Chapter 6 SOFT INFORMATION, SELF-CORRECTION, AND BELIEF CHANGE

6.1 From knowledge to belief as a trigger for actions

While the best available information and knowledge are important to agency, it is also clear that our actions are often driven by less demanding attitudes of belief. I am riding my bicycle this evening because I believe that it will get me home, even though my epistemic range includes worlds where the great San Andreas earthquake finally happens. And more generally, decision theory is about choice and action on the basis of beliefs, since waiting for knowledge may last forever. Thus, our next step in the logical dynamics of rational agency is the study of beliefs. In what follows, we will not view these notions in deep philosophical terms. Rather think of the simple scenarios in our chapters so far. The cards have been dealt. I know that there are 52 of them, and I know their colors. But I have more fleeting beliefs about who holds which card, or about how the other agents will play. [79]

Hard versus soft information With this distinction in attitude comes a further dynamics. A public announcement *!P* of a fact *P* was an event of *hard information*, which changes irrevocably what I know. If I see the Ace of Spades played on the table, I come to know that no one holds it any more. This is the trigger that drove our dynamic epistemic logics in Chapters 2 and 3. Such events of hard information may also change our current beliefs – and we will find a complete logical system for this. But next, there are also events of *soft information*, which affect my current beliefs without affecting my knowledge about the cards. 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. We must also describe events like this, and indeed, we will provide them with a semantics in terms of plausibility orderings suitable for dynamic logic analysis in the same style that we had so far. Now, this process of belief change is usually considered the domain of *AGM*-style 'belief revision theory' (Gaerdenfors & Rott 1995), and we will discuss later how the two styles of analysis are related.

[79] Of course, I could even be wrong about the cards (perhaps the Devil added his visiting card) – but this worry seems morbid, and not useful in investigating normal information flow.

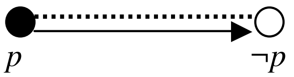
The tandem of jumping ahead and self-correction Here is what is most important to me in this chapter from the standpoint of rational agency. As stated, as acting agents, we are bound to form beliefs that go beyond the hard information that we have. And this is not a concession to human frailty or to our mercurial nature. It is rather the essence of creativity, jumping ahead to conclusions we are not really entitled to, and basing our beliefs and actions on them. But there is another side to this coin, that I would dub our capacity for *self-correction*, or if you wish, for *learning*. We have an amazing capacity for standing up after we have fallen informationally, and to me, rationality is displayed at its best in intelligent responses to new evidence that contradicts what we thought so far. What new beliefs do we form, and what amended actions result? I see this as a necessary *pair of skills*: jumping to conclusions (i.e., beliefs) and correcting ourselves in times of trouble. And the hallmark of a rational agent is to be good at both: it is easy to prove one theorem after another, it is hard to revise your theory when your theory has come crashing down. So, in pursuing the dynamic logics of this chapter, I am trying to chart this second skill.

6.2 Static logic of knowledge and belief

Knowledge and belief have been studied together ever since Plato proposed his equation of knowledge with ‘true belief that is justified’, and much of epistemology today is still about finding a mysterious ‘fourth ingredient’ that would make the equation valid. Without resolving such issues here, how should we think of knowledge versus belief semantically?

Reinterpreting PAL One easy route merely reinterprets dynamic-epistemic logic as we had it so far. We just read the earlier *K*-operators as beliefs, once again viewed as universal quantifiers over the accessible range, and relax the constraints on the accessibility relation to, say, none whatsoever, to stay as general as possible. One immediate test for such an approach is that it must be possible for beliefs to be *wrong*:

Example A mistaken belief.
Consider the following model with two worlds that are epistemically accessible to each other, but the pointed arrow is the only belief relation. Here, in the actual black world to the left, the proposition *p* is true, but the agent mistakenly believes that $\neg p$:



■

With this view of doxastic modalities, which is close to the original approach in Hintikka 1962, the machinery of *DEL* works exactly as before. But there is a problem,

Example, continued

Consider a public announcement *!p* of the true fact *p*. The *PAL* result is the one-world model where *p* holds, with the inherited *empty* doxastic accessibility relation. But on the universal quantifier reading of belief, this means the following: the agent believes that *p*, but also that ¬*p*, in fact $B\bot$ is true at such an end-point. ■

In this way, agents who have their beliefs contradicted are shattered and start believing anything. While this may be true for some people sometimes, such a collapse does not sound right in general, and hence we change the semantics in a more revealing manner.

World comparison by plausibility A richer view of belief follows the intuition that an agent believes the things that are true, not in all of her epistemically accessible worlds, but only in those that are 'best' or 'most relevant' worlds to her. I believe that my bicycle will get me home on time, even though I do not know stricto sensu that it will not suddenly disappear in an earthquake chasm. But the worlds where it stays on the road are more plausible than those where it drops down, and among the former, those where it arrives on time are more plausible than those where it does not. Static models for this setting are easily defined:

Definition Epistemic-doxastic models.

Epistemic-doxastic models are structures $\boldsymbol{M} = (W, \{\sim_i\}_{i\in I}, \{\leq_{i,\,s}\}_{i\in I}, V)$ where the relations $\sim_I$ stand for epistemic accessibility, and the $\leq_{i,\,s}$ are ternary comparison relations for agents read as follows, $\leq_{i,\,s} xy$ if, in world *s*, agent *i* considers *x* at least as plausible as *y*. ■

Models like this occur in the work of Lewis in conditional logic, all the way to Shoham 1988 on preference relations in AI, and the 'graded models' of Spohn 1988. One can impose several conditions on the plausibility relations, depending on their intuitive reading. The minimum found in Burgess 1981 is *reflexivity* and *transitivity*, while Lewis 1973 also imposes *connectedness*: for all worlds s, t, either $s \leq t$ or $t \leq s$. The latter yields the well-known geometrical systems of 'nested spheres' known from conditional logic. [80]

As before with epistemic models, our dynamic analysis works largely independently from such design decisions, important though they may be to specific applications. In particular, working with connected orders yields simply visualizable pictures of a line of 'equi-plausibility clusters', in which there are only three options for worlds *s, t*:

either *strict precedence* $s < t$ *or* $t < s$, or *equi-plausibility* $s \leq t \wedge t \leq s$.

While this is attractive, there are also settings where we want to allow a fourth option of *incomparability:* $\neg\, s \leq t \wedge \neg\, t \leq s$. This happens when comparing worlds according to conflicting criteria – and sometimes, the latter partial orders are just the mathematically more elegant and perspicuous approach (Andréka, Ryan, Schobbens 2002).

Languages and logics One can interpret many logical operators in this richer comparative order structure. In what follows, we choose intuitive 'maximality' formulations for belief $B_i\varphi$, even though these must be modified somewhat in models allowing infinite descent in the ordering. [81] First of all, there is plain belief, whose modality is interpreted as follows. For convenience, we will drop subscripts henceforth where they do not add insight:

Definition Belief as truth in the most plausible worlds.
In epistemic-doxastic models, knowledge is interpreted as usual, while we now say that $\boldsymbol{M}, s \models B_i\varphi$ iff $\boldsymbol{M}, t \models \varphi$ for all worlds t that are maximal in the ordering $\lambda xy. \leq_{i, s} xy$. ■

But as we shall soon see later, just absolute belief does not suffice. The more general notion needed for reasoning about information flow and action is *conditional belief.* We write this notion as follows: $B^{\psi}\varphi$, with the intuitive reading that, conditional on ψ, the agent believes that φ. This is very clause to standard conditional semantics:

[80] The natural *strict variant* of these orderings is defined as follows: $s < t$ iff $s \leq t \wedge \neg\, t \leq s$.

[81] We consider such changes to infinite models an orthogonal issue to the main thrust of what is done in this chapter, and will only refer to more general semantic formulations occasionally.

Definition Conditional beliefs as plausibility conditionals.
In epistemic-doxastic models, $\boldsymbol{M}, s \models B^{\psi}\varphi$ iff $\boldsymbol{M}, t \models \varphi$ for all worlds t which are minimal for the ordering $\lambda xy. \leq_{i,s} xy$ in the set $\{u \mid \boldsymbol{M}, u \models \psi\}$. ■

Absolute belief $B\varphi$ is a special case of this: $B^{True}\varphi$. It can be shown that conditional belief is not definable in terms of absolute belief, so we have a genuine language extension. [82]

Digression on conditionals As with epistemic notions in Chapters 2, 3, conditional beliefs *pre-encode* beliefs that we would have if we were to learn certain things – though we can be more precise in our later dynamic sections. [83] The formal analogy with conditionals is this. A conditional $C \Rightarrow D$ says that D is true in the minimal or 'closest' worlds where C is true, as measured by some comparison order on worlds. This is exactly the above clause. Thus, results from conditional logic apply. For instance, on models with reflexive transitive plausibility orderings, we have this completeness theorem (Burgess 1981, Veltman 1985):

Theorem The logic of $B^{\psi}\varphi$ is axiomatized by the laws of propositional logic plus obvious transcriptions of the following principles of conditional logic:
(a) $\varphi \Rightarrow \varphi$, *(b)* $\varphi \Rightarrow \psi$ *implies* $\varphi \Rightarrow \psi \vee \chi$, *(c)* $\varphi \Rightarrow \psi$, $\varphi \Rightarrow \chi$ *imply* $\varphi \Rightarrow \psi \wedge \chi$,
(d) $\varphi \Rightarrow \psi$, $\chi \Rightarrow \psi$ *imply* $(\varphi \vee \chi) \Rightarrow \psi$, *(e)* $\varphi \Rightarrow \psi$, $\varphi \Rightarrow \chi$ *imply* $(\varphi \wedge \psi) \Rightarrow \chi$.

Richer modal languages One can also interpret richer modal languages on epistemic-doxastic models. For instance, the idea of a 'best' world induces a binary relation '*best*' between worlds s and t, defined as "t is maximal in $\lambda xy. \leq_s xy$". One could introduce an

[82] Likewise, the binary quantifier *"Most A are B"* is not definable in first-order logic extended with just a unary quantifier *"Most objects in the universe are B" (cf.* Peters & Westerståhl 2005).

[83] A conditional belief $B^{\psi}\varphi$ does not quite say what we would believe if we learnt the antecedent. For, the action of learning the antecedent ψ changes the current model $\boldsymbol{M}$, and hence the truth value of the consequent φ might change, as the modalities in φ now range over different worlds in $M|\psi$. Similar phenomena occurred with epistemic statements after communication in Chapter 3, and in logic in general. E.g., the relativized quantifier in "All mothers have daughters" does not say that, if we relativize to the subset of mothers, all of them have daughters who are mothers themselves.

explicit modality for this and make the above belief modality definable. [84] More generally, one can define conditional belief with explicit modal operators *[best ψ]φ*. More powerful modal languages of this kind occur in modal logics for describing games with preference relations, which are akin to plausibility relations (cf. Chapters 8 and 9 below). We will consider further extensions of our language later on, motivated by dynamic considerations.

Epistemic-doxastic logics In line with the general approach in this book, we will not pursue completeness theorems for static logics of knowledge and belief per se. But to ease what follows, this chapter makes one semantic simplification which reflects immediately in the logic. Henceforth, we assume that epistemic accessibility is an *equivalence relation*, and plausibility a *pre-order over the equivalence classes*, the same as viewed from any world inside such a class. This will have the immediate effect of making the following valid:

$B\varphi \rightarrow KB\varphi$ *Epistemic-Doxastic Introspection*

While we admit that this is a strong assumption (though one often adopted in the literature), it does help focus on the main ideas of the dynamics which we will investigate now.

6.3 Belief change under hard information

Now we are in a position to present our first dynamic logic of belief revision. It puts together the logic *PAL* for public announcements *!P* of true propositions *P* with our static models for conditional belief, following exactly the same methodology as earlier chapters. This will allow us to move faster with stating the results of our analysis, since we need not repeat the general logical points that were already explained in Chapters 2, 3.

A complete axiomatic system For a start, we must locate the key recursion axiom for the new beliefs, something which can be done easily, using update pictures as before:

Fact The following formula is valid for beliefs after hard information:

$[!P]B\varphi \leftrightarrow (P \rightarrow B^P([!P]\varphi)$.

This is like the *PAL* reduction law for knowledge under public announcement, but note the conditional belief in the consequent, which cannot be mimicked by a conditional absolute

[84] This is like directly describing ‘selection functions’ in conditional logic.

belief of the form $B(P \rightarrow$. But of course, to keep the complete dynamic language in harmony, this principle is not enough. We need to know, not just which beliefs are formed after new information, but which conditional beliefs are formed. This point is overlooked in classical belief revision theory, where the emphasis was on describing how new beliefs are formed: which means that one gets stuck in one round, since the new belief state does not pre-encode any further information about what happens in the next round of revision. This so-called *'Iteration Problem'* cannot arise in a systematic logical set-up.

So, what is the stable recursion principle for change in both conditional and absolute beliefs under hard information? In principle, there might be an infinite regress here toward 'conditional conditional beliefs', but in fact, there is not:

Theorem The logic of conditional belief under public announcements is axiomatized completely by (a) any complete static logic for the model class chosen,
(b) the *PAL* recursion axioms for atomic facts and Boolean operations,
(c) the following new recursion axiom for conditional beliefs:
$[!P]B^{\psi}\varphi \leftrightarrow (P \rightarrow B^{P \wedge [!P]\psi} [!P]\varphi).$

Proof First we check the soundness of the new axiom. On the left hand side, it says that in the new model $(\boldsymbol{M}|P, s)$, φ is true in the best ψ-worlds. With the usual precondition for the announcement, on the right-hand side, it says that in $(\boldsymbol{M}, s)$, the best worlds that are P now and will become ψ after announcing that P, will also become φ after announcing P. This is indeed equivalent. The remainder of the proof is our earlier stepwise reduction analysis, noting that the above axiom is recursive, pushing announcement modalities inside. ■

To get the combined version with knowledge, we just combine with the *PAL* axioms.

Pitfalls of update: clarifying the Ramsey Test Our dynamic logic sharpens up the usual discussion of the famous *Ramsey Test*, which says this: "A conditional proposition $A \Rightarrow B$ is true, if, after adding A to your current stock of beliefs, the minimal revision to make the result consistent implies that B." In our current perspective, this passage is ambiguous, since B need no longer mean the same thing after the described change has taken place.

That is why the above recursion axiom would carefully distinguish between propositions φ before an update with the antecedent A and what happens to them after, as in $[!A]\varphi$.

Even so, it is interesting to look at the special case of *factual propositions* φ without epistemic or doxastic operators (cf. Chapter 3), which do not change their truth value under announcement. In that case the above two axioms become, with Q, R factual propositions:

$$[!P]BQ \leftrightarrow (P \rightarrow B^{P}Q)$$
$$[!P]B^{R}Q \leftrightarrow (P \rightarrow B^{P \wedge R}Q)$$

This of course, is much closer to linking conditional assertions and update modalities. ■

Belief change under hard update is not yet genuine belief revision in the usual sense, which may also be triggered by weaker incoming information (Section 6.4 below). Nevertheless, we pursue it a bit further by itself, since it is linked to two important themes in the study of rational agency: variety of *consequence relations*, and variety of *epistemic attitudes*.

Update versus inference: non-monotonic logic Update of beliefs under hard information is also an alternative to so-called ‘nonstandard notions of consequence’. Here is a brief illustration (details are in van Benthem 2008). Classical consequence from premises $\boldsymbol{P}$ to conclusion C says all models of $\boldsymbol{P}$ are models for C. Now McCarthy 1980 pointed out that problem solving and planning go beyond this, getting more out of premises by zooming in on the most 'congenial' models. A *circumscriptive* consequence from $\boldsymbol{P}$ to C says that

C is true in all the *minimal* models for $\boldsymbol{P}$

Here, minimality refers to a comparison order ≤ for models: inclusion of object domains, or of denotations for specified predicates, and so on. The general idea is minimization over any reflexive transitive order of 'relative plausibility' (Shoham 1988), and there is by now a rich theory of non-monotonic consequence relations. Now, this is precisely our framework of plausibility models, and we think one might rethink the original motivation of this area. We are given some initial information in a puzzle or a game, and need to find the true situation, as new information comes in. The striking phenomenon in such scenarios is *not inference at all*, but rather our receiving that information, and our subsequent responses:

We are playing the board game "Kings and Cardinals" (the board is an object of public observation having 'monasteries' and 'advisors' placed here and there. I look at the cards in my hand (a private observation), and also at the map of medieval Europe on the board. Right now, I know certain things about the outcome of the game, while I believe more than what I strictly know, based on my expectations about cards that the other players hold, or their temperaments: timid, bluffing, etcetera. Now, new information comes in: say, you select a new country on the map and place some counters there. This observation changes my current information state. I know more now, and additionally, the observation may even speed along further beliefs of mine: you are trying to build a trade route from Burgundy to Bohemia. Of course, these current beliefs may be refuted by further moves of yours, unlike the hard indefeasible knowledge which I have obtained about what's on the board.

Solving puzzles and playing games is all about information update and belief change. Nonmonotonic logics have such processes in the background, but they leave them *implicit*. But making them explicit is precisely the point of our dynamic logics. To me, circumscriptive inference is about belief formation, and our logics do more justice to the original intuitions.

Dynamic consequence on a classical base In fact, our logic suggests two kinds of dynamic consequence, depending on what holds once the premises are processed. First, *knowledge* may result, as in the dynamic inference of Chapter 3, and we get classical consequence at least for factual assertions. [85] Alternatively, *belief* may result, and we go to McCarthy's minimal worlds in the order. Thus, what is usually cast as notions of consequence

$$P_1, \ldots, P_k \Rightarrow \varphi$$

even gets several dynamic variants definable in our language:

either $[!P_1] \ldots [!P_k]\, K\varphi$ [86] or $[!P_1] \ldots [!P_k]\, B\varphi$

whose behaviour is fully captured by our complete logic. I think it is this diversity of responses to various sorts of incoming information which truly explains the modern galaxies of 'notions of consequence', where different styles live together. Moreover, and

[85] Factual assertions seem all that are considered in accounts of nonstandard consequence relations. But as we saw in Chapter 3, structural rules get dynamic twists when we consider the full language.

[86] This outcome should of course be stated as *common knowledge* in the multi-agent case.

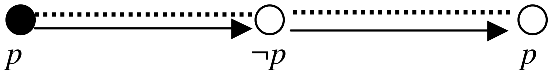
this is a truly deviant point of view, once these events have been made explicit, the dynamic logic just works with a classical notion of consequence. In a slogan:

Non-monotonic logic is monotonic dynamic logic of belief change.

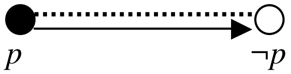
Language extensions: a richer repertoire of attitudes Next, we consider another striking feature of our logic. The above setting may seem simple, but it hides tricky scenarios:

Example Misleading with the truth.

Consider a model where an agent believes that p, which is indeed true in the actual world to the far left, but for ‘the wrong reason’, viz. she thinks the most plausible world is the one to the far right. For convenience, assume each world also verifies a unique proposition letter.



Now giving the true information that we are not in the final world *(‘$\neg q_3$’)* updates to



in which the agent believes mistakenly that $\neg p$. ■

Observations like this have been made in philosophy, computer science, and game theory.

Agents have a plethora of attitudes In response to this, it makes sense to consider an alternative view of what we have been doing in the first place in this chapter. So far, we have assumed that knowledge and belief as formalized in the above are the only relevant attitudes that agents can have. But this is largely an inheritance from the tradition in the field. Stepping back, what seems much more likely is that agents have a rich repertoire of attitudes concerning information and action, witness the many terms in natural language with an epistemic or doxastic ring: being certain, being convinced, assuming, etcetera. [87]

Safe belief Among all possible options in this plethora of epistemic-doxastic attitudes, it makes particular sense to define the following new notion, intermediate between knowledge and belief, that has stability under new true information:

[87] Krista Lawlor has pointed me also at the richer repertoire found in pre-modern epistemology.

Definition Safe belief.

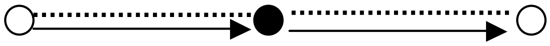
The modality of safe belief $B^+\varphi$ is simply defined as follows: $\boldsymbol{M}, s \mid= B^+\varphi$ iff for all worlds t in the epistemic range of s with $t \geq s$, $\boldsymbol{M}, t \mid= \varphi$. In words, φ is true in all epistemically accessible worlds that are at least as plausible as the current one. [88] ■

This modality is clearly stable under hard information, at least for factual assertions φ which do not change their truth value as the model changes. [89] And indeed, it makes a lot of sense, since it is the obvious *universal base modality [≤]φ for the plausibility ordering*! This modality has been proposed by Boutilier 1994, Halpern 1997 (cf. Shoham & Leyton-Brown 2008), and Baltag & Smets 2006 in epistemology (who credit the idea to Stalnaker), while they also occur prominently in our Chapter 8 on dynamic preference logic.

In what follows, we make safe belief part of the static doxastic language – as a pilot for a richer theory of attitudes in the background. Pictorially, one can think of this as follows:

Example Three degrees of doxastic strength.

Consider this picture, now with the actual world in the middle:



$K\varphi$ describes what we know: φ must be true in all worlds in the epistemic range, less or more plausible than the current one. $B^+\varphi$ describes our safe beliefs in further investigation: φ is true in all worlds from the middle toward the right. Finally, $B\varphi$ describes the most fragile thing: our beliefs as true in all worlds in current topmost position on the right. ■

In addition, safe belief simplifies many things, if only as a technical heuristic device:

Fact The following assertions hold on finite epistemic connected plausibility models:

(a) Safe belief can define its own conditional variant.

(b) Safe belief can define conditional belief.

[88] Note how safe belief uses an *intersection* of two relations: one for epistemic accessibility and one for plausibility. We could also 'decouple' this entanglement, and introduce a similar modality for pure plausibility, but we will ignore these variations here, technically useful as they are.

[89] Note here that new true information will never remove the actual world: our vantage point.

Proof (a) is obvious, since we can conditionalize to $B^+(A \rightarrow \varphi)$ just as with any standard universal modality. (b) uses the observation that on finite connected plausibility models, using also the modality $<B^+>$, which is the existential dual of safe belief:

Claim Conditional belief $B^\psi\varphi$ is equivalent to the iterated modal statement

$$B^+((\psi \wedge \varphi) \rightarrow < B^+>(\psi \wedge \varphi \wedge B^+ (\psi \rightarrow \varphi))).$$

This claim is not for the faint-hearted, but it can be proved with a little puzzling. [90] ■

Safe belief also has some less obvious features. For instance, since its accessibility relation is transitive, it satisfies Positive Introspection, but since that relation is not Euclidean, it fails to satisfy Negative Introspection. The reason is that safe belief mixes purely epistemic information with *procedural information* as discussed briefly in Chapter 3 (cf. also Chapter 11 below). To us, this merely means that, once we admit that agents can have a richer repertoire of doxastic-epistemic attitudes than *K* and *B*, 'omnibus intuitions' concerning axioms to be satisfied are not very helpful in understanding the full semantic picture.

Finally, we turn to dynamics under hard information, i.e., our key recursion axiom:

Theorem The complete logic of belief change under hard information is the one whose principles were stated before, plus the following recursion axiom for safe belief:

$$[!P]\, B^+\varphi \leftrightarrow (P \rightarrow B^+(P \rightarrow [!P]\varphi).$$ [91]

6.4 Radical belief change under soft information

Soft information and plausibility change Belief change as described so far is a 'hybrid': we saw how a 'soft' attitude changes under hard information. The more general scenario would be that an agent is aware of being subject to continuous belief changes, and hence,

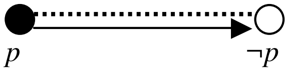
[90] The result generalizes to other models, and the given modal translation is *itself* a good candidate for lifting the maximality account of conditional belief to infinite models, as well as non-connected ones. Alternative versions would use modalities for the strict ordering corresponding to reflexive plausibility $\leq$ to define maximal ψ-worlds directly in the format $\psi \wedge \neg<<>\psi$: cf. Girard 2008.

[91] We leave it to the reader to check that this new axiom for safe belief under hard information automatically derives the one given for conditional belief, if one uses the above modal definition.

that she also takes the incoming signals in a softer manner, without throwing away options forever. But then, public announcement is too strong:

Example No way back.

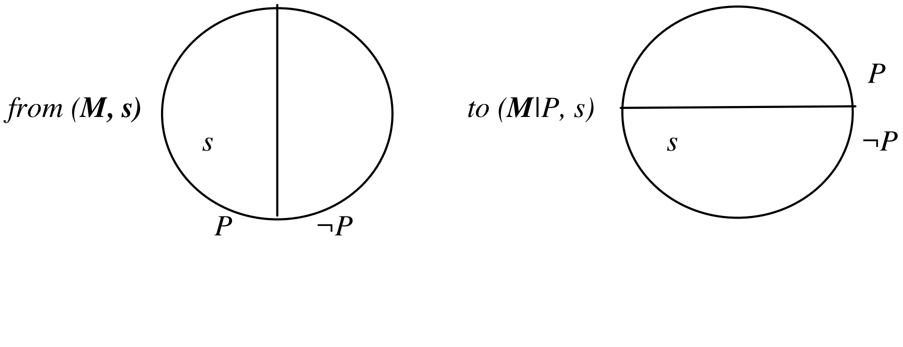
Consider the earlier model where the agent believed that $\neg p$, though p was in fact the case:



Publicly announcing p removes the $\neg p$-world, making later belief revision impossible. ■

What we need instead is a mechanism which just makes the incoming information P more plausible, without burning our ships behind us. An example are the conditional *default rules* $A \Rightarrow B$ in Veltman 1996. Accepting a default rule does not say that all A-worlds must now be B-worlds. It rather says that the counter-examples, i.e., the $A \wedge \neg B$-worlds, are now less plausible until further notice. This 'soft information' does not eliminate worlds, it rather changes their ordering. More precisely, a triggering event which makes us believe that P need only *rearrange worlds* making the most plausible ones P: by 'promotion' rather than elimination of worlds. Thus, on the earlier models $\boldsymbol{M} = (W, \sim_i, \leq_i, V)$, we change the relations $\leq_i$, rather than the domain of worlds W or the epistemic accessibilities $\sim_i$. Many such instructions for plausibility change have long existed in the semantics of belief revision (Grove 1988, Rott 2006) as different *policies* that agents might adopt toward the incoming information, and we will now show how our dynamic logics deal with them. [92]

Radical revision One very strong policy is like a radical social revolution where some underclass P now becomes the upper class. In a picture, we get this reversal:



[92] Alternatively, in formal learning theory (Kelly 1996), these are different *learning strategies*.

Definition Radical, or lexicographic upgrade.

A *lexicographic upgrade* $\Uparrow P$ changes the current ordering ≤ between worlds in $(\boldsymbol{M}, s)$ to a new model $(\boldsymbol{M}\Uparrow P, s)$ as follows: all P-worlds in the current model become better than all $\neg P$-worlds, while, within those two zones, the old plausibility ordering remains. [93] ■

With this definition in place, our earlier methodology applies. As for public announcement, we introduce a corresponding upgrade modality into our dynamic doxastic language:

$$\boldsymbol{M}, s \models [\Uparrow P]\phi \quad \text{iff} \quad \boldsymbol{M}\Uparrow P, s \models \phi$$

Here is a complete account of how agents' beliefs change under soft information, in terms of the key recursion axiom for changes in conditional belief under radical revision:

Theorem The dynamic logic of lexicographic upgrade is axiomatized completely by

(a) any complete axiom system for conditional belief on the static models, plus

(b) the following recursion axioms:

$$\begin{array}{lll}
[\Uparrow P]\, q & \leftrightarrow & q \qquad \text{for all atomic proposition letters } q \\
[\Uparrow P]\, \neg\phi & \leftrightarrow & \neg[\Uparrow P]\phi \\
[\Uparrow P]\, (\phi\wedge\psi) & \leftrightarrow & [\Uparrow P]\phi \wedge [\Uparrow P]\psi \\
[\Uparrow P]\, K\phi & \leftrightarrow & K[\Uparrow P]\phi \\
[\Uparrow P]\, B^{\psi}\phi & \leftrightarrow & (\Diamond(P \wedge [\Uparrow P]\psi) \wedge B^{P \wedge [\Uparrow P]\psi}\, [\Uparrow P]\phi) \\
 & & \vee\, (\neg\Diamond(P \wedge [\Uparrow P]\psi) \wedge B^{[\Uparrow P]\psi}\, [\Uparrow P]\phi)
\end{array}$$ [94]

Proof The first four axioms are simpler than those for *PAL*, since there is no precondition for $\Uparrow P$ as there was for $!P$. The first axiom says that upgrade does not change truth values of atomic facts. The second says that upgrade is a function on models, the third is a general law of modality, and the fourth says no change takes place in epistemic accessibility.

The fifth axiom is the locus where we see the specific change in the plausibility ordering. It looks forbidding, but it is really not hard to grasp. The left-hand side says that, after the P-upgrade, all best ψ-worlds satisfy φ. On the right-hand side, there is a case distinction.

[93] This is just the lexicographic policy for relational belief revision of Rott 2006.

[94] Here, as in Chapter 2, '$\Diamond$' is the dual *existential epistemic modality* $\neg K\neg$.

Case (1): there are accessible P-worlds in the original model $\boldsymbol{M}$ that become ψ after the upgrade. Then lexicographic reordering $\Uparrow P$ makes the best of these worlds in $\boldsymbol{M}$ the best ones over-all in $\boldsymbol{M}\Uparrow P$ to satisfy ψ. Now, in the original model $\boldsymbol{M}$ – viz. its epistemic component visible from the current world s – the worlds of Case 1 are just those satisfying the formula $P \wedge [\Uparrow P]\psi$. Therefore, the formula $B^{P \wedge [\Uparrow P]\psi} [\Uparrow P]\phi$ says that the best among these in $\boldsymbol{M}$ will indeed satisfy φ after the upgrade. And these best worlds are the same as those described earlier, as lexicographic reordering does not change the ordering of worlds inside the P-area. Case (2): no P-worlds in the original $\boldsymbol{M}$ become ψ after upgrade. Then the lexicographic reordering $\Uparrow P$ makes the best worlds satisfying ψ after the upgrade just the same best worlds over-all as before that satisfied $[\Uparrow P]\psi$. Here, the relevant formula $B^{[\Uparrow P]\psi} [\Uparrow P]\phi$ in the reduction axiom says that the best worlds become φ after upgrade.

The rest of the proof is the reduction argument of Chapter 3. More details on this result as well as the others in this chapter are in van Benthem 2007, van Benthem & Liu 2007. ■

The final equivalence describes which conditional beliefs agents have after soft upgrade. This may look daunting, but try to read the principles of some default logics existing today! Also, recall the earlier point that we need to describe how conditional beliefs change, rather than absolute beliefs, to avoid getting trapped in the 'Iteration Problem' of belief revision.

Special cases Moreover, looking at special cases may help. First, consider unconditional beliefs $B\phi$. Conditioning on 'True', the key recursion axiom simplifies to:

$$([\Uparrow P]\, B\phi \;\leftrightarrow\; (<>P \wedge B^{P}[\Uparrow P]\phi) \vee (\neg <>P \wedge B[\Uparrow P]\phi\,)$$

Next, consider the case of *factual propositions*, which do not change their truth values under update or upgrade. In this case, our crucial recursion axiom simplifies to

$$[\Uparrow P]\, B^{\psi}\phi \;\leftrightarrow\; (<>(P \wedge \psi) \wedge B^{P \wedge \psi}\phi) \;\vee (\neg <>(P \wedge \psi) \wedge B^{\psi}\phi$$ [95] ■

Safe belief once more As a final simplification, recall the earlier notion of safe belief, which can define conditional belief. We can also derive the above from the following:

[95] This is the sense in which a 'Ramsey Test' holds for our logic, but we will not pursue this here.

Fact The following recursion axiom is valid for safe belief under radical revision:

$$[\Uparrow P]\, B^{+}\phi \;\leftrightarrow\; (P \wedge B^{+}(P \rightarrow [\Uparrow P]\phi)) \vee (\neg P \wedge B^{+}(\neg P \rightarrow [\Uparrow P]\phi) \wedge K(P \rightarrow [\Uparrow P]\phi)).$$

Before considering other policies, here are a few further issues high-lighted by our system, which have been introduced before at several places.

Static pre-encoding Our compositional reduction says that any statement about effects of hard or soft information is already 'encoded' in the initial model, before any events have taken place. We phrased this before as: 'the epistemic present contains the epistemic future'. We have used this line here to design appropriate static languages. Technically, this involves a new form of closure, beyond the syntactic *relativization* discussed in Chapter 3. We now also need closure under syntactic *substitutions* of defined predicates for old ones.

We conclude this section with two obvious comparisons between our dynamic logic of belief and other approaches: standard *belief revision theory* and *non-monotonic logics*.

A brief comparison with AGM The best-known account of belief revision so far is *AGM* theory (Gaerdenfors 1988, Gaerdenfors & Rott 1995), and the reader may want a word on the connection. In contrast with the above 'concrete' plausibility reordering for which we found a complete logic by *DEL* techniques, *AGM* analyzes belief change without any fixed mechanism, placing abstract postulates on the process. And there are further differences. *AGM* deals with single agents and factual information only, while *DEL* is about interaction between agents, typically including higher-order information about what others believe. And finally, *DEL* analyzes explicit triggers for belief change, from public announcements to complex informational events. By contrast, *AGM* theory takes three abstract operations *+A* ('update'), **A* ('revision'), *–A* ('contraction') whose completeness as a repertoire of actions is left open. Even so, the *AGM postulates* constrain all reasonable belief revision rules, so should they apply to lexicographic upgrade? Actually, we find clear differences. For instance, the *'Success Postulate'* says that all new information comes to be believed. This is indeed a consequence of our axioms for factual propositions, [96] but it fails at once for complex doxastic propositions. The reason is just as with *PAL*: Moore-type examples

[96] 'Success' may then be formally derived from the recursion axiom for atomic statements.

may be true, but they cannot be believed after their announcement. The reason why this did not show in belief revision theory is that its main intuitions have been developed for factual assertions. Things are even more complicated with the sophisticated *AGM* postulates for conjunctive assertions. These mix two issues that we have carefully distinguished earlier: announcing a conjunction of propositions, and announcing two propositions successively.

Instead of comparing things in detail, we just make an observation about *iterated* scenarios. In *PAL*, successive announcements can be compressed by the law

$$[!P][!Q]\varphi \leftrightarrow [!(P \wedge ([!P]Q))]\varphi$$

Is there a similar 'compression law' for relation change and belief revision? Note that two successive steps $\Uparrow P;\ \Uparrow Q$ rearrange the model as follows. First *P*-worlds come on top of the $\neg P$-ones, then we do the same with the *Q*-worlds. The result is the following order pattern:

$$\boxed{PQ \quad \geq \quad \neg PQ \quad \geq \quad P\neg Q \quad \geq \quad \neg P\neg Q}$$

No single upgrade does this, and no iteration law compresses the effect of two revision steps to just one with the same consequences for conditional belief. Why should there be?

Beyond circumscription We conclude with a comparison with another broad paradigm, that of consequence relations. We saw before that belief change after hard information is an alternative take on circumscription and related forms of *non-monotonic consequence*:

$$P_1, \dots, P_k \Rightarrow_{circ\text{-}hard} \varphi \quad \text{iff} \quad [!P_1] \dots [!P_k]\, B\varphi$$

But our logic for soft information even suggests new consequence relations of this kind:

$$P_1, \dots, P_k \Rightarrow_{circ\text{-}soft} \varphi \quad \text{iff} \quad [\Uparrow P_1] \dots [\Uparrow P_k]\, B\varphi$$

These two consequence relations are definitely not the same.

Fact For factual assertions P, Q, (i) $P, Q \Rightarrow_{circ\text{-}hard} P$, (ii) not $P, Q \Rightarrow_{circ\text{-}soft} P$.

Proof (i) Successive hard updates yield subsets of the *P*-worlds. (ii) The last upgrade with *Q* may have demoted all *P*-worlds from their former top positions. ■

Thus, we have an interesting interplay between logical dynamics of belief change and the design of new non-monotonic consequence relations.

Open Problem: What are complete sets of structural rules for these consequence relations?

6.6 Dealing with other revision policies, and general formats

Conservative upgrade Radical revision was our pilot case for belief change, but its specific plausibility change is just one way of taking soft information. For instance, a more conservative policy, aiming for 'just believing' the new proposition, puts not *all P*-worlds on top qua plausibility, but just *the most plausible P-worlds*. 'After the revolution', this policy co-opts just the leaders of the underclass, not all of them – the sage advice that Macchiavelli gave to rulers pondering what to do with the mob outside of their palace.

Definition Conservative plausibility change.
The operation $\uparrow P$ replaces the current ordering relation ≤ in a model $\boldsymbol{M}$ by the following: the best P-worlds come on top, but apart from that, the old ordering remains. ■

Technically, we can view $\uparrow P$ as a special case of radical revision: $\Uparrow(best(P))$, assuming we can define the latter in our static language. But it seems of interest to analyze it per se. In fact, our earlier methodology again produces a matching dynamic logic.

Theorem The dynamic logic of conservative upgrade is axiomatized completely by

(a) a complete axiom system for conditional belief on the static models, and

(b) the following reduction axioms:

$[\uparrow P]\, q \quad \leftrightarrow \quad q$ for all atomic proposition letters q

$[\uparrow P]\, \neg\phi \quad \leftrightarrow \quad \neg[\uparrow P]\phi$

$[\uparrow P]\, (\phi\wedge\psi) \quad \leftrightarrow \quad [\uparrow P]\phi \wedge [\uparrow P]\psi$

$[\uparrow P]\, K\phi \quad \leftrightarrow \quad K[\uparrow P]\phi$

$[\uparrow P]\, B^{\psi}\phi \quad \leftrightarrow \quad (B^{P}\neg[\uparrow P]\psi \wedge B^{[\uparrow P]\psi}\, [\uparrow P]\phi) \vee$

$\qquad (\neg B^{P}\neg[\uparrow P]\psi \wedge B^{P\, \wedge\, [\uparrow P]\psi}\, [\uparrow P]\phi)$

We leave a proof to the reader. Of course, one can also combine this logic with the earlier one, to describe combinations of different sorts of revising behaviour, as in $[\Uparrow][\uparrow]\varphi$.

Policies Many further possible changes are possible in a plausibility ordering in response to an incoming signal. This reflects the host of ‘belief revision policies’ in the literature: Rott 2006 has 27, and counting… Relation change was already proposed as a general theme in van Benthem, van Eijck & Frolova 1993, calling for a dynamification of preference logic. The same is true for defaults, commands (Yamada 2006), and other areas where plausibility or preference can change, as we shall see in Chapter 8. Our approach suggests that we can take any reasonable definition of plausibility change, search for a matching recursion axiom, and then write the complete dynamic logic. But how general is this skill? [97]

Relation transformers in dynamic logic One general viewpoint works by inspection of the style of definition in the above examples. For instance, it is easy the following

Fact Radical upgrade $\Uparrow P$ is definable as a program in propositional dynamic logic.

Proof The format is as follows, with ‘T’ the universal relation between all worlds:

$$\Uparrow P(R) := (?P; T ; ?\neg P) \cup (?P ; R; ?P) \cup (?\neg P ; R; ?\neg P) \qquad \blacksquare$$

Van Benthem & Liu 2007 then introduce the following format.

Definition *PDL*-format for relation transformers.
A definition for a new relation *R* on models is in *PDL-format* if it can be stated in terms of the old relation*, union, composition*, and *tests*. ■

A further example is a ‘suggestion’ *#P* which merely takes out *R*-pairs with ‘¬*P* over *P*’:

$$\#P(R) = (?P; R) \cup (R; ?\neg P)$$

This format generalizes our earlier procedure with recursion axioms considerably:

[97] One might question what this diversity means. Maybe ‘policy’ is the wrong term, as it suggests a persistent habit of an agent over time, like being eager, or stubborn. But our events really describe local responses to particular inputs. Speech act theories have a nice distinction between incoming information per se (what is said) and the *uptake,* the way in which the recipient reacts to them. In that sense, the 'softness' of our scenarios might be in the response, rather than in the signal itself.

Theorem For each relation change defined in *PDL*-format, there is a complete set of recursion axioms which can be derived via an effective procedure.

Proof Here are two examples of computing modalities for the new relation after the model change, using the recursive program axioms of *PDL*. Note how the second calculation uses the existential epistemic modality $\Diamond$ for the occurrence of the universal relation:

$$<\#P(R)><R>\varphi \leftrightarrow <(?P;\, R) \cup (R;\, ?\neg P)>\varphi \leftrightarrow <(?P;\, R)>\varphi \vee <(R;\, ?\neg P)>\varphi$$
$$\leftrightarrow <(?P><R>\varphi \vee <R><?\neg P>\varphi \leftrightarrow (P \wedge <R>\varphi) \vee <R>(\neg P \wedge \varphi).$$

$$<\Uparrow P(R)>\varphi \leftrightarrow <(?P;\, T\, ;\, ?\neg P) \cup (?P\, ;\, R;\, ?P) \cup (?\neg P\, ;\, R;\, ?\neg P)>\varphi$$
$$\leftrightarrow <(?P;\, T\, ;\, ?\neg P)>\varphi \vee <(?P\, ;\, R;\, ?P)>\varphi \vee <(?\neg P\, ;\, R;\, ?\neg P)>\varphi$$
$$\leftrightarrow <?P><T><?\neg P)>\varphi \vee <?P><R><?P)>\varphi \vee <?\neg P><R><?\neg P)>\varphi$$
$$\leftrightarrow (P \wedge \Diamond(\neg P \wedge \varphi)) \vee (P \wedge <R>(P \wedge \varphi)) \vee (\neg P \wedge <R>(\neg P \wedge \varphi)).$$

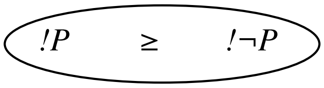
This gives uniformity behind earlier cases. For instance, the latter easily transforms into an axiom for safe belief after radical upgrade $\Uparrow P$, equivalent to the one we gave before. ■

Event models as triggers Another way of achieving generality uses the format of *DEL* with event models, as developed in Chapter 4. In dynamic epistemic logic of this sort, triggers for information change can be much more complex than public announcements, or the few specific policies that we have discussed. While the motivation for this came from partial observation, the technique also applies to receiving signals with different strengths. Here is the idea from Baltag & Smets 2006 (with a precursor in Aucher 2004):

Definition Plausibility event models.

Plausibility event models are event models just as in Chapter 4, but now expanded with an additional plausibility relation over their epistemic equivalence classes. ■

For example, think of radical upgrade $\Uparrow P$ as follows now: we do not throw away worlds, so we need 'signals' *!P* and *!¬P* with the obvious preconditions *P, ¬P*. But we now say that *!P* is more plausible than *!¬P*, relocating the revision policy in the nature of the input:



'One Rule To Rule Them All' Next we need an update rule for products ***M x E***. Here it is, squarely placing the emphasis on the last event observed (a motivation will follow soon):

Definition Priority Update.

Consider an epistemic plausibility model *(**M**, s)* and a plausibility event model *(**E**, e)*. The product model *(**M x E**, (s, e))* is defined entirely as in Chapter 4, with the following new rule for the plausibility relation, with < the strict version of the relation:

$$(s, e) \leq (t, f) \text{ iff } (s \leq t \;\&\; e \leq f) \vee e < f. \qquad \blacksquare$$

It is easy to see that this rule fits our description of radical upgrade. If the newly important predicate *P* induces a preference between worlds, then that takes precedence: otherwise, we go by the old plausibility ordering. More generally, this rule places very heavy weight on the last observation made, or signal received. This may seem strange at first sight, but it is in line with belief revision theory, where receiving just one signal **P* leads me to believe that *P*, even if all of my life so far, I had been receiving strong evidence against *P*. It is also in line with 'Jeffrey Update' in probability, where we impose some new probability for a proposition, while adjusting all other probabilities proportionally (Halpern 2003). [98] [99]

Theorem The dynamic logic of priority update is axiomatizable completely.

Proof As before, it suffices to state the crucial recursion axioms reflecting the above rule. We display just one case, for the relation of safe belief, in existential format:

$$<\boldsymbol{E}, e><\leq>\varphi \;\leftrightarrow\; (PRE_e \wedge (\bigvee_{e \leq f \, in \, \boldsymbol{E}} <\leq><\boldsymbol{E}, f>\varphi \vee \; (\bigvee_{e < f \, in \, \boldsymbol{E}} \Diamond<\boldsymbol{E}, f> \varphi))$$

where ◇ is again the existential epistemic modality. ■

[98] There may be a worry here that this shifts from *DEL's pre-condition* analysis to a forward style of thinking in terms of *post-conditions*: cf. Chapter 3, but we will not pursue this possible objection.

[99] As in Chapter 4, product update with event models generalizes easily to allow for *real world change*, thus taking on board the non-*AGM* style Katsuno-Mendelzon sense of 'temporal update'.

What this does is *shift the locus of description*. Instead of many policies for processing an input signal, each with their own logical axioms, we now put the policy inside the input argument ***E***. Of course, this has some artificial features: for instance, the new models are much more abstract than event models as originally motivated in Chapter 4. Also, even to describe simple policies like the earlier conservative upgrade, the language of these event models will have to be extended, to allow for event preconditions of the form '*most-plausible(P)*'. But the benefit is also clear: belief change now works with just one update rule, and hence the common objection that belief revision theory is 'non-logical' and 'messy' for its non-deterministic character, viz. a proliferation of policies, evaporates.

Digression: abrupt revision versus slow learning An update rule which places this much emphasis on the last signal is of course very special. In Chapter 10, we will do an additional analysis in terms of social choice between 'old and new signals' bringing this out. Indeed, in theories of learning, there are also slower ways of merging old with new information, in the gentler manner of inductive logic, so that one new observation of *P* gets some force, but without immediately overriding all our experience so far. This theme will return in our discussion of a spectrum of probabilistic update rules in Chapter 7, and again with 'score-based rules' for preference dynamics in Chapter 8.

It is not entirely clear how the two given formats: *PDL*-style definitions with computation of new modalities, and event models with Priority Update. [100] But either way, it will be clear that the account of belief change in this chapter is much more general than might have appeared from our two very specific examples of radical and conservative policies.

6.7 Belief revision postulates as modal frame correspondences

Finally, what about the postulational approach to belief revision? *AGM* theory advocates no specific mechanism for relation change, but its postulates constrain the family of options. A corresponding modal style way of thinking is 'dynamic doxastic logic' *DDL* (Segerberg 1995). This abstract framework merely assumes some relation change on the current model: functional, or non-deterministic relational. The main operator looks like this:

[100] The dissertation Liu 2008 has a first discussion plus some formal observations.

Definition Abstract modal logic of model change.

Let $\boldsymbol{M}$ be a model, $[[P]]$ the set of worlds in $\mathbf{M}$ satisfying P, and $\boldsymbol{M}{*}[[P]]$ some new model. For the matching modal operator, we set $\boldsymbol{M}, s \models [{*}P]\varphi$ iff $\boldsymbol{M}{*}[[P]], s \models \varphi$. ■

DDL uses models that resemble Lewis sphere systems for conditional logic, or generalized neighbourhood versions (cf. Girard 2008). The axioms of the minimal modal logic *K* are valid on these models, and on top of that, additional axioms constrain relation changes that correspond to 'bona fide' belief revision policies. In the limit, a particular set of axioms might even determine one particular revision policy. We will show how this ties up with our earlier approach, by reversing the perspective in terms of *frame correspondence*.

Usually, frame correspondences serve to analyze the semantic content of given axioms in a static modal language. But one can just as well take the above functional framework of arbitrary relation changing operations ♥P over models consisting of worlds and a ternary comparison relation $\leq_s xy$. ♥P takes any model $\boldsymbol{M}$ and a set of worlds P in it, [101] and yields a new model $\boldsymbol{M}$♥P with the same set of worlds but some possibly changed relation $\leq_s$. Axioms may then constrain this. In fact, we saw this style of analysis in Chapter 3, when capturing *PAL*-style eliminative update as essentially the only model-changing operator satisfying the recursion axioms for knowledge, plus for the existential modality. Even more abstract spaces of models can be used here to provide the general background for analyzing the content of dynamic axioms, but our setting suffices to make our main points.

Analyzing a few AGM postulates For a start, the postulate of 'Success' says something weak, which holds for both earlier operations $\Uparrow P$ and $\uparrow P$: [102]

[101] Here we have dropped the above double denotation brackets $[[P]]$ for convenience.

[102] A technical clarification. Standard frame correspondences come in the following format. The modal *K4*-axiom $\Box p \rightarrow \Box\Box p$ is true at world s in frame $\boldsymbol{F} = (W, R)$ iff the relation R is transitive at s: i.e., $\boldsymbol{F}, s \models \forall y(Rxy \rightarrow \forall z(Ryz \rightarrow Rxz))$. 'Frame truth' means a formula is true under all valuations on frame $\boldsymbol{F}$ for its proposition letters. Thus, it does not matter whether we use a formula $\Box p \rightarrow \Box\Box p$ or the schema $\Box\varphi \rightarrow \Box\Box\varphi$. Not so for *PAL* and *DEL*, given the earlier difference between plain validity and schematic validity. In the following proofs we use proposition letters.

Fact The formula $[\heartsuit p]Bp$ says that the best worlds in $\boldsymbol{M}\heartsuit p$ are all in p.

This trivial observation needs no proof. But actually, we might demand something much stronger on relation change for belief revision, viz. the best worlds in $\boldsymbol{M}\heartsuit p$ are precisely the best p-worlds in $\boldsymbol{M}$ (*UC*). This, too, can be expressed. But we need a stronger Ramsey-style dynamic formula, involving two different proposition letters p and q:

Fact The formula $B^{p}q \leftrightarrow [\heartsuit p]Bq$ expresses *UC*.

But this preoccupation with the 'Upper Classes' still fails to constrain the total relation change. For that, as emphasized before, we really need to look at the social order in all classes after the Revolution, i.e., at conditional beliefs following relation upgrade.

As a deeper illustration, then, we consider the crucial reduction axiom for $\Uparrow P$, now using proposition letters instead of schematic variables for arbitrary formulas. As these refer to bare sets, we suppress the earlier dynamic modalities $[\Uparrow P]\psi$ which kept track of possible 'transfer effects'. The following shows this determines lexicographic reordering of models completely: a show-case for our correspondence take on postulational belief revision:

Theorem The formula $[\heartsuit p]\, B^{r} q \leftrightarrow (E(p \wedge r) \wedge B^{p \wedge r} q) \vee (\neg E(p \wedge r) \wedge B^{r} q)$
holds in a universe of frames iff the operation $\heartsuit p$ is lexicographic upgrade.

Proof Let $\leq_s xy$ in $\boldsymbol{M}\heartsuit p$. We show that $\leq_s$ is the relation produced by lexicographic upgrade. Let r be the set $\{x, y\}$ and $q = \{x\}$. Then the left-hand side of our axiom is true. There are two cases on the right-hand side. *Case 1*: one of x, y is in p, and hence $p \wedge r = \{x, y\}$ (1.1) or $\{y\}$ (1.2) or $\{x\}$ (1.3). Moreover, $B^{p \wedge r} q$ holds in $\boldsymbol{M}$ at s. If (1.1), we have $\leq_s xy$ in $\boldsymbol{M}$. If (1.2), we must have $y=x$, and again $\leq_s xy$ in $\boldsymbol{M}$. Case (1.3) can only occur when $x \in p$ and $y \notin p$. Thus, all new relational pairs in $\boldsymbol{M}\heartsuit p$ satisfy the description of the lexicographic reordering. *Case 2* is when we have $\neg E(p \wedge r)$ and none of x, y are in p. This can be analyzed analogously, using the truth of the disjunct $B^{r} q$.

Conversely, we show that all pairs satisfying the description of lexicographic upgrade do make it into the new order. Here is one example; the other case is similar. Suppose that $x \in p$

while $y \notin p$. Then $p \wedge r = \{x\}$. Next, set $r = \{x, y\}$ and $q = \{x\}$. Then we have $B(q\, |r)$ for trivial reasons. The left-hand side formula $[\heartsuit p]\ B^{r} q$ is then also true, since our axiom is supposed to hold for any interpretation of the proposition letters q, r – and it tells us that, in the model $\boldsymbol{M}\heartsuit p$, the best worlds in $\{x, y\}$ are in $\{x\}$: i.e., $\leq_s xy$. ■

This generalizes to abstract universes of plausibility models and transitions between them, with second-order quantifiers ranging over sets of worlds inside and across models. [103]

Further *AGM*-postulates are also modal principles to be analyzed through correspondence, with one new twist. In general, they interleave two abstract operations that change models: *update !P* and *upgrade* $\heartsuit P$, leading to mixed principles such as

(a) $[\heartsuit(p \wedge q)]Br \rightarrow [!q][\heartsuit p]Br$

(b) $([\heartsuit p]Eq \wedge [!q][\heartsuit p]Br) \rightarrow [\heartsuit(p \wedge q)]Br$

Such principles constrain simultaneous choice of two abstract model changing operations, for update and for upgrade. We do not pursue this generalization, but we have shown that there is a viable correspondence theory for languages with model-changing modalities.

6.8 Conclusion

This chapter has realized the second stage of our logical analysis of agency, extending the dynamic approach for knowledge to belief. The result is one merged theory of information update and belief revision, which uses standard modal techniques, the 'lingua franca' of our field. Moreover, we can now freely transfer issues and results between the two research areas – provided we see through superficial differences in 'lifestyle' in *AGM* and *DEL.*

6.9 Appendix: further issues and open problems

Variations on the static base logic We assumed that agents have epistemic introspection of their plausibility order. If we drop this simplifying assumption, we need genuine ternary world-dependent plausibility relations, as in conditional logic. What do our *{K, B}*-based systems look like then? Also, the discussion of safe belief suggests another set-up with just

[103] In fact, the above arguments then work uniformly by Sahlqvist substitution techniques.

one primitive plausibility pre-order ≤, where we define knowledge as truth in all worlds, whether less or more plausible. What happens when we switch to the latter scheme?

Common belief We have not analyzed 'common belief' in this chapter, the natural counterpart to the earlier common knowledge. While the definition is standard, using the fixed-point equation $CB_G\varphi \leftrightarrow \bigwedge_{i\in G} B_i (\varphi \wedge CB_G\varphi)$, providing a complete axiom system would call for a combination of relation change with the *PDL* techniques of Section 4.6.

Many agents and merging toward social beliefs Standard belief revision policies describe what a single agent does when confronted with surprising facts. But beliefs tend to change because *other agents* are involved, often even contradicting us, and hence we want a generalization to interactive settings where agents are confronted with information from other sources, which need to be integrated into one new plausibility ordering. In that case, we must analyze events of *belief merge* (Maynard-Reid & Shoham 1998), and more general '*judgment aggregation*' (List & Pettit 2004). Construed either way, we need to see how groups acquire beliefs after upgrade. We will discuss these issues in Chapter 10.

Agent diversity Different policies for belief revision *parametrize* agents, say, into different sort of 'learners'. But more general diversity of agents is a fact of life mentioned before. So far, we have found several ways of parametrizing agents' powers: powers of observation in Chapter 4, powers of inference in Chapter 5, while van Benthem & Liu 2004, Liu 2008 parametrize powers of memory. One open problem is determining how all these methods fit into one coherent realistic picture of logical agents, since their methods look dissimilar.

Temporal perspective *DEL* and *AGM* are in the same boat with respect to further issues. We have already observed in our brief discussion of *protocols* (cf. Chapters 3, 4) that informational processes involve both the temporal past, i.e., the history of what has happened so far, and the temporal future. Likewise, our beliefs about an agent will typically depend on hypotheses about its long-term future behavior. This brings us once more to the realm of *epistemic temporal logics*, which will be investigated in great detail in Chapters 9, 11 connecting up logics of belief change with mathematical learning theory.

'Backward' versus 'forward' in update Recall also a basic contrast from earlier chapters. Like many logics in the temporal tradition, *AGM* is 'forward-looking'. Unlike *DEL*, it derives new states not from events informing us about the current setting, but from goal-oriented commands of the *STIT*-type (Belnap et al. 2001): “see to it that φ comes about” (that is, in the present setting: ‘come to believe that *P*’, ‘join the Believers’). In the latter style, one does not tell the agent exactly how to do this. As long as propositions *P* are factual, and hence time-invariant, this difference does not matter much, and hence *AGM* and *DEL* can see eye to eye. But once they may contain epistemic operators, the future-oriented approach is much harder to formulate, and starts feeling like ‘wishful thinking’: what does it mean to execute a command ‘See to it that this agent does not know that φ’?

We will pursue this contrast in methodology, too, in Chapter 11, as epistemic temporal process trees are a setting where future-oriented commands make sense, since the space of possible updates is already there. [104] This is the ‘Grand Stage’ of a temporal universe already containing all histories that are possible lines of investigation. An update *!P* is then an instruction to make a *minimal* move to some already available future state where one knows that *P*, and likewise for belief change. On such a Grand Stage, our systems of *DEL* and now also *‘DDL’* are a sort of ‘mini process algebra’ of successive model construction.

[104] Interestingly, my own first modal analysis of *AGM* in van Benthem 1989 (a Logic Colloquium lecture from 1987 reacting to Gärdenfors' early work) works with three modalities *[+P], [–P],* and *[*P],* for update, contraction, and revision, over a temporal universe of growing information stages.