

## ResearchSpace@Auckland

### Version

This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 <http://www.niso.org/publications/rp/>

### Suggested Reference

Girard, P., & Rott, H. (2014). Belief Revision and Dynamic Logic. In A. Baltag, & S. Smets (Eds.), *Johan van Benthem on Logic and Information Dynamics* (Vol. 5, pp. 203-233). Springer International Publishing. doi: [10.1007/978-3-319-06025-5\\_8](https://doi.org/10.1007/978-3-319-06025-5_8)

### Copyright

The final publication is available at Springer via [http://dx.doi.org/10.1007/978-3-319-06025-5\\_8](http://dx.doi.org/10.1007/978-3-319-06025-5_8)

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

<http://www.springer.com/gp/open-access/authors-rights/self-archiving-policy/2124>

<https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm>

# Belief revision and dynamic logic

Patrick Girard and Hans Rott

**Abstract** We explore belief change policies in a modal dynamic logic that explicitly delineates knowledge, belief, plausibility and the dynamics of these notions. Taking a Kripke semantics counterpart to Grove semantics for AGM as a starting point, we analyse belief in a basic modal language containing epistemic and doxastic modalities. We critically discuss some philosophical presuppositions underlying various modelling assumptions commonly made in the literature, such as the limit assumption and negative introspection for knowledge. Finally, we introduce in the language a general dynamic mechanism and define various policies of iterated belief expansion, revision, contraction and two-dimensional belief change operations.

**Key words:** Belief revision, AGM, dynamic epistemic logic, propositional dynamic logic, epistemic doxastic logic

## 1 Introduction

The history of belief revision is one in which researchers from various fields have tackled the same problem from different perspectives, but it originated from philosophy. After earlier work of Isaac Levi and William Harper, Carlos Alchourrón, Peter Gärdenfors and David Makinson (often referred to by the acronym “AGM”) initiated the formal study of belief change operations in the 1980s. They analyzed

---

P. Girard (✉)

Department of Philosophy, The University of Auckland, 18 Symonds Street, Private Bag 92019, Auckland 1142, New Zealand. e-mail: p.girard@auckland.ac.nz

H. Rott

Department of Philosophy, University of Regensburg, 93040 Regensburg, Germany. e-mail: hans.rott@ur.de

belief change using three kinds of models: partial meet contractions and revisions (in terms of maximal non-implying sets, [1]), safe contractions and revisions (in terms of minimal implying sets of sentences, [2]), and entrenchment-based contractions and revisions (based on the comparative retractability of sentences, [24]). Adam Grove [28] provided a possible worlds semantics for partial meet contraction. Iterated belief change were addressed in the 1990s, with important contributions by Boutilier [16], Darwiche and Pearl [17] and Nayak [37]. More recently, van Benthem [9], and Baltag and Smets [4, 5] modelled epistemic and doxastic states and their transformations on simple relational structures, introducing a new complexity due to the analysis of higher-order belief and knowledge. The relevant problems were attacked in a plethora of ways that may sometimes be differentiated by the degree of abstraction adopted.<sup>1</sup> Influenced by van Benthem, we choose our level of abstraction in line with a modular and minimalist attitude. This means that we only assume constraints on models when they become necessary, with philosophical awareness, and we use formal languages as simple as we can.

- **Minimality:** We try to keep the assumptions on models to be minimal, and we try to keep our basic object language as simple as possible.
- **Modularity:** We analyse the key notions of knowledge and belief in terms of two primitive relations representing indistinguishability and plausibility. We also break down complex dynamic notions into simpler parts using the programmatic PDL language.

## 2 Grove systems of spheres

In retrospect it looks as if AGM followed a purely syntactic approach to belief revision. We think that this is not quite correct, for both belief sets and inputs for belief change were essentially individuated only up to logical equivalence. Still it was a very important step for the program of belief revision when Adam Grove [28] provided a possible worlds semantics for AGM's partial meet contractions. He used systems of spheres modelling for belief change, closely following the seminal work of Lewis [36] on counterfactuals. We summarise Grove's semantics in some detail, and later show how to adapt the notation and interpretation to our needs.

Given a language  $\mathcal{L}$ , define a *theory* over  $\mathcal{L}$  as a set of  $\mathcal{L}$ -sentences that are logically closed, i.e., for which  $\text{cl}(T) \subseteq T$ . Here  $\text{cl}(\cdot)$  is an operation that forms the logical closure of a set of sentences from  $\mathcal{L}$ . A possible world for  $\mathcal{L}$  is an entity

---

<sup>1</sup> This caused in turn a proliferation of acronyms such as DEL for 'dynamic epistemic logic' which limits our freedom in using those that would naturally arise with our terminology throughout the paper. We thus use EDL to stand for 'epistemic doxastic logic'. The reader should try not to get confused by this choice.

that assigns, for each atom of  $\mathcal{L}$ , a truth value 1 (for “true”) or 0 (for “false”).<sup>2</sup> We denote by  $W$  the set of all possible worlds, and by  $[[\varphi]]$  or  $[[T]]$  the set of all possible worlds that satisfy a sentence  $\varphi$  or a set of sentences  $T$ . For a set of worlds  $V$ , we denote by  $\text{Theory}(V)$  the set of sentences true at each world in  $V$ .

A *Grove system of spheres* centered around a set  $\mathcal{B}$  of possible worlds for  $\mathcal{L}$  is a set  $\mathcal{S}_{\mathcal{B}} = \{S \mid S \subseteq W\}$  that satisfies the following conditions:<sup>3</sup>

1.  $\mathcal{S}_{\mathcal{B}}$  is totally ordered by set-containment  $\subseteq$ ,
2.  $\mathcal{B} \in \mathcal{S}_{\mathcal{B}}$ ,
3.  $\mathcal{B} \subseteq U$  for every  $U \in \mathcal{S}_{\mathcal{B}}$ ,
4. *The limit assumption:* For any sentence  $\varphi$ , if there is a sphere  $S \in \mathcal{S}_{\mathcal{B}}$  which intersects  $[[\varphi]]$ , then there is a smallest sphere  $S' \in \mathcal{S}_{\mathcal{B}}$  which intersects  $[[\varphi]]$ .

A Grove system of spheres is called

5a. *universal* if in addition,  $W = \bigcup \mathcal{S}_{\mathcal{B}}$ ; or

5b. *strongly universal* if in addition,  $W \in \mathcal{S}_{\mathcal{B}}$ .

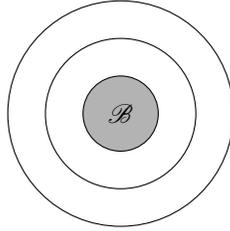
The term “universal” is taken over from Lewis [36, p. 16]. Grove originally required his doxastic systems of spheres to be strongly universal. We choose to deviate from this by requiring not even universality. We do not want to rule out that an agent may consider some metaphysically possible worlds inconceivable.

Belief in a Grovean sphere system is identified with the innermost sphere  $\mathcal{B}$  of  $\mathcal{S}_{\mathcal{B}}$ .<sup>4</sup> More precisely, a belief set in Grove spheres is identified with the set of all sentences  $\text{Theory}(\mathcal{B})$  that are true throughout the innermost sphere  $\mathcal{B}$  in  $\mathcal{S}_{\mathcal{B}}$ . The belief set  $\text{Theory}(\mathcal{B})$  is the one to be revised with the AGM operators of expansion, contraction and revision, which we will discuss below. We represent a Grove system of spheres centered around a set  $\mathcal{B}$  in the following way, shading the innermost sphere as characterizing the belief set  $\mathcal{B}$ :

<sup>2</sup> Grove [28] himself used maximal consistent sets of sentences rather than possible worlds, but this difference need not concern us here.

<sup>3</sup> We choose the notation  $\mathcal{B}$  as a mnemonic device for ‘belief’.

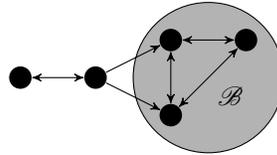
<sup>4</sup> Grove does not talk about beliefs, but about theories, and his semantics is about theory change broadly construed. However, his semantics is directly tailored to accommodate the AGM postulates, so we focus exclusively on a doxastic interpretation of his system.



The sphere formulation is in no way necessary, as Grove notes: “a system of spheres is really an ordering on the set of worlds” ([28], p. 160). To see this, we can reformulate Grove spheres in terms of a relation  $\leq$  on  $W$ , which we call a *Grove relation*. A Grove relation centered around a set  $\mathcal{B} \subseteq W$  is a relation  $\leq_{\mathcal{B}}$  on  $W$  that satisfies the following conditions<sup>5</sup>:

1.  $\leq_{\mathcal{B}}$  is connected and transitive,
2.  $x \in W$  is  $\leq_{\mathcal{B}}$ -maximal iff  $x \in \mathcal{B}$ .
3. *The limit assumption for  $\leq_{\mathcal{B}}$* : For any sentence  $\varphi$ , if there are any worlds at which  $\varphi$  is true, there are  $\leq_{\mathcal{B}}$ -maximal worlds  $w$  at which  $\varphi$  is true, so  $\{x \in \llbracket \varphi \rrbracket \mid y \leq_{\mathcal{B}} x \text{ for all } y \in \llbracket \varphi \rrbracket\}$  is non-empty.

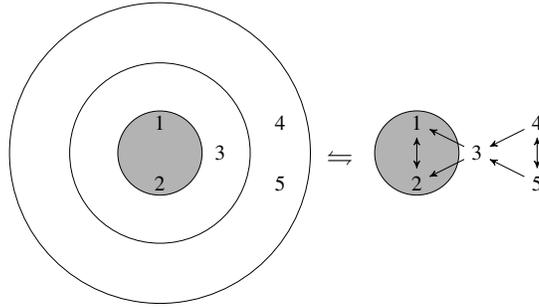
In a Grove relation, the set  $\{w \in W \mid v \leq_{\mathcal{B}} w \text{ for all } v \in W\}$  is the belief set, identified by the maximal elements in  $\leq_{\mathcal{B}}$ . We again refer to that set with  $\mathcal{B}$ . We represent Grove relations in the following way:



We have not drawn all arrows here, assuming that it is easy to see how the relations are transitively closed. (Later we shall be even more economical in our use of arrows).

Formally, the difference between Grove spheres and relations is almost only one of taste, just requiring a Gestalt switch of the following kind:

<sup>5</sup> Unlike Grove, we read  $x \leq_{\mathcal{B}} y$  as “y is more plausible than x according to  $\leq$ ” and talk of *maximal* worlds instead of *minimal* worlds. Katsuno and Mendelzon [31] is a seminal reference for the use of ordering semantics within the belief revision community.



We can be more specific about connection between Grove spheres and orderings: A Grove ordering  $\leq_{\mathcal{B}}$  is obtained from a system of spheres  $\mathcal{S}_{\mathcal{B}}$  by defining  $v \leq_{\mathcal{B}} w$  (read: “ $v$  is at most as plausible as  $w$ ”) iff for all  $S \in \mathcal{S}_{\mathcal{B}}$  such that  $v \in S$  it also holds that  $w \in S$ ; the field  $W_{\mathcal{B}}$  of  $\leq_{\mathcal{B}}$  is  $\bigcup \mathcal{S}_{\mathcal{B}}$ . Conversely, a system of Grove spheres  $\mathcal{S}_{\mathcal{B}}$  is obtained from a Grove ordering  $\leq_{\mathcal{B}}$  of  $W$  by collecting all sets  $S$  of the form  $S_w = \{v \in W : w \leq_{\mathcal{B}} v\}$ .

This modelling importantly allows for *ties* between the plausibilities of possible worlds, and this can of course be expressed in both modeling. In the systems of spheres modelling, this means that for two neighbouring spheres  $S$  and  $S'$  in  $\mathcal{S}_{\mathcal{B}}$  with  $S \subseteq S'$ , the set  $S' - S$  has in general more than just one element. Expressed in orderings, this means that for many worlds  $w$  there may be any number of distinct worlds  $v$  such that both  $v \leq_{\mathcal{B}} w$  and  $w \leq_{\mathcal{B}} v$ . The correspondence between systems of spheres and weak orderings is not perfect, however. The (strong) universality of a system of spheres  $\mathcal{S}_{\mathcal{B}}$  cannot be expressed by a weak ordering alone. This is why we need to specify the field  $W_{\mathcal{B}}$  of  $\leq_{\mathcal{B}}$ . We also take into account *inconsistent* belief states, represented by sphere systems  $\mathcal{S}_{\mathcal{B}}$  that contain the empty sphere  $\emptyset$  as an element. But if a sphere in  $\mathcal{S}_{\mathcal{B}}$  is empty, it generates the same weak ordering  $\leq_{\mathcal{B}}$ , in the sense just explained, as the sphere system  $\mathcal{S}_{\mathcal{B}} - \{\emptyset\}$ . We take the case of empty innermost spheres seriously, because we want to address belief expansion in Section 5.2. This requires representing trivialisation under expansion with information inconsistent with current beliefs. As the equivalent of empty spheres is not available in ordering semantics, we employ domain restriction in order to achieve similar results.

We need to comment on the interpretation of Grove spheres, or Grove orderings, understood as semantics for the classical AGM style belief revision. Semantically, a whole Grovean system of spheres (or a Grove ordering, or any other model for AGM-style one-shot belief revision) *represents* or, loosely speaking, *is* a single agent’s belief state. In some way (though it is hard to say in what way exactly), it encodes a first-person point of view: the set of worlds considered (doxastically) possible by the agent, and their comparative plausibilities as judged by the agent. Like the AGM approach, this modelling does not make room for *belief at a certain possible world*, it only encodes the beliefs and conditional beliefs as they are present in a single agent at the actual world (“here and now”).

Another philosophical point: one should *not* identify knowledge with the “outer belief modality” in Grove spheres, i.e., with “irrevocable belief” in the sense of Segerberg [46]. Knowledge is not the same as irrevocable belief. While knowledge implies truth, irrevocable belief need not do so. Baltag and Smets’s [4, p. 16] claim that the idea of identifying knowledge and irrevocable belief “can be traced back to Stalnaker [47]” is not quite correct; Stalnaker just defines  $\Box\varphi$  as  $\neg\varphi > \varphi$ , without any epistemic interpretation.<sup>6</sup> Segerberg [46] uses this reading only “unofficially” and alternatively speaks of *doxastic commitment*. Leitgeb and Segerberg [33, p. 176] allow themselves some philosophical looseness, too, and use the slogan *K stands for “knowledge”, not for ‘knowledge’*. Our feeling is that this identification is made only for the sake of convenience, because it is then easy to define knowledge in terms of (conditional) belief—something that epistemologists were never able to achieve—, and it is easy to argue for positive and negative introspection concerning knowledge—something that epistemologists have never wanted to achieve. If agents were infallible, then irrevocable belief would come close to knowledge, because then irrevocable belief would be guaranteed to be true. But human agents are not infallible, even their most deeply rooted beliefs may turn out to be false. We will return to this topic in Section 3.

Finally, even though Grove spheres are inspired by Lewis systems of spheres ([36]), Lewis explicitly rejects the limit assumption. Lewis’ argument is about similarity relations on worlds underlying counterfactual reasoning, but similar worries hold for plausibility orders: there are no most similar worlds in which Franck Ribéry is taller than 180 cm. Lewis’ semantics for counterfactuals does not depend on the limit assumption. Likewise, we can define belief and conditional belief using sequences of increasingly plausible worlds that do not require the limit assumption. Furthermore, there are AGM-inspired doxastic transformations of expansion, revision and contraction that do not require the limit assumption. Hence, following our minimalist attitude, we refrain to assume the limit assumption until necessary.<sup>7</sup>

---

<sup>6</sup> Alexandru Baltag (p.c.) has reminded us that Stalnaker [47, p. 102] endorses a condition akin to universality (see Stalnaker’s semantic condition (2)). Thus irrevocable belief must be true, and our argument against its identification with knowledge is no longer applicable. This is correct, but endorsing universality in this context means arguing that only metaphysical necessities can be known, since no metaphysically possible world gets epistemically excluded. For this reason, we strongly prefer not to require universality.

<sup>7</sup> Our reservation to assume the limit assumption from the outset is not only driven by philosophical concerns. To find an adequate axiomatisation of the limit assumption in our framework is not easy, and presently we can only make an informed conjecture that, in the context of the other axioms we are using, the axiom known as the ‘Löb’s axiom’ is exactly what we need. We will come back to this point below.

### 3 Epistemic Doxastic Logic

In contrast to the use of Grove models within the AGM paradigm, we will model knowledge and belief with Kripke structures. Thus what we call ‘epistemic doxastic logic’ in this paper (EDL for short) stands in the tradition of epistemic logic going back at least as far as the seminal work of Hintikka [30]. More recently, epistemic logic has been dynamified in the Dutch tradition and is now referred to as ‘dynamic epistemic logic’ (DEL for short). The acronym DEL is sometimes used more generally to also include doxastic transformations (see for instance van Benthem [20]). We use DEL in this general sense, and we reserve the acronym EDL for our own version of DEL in this paper. For us, EDL is the static core of our analysis of knowledge and belief. We will introduce dynamics in the next section and adapt our terminology accordingly.

In EDL, all theorizing begins with knowledge-at-a-possible-world- $w$  or belief-at-a-possible-world- $w$ . So belief and knowledge are “local” in this way. A whole EDL model does *not* represent some agent’s or several agents’ doxastic and epistemic states. The worlds in an EDL model are ways our actual world might be, metaphysically, *not* epistemically speaking.

One should not get confused by some “global” structures within EDL models. It might look strange at first sight that *within a cell of the epistemic partition of the worlds*, there is only one doxastic plausibility relation. In the local interpretation, which we believe is the correct one, it would be more precise to say that within such a cell, every world has the same plausibility relation assigned to it, and that it is only a matter of notational convenience that this relation is assigned to the cell rather than to every world within the cell. If possible worlds are considered as indices of evaluation, we might say that in our modelling, the doxastic relation is really just as *indexical* as the epistemic relation.<sup>8</sup>

That the plausibility relation is the same at each world within a cell, however has a very good justification. A plausibility relation is a formal encoding of the agent’s doxastic state. The fundamental assumption is that agents fully know their own doxastic attitudes.<sup>9</sup> There has to be an equally fundamental EDL axiom that captures this assumption. If we had just beliefs, the axiom should be something like  $B\phi \rightarrow KB\phi$ . But since we work with plausibility relations that allow us to express something like degrees of beliefs (entrenchments), we capture that agents are fully aware of their degrees of belief.<sup>10</sup>

<sup>8</sup> The indexical stance is represented in the models of Board [14, p. 60–61], van Ditmarsch [19, p. 237], and van Benthem [9, p. 138], while Baltag and Smets [3, p. 12] and [5, pp. 17–23] prefer a presentation in terms of global plausibility relations.

<sup>9</sup> Similar ideas of extending positive and negative introspection assumptions about belief were advanced by Stalnaker [48, p. 145] [49, p. 189] and Board [14, pp. 60–61]. See Demey [18, p. 387] on “uniform” epistemic plausibility models.

<sup>10</sup> We will return to this point later, see the comment on definitions (2) – (4) in Section 3.2 below.

Such a fundamental axiom is an expression of full (epistemic) introspection of doxastic attitudes. By viewing a global relation *within a cell of the knowledge partition* just as an abbreviated way of specifying that this very relation is assigned to each world within the cell, we can justify the use of global relations (restricted to cells) by the above introspection principles, while still staying firmly on the ground of the local tradition of DEL.

It does not make sense to look at an EDL model and ask what an agent believes or knows *in that model*. One can only ask what an agent believes or knows *at a certain world* in that model. The situation changes in pointed models that come with a distinguished world. This is why Baltag, van Ditmarsch and Moss [6], for instance, define an ‘epistemic state’ (p. 387) or a ‘doxastic state’ (p. 397) as tuples  $(M, s)$ , where  $M$  is a (relational/plausibility) model and  $s$  is a state or world.<sup>11</sup> Unfortunately, the authors do not comment at all on why they add a distinguished world. Summing up, a pointed model represents the actual world, with all the actual beliefs being in turn represented in terms of possible (conceivable) worlds. It thus makes sense to ask what an agent believes or knows *in a pointed model*.

If our interpretation is right, it hardly makes sense to drop worlds from the model as a result of some doxastic action. Worlds in EDL models are metaphysically possible worlds, not doxastically or epistemically possible worlds, as in Grove models. But as EDL theorists point out, a doxastic or epistemic action normally changes what is true in a world like any other action. Unlike the case of non-doxastic or non-epistemic actions, the effects of doxastic and epistemic actions can be represented in the model. The corresponding model transformation consists in manipulating the relevant relations between possible worlds, that is, in actions like cutting links, refining partitions, or shifting plausibilities.<sup>12</sup> This tension between world-elimination and link-cutting is a salient one for us. We will come back to it in Section 4.1.

The established tradition in modal logic as applied in computer science and game theory is to model knowledge by an equivalence relation that represents *indistinguishability for the agent*. The idea is that an agent knows that  $\varphi$  if and only if  $\varphi$  is true in all possible worlds that the agent cannot distinguish from the real world. We should like to emphasize that this is decidedly *not* the ordinary notion of knowledge. For knowledge in anything close to the ordinary, general sense, and the sense studied by epistemologists, negative introspection is a paradigmatic *non-theorem* rather than a theorem for epistemic logic.<sup>13</sup> More often than we like, we fail to know because we are wrong. In many such cases, we believe that we know  $p$ , but this belief is wrong. Thus, we don’t know (because  $p$  is false), but don’t know that we don’t know (an exemplification of an important kind of unknown unknowns). It is abundantly clear that the Brouwerian principle  $\neg\varphi \rightarrow K\neg K\varphi$  and the interaction

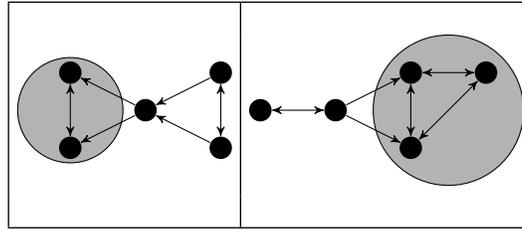
<sup>11</sup> Similarly for van Ditmarsch’s [19, p. 237] ‘doxastic-epistemic state’.

<sup>12</sup> Perhaps making some worlds “infinitely implausible”.

<sup>13</sup> Even positive introspection for knowledge, the so-called KK thesis, has been much contested, especially in the light of the success of externalist and reliabilist accounts in epistemology. Very recently, some new defenders of KK have entered the scene, see Okasha [38] and Greco [27].

principle  $BK\phi \rightarrow K\phi$  are invalid for the ordinary, general notion of knowledge,<sup>14</sup> but they come out as valid according to the standard DEL semantics. The notion of knowledge that is modelled by S5 structures is the knowledge of agents that are infallible. But humans are not. We can justify taking such structures only by restricting ourselves to specific domains or contexts in which agents do not make any mistakes, i.e., in which their doxastic possibilities do not rule out the actual world. Such contexts are provided by certain fields of research in computer science and game theory. We pretend that we are working in some such context and just ignore this problem as a matter of idealisation.

We thus conceive of the combination of the epistemic and doxastic relations in the following way. The epistemic relation creates a partition of the domain  $W$ , and each equivalence class of the partition contains a Grove relation (or a system of spheres). That we only have a single Grove relation (or a single system of spheres) for each epistemic indistinguishability class reflects the idea that an agent has the same doxastic state in each possible world within a cell of the knowledge partition. So we assume that agents are fully knowledgeable about (have full introspection concerning) their own doxastic states. If they were not, we would need to assign Grove spheres to each world individually. We represent the doxastic epistemic structure of a single agent as follows:



The domain is partitioned in two equivalence classes, and each “cell” contains a Grove relation, with maximal elements shaded in gray. The maximal elements of each cell are the states at which beliefs at a world in that very cell are to be evaluated. All worlds within a cell refer to the same belief set. So we picture Grovean sphere systems only within the cells of the knowledge partition (and each such cell contains exactly one such sphere). We impose this as a (the one standard) constraint on the interaction between the doxastic and epistemic relation, namely that the former is a subrelation of the latter:  $\leq \subseteq \sim$ .

<sup>14</sup> Lenzen [34] and, following Lenzen, Stalnaker [49] offer very strong arguments in favour of defining  $B\phi$  as  $\neg K\neg K\phi$ . In the light of this definition, Brouwer’s principle, which corresponds to the symmetry of the accessibility relation, just *means*  $B\phi \rightarrow \phi$ , i.e., infallibility! Lenzen [34, p. 43], Lamarre and Shoham [32, pp. 415, 420] and Stalnaker [49, p. 179] advocate the principle of strong belief  $B\phi \rightarrow BK\phi$  (the term “strong belief” is Stalnaker’s, Lamarre and Shoham use the term “certainty”). Taken together with the strong belief principle, the interaction principle  $BK\phi \rightarrow K\phi$  implies the undesirable  $B\phi \rightarrow K\phi$ . – In his attempt to maintain negative introspection for knowledge, Halpern [29] proposes to restrict the principle that knowledge implies belief to nonmodal sentences.

The fact that we have Grove relations  $\leq$  within each knowledge cell means that the domain of such Grove relations is not the whole of  $W$  (i.e., that the relevant Grove systems are not universal). The structural properties of Grove relations are all restricted to each individual indistinguishability cell. There are no plausibility comparison across cells. Hence, if we want to stick to the single relation modelling, we have to use more complicated structures than Grove orderings.

A *generalized Grove relation* is a reflexive and transitive relation  $\leq$  over  $W$  such that the relation  $\sim$ , defined by

$$u \sim v \quad \text{if and only if} \quad \text{either } u \leq v \text{ or } v \leq u$$

is an equivalence relation.<sup>15</sup> Notice that we do not assume the limit assumption for the generalized Grove relation. This definition derives the epistemic relation from the doxastic relation: Indistinguishability means comparability in terms of plausibility. The relation containment of  $\leq$  inside  $\sim$  is satisfied by definition. It is not required that generalized Grove relations are connected over the whole of  $W$ . They may have, and typically do have many belief sets (sets of doxastically possible worlds) on which systems of spheres are centered—one such structure in each cell of the partition. So  $\leq$  never makes any plausibility comparisons across cells. But it is easily verified that each cell with respect to  $\sim$  is a Grove relation – without the limit assumption.

### 3.1 EDL *Language and Semantics*

We define *epistemic doxastic models* (for single agents), or EDL models for short, as structures  $M = \langle W, \sim, \leq, <, V \rangle$ , with  $W$  a non-empty set of worlds, an  $V$  a valuation function assigning sets of possible worlds to propositional variables. We take  $\leq$  to be a generalised Grove relation. We define  $\sim$  as above, and take  $<$  as the strict subrelation of  $\leq$  in the usual way:  $w < v$  iff  $w \leq v$  and  $v \not\leq w$ .<sup>16</sup> We read  $w \leq v$  as ‘ $v$  is at least as plausible as  $w$ ’, and  $w < v$  as ‘ $v$  is (strictly) more plausible than  $w$ ’. To talk about EDL models, we use a basic EDL modal language with three modalities corresponding to the three accessibility relations:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid [\sim]\varphi \mid [\leq]\varphi \mid [<]\varphi$$

<sup>15</sup> Given the transitivity of  $\leq$ , a condition equivalent to the transitivity of  $\sim$  thus defined is the requirement of weak connectedness of  $\leq$  both forwards and backwards, also known as “no branching of  $\leq$  to the left or to the right”:

If  $w \leq u$  and  $w \leq v$ , then either  $u \leq v$  or  $v \leq u$  (NBR)

If  $u \leq w$  and  $v \leq w$ , then either  $u \leq v$  or  $v \leq u$  (NBL)

This requirement is used by Baltag et al. [6, p. 396].

<sup>16</sup> The presence of  $\sim$  and  $<$  in models are thus redundant, as they are definable in terms of  $\leq$  in models, but we keep them for reasons that will become clear later.

As usual, we define dual diamond operators as, for example,  $\langle \sim \rangle \varphi := -[\sim]\neg\varphi$ .

For the interpretation of the EDL language, we extend the valuation  $V$  to a valuation  $\llbracket \cdot \rrbracket^M$  assigning semantic values, or sets of possible worlds, to the sentences of the EDL language. Hence, in each epistemic doxastic model  $M = \langle W, \sim, \leq, <, V \rangle$ , semantic values  $\llbracket \varphi \rrbracket^M \subseteq W$  are given by:

$$\begin{aligned} \llbracket p \rrbracket^M &= V(p) \\ \llbracket \neg\varphi \rrbracket^M &= W \setminus \llbracket \varphi \rrbracket^M \\ \llbracket (\varphi \wedge \psi) \rrbracket^M &= \llbracket \varphi \rrbracket^M \cap \llbracket \psi \rrbracket^M \\ \llbracket [\sim]\varphi \rrbracket^M &= \{w \in W \mid \text{if } w \sim v, \text{ then } v \in \llbracket \varphi \rrbracket^M, \text{ for every } v \in W\} \\ \llbracket [\leq]\varphi \rrbracket^M &= \{w \in W \mid \text{if } w \leq v, \text{ then } v \in \llbracket \varphi \rrbracket^M, \text{ for every } v \in W\} \\ \llbracket [<]\varphi \rrbracket^M &= \{w \in W \mid \text{if } w < v, \text{ then } v \in \llbracket \varphi \rrbracket^M, \text{ for every } v \in W\} \end{aligned}$$

For convenience, we sometimes use the more common notation  $M, w \models \varphi$  instead of  $w \in \llbracket \varphi \rrbracket^M$ .

### 3.2 Knowledge and Belief in EDL

For reasons of simplicity and continuity with much of the literature, we assume that agents know that  $\varphi$  in a world  $w$  just in case  $\varphi$  is true in all worlds  $v$  that they cannot distinguish from  $w$ :

$$M, w \models K\varphi \quad \text{iff} \quad \text{for all } v \text{ such that } v \sim w, M, v \models \varphi.$$

In other words,

$$K\varphi := [\sim]\varphi \tag{1}$$

We thus have a semantics for  $K\varphi$  that is widely accepted in computer science and game theory. As we explained above, we have philosophical scruples about it, but we wish to keep the focus of this paper on belief.

So much for the epistemic party. For doxastic operators, we need to do more work. Our strategy is to derive the analysis of (various kinds of) belief by using the more primitive modalities  $[\leq]$  and  $[<]$ . Following a strategy that can be traced back to at least Boutilier [15, p. 44], we have an intended semantics for belief that is in line with the AGM tradition:

$$M, w \models B\varphi \quad \text{iff} \quad \text{there is a } u \text{ such that } u \sim w \text{ and for all } v \text{ with } u \leq v, M, v \models \varphi.$$

Notice the epistemic constraint  $v \sim w$  in the semantic definition of  $B\varphi$ , in order to guarantee that beliefs are independently evaluated in each class of the epistemic

partition. Given our assumption of connectedness inside each epistemic class, this semantics says that at some point along the plausibility order,  $\varphi$  is true for every world at least as plausible. We can explicitly define this notion of belief in the EDL language:

$$B\varphi := \langle \sim \rangle [\leq] \varphi \quad (2)$$

The right-hand-side of Equation 2 says that some worlds among the epistemically indistinguishable worlds have  $[\leq]\varphi$ , which is precisely the semantics of  $B\varphi$ . Assuming the limit assumption would guarantee that  $\varphi$  is true in all maximal worlds in the model, which is the more common definition of belief in the AGM tradition.

We can also express that a belief in  $\varphi$  is stronger or more entrenched than a belief in  $\psi$ , which we denote by  $B(\psi \prec \varphi)$ .

$$B(\psi \prec \varphi) := \langle \sim \rangle (\neg \psi \wedge [\leq] \varphi) \quad (3)$$

If we set  $\psi = \perp$  in definition (3), we recover definition (2). Another doxastic notion that we can define is conditional belief of  $\varphi$  given  $\psi$  (cf., for instance, [4, 5, 14, 12]):

$$B(\varphi | \psi) := \langle \sim \rangle \psi \rightarrow \langle \sim \rangle (\psi \wedge [\leq] (\psi \rightarrow \varphi)) \quad (4)$$

If we set  $\psi = \top$  in definition (4), we again recover exactly definition (2). Notice that we didn't need to use the strict modality  $[<]\varphi$  to define belief so far. But if we were to work over partial orders, we could use it to define belief, comparative entrenchment of belief and conditional belief in the following way:

$$B\varphi := [\sim] (\neg \varphi \rightarrow \langle < \rangle (\varphi \wedge [<] \varphi)) \quad (5)$$

$$B(\psi \prec \varphi) := [\sim] (\neg \varphi \rightarrow \langle < \rangle ((\varphi \wedge \neg \psi) \wedge [<] \varphi)) \quad (6)$$

$$B(\varphi | \psi) := [\sim] ((\psi \wedge \neg \varphi) \rightarrow \langle < \rangle ((\psi \wedge \varphi) \wedge [<] (\psi \rightarrow \varphi))) \quad (7)$$

Again, if we set  $\psi = \perp$  in definition (6) or  $\psi = \top$  in definition (7), we recover definition (5), as expected.

Finally, if we set  $\psi = \neg \varphi$  in definition (7), we get the notion of irrevocable belief (Segerberg [46]), i.e., belief that is sustained even in the face of contradicting evidence. Since  $\langle < \rangle (\varphi \wedge \neg \varphi)$  cannot be true anywhere,  $B(\varphi | \neg \varphi)$  reduces to  $[\sim] \varphi$ . But this means that irrevocable belief reduces to knowledge-as-indistinguishability. In contrast to Baltag and Smets [4, pp. 14–15][5, p. 28], we have argued against such an identification on Section 2, because even the strongest beliefs may be wrong. But with the means used in this paper we cannot express the difference.<sup>17</sup>

<sup>17</sup> Here are a few hints how this situation could get remedied. The picture is basically that within each indistinguishability cell ( $\sim$ -cell), there is a single system of spheres  $\$$  that need not exhaust

We can now address the idea that agents have full introspection of their doxastic states. This is in fact built firmly into our notions of belief. Notice that the definitions (2)–(7) of belief have either  $\langle \sim \rangle$  or  $[\sim]$  as their main operator.<sup>18</sup> Given our semantics for the indistinguishability relations, it is clear that we get the characteristic S5 axioms of positive and negative introspection with respect to  $[\sim]$ . But given definition (1), this means that if the agent has certain beliefs at a world  $w$  that can be expressed by  $B\phi$ ,  $B(\psi \prec \phi)$  or  $B(\phi | \psi)$ , then, by the very definition of these expressions,  $KB\phi$ ,  $KB(\psi \prec \phi)$  or  $KB(\phi | \psi)$ , respectively, are also true at  $w$ .<sup>19</sup>

One advantage of our modular and minimalist approach is that we can express precisely what we mean when we talk of knowledge and (comparative firmness of) belief. We are very economical with the assumptions we impose on our models, and we relegate them to explicit definitions in our language. We refrain as much as we can to use background assumptions.

---

this cell. In order to characterize  $\bigcup \mathcal{S}_{\emptyset}$ , we suggest to extend epistemic doxastic models by a new relation  $\rightsquigarrow$  that helps representing non-universal systems of spheres (see Section 2—but now everything happens within every single  $\sim$ -cell).  $\rightsquigarrow$  should be a serial, transitive and Euclidean sub-relation of the global indistinguishability relation  $\sim$  that specifies a unique set of “conceivable” worlds within each set of indistinguishable worlds. Intuitively,  $u \rightsquigarrow v$  for worlds  $u$  and  $v$  means that  $u \sim v$  and  $v$  is within the relevant  $\sim$ -cell’s system of spheres. Thus, if  $u \sim v$  and  $u \rightsquigarrow w$ , then also  $v \rightsquigarrow w$ . We would also need to harmonise  $\rightsquigarrow$  with  $<$  (and  $\leq$ ), by conditions like ‘If  $u \sim v$  and  $w \rightsquigarrow u$  but not  $w \rightsquigarrow v$ , then  $v < u$ ’ and ‘If  $u \sim v$  and there is no  $w$  such that  $w \rightsquigarrow u$  or  $w \rightsquigarrow v$ , then neither  $u < v$  nor  $v < u$ .’ We would then use  $\langle \rightsquigarrow \rangle$  and  $[\rightsquigarrow]$  rather than  $\langle \sim \rangle$  and  $[\sim]$  in the definitions (4) and (7) of conditional belief. Correspondingly, the notion of irrevocable belief would reduce to  $[\rightsquigarrow]\phi$  rather than  $[\sim]\phi$ . Knowledge that  $\phi$  (in the indistinguishability sense) would then imply irrevocable belief that  $\phi$ , but not vice versa, as desired.

<sup>18</sup> Definition (4) is an exception. But even on this definition, we have  $B(\phi | \psi) \rightarrow KB(\phi | \psi)$ . It is easy to see this. Assume that  $M, w \models B(\phi | \psi)$  for some  $w$ . For  $M, w \models KB(\phi | \psi)$ , we need to show that  $M, v \models B(\phi | \psi)$  for all  $v$  such that  $w \sim v$ . By definition,  $M, w \models B(\phi | \psi)$  means that either  $M, w \models [\sim]\neg\psi$  or  $M, w \models \langle \sim \rangle(\psi \wedge [\leq](\psi \rightarrow \phi))$ . But the truth value of both of these sentences are independent of the world  $v$  of evaluation, as long as  $w \sim v$ . So either  $M, v \models [\sim]\neg\psi$  or  $M, v \models \langle \sim \rangle(\psi \wedge [\leq](\psi \rightarrow \phi))$ , and thus  $M, v \models B(\phi | \psi)$ , as desired.

<sup>19</sup> The modality ‘ $[\leq]$ ’ is referred to as “knowledge” by Lamarre and Shoham [32, p. 418], as “knowledge according to the defeasibility analysis” by Stalnaker [49, Section 6], and as “safe belief”, “defeasible knowledge” and “Stalnaker knowledge” by Baltag and Smets [5, see in particular pp. 27-32]. In contrast to  $K\phi$  and  $B\phi$ , the truth value of  $[\leq]\phi$  is in general not constant within a  $\sim$ -cell. The early paper of Lamarre and Shoham is interesting: It disavows negative introspection for knowledge and finds strong belief (“certainty”) that  $\phi$  to be equivalent with  $\neg K\neg K\phi$ —points we acclaim from a philosophical perspective. But it also finds knowledge that  $\phi$  to be equivalent with the conditional belief  $B(\phi | \neg\phi)$ —a result we object to from a philosophical perspective. This unexpected conjunction is connected with the fact that Lamarre and Shoham let not only knowledge, but also conditional belief and conditional certainty vary from world to world, and thus disavow negative introspection for conditional belief and conditional certainty, too.

### 3.3 Axiomatisation of EDL

The axiomatisation of *static* EDL is based on standard propositional logic. Figure 1 shows the axioms for the static part of our modal logic in building blocks. We have the standard set of S5 axioms for the knowledge-as-indistinguishability modality  $[\sim]$ . For the non-strict plausibility relation  $[\leq]$  which takes  $\leq$  as an accessibility relation, we have the S4.3 axioms that correspond to connected relations. For strict plausibility  $[\prec]$ , we have K4 plus (Mod $\prec$ ). This latter axiom is interesting. As far as we know, it has not been used in epistemic or doxastic logics so far, but van Benthem [7, p. 200] has stated and analyzed it early on in the context of temporal logic. Being a Sahlqvist formula (cf. Blackburn, de Rijke and Venema [13]), (Mod $\prec$ ) enforces the modularity (or ‘almost-connectedness’ or ‘virtual connectivity’<sup>20</sup>) to the right of the strict plausibility relation  $\prec$ : If  $u \prec v$ ,  $u \prec w$ ,  $u \prec z$  and  $v \prec w$ , then either  $v \prec z$  or  $z \prec w$ . Finally, we need interaction principles between  $[\sim]$ ,  $[\leq]$  and  $[\prec]$ . These principles are there to counteract the modal undefinability of  $\prec$  in terms of  $\leq$ , as has been noted in van Benthem, Girard and Roy [10]. They guarantee that the relation  $\prec$  is *adequate* under bulldozing (cf. Segerberg [45]) of the canonical model, so that  $w \prec v$  iff  $w \leq v$  and  $v \not\leq w$ . In a similar fashion, it is well-known that we cannot modally express that  $\sim$  is the same as  $\leq \cup \leq^{-1}$ , but the canonical model can be adapted accordingly. Since the interaction principles are fairly strong, we are not claiming that our axiomatisation is free of redundancies. We prefer to have fully independent axiomatizations for each of our modal operators instead.

We have not assumed the limit assumption up until now, and we are still inclined against endorsing it. However, should one insist to include it, we suggest to add the so-called Löb axiom (Löb $\prec$ ), which corresponds to transitivity and converse well-foundedness (and thus irreflexivity), as an optional extra. It is known to exclude infinite chains and so is the natural counterpart to the limit assumption in ordering semantics.<sup>21</sup> While the limit assumption is not important in the static contexts of (conditional and unconditional) belief, it will turn out to be necessary for many important belief change operations on epistemic doxastic models.

<sup>20</sup> The most descriptive term ‘modularity’ was suggested by Ginsberg [25, p. 49]; ‘almost-connectedness’ is due to van Benthem [7, p. 194], [8, p. 232], ‘virtual connectivity’ to Alchourón and Makinson [2, p. 415]. Notice that there is also a different sense of ‘almost-connectedness’ in the literature (see Doble et al. [21]).

<sup>21</sup> See Blackburn, de Rijke and Venema [13, pp. 130–132]. It would be nice to have a more compact axiomatisation of K4 plus (Mod $\prec$ ) and (Löb $\prec$ ). At this point, we can only conjecture that adding Löb is sufficient to get a complete axiomatisation with the limit assumption, but we have to leave open the problem of showing the logic to be (weakly) complete with respect to the relevant class of frames.

Modal logic for $\sim$ , $\leq$ and $\langle$ , and their interaction	
(Nec $\sim$ )	$\vdash \varphi \Rightarrow \vdash [\sim]\varphi$
(K $\sim$ )	$[\sim](\varphi \rightarrow \psi) \rightarrow ([\sim]\varphi \rightarrow [\sim]\psi)$
(T $\sim$ )	$[\sim]\varphi \rightarrow \varphi$
(4 $\sim$ )	$[\sim]\varphi \rightarrow [\sim][\sim]\varphi$
(5 $\sim$ )	$\neg[\sim]\varphi \rightarrow [\sim]\neg[\sim]\varphi$
(Nec $\leq$ )	$\vdash \varphi \Rightarrow \vdash [\leq]\varphi$
(K $\leq$ )	$[\leq](\varphi \rightarrow \psi) \rightarrow ([\leq]\varphi \rightarrow [\leq]\psi)$
(T $\leq$ )	$[\leq]\varphi \rightarrow \varphi$
(4 $\leq$ )	$[\leq]\varphi \rightarrow [\leq][\leq]\varphi$
(.3 $\leq$ )	$[\leq]([\leq]\varphi \rightarrow \psi) \vee [\leq]([\leq]\psi \rightarrow \varphi)$
(.3 $\leq$ ) $\diamond$	$(\langle \leq \rangle \varphi \wedge \langle \leq \rangle \psi) \rightarrow (\langle \leq \rangle (\varphi \wedge \langle \leq \rangle \psi) \vee \langle \leq \rangle (\varphi \wedge \psi) \vee \langle \leq \rangle (\psi \wedge \langle \leq \rangle \varphi))$
(Nec $\langle$ )	$\vdash \varphi \Rightarrow \vdash [\langle]\varphi$
(K $\langle$ )	$[\langle](\varphi \rightarrow \psi) \rightarrow ([\langle]\varphi \rightarrow [\langle]\psi)$
(4 $\langle$ )	$[\langle]\varphi \rightarrow [\langle][\langle]\varphi$
(Mod $\langle$ )	$([\langle](\varphi \vee [\langle]\chi) \wedge [\langle](\chi \vee [\langle]\psi) \wedge \neg[\langle]\chi) \rightarrow [\langle](\varphi \vee [\langle]\psi)$
(Mod $\langle$ ) $\diamond$	$(\langle \langle \rangle (\varphi \wedge \langle \rangle \psi) \wedge \langle \rangle \chi) \rightarrow (\langle \langle \rangle (\varphi \wedge \langle \rangle \chi) \vee \langle \rangle (\chi \wedge \langle \rangle \psi))$
(Löb $\langle$ )	$[\langle]([\langle]\varphi \rightarrow \varphi) \rightarrow [\langle]\varphi$
(Löb $\langle$ ) $\diamond$	$\langle \rangle \varphi \rightarrow \langle \rangle (\varphi \wedge \langle \rangle \varphi)$
( $\sim \leq$ )	$[\sim]\varphi \rightarrow [\leq]\varphi$
( $\leq \langle$ )	$[\leq]\varphi \rightarrow [\langle]\varphi$
( $\langle \leq 1$ )	$[\langle]\varphi \rightarrow [\leq][\langle]\varphi \quad (\langle \leq 1 \rangle) \quad \langle \leq \rangle \langle \rangle \varphi \rightarrow \langle \rangle \varphi$
( $\langle \leq 2$ )	$[\langle]\varphi \rightarrow [\langle][\leq]\varphi \quad (\langle \leq 2 \rangle) \quad \langle \rangle \langle \leq \rangle \varphi \rightarrow \langle \rangle \varphi$
(QuAdeq)	$([\langle]\psi \wedge [\leq](\psi \vee [\leq]\varphi)) \rightarrow ([\leq]\psi \vee \varphi)$
(QuAdeq) $\diamond$	$(\varphi \wedge \langle \leq \rangle \psi) \rightarrow (\langle \rangle \psi \vee \langle \leq \rangle (\psi \wedge \langle \leq \rangle \varphi))$

**Fig. 1** Axiomatisation of static EDL. For the reader's convenience, we have added a few variants using the possibility operator  $\diamond$ , since these are sometimes more intuitive.

## 4 Epistemic Doxastic PDL Logic

Epistemic doxastic PDL logic, EDPDL for short, is a variant of the now well-established PDL logic (propositional dynamic logic), whose first interpretation over relational structures can be found in Pratt [41], and further elaborated in Fischer and Ladner [23]. The original purpose of PDL was to provide a logic of programs in a modal framework, taking programs as modal operators or binary relations between states (transitions between states of a machine). The interpretation of PDL modalities  $\langle \pi \rangle \varphi$  according to [23] is: ' $\pi$  can terminate with  $\varphi$  holding on termination'. However, this interpretation is not what we are after. The part of PDL that is relevant to us is the modal calculus of relation combination, which we exploit to for-

malise belief change. Formally, we extend the language of EDL by adding the PDL operations of test, choice and composition to the basic ingredients of our language:

$$\begin{aligned}\pi &::= \sim \mid \leq \mid < \mid \varphi? \mid \pi \cup \pi' \mid \pi; \pi' \\ \varphi &::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid [\pi]\varphi\end{aligned}$$

As usual, we define dual box operators  $[\pi]\varphi := \neg\langle\pi\rangle\neg\varphi$  for each program  $\pi$ . In this notation, the special modalities  $[\pi]\varphi$  with  $\pi \in \{\sim, \leq, <\}$  are just the basic modalities  $[\sim]\varphi$ ,  $[\leq]\varphi$  and  $[<]\varphi$  of the previous section. In each epistemic doxastic model  $M = \langle W, \sim, \leq, <, V \rangle$ , semantic values  $\llbracket\varphi\rrbracket^M \subseteq W$  and  $\llbracket[\pi]\varphi\rrbracket^M \subseteq W^2$  are given by:

$$\begin{aligned}\llbracket p \rrbracket^M &= V(p) \\ \llbracket \neg\varphi \rrbracket^M &= W \setminus \llbracket \varphi \rrbracket^M \\ \llbracket (\varphi \wedge \psi) \rrbracket^M &= \llbracket \varphi \rrbracket^M \cap \llbracket \psi \rrbracket^M \\ \llbracket \langle\pi\rangle\varphi \rrbracket^M &= \{u \in W \mid u \llbracket [\pi] \rrbracket^M v \text{ and } v \in \llbracket \varphi \rrbracket^M, \text{ for some } v \in W\} \\ \llbracket [\sim] \rrbracket^M &= \sim \\ \llbracket [\leq] \rrbracket^M &= \leq \\ \llbracket [<] \rrbracket^M &= < \\ \llbracket [\varphi?] \rrbracket^M &= \{\langle u, u \rangle \mid u \in \llbracket \varphi \rrbracket^M\} \\ \llbracket [\pi_1; \pi_2] \rrbracket^M &= \{\langle u, v \rangle \mid u \llbracket [\pi_1] \rrbracket^M w \text{ and } w \llbracket [\pi_2] \rrbracket^M v, \text{ for some } w \in W\} \\ \llbracket [\pi_1 \cup \pi_2] \rrbracket^M &= \llbracket [\pi_1] \rrbracket^M \cup \llbracket [\pi_2] \rrbracket^M\end{aligned}$$

As usual, we also write  $u \llbracket [\pi] \rrbracket^M v$  for  $\langle u, v \rangle \in \llbracket [\pi] \rrbracket^M$  and  $M, u \models \varphi$  for  $u \in \llbracket \varphi \rrbracket^M$ . Incorporating PDL in our axiomatisation is simple, especially since we are only appealing to the fragment of PDL without the Kleene star. In line with our modular approach, the PDL operators or test, choice and composition are recursively introduced (Figure 2).

PDL	
13.	$\vdash \varphi \Rightarrow \vdash [\pi]\varphi$
14.	$\vdash [\pi](\varphi \rightarrow \psi) \rightarrow ([\pi]\varphi \rightarrow [\pi]\psi)$
15.	$\vdash \langle\pi\rangle\varphi \leftrightarrow \neg[\pi]\neg\varphi$
16.	$[\pi_1; \pi_2]\varphi \leftrightarrow [\pi_1][\pi_2]\varphi$
17.	$[\pi_1 \cup \pi_2]\varphi \leftrightarrow ([\pi_1]\varphi \wedge [\pi_2]\varphi)$

**Fig. 2** Axiomatization of general PDL part

### 4.1 Doxastic PDL Transformations

To formalise belief change, we use a special case of PDL-transformations as defined in Girard, Seligman and Liu [26]. This latter paper was directly motivated by Fact 19 from van Benthem and Liu [11]: every relation-changing operation that is definable in PDL without iteration has a complete set of reduction axioms in dynamic epistemic logic. We fully exploit this idea in the remainder of the paper. For the details of the general case of PDL-transformations, the reader should consult Section 1 of [26]. We give here a self-contained specification of the special case of PDL-transformations required for our purposes. Basically, a *doxastic* PDL-transformation  $\Lambda$  is a transformation on models that has two components: 1) a domain restriction provided by some sentence denoted  $|\Lambda|$ ,<sup>22</sup> and 2) PDL-definable transformations of the relations  $\sim$ ,  $\leq$  and  $<$ . Even though we feel uncomfortable about world-elimination, as we already explained, we will use domain restrictions to differentiate between expansion and revision. We are very much aware of philosophical difficulties that may ensue, and will try to treat them with care.

Given a model  $M = \langle W, \sim, \leq, <, V \rangle$  and a PDL-transformation  $\Lambda$ , the result of transforming  $M$  with  $\Lambda$  is the model  $\Lambda M = \langle \Lambda W, \Lambda(\sim), \Lambda(\leq), \Lambda(<), \Lambda V \rangle$ .  $\Lambda M$  is computed by setting  $\Lambda W = \llbracket |\Lambda| \rrbracket^M$  and taking all relations  $\Lambda(\sim)$ ,  $\Lambda(\leq)$  and  $\Lambda(<)$  that are defined explicitly for each particular program to be restricted to  $(\Lambda W)^2$ . Likewise the valuation  $\Lambda V$  is simply the valuation  $V$  restricted to the domain  $\Lambda W$ .

We also define computable translations  $\varphi^\Lambda$  of sentences corresponding to PDL-transformations on model:

$$\begin{aligned}
 p^\Lambda &= p & \sim^\Lambda &= \Lambda(\sim); |\Lambda|? \\
 (-\varphi)^\Lambda &= -\varphi^\Lambda & \leq^\Lambda &= \Lambda(\leq); |\Lambda|? \\
 (\varphi \wedge \psi)^\Lambda &= (\varphi^\Lambda \wedge \psi^\Lambda) & <^\Lambda &= \Lambda(<); |\Lambda|? \\
 (\langle \pi \rangle \varphi)^\Lambda &= \langle \pi^\Lambda \rangle \varphi^\Lambda & (\varphi?)^\Lambda &= (\varphi^\Lambda)? \\
 & & (\pi_1; \pi_2)^\Lambda &= \pi_1^\Lambda; \pi_2^\Lambda \\
 & & (\pi_1 \cup \pi_2)^\Lambda &= \pi_1^\Lambda \cup \pi_2^\Lambda
 \end{aligned}$$

As demonstrated in [26], the translation  $\varphi^\Lambda$  guarantees that the following lemma holds:

**Lemma 1.** *For each state  $u$  of  $\Lambda M$  and  $v$  of  $M$ , and for each sentence  $\varphi$ ,*

$$\begin{aligned}
 M, u \models \varphi^\Lambda & \text{ iff } \Lambda M, u \models \varphi, & \text{ and} \\
 u \llbracket \pi^\Lambda \rrbracket^M v & \text{ iff } v \in \Lambda W \text{ and } u \llbracket \pi \rrbracket^{\Lambda M} v.
 \end{aligned}$$

We represent PDL-transformations in the following way:

<sup>22</sup> The accustomed reader will recognise this as something very much like public announcement.

Name of a PDL-transformation	
$\Lambda$	$ \Lambda $
	$\sim := \Lambda(\sim)$
	$\leq := \Lambda(\leq)$
	$< := \Lambda(<)$

A well-known special case of a PDL-transformation is the public announcement of a sentence  $\varphi$ , first studied by Plaza in 1989 and now republished in [40]:

Public Announcement	
$\varphi!$	$\varphi$
	$\sim := \sim$
	$\leq := \leq$
	$< := <$

In this representation, ' $\sim := \sim$ ' means that the relation  $\sim$  is assigned to its restriction to the new domain in the new model, i.e.,  $\varphi!(\sim) = \sim \cap (\varphi!W)^2 = \sim \cap ((\llbracket \varphi \rrbracket^M)^2)$ . Thus, all a public announcement does is to restrict the domain by eliminating  $\neg\varphi$ -worlds. All relations are kept as they were, but restricted to the new domain. To ease notation, we omit writing identity assignment such as ' $\sim := \sim$ ' in transformations. We also omit writing the domain restriction when  $|\Lambda| = \top$ . Thus the public announcement transformation can be succinctly written as:

Public Announcement	
$\varphi!$	$\varphi$

With this established, we expand our language of doxastic epistemic PDL logic with modalities  $[\Lambda]\varphi$  for each PDL-transformation  $\Lambda$  with the following semantics:

$$M, w \models [\Lambda]\varphi \quad \text{iff} \quad \Lambda M, w \models \varphi.$$

We stipulate that  $M, w \models [\Lambda]\varphi$  is vacuously true in case  $M, w \not\models |\Lambda|$ .

A technical difficulty with PDL-transformations is that they may not always transform models into new models of the right kind. A doxastic epistemic model  $M$  may be transformed by a PDL-transformation  $\Lambda$  in such a way that  $\Lambda M$  is not a doxastic epistemic model. To avoid this issue, we do not accept any possible doxastic PDL-transformation, but instead provide a class of doxastic PDL-transformations that are *proper*, in the sense that they always return doxastic epistemic model.

## 5 AGM operations

In this section we provide a class of PDL-transformations which are candidates of unrestricted transformations in the style of Alchourrón, Gärdenfors and Makinson (compare [1, 2, 17, 24, 28, 31]); we refer to them as “AGM operations” or “doxastic transformations”. We base our selection on those given in Rott [43].<sup>23</sup> We first study the operations of *belief expansion* and *belief revision*, which can be nicely treated in EDPDL logic. We then move to a study of the operation of *belief contraction*. Our selection is partly for the sake of exposition, but we include standard DEL doxastic change operations to be found in recent work such as van Benthem [9] or Baltag and Smets [5].

### 5.1 Expansion and Revision

The operations of expansion and revision are about adding beliefs to belief states. We start with three types of doxastic change, that we categorise as conservative, radical and moderate. A conservative doxastic transformation by  $\varphi$  is one that only shifts around maximal  $\varphi$ -worlds (or, in the case of contraction,  $\neg\varphi$ -worlds), and leaves the ordering between the other worlds intact. A moderate doxastic transformation by  $\varphi$  is one that shifts around all  $\varphi$ -worlds (or, in the case of contraction,  $\neg\varphi$ -worlds) in a uniform way. A radical doxastic transformation by  $\varphi$  is one that only preserves  $\varphi$ -worlds (or, in the case of contraction, almost only  $\neg\varphi$ -worlds).

We use the following abbreviations:

$$\begin{aligned} \sim^\varphi & ::= (\varphi?; \sim; \varphi?) \\ \leq^\varphi & ::= (\varphi?; \leq; \varphi?) \\ <^\varphi & ::= (\varphi?; <; \varphi?) \\ \max(\varphi) & ::= (\varphi \wedge [<]\neg\varphi) \end{aligned}$$

Notice that the sentence  $\max(\varphi)$  is only true in the most plausible  $\varphi$ -worlds, i.e.,  $\varphi$ -worlds such that all worlds more plausible, if there are any, satisfy  $\neg\varphi$ .<sup>24</sup>

<sup>23</sup> Any PDL transformation which outputs a Grove relation would be formally legitimate. To categorise this general class of transformations is still an open problem in GDDL, and we will not address it here, as our main concern is with AGM motivated transformations.

<sup>24</sup> Slightly abusing the term “maximality”, one could also experiment with putting  $\max(\varphi) ::= (\varphi \wedge ([<]\neg\varphi \vee [\leq]\varphi))$ , but we will not pursue this idea in the present paper.

## 5.2 Expansion

We start with expansion. Generally speaking, expanding one's beliefs with  $\varphi$  is to start believing  $\varphi$  without caring about consistency. If  $\varphi$  is not consistent with what the agent believes, then her beliefs trivialise and she now believes  $\perp$ . But it is important to note that her belief *state* does *not* trivialise. In the semantics using Grovean systems of spheres, we can represent this nicely by adding an empty sphere to  $\mathcal{S}_\emptyset$ . Unfortunately, no such picture is possible with the pure ordering approach. We have to stipulate here that if  $\varphi$  is inconsistent with the beliefs supported by  $\leq$ , which we write as  $\neg[\sim]\langle\leq\rangle\varphi$ , then the belief set that results from the expansion is the trivial one,  $\text{cl}(\perp)$ . We achieve this by introducing the domain restriction  $|\llbracket[\sim]\langle\leq\rangle\varphi\rrbracket|$  in expansion transformations. Notice that this restriction doesn't really restrict the domain, because of our underlying assumption that the plausibility order is uniform inside epistemic classes. So either all worlds in a class satisfy  $[\sim]\langle\leq\rangle\varphi$ , or none do. In the latter case, the agent ends-up believing  $\perp$ .

We first look at conservative expansion. Conservative expansion by  $\varphi$  reorders maximal  $\varphi$ -worlds and leave the rest of the model intact. That is, the order stays intact among the worlds that are not maximal  $\varphi$ -worlds (those that either do not satisfy  $\varphi$  or for which there are strictly more plausible worlds that satisfy  $\varphi$ ), and it makes every maximal  $\varphi$ -world equally plausible to each other as well as strictly more plausible than any other world:

Conservative Expansion	
	$[\sim]\langle\leq\rangle\varphi$
CE $\varphi$	$\leq := \leq^{\neg\text{max}(\varphi)} \cup (\sim; \text{max}(\varphi)?)$
	$< := <^{\neg\text{max}(\varphi)} \cup (-\text{max}(\varphi)?; \sim; \text{max}(\varphi)?)$

To say that beliefs of agents trivialise under expansion with information  $\varphi$  that is inconsistent with their beliefs amounts to saying that  $M, w \models [\text{CE}\varphi]\mathcal{B}\perp$  in case  $M, w \not\models [\sim]\langle\leq\rangle\varphi$ . Notice that conservative expansion is not successful if we do not assume the limit assumption. If there are no maximal  $\varphi$ -worlds, then nothing happens to the doxastic structure.

Second, consider the operation of moderate expansion, which differs from the conservative operation by reordering all  $\varphi$ -worlds instead of only the maximal ones. Hence, moderate expansion preserves the order among the  $\varphi$ -worlds and among the  $\neg\varphi$ -world, and makes every  $\varphi$ -world strictly more plausible than every  $\neg\varphi$ -world. In this case, the old maximal  $\varphi$ -worlds become the most plausible ones overall. Formally, moderate expansion ME $\varphi$  is defined by:

Moderate expansion	
	$[\sim]\langle\leq\rangle\varphi$
ME $\varphi$	$\leq := \leq^\varphi \cup \leq^{\neg\varphi} \cup (-\varphi?; \sim; \varphi?)$
	$< := <^\varphi \cup <^{\neg\varphi} \cup (-\varphi?; \sim; \varphi?)$

Finally, radical expansion is an action which reduces to a domain restriction. If  $\varphi$  is consistent with the agent's beliefs, then we only keep the  $\varphi$ -worlds. Thus, if there are maximal worlds that are  $\varphi$ -worlds, radical expansion deletes all  $\neg\varphi$ -worlds; otherwise beliefs trivialise. This is all that is required, so relations are simply restricted as they were to the new domain. Formally, radical expansion RE $\varphi$  is defined by:

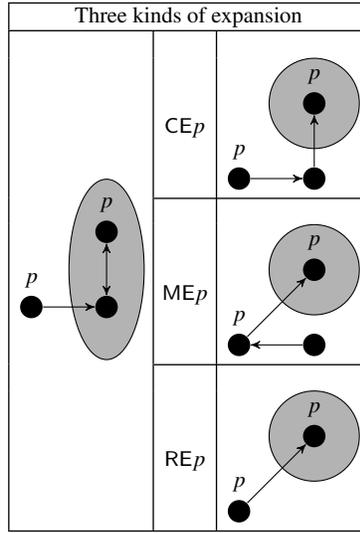
Radical expansion	
RE $\varphi$	$\varphi \wedge [\sim]\langle\leq\rangle\varphi$

Figure 3 displays the three kinds of expansion acting on the same model. Every expansion returns the same set of maximal states. The difference is in the ordering of the remaining worlds. We have chosen a model in which some  $p$ -worlds are among the maximal worlds. Radical expansion restricts the domain to  $p$ -worlds, exemplifying the way in which it is radical compared to the other ones.

Now that we have precise definitions of expansion as doxastic transformation, we can specify distinguished modalities for each of them: [CE $\varphi$ ], [ME $\varphi$ ], and [RE $\varphi$ ]. So for instance, the sentence [CE $\varphi$ ]B $\psi$  says that  $\psi$  is believed after moderately expanding with  $\varphi$ . As we have no restriction on iterations of doxastic actions, we can also express and analyse complex sentences such as the validities [CE $p$ ][RE $p$ ]B $\psi \leftrightarrow$  [RE $p$ ]B $\psi$  and [RE $p$ ][RE $q$ ] $\psi \leftrightarrow$  [RE( $p \wedge q$ )] $\psi$ . In a multi-agent language, we could also analyse higher-order beliefs about doxastic change. For instance, with  $s =$  "Robert is a spy" and  $l =$  "Robert is a liar", the sentence  $B_r[\text{RE}_b s]B_b l$  expresses that "Robert believes that Bernadette believes that Robert is a liar after radically expanding her belief by the fact that Robert is a spy".

### 5.3 Revision

Revision is exactly like expansion, except that agents do not get trivial beliefs when revising with information that was not consistent with their initial beliefs. Thus the only difference between revision and expansion is in the way the beliefs are retrieved from a plausibility ordering. For revisions, the standard rules apply, and thus an agent simply cannot have an inconsistent belief set! We can accommodate this nicely with conservative and moderate revision, but radical revision is problematic. We can get two interpretations, but neither works properly as a revision. We start with conservative and moderate revision:



**Fig. 3** Three kinds of expansion by  $p$ . Models are closed under transitivity and reflexivity.

Conservative Revision	
$CR\varphi$	$\leq := \leq^{\neg \max(\varphi)} \cup (\sim; \max(\varphi)?)$ $< := <^{\neg \max(\varphi)} \cup (\neg \max(\varphi)?; \sim; \max(\varphi)?)$
Moderate Revision	
$MR\varphi$	$\leq := \leq^\varphi \cup \leq^{\neg \varphi} \cup (\neg \varphi?; \sim; \varphi?)$ $< := <^\varphi \cup <^{\neg \varphi} \cup (\neg \varphi?; \sim; \varphi?)$

Moderate revision precisely corresponds to the *lexicographic upgrade*, and conservative revision to the *elite change* of van Benthem [9, p. 141]. We will see in Section 5.4 that these operations can be regarded as the natural limiting cases of a common idea (viz., that of bounded revision). However, there is an important difference. While moderate revision has no need whatsoever for the limit assumption, conservative revision needs it badly. Like conservative expansion, conservative revision might not be successful without the limit assumption. If it is not met, then there may not be any maximal  $\varphi$ -worlds and conservative revision may not effect anything.

We can give two interpretations of radical revision as described in Rott [43]. An important aspect of radical revision by  $\varphi$  is that  $\neg\varphi$ -worlds can never be recovered. One way of incorporating this is by using a domain restriction  $|\varphi|$  that removes  $\neg\varphi$ -worlds from the model altogether. Another way is to have no domain restriction, like in conservative and moderate revision, but cut every link between  $\varphi$  and  $\neg\varphi$ -worlds.

The first approach, in which we guarantee irrevocable revision with a domain restriction, is the following doxastic transformation:

Radical Revision, version 1	
RR $\varphi$	$\varphi$

This version of radical revision by  $\varphi$  is the same as public announcement of  $\varphi$  as we've analysed above. In the terminology of van Benthem [9], radical revision is a change under *hard information*, whereas conservative and moderate revisions are two alternatives of change under *soft information*.

The way in which this version of radical revision captures the irrevocability of  $\varphi$  is by deleting all the  $\neg\varphi$ -worlds. This is indeed a radical way of guaranteeing that  $\neg\varphi$ -worlds cannot be recovered. This approach also captures success of revision for Boolean sentences, but not without a price. The price to pay is that beliefs of agents trivialise in  $\neg\varphi$ -worlds when radically revising by  $\varphi$ . Indeed, assume that  $M, w \models \neg\varphi$  in some model  $M$ . Because of the domain restriction  $|\varphi|$ , we get that  $M, w \models [\text{RR}\varphi]B\perp$ . Success comes at the price of triviality, which is not what revision operators have been invented for.

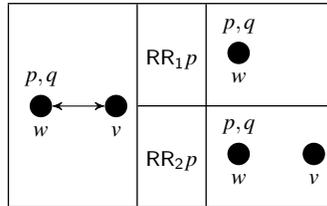
One way to make sure that belief sets do not trivialise under revision is to avoid restricting the domain, as in conservative and moderate revision, with the following transformation:

Radical Revision, version 2	
	$\sim := \sim^\varphi \cup \sim^{\neg\varphi}$
RR $\varphi$	$\leq := \leq^\varphi \cup \leq^{\neg\varphi}$
	$< := <^\varphi \cup <^{\neg\varphi}$

The way in which this version of radical revision is irrevocable comes from the definition of our epistemic relation. We did not introduce a free transition as a basic program. We have been using the relation  $\sim$  instead. For instance, in moderate revision, we can make sure that all  $\varphi$ -worlds become more plausible than  $\neg\varphi$ -worlds by re-defining  $<$  as  $<^\varphi \cup <^{\neg\varphi} \cup (-\varphi?; \sim; \varphi?)$ , where  $\sim$  is used to create new plausibility links in  $(-\varphi?; \sim; \varphi?)$ . Now, in our second interpretation of radical revision, once we cut links between  $\varphi$  and  $\neg\varphi$ -worlds, these links are no longer recoverable! This is nice, but also comes with its own cost, again when radical revision is evaluated in  $\neg\varphi$ -worlds.<sup>25</sup>

<sup>25</sup> For complex sentences  $\varphi$  that involve doxastic operators, it is possible that  $\varphi$  becomes true again at a  $\neg\varphi$ -world after other doxastic transformations. The old  $\varphi$ -worlds are irrevocable. But it is the worlds that are irrevocable, not sentences. Only Boolean sentences (those without modalities) are truly irrevocable. This is related to the consideration of the AGM success postulate: Only Boolean sentences are guaranteed to be successful. If one revises by a sentence that says “ $p$  is true but you don't know it”, then one does not get success.

Let us highlight the problem about interpreting (possibly untruthful) public announcement and radical revision with the help of an example. Take a very simple model with two epistemically indistinguishable and equiplausible worlds  $w$  and  $v$ , with  $p$  and  $q$  true at  $w$ , but false at  $v$  (Figure 4). Consider a radical revision by  $p$ . If radical revision goes by domain restriction,  $v$  simply vanishes, and we get that both  $[\text{RR}p]Bp$  and  $K[\text{RR}p]Bp$  are true at  $w$ . However, if radical revision goes by a “link-cutting” action then again  $[\text{RR}p]Bp$  is true at  $w$ , but false at  $v$ —surprisingly, even  $[\text{RR}p]B\neg p$  is true at  $v$ . Hence, since  $v \sim w$ ,  $K[\text{RR}p]Bp$  is false at  $w$ . Suppose that in the initial situation the agent is *actually* located at  $v$ , but cannot distinguish  $v$  from  $w$ . So all the agent knows or believes at the beginning is  $p \leftrightarrow q$ . Now the agent receives a public (but not truthful!) announcement that  $p$  and as a result performs some kind of radical revision on  $p$ . What would happen? Metaphysically, world  $v$  would not cease to exist—this is what makes domain restriction strange.<sup>26</sup> The agent would (wrongly) believe that  $p$ , but of course she would not know that  $p$ . But she would not know or believe that  $\neg p$  either – this is what makes link-cutting problematic. Intuitively, while the beliefs have changed—and this is why  $[\text{RR}p]Bp$  should come out true at  $v$ —, the knowledge has not increased. The agent located at  $v$  is still not able to epistemically distinguish her world from  $w$ . None of our modelings has this option.



**Fig. 4** Small example illustrating a problem with radical revision. Models are closed under reflexivity.

The versions of revision we have been investigating are illustrated as operating on the same initial model in Figure 5.

<sup>26</sup> How can we evaluate a sentence at a world  $v$  which has vanished in the course of the evaluation? Above, we have stipulated that  $M, w \models [\Lambda]\varphi$  to be vacuously true in case  $M, w \not\models [\Lambda]$ , in order to avoid facing the main clause  $\Lambda M, w \models \varphi$  when  $w$  fails to be in  $\Lambda W$  of  $\Lambda M$ . But evidently, this is not a solution to the problem of untruthful public announcements or radical revisions.

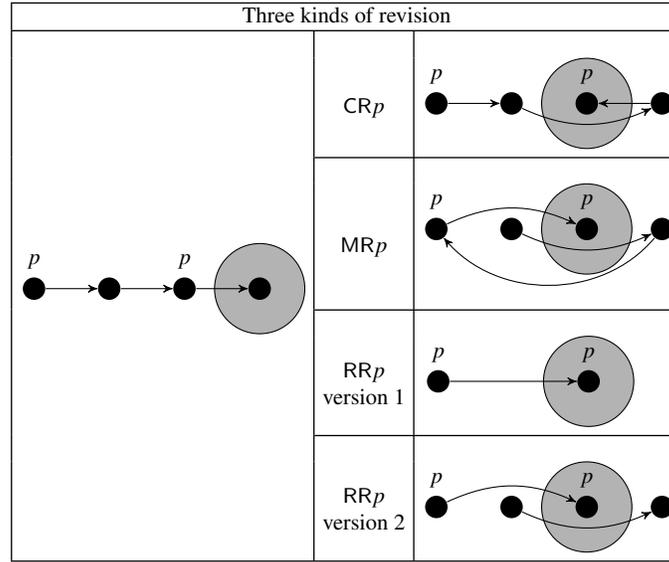


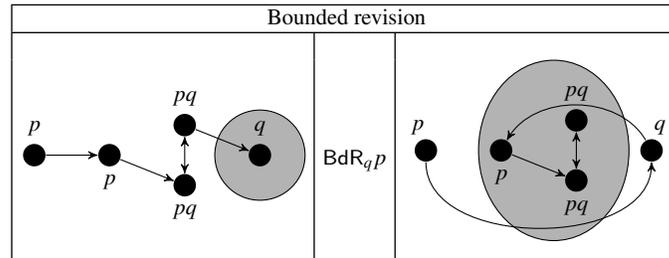
Fig. 5 Three kinds of revision by  $p$ . Models are closed under transitivity and reflexivity.

### 5.4 Two-dimensional belief change operators

We continue our brief overview of revision operations in the framework of EDL with two-dimensional change operations in the sense of Rott [44]. These models are meant to increase the expressive power of purely qualitative, relational, thus non-numerical models for belief change. The extent to which an input sentence  $\varphi$  is accepted, is specified by a reference sentence  $\psi$ . The first two-dimensional belief change operation we consider is *bounded revision*. The idea of bounded revision is to accept  $\varphi$  as long as  $\psi$  holds along with  $\varphi$ —and just a little longer. Bounded revision satisfies (generalizations of) the semantically motivated postulates of Darwiche and Pearl [17], as well as a “Same beliefs condition” according to which the posterior *beliefs* of the agent should not depend on the reference sentence (although the posterior *belief state* does). For further motivation we refer to [44]. Bounded revision  $BdR_\psi\varphi$  is defined by:

Bounded Revision	
$BdR_\psi\varphi$	$\leq := \leq\varphi \wedge [<](\varphi \rightarrow \psi) \cup \leq \neg(\varphi \wedge [<](\varphi \rightarrow \psi)) \cup$ $(\neg(\varphi \wedge [<](\varphi \rightarrow \psi)))?; \sim; (\varphi \wedge [<](\varphi \rightarrow \psi))?$
	$< := <\varphi \wedge [<](\varphi \rightarrow \psi) \cup < \neg(\varphi \wedge [<](\varphi \rightarrow \psi)) \cup$ $(\neg(\varphi \wedge [<](\varphi \rightarrow \psi)))?; \sim; (\varphi \wedge [<](\varphi \rightarrow \psi))?$

It is not difficult to verify that bounded revision reduces to the unary operation of conservative revision if the reference sentence  $\psi$  is fixed to  $\perp$ , and that it reduces to the unary operation of moderate revision if the reference sentence  $\psi$  is fixed to  $\top$ . In general, bounded revision requires the limit assumption, since for instance, if  $\varphi$  and  $\psi$  are inconsistent with each other, minimal  $\varphi$ -worlds are needed to ensure the success of the revision operation. By a deliberate choice of the reference sentence  $\psi$ , however, one may in many cases make sure that there is a broad enough range of worlds that satisfy  $\varphi \wedge [\prec](\varphi \rightarrow \psi)$ , and then the operation performs well even if the model does not satisfy the limit assumption. Figure 6 gives an illustration of bounded revision in a finite model.



**Fig. 6** Bounded revision by  $p$  with respect to reference sentence  $q$ . Models are closed under transitivity and reflexivity.

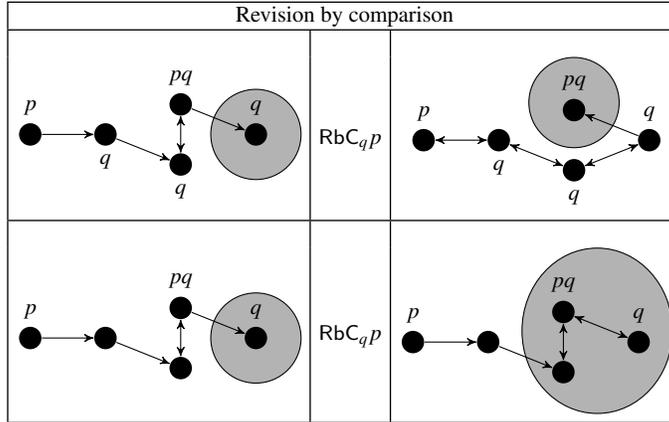
Another interesting two-dimensional operation is *revision by comparison* (Fermé and Rott [22]). It is motivated by the same concerns as bounded revision. But while the idea of bounded revision is to accept  $\varphi$  as long as  $\psi$  holds along with it (and a little longer), revision by comparison accepts  $\varphi$  with a *strength* that at least equals that of the acceptance of  $\psi$ . In contrast to bounded revision, revision by comparison does not satisfy the Darwiche-Pearl postulates. In its intended cases of application, it is a revision operation, but it can also have the effects of a contraction operation (see Section 5.5): If the reference sentence is too weak (more precisely, if the input sentence is at least as surprising as the negation of the reference sentence), then the revision fails, and instead a severe contraction with respect to the reference sentence is performed, provided that there are maximal nonmodels of the reference sentence<sup>27</sup>. In at least one way of presenting it (namely by manipulations of prioritized belief bases, cf. Rott [43]), revision by comparison is an extremely natural belief change operation.

The definition of revision by comparison given in Fermé and Rott [22, p. 14] can be represented in our framework as follows:

<sup>27</sup> The limit assumption guarantees this. If, however, the limit assumption is not satisfied, revision by comparison as defined below may fail to make the input sentence at least as firmly accepted as the reference sentence.

Revision by comparison	
$\text{RbC}_{\psi\varphi}$	$\leq := \leq^{\varphi} \cup (-[\leq]\psi?; \leq) \cup (-\varphi?; \sim; [<]\psi?)$
	$< := <^{\varphi} \cup (-[<]\psi?; <) \cup (-\varphi?; \sim; (\varphi \wedge [\leq]\psi?)$

Figure 7 gives an illustration of revision by comparison in a finite model. We are representing two cases, the successful one in which the input sentence gets accepted, and the unsuccessful one in which the reference sentence gets withdrawn.



**Fig. 7** Revision by comparison by  $p$  with respect to stronger and weaker reference sentences  $q$ : the successful case of a revision and the unsuccessful case reducing to a severe withdrawal of  $q$ . Models are closed under transitivity and reflexivity.

If we set the reference sentence to  $\top$ , then revision by comparison reduces to a unary revision operation that is more radical than moderate revision but somewhat less radical than the revision operations we have called radical:

Radical Revision, version 3	
$\text{RR}_3\varphi$	$\leq := \leq^{\varphi} \cup (-\varphi?; \sim)$
	$< := <^{\varphi} \cup (-\varphi?; \sim; \varphi?)$

We deal with another interesting limiting case of revision by comparison obtained by setting the input sentence to  $\perp$  in the next section.<sup>28</sup>

<sup>28</sup> Also compare Rott [42].

## 5.5 Contraction

The operation of contraction is about withdrawing beliefs from belief states. The main idea is that a contraction by  $\varphi$  is effected by promoting the maximal  $\neg\varphi$ -worlds, and possibly some more worlds, to the ranks of the maximal worlds (i.e., the maximal  $\top$ -worlds).

To be successful, each of the following operations requires the use of  $\max(\neg\varphi)$ , and most of them require the use of  $\max(\top)$  as well. Thus the difficulty with belief contraction is that we need to identify maximal states: the states where  $[\langle] \varphi$  or, respectively,  $[\langle] \perp$  hold. But without something like the limit assumption, there is no guarantee that maximal states exist in models. We can still define the operations with PDL-transformations, as we did for all other operations, but unless we have some means of ensuring the existence of maximal worlds, contraction even with atomic information might not be successful. Now, the way we have proposed to get the limit assumption is by introducing the Löb axiom. We know that our logic is sound over the appropriate class of frames in which there are maximal worlds, but we do not know how to prove completeness. Setting this technical question aside for future research, we proceed in this section assuming that models always have maximal worlds, which we identify as those worlds in which  $\max(\neg\varphi)$  or  $[\langle] \perp$  is true.

Our first contraction operation is very simple. It has been studied by various authors and is perhaps best known under the names *severe withdrawal* (Pagnucco and Rott [39]) and *mild contraction* (Levi [35]). The incisions into sets of beliefs induced by severe withdrawal are substantially greater than those induced by (iterable generalisations of) AGM contraction functions.

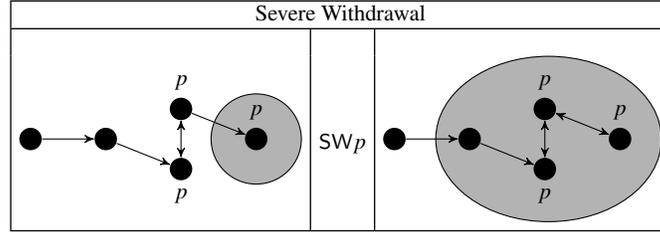
Severe withdrawal	
SW $\varphi$	$\leq := \leq \cup (\sim; [\langle] \varphi?)$ $< := (\neg[\langle] \varphi?; <)$

Notice that severe withdrawal has no need of identifying maximal  $\top$ -worlds. But if the limit assumption is not met, then there may be no maximal  $\neg\varphi$ -worlds and severe withdrawal may weaken the beliefs of the agent without getting rid of  $\varphi$ . Figure 8 gives an illustration of a successful severe withdrawal:

It is easy to check that a revision by comparison  $\text{RbC}_\psi \varphi$  reduces to a severe withdrawal of the reference sentence,  $\text{SW}\psi$ , if we substitute  $\perp$  for  $\varphi$ .<sup>29</sup>

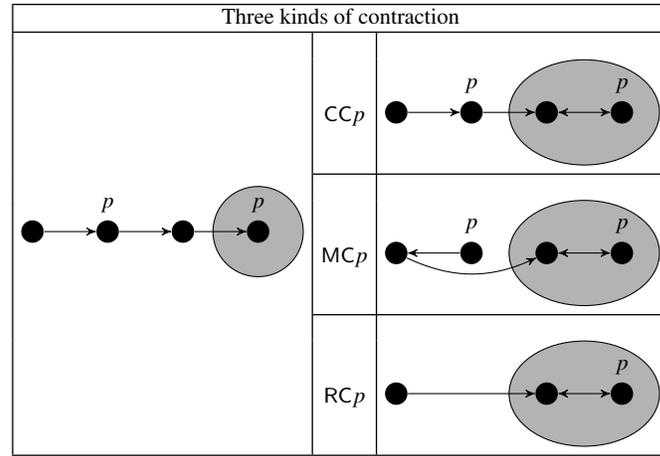
The following three kinds of contraction are modelled in analogy to conservative, moderate and radical revision. In line with the basic AGM theory, the way the corresponding contraction operations proceed is by putting the maximal  $\neg\varphi$ -worlds and

<sup>29</sup> Notice that the transformation  $(\neg[\leq] \psi?; \leq) \cup (\sim; [\langle] \psi?)$  is identical to the transformation  $\leq \cup (\sim; [\langle] \psi?)$ , because  $M, w \models [\leq] \psi$  and  $w \leq v$  taken together imply  $M, v \models [\langle] \psi$ .



**Fig. 8** Severe withdrawal of  $p$ . Models are closed under transitivity and reflexivity.

the maximal  $\top$ -worlds on a par, in a maximal position. We now move to the investigation of conservative, moderate and radical contraction.



**Fig. 9** Four kinds of contractions with respect to  $p$ . Models are closed under transitivity and reflexivity.

Conservative contraction, like conservative revision above, keeps most of the structure intact and reorders maximal worlds. First, the order is preserved among the non-maximal  $\neg\varphi$ -worlds. Second, the maximal  $\neg\varphi$ -worlds are upgraded on top of non-maximal  $\varphi$ -worlds and made as plausible as the maximal  $\top$ -worlds. Formally, conservative contraction is the following doxastic transformation:

Conservative contraction	
CC $\varphi$	$\leq := \leq^{\neg\max(\neg\varphi)} \cup (\sim; \max(\neg\varphi)?) \cup (\sim; \max(\top)?)$ $< := <^{\neg\max(\neg\varphi)} \cup ((\neg\max(\neg\varphi) \wedge \neg\max(\top))?) ; \sim; \max(\neg\varphi)?)$

Moderate contraction is defined in analogy to moderate revision, but it is hard to come up with a motivation for it. Why should the idea of being open-minded about  $\varphi$  result in a belief state that gives a lot of credit to  $\neg\varphi$ ? We present it for reasons of uniformity [43].

Moderate contraction	
MC $\varphi$	$\leq := \leq^{\neg\varphi} \cup \leq^{\varphi} \cup ((\varphi \wedge \neg\max(\top))?) ; \sim ; \neg\varphi?$ $\cup (\sim ; \max(\top)?) \cup (\sim ; \max(\neg\varphi)?)$
	$< := <^{\varphi} \cup <^{\neg\varphi} \cup ((\varphi \wedge \neg\max(\top))?) ; \sim ; \neg\varphi?$ $\cup ((\neg\max(\neg\varphi) \wedge \neg\max(\top))?) ; \sim ; \max(\top)?)$

We finally turn to radical contraction to which similar, and even stronger, cautionary remarks concerning its reasonableness apply.

Radical contraction	
RC $\varphi$	$ \neg\varphi \vee \max(\top) $
	$\leq := \leq^{\neg\varphi} \cup (\sim ; \max(\neg\varphi)?) ; \cup (\sim ; \max(\top)?)$ $< := <^{\neg\varphi} \cup ((\neg\varphi \wedge \neg\max(\neg\varphi))?) ; \sim ; \max(\top)?)$

## 6 Conclusion

We have explored AGM belief change policies in a modal dynamic logic that explicitly delineates knowledge, belief, plausibility and the dynamics of these notions. Taking a Kripke semantics counterpart to Grove semantics for AGM as a starting point, we used a basic modal language containing one epistemic modality  $[\sim]\varphi$  and two plausibility modalities  $[\leq]\varphi$  and  $[<]\varphi$ , and defined several notions of belief. We critically discussed the philosophical presuppositions underlying various modelling assumptions commonly made in the literature, such as negative introspection for knowledge and the limit assumption. Then, we introduced PDL-transformations to define various policies of iterated belief expansion, revision, contraction and two-dimensional belief change operations. EDPDL thus formalises our minimalist and modular attitude.

**Acknowledgements** We would like to thank Alexandru Baltag, Johan van Benthem and Michael Hillas for very valuable comments on a previous version of this paper. We dedicate this paper to Johan von Benthem, whom we admire as a researcher and who has been, in many and various ways, a friend and mentor to both of us. Thank you so much, Johan!

## References

1. Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.
2. Carlos E. Alchourrón and David Makinson. On the logic of theory change: Safe contraction. *Studia Logica*, 44:405–422, 1985.
3. Alexandru Baltag and Sonja Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science*, 165:5–21, 2006. Proceedings WoLLIC 2006.
4. Alexandru Baltag and Sonja Smets. The logic of conditional doxastic actions. In Krzysztof R. Apt and Robert van Rooij, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 9–31. Amsterdam University Press, 2008.
5. Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Logic and the Foundations of Game and Decision Theory (LOFT 7)*, volume 3 of *Texts in Logic and Games*, pages 11–58. Amsterdam University Press, 2008.
6. Alexandru Baltag, Hans van Ditmarsch, and Lawrence S. Moss. Epistemic logic and information update. In Pieter Adriaans and Johan van Benthem, editors, *Philosophy of Information*, Handbook of the Philosophy of Science, pages 361–455. Elsevier, 2008.
7. Johan van Benthem. Later than late: On the logical origin of the temporal order. *Pacific Philosophical Quarterly*, 63:193–203, 1982.
8. Johan van Benthem. *The Logic of Time. A Model-Theoretic Investigation into the Varieties of Temporal Antology and Temporal Discourse*. Reidel, 1983.
9. Johan van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-classical Logic*, 17:129–155, 2007.
10. Johan van Benthem, Patrick Girard, and Olivier Roy. Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic*, 38:83–125, 2009.
11. Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 17:157–182, 2007.
12. Johan van Benthem, Olivier Roy, and Sieuwert van Otterloo. Preference logic, conditionals, and solution concepts in games. In Lagerlund, Lindström, and Sliwinski, editors, *Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg*, volume 53. Uppsala Philosophical Studie, 2006.
13. Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge University Press, New York, NY, USA, 2001.
14. Oliver Board. Dynamic interactive epistemology. *Games and Economic Behavior*, 49:49–80, 2004.
15. Craig Boutilier. Unifying default reasoning and belief revision in a modal framework. *Artificial Intelligence*, 68:33 – 85, 1994.
16. Craig Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25:263–305, 1996.
17. Adnan Darwiche and Judea Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89:1–29, 1997.
18. Lorenz Demey. Some remarks on the model theory of epistemic plausibility models. *Journal of Applied Non-Classical Logics*, 21:375–395, 2011.
19. Hans van Ditmarsch. Prolegomena to dynamic logic for belief revision. *Synthese*, 147:229–275, 2005.
20. Hans van Ditmarsch, Wiebe van Der Hoek, and Barteld Kooi. *Dynamic epistemic logic*, volume 1. Springer New York, 2007.
21. Christopher W Doble, Jean-Paul Doignon, Jean-Claude Falmagne, and Peter C Fishburn. Almost connected orders. *Order*, 18:295–311, 2001.
22. Eduardo Fermé and Hans Rott. Revision by comparison. *Artificial Intelligence*, 157:5–47, 2004.

23. Michael J. Fischer and Richard E. Ladner. Propositional dynamic logic of regular programs. *Journal of Computer and System Sciences*, 18:194–211, 1979.
24. Peter Gärdenfors and David Makinson. Revisions of knowledge systems using epistemic entrenchment. In Moshe Vardi, editor, *TARK'88: Proceedings of the Second Conference on Theoretical Aspects of Reasoning About Knowledge*, pages 83–95. Morgan Kaufmann, Los Altos, CA, 1988.
25. Matthew L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30:35–79, 1986.
26. Patrick Girard, Jeremy Seligman, and Fenrong Liu. General dynamic dynamic logic. In Thomas Bolander, Torben Braüner, Silvio Ghilardi, and Lawrence S Moss, editors, *Advances in Modal Logics Volume 9*, pages 239–260, 2012.
27. Daniel Greco. Could KK be OK? *Journal of Philosophy*. forthcoming.
28. Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.
29. Joseph Y. Halpern. Should knowledge entail belief? *Journal of Philosophical Logic*, 25:483–494, 1996.
30. Jaakko Hintikka. *Knowledge and belief*, volume 414. Cornell University Press Ithaca, 1962.
31. Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52:263–294, 1991.
32. Philippe Lamarre and Yoav Shoham. Knowledge, certainty, belief, and conditionalisation (abbreviated version). *Proceedings of KR'94*, pages 415–424, 1994.
33. Hannes Leitgeb and Krister Segerberg. Dynamic doxastic logic: Why, how, and where to? *Synthese*, 155:167–190, 2007.
34. Wolfgang Lenzen. Epistemologische Betrachtungen zu [S4,S5]. *Erkenntnis*, 14:33–56, 1979.
35. Isaac Levi. *Mild Contraction: Evaluating Loss of Information due to Loss of Belief*. Oxford University Press, Oxford, 2004.
36. David K. Lewis. *Counterfactuals*. Oxford, Blackwell, 1973.
37. Abhaya Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.
38. Samir Okasha. On a flawed argument against the KK principle. *Analysis*, 73:80–86, 2013.
39. Maurice Pagnucco and Hans Rott. Severe withdrawal – and recovery. *Journal of Philosophical Logic*, 28:501–547, 1999.
40. Jan Plaza. Logics of public communications. *Synthese*, 158:165–179, 2007.
41. Vaughan R. Pratt. Semantical considerations on Floyd-Hoare logic. Technical report, MIT, Cambridge, MA, USA, 1976.
42. Hans Rott. Revision by comparison as a unifying framework: Severe withdrawal, irrevocable revision and irrefutable revision. *Theoretical Computer Science*, 355:228–242, 2006.
43. Hans Rott. Shifting priorities: Simple representations fortwenty-seven iterated theory change operators. In David Makinson, Jacek Malinowski, and Heinrich Wansing, editors, *Towards Mathematical Philosophy*, volume 28 of *Trends in Logic*, pages 269–296. Springer Netherlands, 2009.
44. Hans Rott. Bounded revision: Two-dimensional belief change between conservative and moderate revision. *Journal of Philosophical Logic*, 41:173–200, 2012.
45. Krister Segerberg. *An essay in classical modal logic*, volume 1 of *Filosofiska studier utgivna av Filosofiska föreningen och Filosofiska institutionen vid Uppsala universitet*. Filosofiska föreningen och Filosofiska institutionen vid Uppsala universitet, 1971.
46. Krister Segerberg. Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39:287–306, 1998.
47. Robert Stalnaker. A theory of conditionals. In Nicholas Rescher, editor, *Studies in Logical Theory*, volume 2 of *APQ Monograph Series*, pages 98–112. Blackwell, Oxford, 1968.
48. Robert Stalnaker. Knowledge, belief and counterfactual reasoning in games. *Economics and Philosophy*, 12:133–163, 1996.
49. Robert Stalnaker. On logics of knowledge and belief. *Philosophical Studies*, 128:169–199, 2006.