that are less so.