Changing for the better : preference dynamics and agent diversity

Liu, F.

Publication date
2008

Citation for published version (APA):
Liu, F. (2008). *Changing for the better : preference dynamics and agent diversity*. ILLC.

Chapter 6

Diversity of Agents and their Interaction

6.1 Diversity inside logical systems

Logical systems seem to prescribe one norm for an "idealized agent". Any discrepancies with actual human behavior are then irrelevant, since the logic is meant to be normative, not descriptive. But logical systems would not be of much appeal if they did not have a plausible link with reality. And this is not just a matter of confronting one ideal norm with one kind of practical behavior. The striking fact is that human and virtual agents are not all the same: actual reasoning takes place in societies of diverse agents.

This diversity shows itself particularly clearly in *epistemic logic*. There have been long debates about the appropriateness of various basic axioms, and they have to do with agents' different powers. In particular, the ubiquitous modal Distribution Axiom has the following epistemic flavor:

6.1.1. Example. Logical omniscience: $K(\varphi \to \psi) \to (K\varphi \to K\psi)$.

Do rational agents always *know the consequences* of what they know? Most philosophers deny this. There have been many attempts at bringing the resulting diversity into the logic as a legitimate feature of agents. Some authors have used "awareness" as a sort of restriction on short-term memory ([FH85]), others have concentrated on the stepwise dynamics of making inferences ([Kon88], [Dun95]). A well-informed up-to-date philosophical summary is found in [Egr04].

The next case for diversity lies in a different power of agents:

6.1.2. Example. Introspection axioms: $K\varphi \to KK\varphi$, $\neg K\varphi \to K\neg K\varphi$.

Do agents *know when they know* (or *do not know*)? Many philosophers doubt this, too. This time, there is a well-established way of incorporating different powers into the logic, using different accessibility relations between possible worlds in Kripke models. Accordingly, we get different modal logics: K, T, $S4$, or $S5$. Each of these modal logics can be thought of as describing one sort of agents. The interesting setting is then one of combinations. E.g., a combined language with two modalities K_1, K_2 describes a two-person society of introspectively different agents! This gives an interestingly different take on current logic combinations ([GS98], [KZ03]): the various ways of forming combined logics, by "fusions" $S5+S4$ or "products" $S5\times S4$, correspond to different assumptions about how the agents *interact* in an abstract sense. Effects may be surprising here. E.g., later on, in our discussion of memory-free agents, we see that knowledge of memory-free agents behaves much like "universal modalities". But in certain modal logic combinations, adding a universal modality drives up complexity, showing how the interplay of more clever and more stupid agents may itself be very complex...

Thus, we have seen how *diversity exists inside standard epistemic logic*, and hence likewise in doxastic logic. The purpose of this chapter is to bring to light some further sources of diversity in existing logics of information. Eventually, we would want to move from complaints about "limitations" and "bounds" to a positive understanding of how societies of diverse agents can perform difficult tasks ([GTtARG99]). In addition to identifying diversity of behavior, this also requires a study of *interactions* between different agents: e.g., how one agent learns the types of the agents she is encountering and makes use of such knowledge in communication. This chapter is structured as follows. Section 6.2 briefly identifies some further parameters of variation for agents beyond the well-known, and somewhat over-worked, concerns of standard epistemic logic. These are: powers of observation, powers of memory, and policies for belief revision. Section 6.3 then looks at dynamic epistemic logics of information update, showing how limited powers of observation for different agents are already accounted for, while we then add some new update systems which also describe varieties of bounded memory. Moving on to correcting beliefs on the basis of new information, Section 6.4 takes a parallel look at dynamic doxastic logics for belief revision, and shows how different revision policies can be dealt with inside one logical system. Section 6.5 is a brief summary of sources of diversity, and a transition to our next topic: that of interaction between different agents. In particular, Section 6.6 discusses several scenarios where different sorts of agent meet, involving identification of types of speaker (liars versus truth-tellers), communication with agents having different introspective powers, and encounters between belief revisers following different policies. We show how these can be dealt with in plausible extensions of dynamic-epistemic and dynamic-doxastic logics. Finally, in Section 6.7, we summarize, and pose some further more ambitious questions.

This chapter is based on existing literature, unpublished work in my Master's Thesis ([Liu04]) plus some new research in the meantime. We will mainly cite the relevant technical results without proof, and put them into a fresh story.

6.2 Sources of diversity

The diversity of logical agents seems to stem from different sources. In what follows, we shall mainly speak about "limitations", even though this is a loaded term suggesting "failure". Of course, the more cheerful reality is that agents have various resources, and they use these positively to perform many difficult tasks, often highly successfully.

Our epistemic axioms point at several "parameters" of variation of agents, and indeed, we already identified two of them:

(a) *inferential/computational power*: making all possible proof steps,

(b) *introspection*: being able to view yourself in "meta-mode".

One further potential parameter relevant to epistemic logic is the "awareness" studied by some authors ([FH85]), which suggests some resource like limited attention span, or short-term memory.

Next, consider modern dynamic logics of information, whose motivation sounds closer to actual cognitive practice. These also turn out to incorporate idealizations that suggest further parametrization for diversity. We start with the case of information update.

Consider the basics of *public announcement logic* (PAL): the event $!\varphi$ in this language means "the fact φ is truthfully announced". PAL considers the epistemic effects these announcement actions bring about. In addition to static epistemic axioms that invite diversity, here is a new relevant issue which merges only in such a dynamic setting. The following principle is crucial to the way PAL analyzes epistemic effects of public assertions, say, in the course of a conversation, or a sequence of experiments with public outcomes:

$$[!\varphi]K_a\psi \leftrightarrow \varphi \rightarrow K_a[!\varphi]\psi \quad \textit{Knowledge Prediction Axiom}$$

But the validity of this axiom presupposes several things, notably *Perfect Observation* and *Perfect Recall* by agents. The event of announcement must be clearly identifiable by all, and moreover, the update induced by the announcement only works well on a unique current information state recording all information received so far. This informal description is made precise in the detailed soundness proof for Knowledge Prediction Axiom in Section 6.3. Also, we will discuss this in the more general framework of "product update" for dynamic epistemic languages ([BMS98]). Thus, we have found two more parameters of diversity in logic. Agents can also differ in their powers of:

(c) *observation*: variety of agents' powers for observing current events,

(d) *memory*: agents may have different memory capacities, e.g., storing only the last k events observed, for some fixed k.

Can one deal with these additional forms of diversity inside the logic? As we will see, dynamic epistemic logic with product update can itself be viewed as a calculus of observational powers. And as to memory, [BL04] has shown how to incorporate this into dynamic epistemic logic (DEL) for memory-free agents, and we will extend their style of analysis below to arbitrary finite memory bounds.

The above four aspects are not the only places where diversity resides. Yet another source lies in *belief revision theory* ([AGM85]). Rational agents also revise their beliefs when incoming information contradicts what they believed so far. This scenario is different from the preceding one, as has been pointed out from the start in this area ([GR95]). Even for agents without limitations of the earlier sorts, there is now another legitimate source of diversity, viz. their 'learning habits' that create diversity:

(e) *revision policies*: varying from conservative to radical revision.

Different agents may react differently towards new information: some behave conservatively and try to keep their original beliefs as much as possible, others may be radical, easily accepting new information without much deliberation. However, these policies are not explicitly part of belief revision theory, except for some later manifestations ([Was00]). We will show in this chapter, following [Liu04], [BL07], how they can be brought explicitly into dynamic logic as well.

This concludes the list of parameters of diversity that we see in current dynamic-epistemic and dynamic-doxastic logics. It is important to mention that acknowledging this diversity inside logical systems is not a concession to the ugliness of reality. It is rather an attempt to get to grips with the most striking aspect of human cognition: despite our differences and limitations, societies of agents like us manage to cooperate in highly successful ways! Logic should not ignore this, but rather model it and help explain it. This chapter is a modest attempt at systematization toward this goal.

6.3 Dynamic logics of information update

Preliminaries in dynamic epistemic logic

To model knowledge change due to incoming information, a powerful current mechanism is dynamic epistemic logic, which has been developed intensively by [Pla89], [Ben96], [BMS98], [Ger99], [DHK07], etc. Since our discussions in this chapter will be based on DEL, we briefly recall its basic ideas and techniques.

6.3.1. Definition. An *epistemic model* is a tuple $\mathcal{M} = (S, \{\sim_a \mid a \in G\}, V)$ [1] such that S is a non-empty set of states, G is a group of agents, each $\sim_a$ is a

[1] We will sloppily write $\mathcal{M} = (S, \sim_a, V)$ when G is clear from the context.

binary epistemic equivalence relation, V is a map assigning to each propositional variable p in Φ a subset $V(p)$ of S.

We also have explicit models for our special citizens, the 'events'. Abstractly speaking, it has a similar structure as the epistemic model. Recall Definition 2.5.4 from Chapter 2.

The dynamic epistemic language is an extension of the one for standard epistemic logic. It is defined as follows

6.3.2. Definition. Let a finite set of propositional variables Φ, a finite set of agents G, and a finite set of events E be given. The *dynamic epistemic language* is defined by

$$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \mid [\mathcal{E}, e]\varphi$$

where $p \in \Phi$, $a \in G$, and $e \in E$.

As usual, $K_a\varphi$ stands for 'agent a knows that φ'. There are also new well-formed formulas of the type $[\mathcal{E}, e]\varphi$, which intuitively mean 'after event e takes place, φ will hold'. Here the $[\mathcal{E}, e]$ act as dynamic modalities. Thus, the expressiveness of the language is expanded in comparison with that of epistemic logic. One could also add the usual program operations of composition, choice, and iteration from propositional dynamic logic to the event vocabulary to deal with more complex situations like two events happening in sequence, choice of two possible events, and events taking place repeatedly. However in the current context, we will only consider a language without these operations.

6.3.3. Definition. Given an epistemic model $\mathcal{M} = (S, \{\sim_a \mid a \in G\}, V)$, we define $\mathcal{M}, s \models \varphi$ (formula φ *is true in* $\mathcal{M}$ at s) by induction on φ:

1. $\mathcal{M}, s \models \top$ always
2. $\mathcal{M}, s \models p$ iff $s \in V(p)$
3. $\mathcal{M}, s \models \neg\varphi$ iff not $\mathcal{M}, s \models \varphi$
4. $\mathcal{M}, s \models \varphi \wedge \psi$ iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}, s \models \psi$
5. $\mathcal{M}, s \models K_a\varphi$ iff for all $t : s \sim_a t$ implies $\mathcal{M}, t \models \varphi$.

In order to define the truth condition for the new formulas of the form $[\mathcal{E}, e]\varphi$, we need to define the product update model, again recall Definition 2.5.5 from Chapter 2. We can then add one more item for the truth definition of the formulas $[\mathcal{E}, e]\varphi$ to the above Definition 6.3.3:

6. $\mathcal{M}$, $s \models [\mathcal{E}, e]\varphi$ iff $\mathcal{M}, s \models PRE(e)$ implies $\mathcal{M}\times\mathcal{E}, (s, e) \models \varphi$.

Next, so called *reduction axioms* in DEL play an important role in encoding the epistemic changes. In particular, the following principle describes knowledge change of agents following some observed event in terms of what they knew before that event takes place:

$$[\mathcal{E}, e]K_a\varphi \leftrightarrow PRE(e) \rightarrow \bigwedge_{f\in\mathcal{E}}\{K_a[\mathcal{E}, f]\varphi : e \sim_a f\}.$$

Intuitively, after an event e takes place the agent a knows φ, is equivalent to saying that if the event e can take place, a knows beforehand that after e (or any other event f which a can not distinguish from e) happens φ will hold. Such a principle is of importance in that it allows us to relate our knowledge after an action takes place to our knowledge beforehand, which plays a crucial role in communication and general interaction.

This concludes our brief review of dynamic epistemic logic. We are ready to move to more complex situations where different agents live and interact. Public announcement logic is the simplest logic which is relevant here, as it describes agents who communicate via public assertions. This is the special case of DEL in the sense that the event model contains just one single event. The precondition of $!\varphi$ boils down to the fact that φ is true, as we will see in the formulas in the next section. In this chapter, for easy understanding, we use simple variants of PAL to motivate our claims, though we also consider a few scenarios using full-fledged DEL with a general mechanism of product update.

Public announcement, observation, and memory

First, we recall the complete axiom system for public announcement.

6.3.4. Theorem. ([*Pla89*][*Ger99*]). *PAL is axiomatized completely by the usual laws of epistemic logic plus the following reduction axioms:*

$(!p)$. $[!\varphi]p \leftrightarrow \varphi \rightarrow p$ *for atomic facts* p

$(!\neg)$. $[!\varphi]\neg\psi \leftrightarrow \varphi \rightarrow \neg[!\varphi]\psi$

$(!\wedge)$. $[!\varphi](\psi \wedge \chi) \leftrightarrow [!\varphi]\psi \wedge [!\varphi]\chi$

$(!K)$. $[!\varphi]K_a\psi \leftrightarrow \varphi \rightarrow K_a[!\varphi]\psi$.

Next, to introduce variety in *observation*, we need to assume a set of possible announcements $!\varphi$, $!\psi$, ... where an agent a need not be able to distinguish all of them. This uncertainty can be modelled by a simple event model with equivalence relation $\sim_a$ between statements which a cannot distinguish. The following example illustrates the difference in agents' powers of observation:

6.3.5. Example. Two agents a and b are traveling in Amsterdam and they want to visit the Van Gogh Museum. But they do not know whether Tram Line 5 goes there. A policeman said 'Tram Line 5 goes to the Van Gogh Museum'. a heard it, but b did not, as she was attracted by a Street musician who was playing her favorite song. So a learned something new, but b did not. Taking p to denote 'Tram Line 5 goes to the Van Gogh Museum', the state model and event model are depicted as follows:

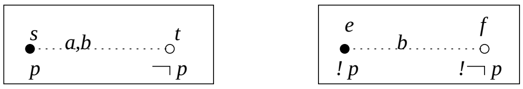

Figure 6.1: State model and event model

The dotted lines express the epistemic uncertainties. The black nodes stand for the actual world and the actual event. The update leads to the following model:

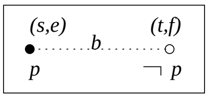

Figure 6.2: b is still uncertain

Note that at this stage a can distinguish between the two possible worlds, but b is still uncertain. There is diversity in observation!

The following principle – a special case of the above general DEL reduction axiom – then describes what agents know on the basis of partial observation:

6.3.6. Fact. The following reduction axiom is valid for agents with limited observation power:

$$[!\varphi]K_a\chi \leftrightarrow (\varphi \rightarrow \bigwedge_{!\psi \sim_a !\varphi} K_a[!\psi]\chi)$$

But there is another natural source of diversity, not dealt with by either PAL or DEL. As we have seen in the previous section, *Perfect Recall* assumes that agents can remember all the events that have happened so far. But in reality agents usually have bounded memory, and they can only remember a fixed number of previous events. It is much harder in PAL to model memory difference because the world elimination update procedure shifts agents to ever more informed states. To show the difficulty, consider the following example concerning *memory-free* agents which only acknowledge distinctions made by the last announcement, having no record of things further back in their past:

6.3.7. EXAMPLE. Memory-free agent a is uncertain about p at first. Then p is announced, and afterwards, an "idle" action Id takes place. Then a should not know p any more since she does not remember anything. But here is what our standard update would do:

According to Definition 2.5.5, the model changes in the following way:

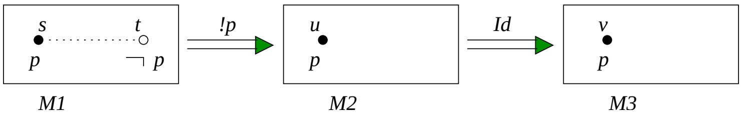

Figure 6.3: Memory-free agent remembers!

There are two possible worlds in the original model $\mathcal{M}_1$, the agent a is uncertain about p. After p is announced, we get $\mathcal{M}_2$. Since p does not hold at the world t, the action $!p$ only executes successfully at the world s, so we have only one world u in the model. Intuitively, after the announcement of p, agent a should now know that p, and indeed this holds in $\mathcal{M}_2$. Next, the Id action happens, which executes successfully everywhere. We get $\mathcal{M}_3$, abbreviating (u, Id) as v. Intuitively, once the action Id has been performed, the memory-free agent a should no longer know whether p, because she already forgot what had happened one step ago, and she should be uncertain again whether p. But in our model sequence, the agent a *knows* p. This is counter-intuitive!

Here is the reason. Standard product update eliminates possible worlds. Therefore, it is impossible to retrieve uncertainty links between worlds that have disappeared. There are several ways of amending this, and two proposals will be presented in detail later in this section. For the moment, we sketch one simple option suggested by [BL04]. First, we need to reformulate PAL update as in [BL07] to never eliminate worlds. The idea is to let announcements $!\varphi$ cut all links between φ-worlds and $\neg\varphi$-worlds, but otherwise, keep all worlds in. In this semantic perspective, the resulting "unreachabilities" between worlds represent the information that agents have so far. One way of describing a memory-restricted agent is then as having forgotten part or all of these "link removals". In the most extreme case, a memory-free agent will only consider distinctions caused by the last announcement – while reinstating all indistinguishability links that had been cut before. (Thus, longer sequences of announcements make no sense for such an agent: it is the last thing said which counts.) In particular, in this update scenario, worlds may also become indistinguishable again: a direct modelling of 'forgetting'. Forgetful agents like this do not satisfy the earlier reduction axiom $(!K)$, as is shown in the following example.

6.3.8. EXAMPLE. Consider the two model changes depicted in Figure 6.4.

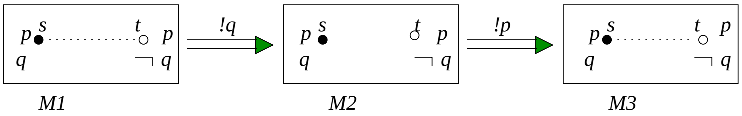

Figure 6.4: Reduction axiom fails

There are two possible worlds, s and t in $\mathcal{M}_1$, p and q hold at s, p and $\neg q$ hold at t. After q is announced, we get a new model $\mathcal{M}_2$, in which there is no uncertainty link between s and t. Then we have $(\mathcal{M}_2, s) \models p \to K_a(p \to q)$, i.e. $(\mathcal{M}_2, s) \models p \to K_a[!p]q$. After that, p is announced, and we have $\mathcal{M}_3 \not\models K_a q$, since the agent forgot $!q$ already. We look back at $\mathcal{M}_2$: $(\mathcal{M}_2, s) \not\models [!p]K_a q$. The reduction axiom does not hold!

With these examples in mind, what is the dynamic epistemic logic of forgetful agents? We will merely discuss a few issues. [BL04] gives the following modified reduction axiom, which trades in a knowledge operator after a dynamic modality for a universal modality $U\varphi$: 'φ is true in all worlds, accessible or not':

$$[\mathcal{E}, e]K_a\varphi \leftrightarrow PRE(e) \to \bigwedge_{e \sim_a f \in \mathcal{E}} U[\mathcal{E}, f]\varphi.$$

This is based on their version of product update which models agents who forget everything except the last event observed by changing the product update rule to this stipulation:

$$(s, e) \sim'_a (t, f) \quad \text{iff} \quad e \sim_a f.$$

Incidentally, to make this work technically, the system also needs a reduction axiom for the universal modality, and it reads as follows:

$$[\mathcal{E}, e]U\varphi \leftrightarrow PRE(e) \to \bigwedge_{e \sim_a f \in \mathcal{E}} U[\mathcal{E}, f]\varphi.$$

Transposed to just the current setting of public announcements (i.e., event models with one publicly observable event), this yields the following principle for forgetful agents:

$$[!\varphi]K_a\psi \leftrightarrow \varphi \to U[!\varphi]\psi.$$

These principles show that it is quite possible to write dynamic-epistemic axioms for agents with bounded memory, in the same style as before. Next, as in [BL07], take the link-cutting variant of public announcements of φ. This amounts to using event models with two events $!\varphi$ and $!\neg\varphi$ which are distinguishable for all agents. Again, the reduction law for forgetful agents follows in a simple manner.

Nevertheless, modeling memory in dynamic epistemic logics raises additional issues, of which we merely mention one. Notice that the preceding K/U equivalence completely obliterates the accessibility structure of the epistemic model that was modified by the last announcement. E.g., the forgetful agent will know that fact q holds after a public announcement of p iff (assuming that p holds) every p-world (whether accessible or not) was a q-world. This may be considered a drawback of the above approach. There appears to be an intuitive difference between (a) forgetting what events took place and (b) what initial situation one started from. An intuitive alternative, suggested by the preceding examples, might let the agent remember the initial model. Here is an illustration of the difference between the two perspectives.

6.3.9. Example. Let the starting model be the following (Figure 6.5), where one world has already become inaccessible (but it might still be accessible via epistemic links for other agents):

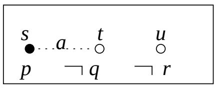

Figure 6.5: Initial model $\mathcal{M}_1$

Announcing $\neg r$ by public link cutting will leave this intact, we get the same picture with actual world s. Announcing $\neg q$ by public link cutting in the radical manner then would give us the following, as shown in Figure 6.6.

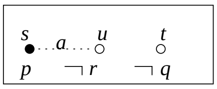

Figure 6.6: Announcing $\neg q$

But if the agent is supposed to remember the initial model, the outcome should be one where she knows that p is the case, see Figure 6.7.

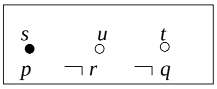

Figure 6.7: If the agent remembers...

Interestingly, implementing the latter less radical view of defective memory means that we have to keep track of *the initial model* $\mathcal{M}_1$, through long sequences of announcements. The reason is that there need not be enough information in $\mathcal{M}_1$'s successive modifications through updates to retrieve its original structure uniquely. Thus, while the behavior of agents with perfect memory may be described by just keeping track of the current epistemic model with all updates performed, the behavior of forgetful agents may require keeping track of a longer history. This may sound paradoxical, but the point is that the latter book-keeping is to be done by the *modeler*, rather than the agent.

We will not formulate reduction axioms for our alternative version of bounded memory here. (Cf. the digression on epistemic temporal logic later in this section for some hints). Even at this somewhat inconclusive stage, however, we see that endowing agents with bounded memory can be achieved in principle.

Our overall conclusion is this: "Logic of public announcement" is actually a family of dynamic epistemic systems, with different update rules depending on the memory type of the agents, and correspondingly, different reduction axioms and reasoning styles.

Adding memory to product update

The previous section shows that the reduction axiom for knowledge under product update fails for memory-free agents. In this section we are going to propose a correct update rule for agents who have a bounded memory for the last observed events. By a *k-memory* agent, we mean an agent that remembers only the last k events before the most recent one. A 0-memory or memory-free agent does not recall anything; a 1-memory agent knows only what she learned from the last two actions, and so on. Modeling this diversity requires some care, witness the Example 6.3.7. As we mentioned, the difficulty there is that eliminating worlds is a form of hard-wired memory: worlds that have been removed do not come back, so one is 'forced to know'. To get this right in a more sensitive manner, we now present two proposals for product update with general memory-free agents. The first source for this is as follows:

6.3.10. Definition. ([Sny04]) Let an epistemic model $\mathcal{M} = (S, \sim_a, V)$ and an event model $\mathcal{E} = (E, \sim_a, PRE)$ be given. The *product update for memory-free agents* is $\mathcal{M} \times \mathcal{E} = (S \otimes E, \sim'_a, V')$ with:

(i) $S \otimes E = \{(s, e) : (s, e) \in S \times E\}$.

(ii) $(s, e) \sim'_a (t, f)$ iff ($\mathcal{M}$, s $\models PRE(e)$ iff $\mathcal{M}, t \models PRE(f)$) and $e \sim_a f$.

(iii) $V'(p) = \{(s, e) \in S \otimes E\colon\ s \in V(p)\}$.

Compared with the standard product update, item (i) in the above definition leaves out the precondition restriction. This keeps all worlds around. Item (ii) then defines the uncertainty relation on all worlds ('active', or not) in the new models. (iii) remains the same, and we will ignore this valuation clause henceforth. To understand this new definition, we look at the example again, now updating models according to the new definition, see Figure 6.8.

This is like Example 6.3.7 – but now, the original state model remains the same. According to Definition 6.3.10, we obtain a different model $\mathcal{M}_2$, abbreviating $(s, !p)$ as u and writing t as v. There is no uncertainty link between them. So

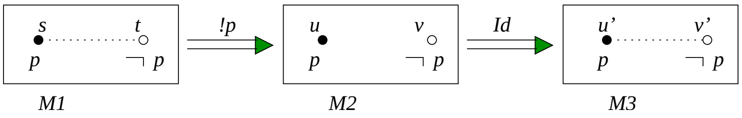

Figure 6.8: How memory-free agents update

the agent a knows that p in $\mathcal{M}_2$. Now the 'idle' identity event Id happens, and we get a new state model $\mathcal{M}_3$, abbreviating (u, Id) as u' and (v, Id) as v'. The agent a is uncertain whether p. This is what we expect for a 0-memory agent. [Sny04] also extended this proposal to the k-memory case.

Here, however, we also put propose an alternative for modelling forgetting, which seems closer to the workings of an actual memory store for agents. We introduce an auxiliary *copy action* $!C$ which always takes an old possible world into the new model with its reflexivity relation. Essentially it puts those worlds which were previously deleted into a stack, and makes sure agents can always retrieve them when needed.

6.3.11. Definition. ([Liu04]) Let an epistemic model $\mathcal{M} = (S, \sim_a, V)$ and an event model $\mathcal{E} = (E, \sim_a, PRE)$ be given. The *product update for memory-free agents* is $\mathcal{M} \times \mathcal{E} = (S \otimes E, \sim'_a, V')$ with:

(i) $S \otimes E = \{(s, e) \in S \times E\colon \mathcal{M}, s \models PRE(e)\}$.

(ii) For $e, f \neq !C$, $(s, e) \sim'_a (t, f)$ iff $e \sim_a f$.

To see how this new proposal works, we go back to the above example, but now update with an additional copy action:

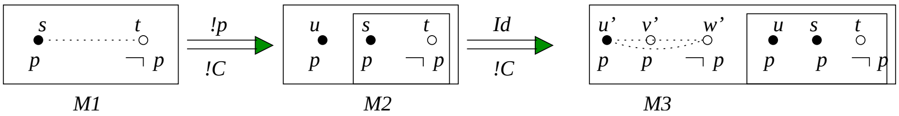

Figure 6.9: Update with copy actions

From the original model, by Definition 6.3.11, we get model $\mathcal{M}_2$, with a new state $(s, !p)$ abbreviated as u and two copied state s and t. The agent a then knows that p. To distinguish the new state and copied state, we put those copied ones in a rectangular box. After the Id action, similarly, we obtain the new model $\mathcal{M}_3$ with new states (u, Id) abbreviated as u', (s, Id) abbreviated as v', and (t, Id) abbreviated as w'. Again, $\mathcal{M}_3$ contains states that are copied from the previous model u, s and t. Again, the agent a is uncertain whether p. This idea is similar to the usual design of operation systems ([SGG03]), where the working memory does the jobs while carrying a stack of old information to be visited when necessary. [Liu04] has a more restrictive variant of the above definition copying worlds only when necessary. This makes the above models less over-loaded.

Extending this approach, we get the following generalized update rule:

6.3.12. Definition. ([Liu04]) Let $\mathcal{M}$ be an epistemic model, $\mathcal{E}_{-k}$ be the k-th event model before the most recent one $\mathcal{E}$. The *product update for k-memory agents* is $\mathcal{M} \times \mathcal{E}_{-k} \times \cdots \times \mathcal{E}_{-1} \times \mathcal{E} = (S \otimes E_{-k} \otimes \cdots \otimes E_{-1} \otimes E, \sim'_a, V')$ with:

(i) $S \otimes E_{-k} \otimes \cdots \otimes E_{-1} \otimes E = \{(s, e_{-k}, \ldots, e_{-1}, e) \in S \times E_{-k} \times \cdots \times E_{-1}$: $\mathcal{M} \otimes \mathcal{E}_{-k} \otimes \cdots \otimes \mathcal{E}_{-1}, (s, e_{-k}, \ldots, e_{-1}) \models PRE(e)\}$.

(ii) For $e_{-k}, \ldots, e_{-1}, e, f_{-k}, \ldots, f_{-1}, f \neq C!$,
$(s, e_{-k}, \ldots, e_{-1}, e) \sim'_a (t, f_{-k}, \ldots, f_{-1}, f)$ iff $e_{-k} \sim_a f_{-k}, \ldots, e_{-1} \sim_a f_{-1}$ and $e \sim_a f$.

Given this update rule, it is straightforward to find a complete dynamic logic in the earlier DEL format, but now for k-memory agents. Here we only consider the case in which $k = 1$, the uncertainty relation in the updated model is the above definition becomes:

For $e_{-1}, e, f_{-1}, f \neq C!, (s, e_{-1}, e) \sim'_a (t, f_{-1}, f)$ iff $e_{-1} \sim_a f_{-1} \& \, e \sim_a f$.

This is to say that a 1-memory agent cannot distinguish between two states in the new updated model, if and only if she cannot distinguish the two events that just took place, and neither the two events that had happened before. The reduction axiom for 1-memory agent is given in the following:

$$[\mathcal{E}, e_{-1}, e]K_a\varphi \leftrightarrow (PRE(e_{-1}) \wedge PRE(e) \rightarrow \bigwedge_{f_{-1}, f \in \mathcal{E}} \{K_a[\mathcal{E}, f_{-1}, f]\varphi : e_{-1} \sim_a f_{-1} \& \, e \sim_a f\}),$$

where e_{-1}, e, f_{-1}, f are not copy actions. Note that we have put two events that are relevant to 1-memory agents into the formula. Since copy actions function independently, we get a reduction axiom that is similar to the one we have for agents with perfect recall.

Of course, this is only the beginning of an array of further questions. In particular, we would like to have a more structured account of memory, as in computer science where we update data or knowledge bases. Update mechanisms are more refined there, referring to memory structure with actions such as information replacement ([Liu04]), where the agent would have a priority order in her database, so that she would know which old information should go to make room for the new. This is one instance of a more "constructive" syntactic approach to update, complementary to our abstract one in terms of model manipulation. Whether our current semantic method or a syntactic one works better for finding agents' parameters of diversity is a question worth investigating.

Digression: Temporal Logic of Forgetful Agents
An alternative, and in some ways more concrete semantic framework for agents with memory bounds are branching tree models for epistemic-temporal logic ([BL04], [BP06]). Nodes in these models are finite sequences of events starting from the root of the tree, and epistemic indistinguishability relations between nodes model what agents have and have not been able to observe. In this setting, the epistemic accessibility relation for a forgetful agent recording just the last event simply becomes this:

$$X \sim_a Y \text{ iff } last(X) = last(Y), \text{ where } last(Z) \text{ is the last event in } Z.$$

As pictured in the following,

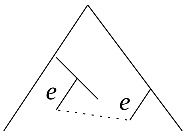

Figure 6.10: Epistemic relations in event trees

Now, the earlier dynamic epistemic reduction axioms become epistemic temporal principles, with indexed modalities $[e], \langle e\rangle$ have their obvious meaning referring to extensions $X^\frown e$ of the current node X. E.g., forgetful agents satisfy the following equivalence:

$$[e]K_a\varphi \leftrightarrow \langle e\rangle\top \to U[e]\varphi.$$

Note again how this trades an epistemic knowledge modality for a universal modality, as in the earlier examples in the previous subsection. The reason is that any node $X^\frown e$ in the temporal tree is epistemically related to any other node $Y^\frown e$. It is now straightforward to find similar principles for the knowledge of agents whose memory retains the last k observed events, as described above.

The total effect of these reduction axioms is as follows. Knowledge modalities are traded in for modal-temporal ones, as the accessibility relation is temporally definable in the model, and hence the epistemic-temporal language reduces to a purely temporal one. [BP06] use this reduction to show that the logic of memory bounded agents is computationally simpler than that of agents with perfect recall. This ends our digression.

This section has identified two new parameters for dynamic updating agents: powers of *observation* and powers of *memory*. *DEL* as it stands already provides a way of modelling the former, while we have shown how it can also be modified to accommodate agents with bounded memory. We consider these two additional phenomena at least as important from an epistemological viewpoint as the usual themes of inferential power and introspection, generated by the earlier static phase of logical theorizing. Of course, as we have noted already, there is no need now to assume that all agents have the same powers. Indeed, our systems can describe the interplay of bounded and idealized agents, including ways in which one might exploit the other.

In the following section, we move to an extension of DEL treating one further crucial aspect of agents' cognitive behavior, when 'things get rough'.

6.4 Diversity in dynamic logics of belief change

Information flow and action based upon it is not always a matter of just smooth update. Another striking phenomenon is the way agents correct themselves when encountering evidence which contradicts their beliefs so far. *Belief revision theory* describes what happens when an agent is confronted with new information which conflicts her earlier beliefs. It has long been acknowledged that there is not one single logical rule for doing this. Indeed, different policies toward revising beliefs, from more 'radical' to more 'conservative' all fall within the compass of the famous AGM postulates.

In this chapter, however, we take another approach inspired by dynamic epistemic logic. First, on the static side, we follow the common idea that beliefs are modelled by so-called *plausibility relations* between worlds, making some epistemically accessible worlds more plausible than others. Agents believe what is true in the most plausible worlds – and the same thinking may also be used to define their conditional beliefs. In this setting, one can then view belief revision on the analogy of the preceding update paradigm, viz. as a mechanism of *change in plausibility relations*. To see this, here is a concrete example of how this can be implemented technically.

Belief revision as changing plausibility relations

One common policy for belief revision works as follows:

6.4.1. Example. ([Ben07a]) ($\Uparrow$) Radical revision
$\Uparrow P$ is an instruction for replacing the current ordering relation $\leq$ between worlds by the following: all P-worlds become better than all $\neg P$-worlds, and within those two zones, the old ordering remains.

Note that the $\neg P$-worlds are not eliminated here: they move downward in plausibility. This reflects the fact that we may change our mind once more on the basis of further information. $\Uparrow P$ is one famous policy for belief revision, corresponding to an 'eager response', or a 'radical revolution', or 'high trust' in the source of the information. But there are many other policies in the literature. Another famous one would just place the best P-worlds on top, leaving the further order unchanged. A more general description of such different policies can be given as definable ways of changing a current plausibility relation ([BL07], [Rot06]). Once we have such a definition for a policy of plausibility change, the corresponding dynamic logic for belief revision can be axiomatized completely in DEL style. Here is the result for the policy of radical revision:

6.4.2. Theorem. ([*Ben07a*]) *The dynamic logic for radical revision* ($\Uparrow$) *is axiomatized completely by an axiom system KD45 on the static models, plus the following reduction axioms*

($\Uparrow p$). $[\Uparrow \varphi]p \leftrightarrow p$

($\Uparrow \neg$). $[\Uparrow \varphi]\neg\psi \leftrightarrow \neg[\Uparrow \varphi]\psi$

($\Uparrow \wedge$). $[\Uparrow \varphi](\psi \wedge \chi) \leftrightarrow [\Uparrow \varphi]\psi \wedge [\Uparrow \varphi]\chi$

($\Uparrow B$). $[\Uparrow\varphi]B\psi \leftrightarrow (E\varphi \wedge B([\Uparrow\varphi]\psi|\varphi)) \vee (\neg E\varphi \wedge B[\Uparrow\varphi]\psi)$

In the last axiom, E is the existential modality, dual to the earlier universal modality U. The symbol | denotes a conditional belief, and it means: 'given that'. Van Benthem's full system also has complete reduction axioms for conditional beliefs, thereby solving the notorious 'Iteration Problem' of AGM theory. This reduction axiom for the new beliefs shows precisely the doxastic effects of the chosen policy.

In the same style, one can also axiomatize other belief revision policies. For instance,'conservative revision' may be defined as follows: $\uparrow\varphi$ replaces the current ordering relation by the following: *the best φ-worlds come on top, but apart from that, the old ordering remains.* [Ben07a] presents a complete set of reduction axioms for this second policy as well. When put together, the result is a dynamic logic of belief revision which describes interactions between agents with different policies, using operator combinations such as, say, $[\Uparrow\varphi][\uparrow\psi]\chi$, which says that after a radical revision with φ followed by a conservative revision with ψ, the proposition χ holds.

All this is still qualitative. But the earlier product update mechanisms also admit of a more refined quantitative version, describing agents' attitudes in a more detailed numerical manner, and allowing for further polices of changing these fine-grained beliefs. In the next subsection, we will briefly show how.

Belief revision as changing plausibility values

Following [Spo88], a κ-ranking function was introduced in [Auc03] to extend DEL with numerical beliefs. A κ-ranking function maps a given set S of possible worlds into the class of numbers up to some maximum Max. The numbers can be thought of as denoting degree of surprise. 0 denotes 'unsurprising', 1 denotes 'somewhat surprising', etc. κ represents a plausibility grading of the possible worlds, in other words, degree of beliefs.

6.4.3. Definition. A *doxastic epistemic model* is a tuple $\mathcal{M} = (S, \sim_a, \kappa_a, V)$, where S, $\sim_a$ and V are defined as usual, and the plausibility function κ_a ranging from 0 to some upper limit Max is defined on all worlds.

6.4.4. Definition. A *doxastic epistemic event model* is a tuple $\mathcal{E} = (E, \sim_a, \kappa^*_a, PRE)$, with E, $\sim_a$ and PRE defined as usual, κ^*_a ranges from 0 to Max, defined on all events.

The κ^*_a-value describes the agent's detailed view on which event is taking place. With plausibilities assigned to states and events, 'graded beliefs' will change via a suitable rule for product update. Here is the quantitative key proposal in [Auc03], the first of its kind in the DEL-style literature:

$$\kappa'_a(s,e) = Cut_{Max}(\kappa_a(s) + \kappa^*_a(e) - \kappa^s_a(\varphi)),$$

where $\varphi = PRE(e)$, $\kappa^s_a(\varphi) = min\{\kappa_a(t) : t \in V(\varphi) \text{ and } t \sim_a s\}$, and

$$Cut_{Max}(x) = \begin{cases} x & \text{if } 0 \leq x \leq Max \\ Max & \text{if } x > Max. \end{cases}$$

While this system looks formidable, a simple more perspicuous version exists. It uses an epistemic-doxastic language with propositional constants to describe the plausibility change ([Liu04]):

6.4.5. Definition. The *epistemic-doxastic language* is defined as

$$\varphi ::= \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \mid q^\delta_a$$

where $p \in \Phi$, a set of propositions, $a \in G$, a set of agents, and δ is a κ-value in $\mathbb{N}$, q^δ_a are a special type of propositional constants.

The interpretation is as usual, but now with the following simple truth condition for the additional propositional constants:

$$(\mathcal{M}, s) \models q^\delta_a \quad \text{iff} \quad \kappa_a(s) \leq \delta.$$

The numerical update mechanism can now be defined quite simply by merely specifying the new κ-value in the product model $\mathcal{M} \times \mathcal{E}$. To keep our discussion simple, we use just the following stipulation:

6.4.6. Definition. (bare addition rule). The new plausibilities for pair-worlds (s, e) in product models are defined by the following rule:

$\kappa'_a(s,e) = \kappa_a(s) + \kappa^*_a(e)$.

In this setting, reduction axioms assume a particularly simple form:

6.4.7. Theorem. ([*Liu04*]) *The complete dynamic logic of plausibility belief revision consists of the key reduction axioms in Theorem 6.3.4 plus the new:*

$$[!\varphi]q_a^\delta \leftrightarrow q_a^{\delta-\kappa_a(!\varphi)}.$$

More generally, different update functions will account for different numerical revision policies. If such an update rule is simply expressible, we can get a complete dynamic logic for it in the style of the preceding result, though mere subtraction may not work anymore.

Additional power of description is provided by yet another device, viz. numerical parameters weighing the contributions of various factors. To illustrate this additional diversity of behavior for agents, we now present an update rule which incorporates further 'degrees of freedom':

6.4.8. Definition. ([Liu04]) Let agent a assign weight λ to world s, and weight μ to the event e. The *plausibility of the new world* (s,e) is calculated by the parametrized rule

$$\kappa'_a(s,e) = \frac{1}{\lambda+\mu}(\lambda\kappa_a(s) + \mu\kappa^*_a(e)) \quad (\natural).$$

Intuitively, κ gives a degree of belief. The two parameters λ and μ express the importance of the state information, and that of the action information, respectively. Their variations then describe a range of various agents. For instance, when μ=0, we get *highly conservative agents*, and the ($\natural$) rule turns into $\kappa'_a(s,e) = \kappa_a(s)$. This means that the agent does not consider the effect of the last-observed event at all. Of course, some normalization is needed here to make sure that the new value is still in $\mathbb{N}$ (cf. [Liu06b]). Similarly, when λ=0, the agents are *highly radical*, and $\kappa'_a(s,e) = \kappa^*_a(e)$. When $\lambda = \mu$, we get *'Middle of the Road agents'* who let plausibility of states and actions play an equally important role in determining the plausibility of the new state. We obtain *conservative agents* when $\lambda > \mu$ and *radical agents* when $\mu > \lambda$. In this manner, we have distinguished five types of agents in dynamic logic. For an even more general view of agents' behavior towards incoming information, see [Liu06b]. Summing up, we may regard our numerical update rule as a refinement of the qualitative dynamic logics for belief change in the previous subsection (cf. [Ben07a] and [BL07]).

6.4.9. Remark. Another relevant comparison is with the probabilistic update semantics proposed in [BGK06]. There the system computes probability values for pairs (s,e) using weighted products of prior world probabilities, occurrence probabilities for the type of event occurring, and observation probabilities describing agents' access to it. We defer a more detailed comparison of our views on agents' processing diversity with qualitative and probabilistic update logics to another occasion.

Some further observations

Our treatment of belief revision provides a simple format of plausibility change, where different policies show naturally in the update rules for either plausibility relations or value constants, and their matching reduction axioms in the dynamic doxastic logic. Moreover, our treatment also goes beyond the standard AGM paradigm, in that more complex event models allow agents to doubt the current information in various ways. Here are a few further issues that come up in this setting, some conceptual, some technical.

First, doubting the current information might also make sense for PAL and DEL scenarios even without belief revision involved. It is easy to achieve this by simply adding further events to an event model, providing, say, a public announcement $!\varphi$ with a counterpart $!\neg\varphi$ with some plausibility value reflecting the strength of the "dissenting voice". Likewise, policies with weights for various factors in update make much sense in recently proposed dynamic logics of probabilistic update (cf. [Auc05], [BGK06]).

Incidentally, this DEL approach via modified event models for different policies may also suggest that we can *relocate* policies from "modified update rules" to "modified event models" with a standard update rule. This has to do with an important more general issue: are we describing single events of update or revision 'locally' without further assumptions about the long-term behavior of the agents involved, or are we witnessing different more 'global' types of agent at work? In the former case, the diversity is in the response, rather than the type. We must leave this issue, and a comparison between the pros and cons of the two stances to another occasion.

Finally, connecting Sections 6.3 and 6.4, revision policies and memory restrictions may not be that disjoint after all. Technically speaking, the update behavior of highly radical agents is similar to that of memory-free agents, as they simply take the new information without considering what happened before (of course, for different reasons). In other words, the event that takes place completely characterizes the "next" epistemic state of the agent. This seems to be related also to notions such as "only knowing" or "minimal knowledge" in [Lev90] and [HJT90]. This final observation also provides a further challenge: viz. unifying some of our parameters of diversity discussed so far.

6.5 From diversity to interaction

We have investigated many different sources of diversity, some visible in static logics, some in dynamic ones. Besides the old parameters from epistemic logic, namely computation and introspection ability, we have added several new aspects, i.e. observation power, memory capacity and revision policy. Our discussion has been mostly in the framework of dynamic epistemic logic and we have shown how it is possible to allow for a characterization of diversity within the logic. To summarize, look at the following diagram consisting of the main components of dynamic epistemic logic:

Static language	*Epistemic model* $\mathcal{M}$
Dynamic language	*Event model* $\mathcal{E}$
Product update	*Model change* $\mathcal{M} \times \mathcal{E}$

In the preceding sections we have shown that the diversity of agents can be explicitly modeled in terms of these logical components. The following table is an outline of the sources we have considered:[2]

Component	Residence	Diversity
$\mathcal{M}$	relations between worlds	introspection
$\mathcal{E}$	relations between actions	observation
$\mathcal{M} \times \mathcal{E}$	update mechanism	memory, revision policy

As we can see from the table, by introducing parameters of variation in each component, we are able to describe diversity of agents inside the logic.

But recognizing and celebrating diversity is only a first step! The next important phenomenon is that diverse agents *interact*, often highly successfully. Describing this interaction raises a whole new set of issues. In particular, our logical systems can describe the behavior of various agents, but they cannot yet state in one single formula "that an agent is of a certain type" or describe what would happen when we encounter those different agents. And as they stand, they are even less equipped to describe the interplay of different agents in a compact illuminating way. Imagine, if you know the type of the agent that you are encountering right now, can you take advantage of that knowledge? Or how could you *learn* about the type of the agent? In the following section, we will explore a few of these issues, and show in how far our current logical framework can handle these phenomena – and what features need to be added.

6.6 Interaction between different agents

Interaction between different agents is a vast area of diverse phenomena, and so, we will only discuss a few scenarios. These will show how the earlier dynamic logics can deal with some crucial aspects - though they also quickly need significant extensions. Our examples cover: reliability of sources (truth-tellers versus liars),

[2] Note that we have not discussed the earlier-mentioned parameter of inferential/computational power for agents. A more syntax-oriented approach to this topic can be find in [AJL06] and [Jag06]. It seems possible to merge the models proposed there with ours, and [Ben08] contains some first proposals for combined inferential and observational updates.

meetings between more or less introspective agents, and interaction between belief revisers following different policies.

‘Living with Liars’: dynamic logics of agent types

In this section we are challenging one of the PAL assumptions, namely, that all the announcements are truthful. What would happen if the announcer is a liar? More generally, can we figure out whether the announcer is a liar or truth-teller? In the following we will focus on such issues and explore how we update our knowledge when encountering people who should be identified first. These questions also bring us to a well-known puzzle about liars and truth-tellers. Here we consider one of its variations, high-lighting the fact that knowing what type of agent you encounter makes life a lot easier:

6.6.1. Example. On a fictional island, inhabitants either always tell the truth, or always lie. A visitor to the island meets two inhabitants, Aurora and Boniface. What the visitor can do is ask questions to discover what he needs to know. His aim is to find out the inhabitants’ type from their statements. The visitor asks a what type she is, but does not hear a’s answer. b then says “a said that she is a liar”. Can you tell who is a liar and who is a truth-teller?

One can try to figure out the answer to the puzzle by intuitive reasoning, but we will give a precise analysis in logical terms in what follows. To describe the situation with the relevant events, the salient fact is the agent-oriented nature of the communication. To bring this out, we first need to extend the language with notation for agent types:

6.6.2. Definition. Take a finite set of propositional variables Φ, and a finite set of agents G. Predicates $L(x)$, $T(x)$ and action terms $!\varphi_a$ are now added. The *dynamic epistemic agent type language* is defined by the rule

$$\varphi := \top \mid p \mid \neg\varphi \mid \varphi \wedge \psi \mid K_a\varphi \mid [\pi]\varphi \mid L(x) \mid T(x)$$
$$\pi := !\varphi_a$$

where $p \in \Phi$, and $a \in G$.

Here $L(a)$ is intended to express ‘agent a is a Liar’, and $T(a)$ expresses ‘agent a is a Truth-teller’. In fact, for the above example, we only need one of these expressions, since the agent is either a liar or a truth-teller. So we can use $\neg L(a)$ to denote ‘agent a is a Truth-teller’. Besides, we also want to express *who* executes some action. Accordingly, $!\varphi_a$ reads intuitively as ‘an announcement of φ performed by agent a’. Next we enrich the structure of our models, to a first approximation, in the following structures with hard-wired known agent types:

6.6.3. Definition. We define new epistemic models as $\mathcal{M} = (S, \{\sim_a \,|a \in G\}, V, L, T)$, where L, T are two types of agents, Liars and Truth-tellers. Moreover, given some suitable event model $\mathcal{E}$, the *truth conditions* for the new well-formed formulas are the following:

1. $\mathcal{M}, s \models T(x)$ iff $x \in T$.

2. $\mathcal{M}, s \models L(x)$ iff $x \in L$.

3. $\mathcal{M}, s \models [!\varphi_a]\psi$ iff ψ holds at the world $(s, !\varphi_a)$ in the product model $\mathcal{M} \otimes \mathcal{E}$.

Clause 1 and 2 are simple, as we only have two types of agents here. In general, there may be a larger set of types $\{L_1, L_2, \ldots L_k\}$, and we would then need to introduce a type function τ such that $\tau(L_i) \subseteq G$, setting $\mathcal{M}, s \models L_i(x)$ iff $x \in \tau(L_i)$. Item 3, however, is incomplete as it stands! This is because we have not given a precise update rule for the new agent-oriented announcements, which would require suitable *preconditions* $\langle !\varphi_a \rangle \top$ for the event of agent a's saying that φ.[3] In order to state useful and precise preconditions, we will definitely need more information about agent types.

Consider the example again. Clearly, the reason why the visitor should first find out who belongs to what type of agent is that it immediately determines the way she judges the incoming information. Here is a general illustration:

Case One: The visitor b does not know whether p is true, but *she knows that the speaker a is a truth-teller.* In fact, p is the case, and a says 'p is the case', after which b updates her knowledge accordingly:

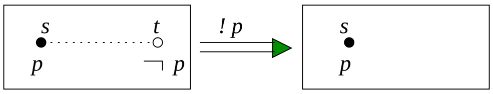

Figure 6.11: Telling the truth

Case Two: Next, the visitor b first does not know if p is true, but *she knows that a is a liar.* Now a says that 'p is not the case'. Agent b updates her knowledge with p instead of $\neg p$, see Figure 6.12:

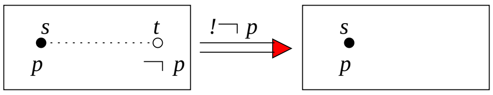

Figure 6.12: Lying

[3] Preconditions for agents' saying certain things may be related to their reliability according to the observing agent. Such a reliability judgment typically need not be publicly known. Thus, diversity of agents leads us to relax another idealization in standard *DEL* as defined earlier, viz. that preconditions of events are common knowledge.

These examples presuppose a definition of agent types, and how they affect preconditions for assertions. In the present scenario, these can be expressed more precisely in the following way:

(1) **truth-teller** $T(a) \to (\langle !\varphi_a \rangle \top \leftrightarrow \varphi)$

(2) **liar** $L(a) \to (\langle !\varphi_a \rangle \top \leftrightarrow \neg\varphi)$

Clause (1) says that a truth teller a can say exactly those things φ that are true. For the liar, this reverses.[4]

Even this simple stipulation has some interesting effects. E.g., no one can say that she is a liar, since our simple logic can formalize a version of the Liar Paradox, as stated in the following fact:

6.6.4. FACT. $\langle !L(a)_a \rangle \top$ does not hold in any case.

Proof. Suppose $\langle !L(a)_a \rangle \top$. There are two cases. Either a is a liar, $L(a)$, or a is a truth-teller, $T(a)$. In the first case, according to (2), we have $\langle !L(a)_a \rangle \top \to \neg L(a)$. Thus, in this case, we get $\neg L(a)$. But if a is a truth-teller, according to (1), we get $\langle !L(a)_a \rangle \top \to L(a)$ – and hence we have $L(a)$. This is a contradiction, and therefore, $\langle !L(a)_a \rangle \top$ does not hold. □

Incidentally, another take on our scenario might make it out to be about just single "lies and truths", rather than long-term liars and truth-tellers. This will not change our analysis here, but it would shift the emphasis in modeling from diversity of agents to what might be called *diversity of signals*. The latter tack is attractive, too, and sometimes simpler - but our main emphasis here is highlighting agent diversity in its own right.

Now as for interaction, we need to describe in general what agents would learn from communication if they knew the type of the other agent. To compute this, we can combine the information about agent types with the general rules of dynamic epistemic logic. For instance, even just minimal modal logic applied to the earlier type definitions yields the following principles:

(3) $K_b T(a) \to K_b(\langle !\varphi_a \rangle \top \leftrightarrow \varphi)$.

(4) $K_b L(a) \to K_b(\langle !\varphi_a \rangle \top \leftrightarrow \neg\varphi)$.

Using also the earlier reduction axioms for knowledge after events have taken place will generate further insights. Here are a few more valid principles about agents' changing knowledge in case a proposition is announced by a source whose type they know:

[4] See [BGP07] for a general account of more realistic conversational scenarios, where the current truth of a proposition need not imply that agents are automatically allowed to say it.

(6) $K_bT(a) \to ([!\varphi_a]K_b\varphi \leftrightarrow K_b[!\varphi_a]\varphi)$.

(7) $K_bL(a) \to ([!\varphi_a]K_b\neg\varphi \leftrightarrow K_b[!\varphi_a]\neg\varphi)$.

Of course, these principles are not yet a full-fledged account of messages. We have analyzed part of the information about the sender, but not yet the fact that it is a message from agent *a to agent b*. For logics of communication with such further aspects from the protocol perspective, we refer to [DW07].

Uncertainty about agent types

Still, the above is not all we need for our Island Puzzle. There, and also in real life, the types of agents encountered may be *unknown*! We need to represent that in our static and dynamic models. There are several ways of doing this. At the very least, the above predicates L, T will no longer be fixed once and for all for agents. They need to be made part of the specification of worlds, or events, so as to allow for uncertainty about them.

One proposal for modeling agent types (cf. [BGK06]) uses *pair events* of the form '(agent type, physical event)', say, "P is said by a truth-teller", or "P is said by a liar". Such abstract events are then epistemically indistinguishable if we can neither tell the agent types apart nor the actual observed events. However, in our analysis of the Island Puzzle, we do not need this rich format yet, since the conversation itself is about the types of agents, which makes things much easier. We therefore stick with a more ad-hoc format.

To model the original epistemic state of the visitor, see Figure 6.13 below. There is no information to indicate who is of what type, therefore, there are 4 possibilities in total, where for example the vertex (1, 1) represents the case in which a and b are both truth-tellers.

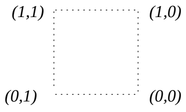

Figure 6.13: Initial model

Again, the dotted line denotes the visitor's uncertainty. Since the visitor does not hear what a says, there is no update for that.[5] Then b says "a said that she is a liar". Since we already noted the general truth that no one can say she is a liar, what b said about a is not true. So we conclude that b is a liar. This reasoning depends on the following principle, which follows from our agent type definition:

(5) $\varphi \wedge \langle !\neg\varphi_a\rangle\top \to L(a)$.

[5] In a more refined multi-agent scenario, there *would* be a product update for this event, as some higher-order knowledge about others changes – but we ignore this aspect here.

Meanwhile, we also know that a must have said that she is a truth-teller, since she was asked what type of agent she is, and there are only two possible answers. In this way, we (or the visitor to the island) split what b said into two statements: 'b is a liar' and 'a is a truth-teller'. To illustrate this more clearly, the update may be carried out in sequence, first with 'b is a liar', see Figure 6.14.

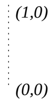

Figure 6.14: After knowing 'b is a liar'

And then with 'a is a truth-teller':

Eventually, we have obtained the required answer: Aurora is a truth-teller, while Boniface is a liar.[6]

Our analysis is in the same spirit as when one tries to figure out what kind of color a card has according to sequential announcements (cf. [Ben06a]). What is new here is that we no longer take any incoming information automatically as truthful. Instead, we first identify the type of agent who makes the statement, then we update our knowledge. Of course, this is only the beginning, since more complex scenarios would involve our updating our ideas about the *degree of reliability* of the source of our information.

The earlier valid principles about agents' changing knowledge when listening to speakers whose types they know easily extends to more complex event models with product events encoding uncertainty about agent types. The earlier general dynamic-epistemic reduction axioms will still work in this setting, when combined with preconditions for the different agent types.

Summing up, we have seen how an adequate account of different sources requires structured communicative events with agents explicitly indicated, explicit representations of agents' types, and a combination of general dynamic-epistemic reasoning principles with specific postulates about types of agent. In such a system, we can derive interesting principles about interaction between different agents. Of course, there are many more types of agent than just Liars and Truth-tellers, and Islands like the above are still logical paradise as compared to the real world. In particular, our views of the reliability of agents may change over time in subtle manners, calling for probabilistic information ([BGK06]). We will leave such further complications to future investigation.

[6] Strictly speaking, this is not quite right, since there is only one event of b's speaking, but we leave the formulation of one single update using our general product update mechanism to the reader.

A meeting between introspective and non-introspective agents

In this subsection we move to the perspective of the *addressee* instead of the *addressor* as investigated in the preceding scenario. Consider the following story.

6.6.5. Example. Two agents are sitting silently on a bench in the park. One of them, a, is non-introspective, but the other: b, is. The complete epistemic situation they find themselves in is depicted below – where the actual world is called s. The agents do not communicate with each other at first. Now, the aim for both of them is to find out which world is the real one as soon as possible. They have only one chance to receive new atomic information from some passer-by, and then they are ready to communicate with each other. What information should they get? What kind of communication should they engage in?

We picture the initial situation in the following diagram. As usual, all worlds are reflexive for each agent, but loops are omitted:

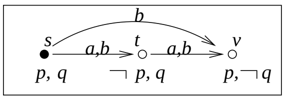

Figure 6.15: Initial situation

Here s is the real world where p and q are true. So, there are two possible atomic announcements one can make, either $!p$ or $!q$. Let's compare what will happen in these two cases. First, when q is truly announced by someone, the new model is pictured in Figure 6.16.

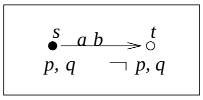

Figure 6.16: Two agents know the same

This new situation is symmetric between both agents. Both a and b are uncertain between s and t, and they do not know that s is the real world. And, given the symmetry, even if they communicate, it does not help, since they both know the same.

By contrast, once the fact p is announced, the new model becomes

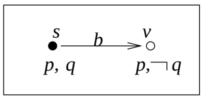

Figure 6.17: Two agents know differently

Here, the effect of this announcement is different for agent a and b! Agent b learns that p, but not that q. But agent a learns that both p and q, since she has no link to the world v. And she knows this is the real situation. Now a can inform b of q, so that b would also know q and realize that s is the real situation. What is going on here? How can the less-introspective agent a learn more? Do intellectuals need help from the man in the street to get their bearings?

We will just analyze what is going on here in terms of straight update. In terms of our epistemic models, a non-introspective agent may have fewer accessibility arrows than a corresponding introspective one, which means she is better informed, even though she does not reflect on this, and may not know everything she knows. Thus, additional information may help her more than her introspective companion.

To model the reasoning in such situations, as we have in the previous subsection, we can introduce agent types $I(a)$ and $NI(a)$ in the language to express that 'a is an introspective agent', and 'a is an anti-introspective agent', respectively, providing type definitions like the following:

(1) $I(a) \to (K_a\varphi \to K_aK_a\varphi)$.

(2) $NI(a) \to (K_a\varphi \to \neg K_aK_a\varphi)$.[7]

Clearly, because of their different introspective abilities, agents a and b may obtain quite different knowledge from what they learn. Intuitively, as we said already, the non-introspective agent even has an advantage in the above initial model, in that the following implication holds:

$$\mathcal{M}, s \models K_b\varphi \to K_a\varphi \quad (*).$$

But it is easy to think of settings where the knowledge of the agents would be incomparable. One can also analyze this type of situation more generically, using reduction axioms for informational events like before, leading to principles describing the interaction of the two agents such as the following:

(3) $NI(a) \wedge K_aI(b) \to ([!\varphi_c]K_b\psi \to K_a[!\varphi_c]\psi)$.

This is the static situation, looking at the agents separately. Of course, our scenario also illustrates another phenomenon, viz. how helpful agents which differ in their capacities may still inform each other, making the group consisting of both

[7] Note that one needs at least a non-normal logic to deal with anti-introspective agents. Since for instance the K-necessity rule $\vdash \varphi$, then $\vdash K_a\varphi$ itself presupposes certain positive introspection. It can lead to a contradiction. Moreover, given the definition (2), it is impossible to assume $K_aK_a\varphi \to K_a\varphi$, since we get $K_aK_a\varphi \to \neg K_a\varphi$ from (2). It would be interesting to investigate how far it is possible to model anti-introspective agents in modal logics.

agents together better informed than its members separately. Thus, our earlier observation that agents with different introspective powers lead to mere sums of modal logics $S4$ or $S5$ becomes just part of a more complex dynamic logic of what happens when they communicate.

Talking with different belief revisors

In our final scenario, we consider both information update and belief revision, and we also allow for diversity of both senders and receivers of information. Can our update models and their logics handle this? The following story is a bit contrived, but it highlights some realistic issues in everyday settings.

6.6.6. Example. Four agents live together, and their types are common knowledge. Agent a is a radical belief revisor, and b a conservative one. Agent c is a very trustworthy person, according to a and b, but d is less so. In the initial situation, there are three possible worlds s (the actual world), t, and v, as pictured in Figure 6.18 below, which also shows the valuation for the proposition letters. As for epistemic or doxastic relations, initially, a and b consider all three worlds possible, and they have the same plausibility ordering over them: v is most plausible, s is least plausible, t is in between. Moreover, c happens to know that p is the case, and d happens to know that q_2 is not the case. One can only speak after the other. Does this matter? Will both orders inform a, b equally well?

The original model may be depicted in Figure 6.18.

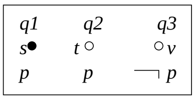

Figure 6.18: The original model: all agents believe the same.

Let us now suppose that d speaks first, truly, and says that p. Because of the different attitudes towards this new information, even though she acknowledges that d might be wrong, the radical (or more trusting) agent a will then change her plausibility ordering over the three worlds, see Figure 6.19.

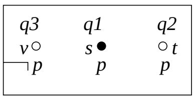

Figure 6.19: Update by a radical agent

In contrast to this, the conservative (or more suspicious) agent b would update his plausibility ordering in the manner depicted in Figure 6.20.

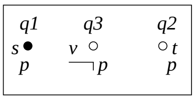

Figure 6.20: Update by conservative agents

We draw these orders separately here, though they will be part of one single total epistemic-doxastic update for the group when d's public announcement takes place. Next, the generally trusted source c tells agents a, b that q_2 is false. The above two models then change into the following new ones:

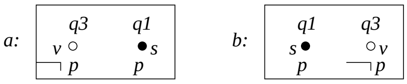

Figure 6.21: The final updates

We see from pictures that $B_a q_1$ and $B_b q_3$. Thus, a has acquired the right belief, but unfortunately, b has not! Thus, different revisors can get different convictions out of witnessing the same events, and indeed, some of them may be misled by correct information into believing false things! There is endless potential here for deceiving other agents – and even 'deception by the truth', which has already been observed by game theorists in the study of signaling games.[8]

Continuing with our example, dynamics of information flow is in principle *order-dependent*. What about the opposite order, where agent c speaks first, and only then the less-trusted d? Look at the original model again. After c's truthful announcement, both a and b will update their model into one with just the two possible worlds s and v of the last picture above. Then, when agent d tells them that p is true, the difference between the revision policies of a and b is immaterial: they can only raise the plausibility of p in one way, putting world s on top. Thus, both a and b acquire the right belief: as $B_a q_1$ and $B_b q_1$ hold.

To analyze this scenario in detail, one can use the machinery of Section 6.5 to express the types of agents qua revision policies (using the dynamic logics for belief revision discussed in Section 6.4), and then describe their interactions using a mixture of these type definitions and the general principles of dynamic epistemic-doxastic logic.

Admittedly, the preceding scenario is a bit contrived. More appealing scenarios of this sort would be variations of Muddy Children, where children revise beliefs rather than just updating knowledge, and where both skeptical and trusting children are around in the garden. In this way, belief revision policies would become concrete objects, whose workings can be determined precisely, and whose

[8] This phenomenon is also discussed in [BBS07] as a motivation for introducing a new epistemic attitude of 'safe belief', intermediate between belief and knowledge.

peculiarities may be exploited in more sophisticated puzzles of communication.

With these three scenarios, our discussion of interaction between diverse agents has come to an end. The main thrust of our investigation has been this. Once we have diverse agents inside one logical system, we can talk about the way they update their information and revise their beliefs. To deal with specific scenarios, we found that we needed the following additional ingredients: (a) more structured views of relevant events, (b) language extensions with types of agent, and their properties, where one may have to distinguish between the sender and the receiver of the information, and (c) mixtures of general dynamic-epistemic reasoning with specific information about agents. All this worked for information update, but we have also indicated how it applies to belief revision, when phrased in a dynamic logic format.

6.7 Conclusion and further challenges

This chapter has presented a more systematic discussion of different sources of diversity for rational agents than is usually found in the literature. More concretely, we showed how such diversity can be encoded in dynamic logics allowing for individual variation among agents. In particular, in the context of knowledge update, we made new proposals for modeling memory capacity, and defined a new version of product update for bounded k-memory agents. Next, in the context of belief revision, we showed how different revision policies can be put into one dynamic logic, allowing for great variation in revision and learning behavior.

Next, we pursued another essential phenomenon. Diversity among single agents is just a first ambition for logical modeling. But clearly, agents should not just 'live apart together'. Thus, we moved to the topic of interaction between agents of different types, discussing several scenarios which may arise then, having to do with different information processing, communication, and achieving of goals, when agents differ in their reliability, introspective powers, or belief revision policies. Our general conclusion was that these phenomena, too, can be modeled in our dynamic logics – but they need to be extended with explicit accounts of agents' types, and more structured informative events.

Even so, all this is only a beginning. There are several questions we would like to explore in the future. First, back to charting the sources of diversity, there remains the issue whether one can have a *general* view of the natural "parameters" that determine differences in behavior of logical agents. Our analysis does not provide such a general account, but at least, it shows more richness and uniformity than earlier ones. Second, even with all these parameters on the map, we have not yet found one framework for all these sources.

One particular area where this is true are agents' limitations in terms of inferential or computational powers. There is a body of work on the latter, witness the survey chapter on ‘Logic and Information’ by [BM07] in the *Handbook of the Philosophy of Information*. In particular, the work of [Dun07], [Ågo04], and [Jag06] in computer science seems relevant here – as in the chapter by [Abr07] on the information content of computation in the same Handbook. Indeed, there are also long-standing connections with discussions of information content in the philosophical literature (cf. [Hin73]). The cited survey chapter discusses attempts at combining inferential diversity with observational and learning diversity as discussed in our chapter. Cf. [VQ07] for some further development.

Our next ambition would be to put all these features together in one plausible computational model of an agent as an information-processing and decision-making device, with modules for perception, memory, and inference which can communicate and share information.

Next, concerning interaction in diverse societies of agents, we have not yet looked at scenarios involving bounded memory – the way game theorists have when they discuss ‘bounded rationality’. Here is where our dynamic epistemic or doxastic logics should meet up with current *game logics*, if we are to describe agents’ longer-term strategies for collaboration, or competition, or more realistically, their frequent mixtures of both... Furthermore, with strategic behavior in the longer-term, our analysis of diversity in single update steps should meet up with temporal epistemic and doxastic logic, as explored in [FHMV95], [PR03], [BP06], and [Bon07].

Even so, we hope that our account of diversity and interaction is of use per se in placing the phenomenon on the map, while it also may provide a fresh look at current logical systems for information update and belief revision. Our cognitive and social reality is that different agents live together, and interact with each other, sometimes with remarkable success. This rich set of phenomena is not just a playground for psychologists or sociologists: it seems to be a legitimate challenge to logicians as well!