Synthese (2007) 155:167–190
Knowledge, Rationality & Action 5–28
DOI 10.1007/s11229-006-9143-8

ORIGINAL PAPER

Dynamic doxastic logic: why, how, and where to?

Hannes Leitgeb · Krister Segerberg

Received: 2 November 2006 / Accepted: 18 December 2006 / Published online: 10 February 2007

Abstract We investigate the research programme of dynamic doxastic logic (DDL) and analyze its underlying methodology. The Ramsey test for conditionals is used to characterize the logical and philosophical differences between two paradigmatic systems, AGM and KGM, which we develop and compare axiomatically and semantically. The importance of Gärdenfors's impossibility result on the Ramsey test is highlighted by a comparison with Arrow's impossibility result on social choice. We end with an outlook on the prospects and the future of DDL.

Keywords Dynamic doxastic logic · Theory change · Belief revision · Belief update · Ramsey test · Arrow's theorem

1 Introduction

A new and extremely influential theory of belief change was introduced by Alchourrón et al. (1985) (from now on AGM) and defended and developed by Gärdenfors (1988) as a formal theory in which belief change on the basis of new evidence is studied axiomatically. In particular, AGM introduced a revision operator $*$, regarded as a mapping from a set of formulæ and a formula to another set of formulæ. The result of revising the so-called belief set K by evidence A, denoted $K * A$, was supposed to be the "minimal mutilation" of K that would include A.

Dynamic doxastic logic (DDL) was introduced with the aim of representing the *meta*-linguistically expressed belief revision operator $*$ as an *object*-linguistic sentence

H. Leitgeb (✉)
Department of Philosophy, University of Bristol, 9 Woodland Road,
Bristol BS8 1TB, UK
e-mail: Hannes.Leitgeb@bristol.ac.uk

K. Segerberg
Filosofiska institutionen, Uppsala Universitet, Box 627 751 26,
Uppsala, Sweden
e-mail: Krister.Segerberg@filosofi.uu.se

operator $[*\_]$ in the style of dynamic modal logic (Segerberg, 1998, 1999; see also van Linder et al., 1995). If ϕ and ψ are formulæ of an object language, then the formula $[*\phi]\psi$ is meant to express that after the revision by ϕ it is the case that ψ, where $[*\phi]\psi$ is located on the same language level as ϕ and ψ. Also other doxastic actions can be studied in terms of the modal-logical systems they give rise to. The main examples are belief contraction and expansion, which are closely related to AGM style belief revision. Other examples are found in a theory developed by Grahne, Katsuno and Mendelzon in response to the AGM paradigm and referred to in this paper as KGM (see in particular Grahne, 1991; Katsuno and Mendelzon, 1992).[1]

In the following we will consider the current "state of the art" of DDL, investigate its underlying motivation and guiding ideas, and look at its future prospects. In Sect. 2 we will analyze the research programme of DDL, and we will enumerate what we think are the main reasons for pursuing it over and above the traditional theory of belief revision. Section 3 starts by focusing on the logical problems that are associated with the Ramsey test for conditionals while highlighting the rôle which these problems have played for the bifurcation of DDL into two logical base systems that correspond to AGM and KGM. After giving a brief sketch of a new result which indicates that Gärdenfors' impossibility result for belief revision with conditionals is logically related to Arrow's impossibility result on social choice, we outline the basic axiomatic and semantic DDL analogues of AGM and KGM in detail and compare the two from the viewpoint of DDL. We close with an outlook in Sect. 4 on the future of DDL (for example, considering its relationship to dynamic epistemic logic (DEL) along the lines of Baltag et al., 1998, and subsequent approaches as van Benthem, forthcoming).

2 Why: object-linguistic vs. meta-linguistic treatments of belief change

Why should we express doxastic actions such as belief revision on the object language level, that is, on the same language level as, for example, alethic modalities in traditional modal logic?

In order to see clearly where the advantages of this strategy lie, let us have a look first at how standard claims about revision can be translated into the language of DDL. Instead of saying meta-linguistically that

$$\theta \in K * \phi,$$

a dynamic object-linguistic sentential operator $[*\_]$ and a static sentential belief operator $\mathbf{B}$ are introduced and applied to two formulæ ϕ and θ:

$$[*\phi]\mathbf{B}\theta$$

According to the intended reading of $[*\_]$ and $\mathbf{B}$, the resulting formula conveys the same information as the meta-linguistic statement; that is, θ is believed after the revision by ϕ. Thus the difference between employing expressions on the object and on the metalanguage level is negligible so far. But now let us assume we do not just intend to study the logic of belief revision itself but also the logical laws and rules that govern our *beliefs about belief revision*. Using the standard syntactic machinery

[1] The KGM-paradigm is called the KM-paradigm in Lindström and Segerberg (2006). However, the more inclusive term employed here seems more appropriate.

of modal logic it is obvious how to do that: for example,

$$\mathbf{B}[*\phi]\mathbf{B}\theta \tag{1}$$

expresses that our cognitive agent—the one whose doxastic states and actions we investigate—believes that after the revision by ϕ she is going to believe that θ. The traditional theory of belief revision on the other hand lacks the syntactic resources to express a corresponding claim. The theory was simply never designed to handle belief sets K' that include sentences of the form '$\theta \in K * \phi$' as *elements* (where K might be identical to K' or distinct from it). Accordingly, instances of nested revision such as

$$[*[*\phi]\mathbf{B}\theta]\mathbf{B}\chi \tag{2}$$

are syntactically unproblematic from the "modal" viewpoint of DDL. What this latter example formula says is: the agent believes that χ is the case after revising her belief state by the information that a revision by ϕ leads to a belief in θ.[2] Again the classic theory of belief revision does not offer the necessary syntactic machinery to deal with sentences of such form. The only way to make such statements accessible to traditional belief revision would be by means of quotational or other syntactic devices, for example, by expressing claim (2) in terms of: $\chi \in K' * \ulcorner \theta \in K * \phi \urcorner$. Here, '$\ulcorner \theta \in K * \phi \urcorner$' would be a singular term denoting a formula, in correspondence with 'ϕ' in '$K * \phi$' being a singular term that denotes a formula. But just as formalizations of modal logic on the basis of sentential operators are preferable to the more complex formalizations in terms of modal predicates as long as quantification over formulæ or propositions is not of central importance, the DDL way of handling such claims seems like a viable alternative.[3]

So we find that the language of DDL is, in an important respect, more expressive than the language of traditional belief revision. In a different respect, the latter turns out to be more expressive than the former and indeed *overly* expressive: this is because standard belief revision uses variables such as 'K' to denote *sets of formul*æ, it employs the binary predicate for *set membership*, and it presupposes predicates for proof-theoretic and semantic concepts such as *consistency* and *logical truth* which are usually defined within the language of set theory. While it is certainly convenient to have these resources at one's linguistic disposal, they leave the proof-theoretic control of the axiomatic system of belief revision severely affected. Moreover, presupposing such set theoretic resources proves to be unnecessary in view of the parsimonious vocabulary of DDL. Basically, the language of propositional logic extended by a few unary or binary sentential operators is all that is needed in order to state the laws of belief revision and of other doxastic operations formally. As well-known results on the metatheory of modal logic show, such languages can express fragments of first- or higher-order logic while still remaining feasible from a proof- and complexity theoretic viewpoint.

[2] We speak here of the revision of a belief *state* rather than of a belief *set*, since the analysis of the doxastic structure of an agent in terms of belief sets K turns out to be too weak for the semantics of DDL. It was one of the insights of research done on iterated belief revision that the doxastic state of an agent should not be considered to be given by a belief set, i.e., by a set of believed formulæ, but rather by a preference ranking of formulæ that encodes both a belief set and a disposition of how to change beliefs in the light of new evidence.

[3] For the development of modal logic in terms of modal predicates and its advantages or disadvantages over the standard operator approach, see Leitgeb (Forthcoming).

Another difference between the language of standard belief revision and the language of DDL lies in the *indexical* character of the latter. Reconsider example (2) from above: $[*[*\phi]\mathbf{B}\theta]\mathbf{B}\chi$ tells us that, in the agent's *present* state of belief, she would believe χ, if it were revised by the information that the revision of her *then present* belief state by ϕ would lead to a belief in θ. All of this can be expressed in DDL without making the reference to the underlying belief states explicit. This is particularly useful in cases where a formula is intended to describe the changing of belief states on the basis of doxastic actions, such that the different parts of the formula refer to different belief states, but where at the same time no further specification of the belief states involved is necessary—for example, if a formula is supposed to occur within a logical law that is meant to hold for *every* belief change whatsoever. In contrast, even if the standard theory of belief revision were extended in a way such that it would be applicable to formulæ like $\chi \in K' * \ulcorner\theta \in K * \phi\urcorner$, one would still be forced to make the reference to belief states explicit; additionally, the formal quotation marks $\ulcorner$ and $\urcorner$ would have to be Quine corners, that is, syntactical operations, rather than proper quotation marks, since we would need to quantify into them. As the great success of possible worlds semantics shows, it proves to be useful to avoid all these problems by treating object language operators such as $[*\_]$ indexically, thereby eliminating one of the argument places of the traditional revision operator $*$, and to push the reference to worlds, accessibility relations, and other semantic objects up to the metalanguage level.

So far we have argued that several of the advantages of the language of modal logic carry over to the language of DDL. But actually the mere fact that DDL is to some extent like modal logic is *in itself* a good reason for pursuing the logical study of doxastic change in this manner. In the last few decades modal logic has developed into a mature logical framework of immense richness, and the possibility of transferring some of its classic metatheoretical results and techniques to the investigation of belief revision, belief update, and the like is certainly attractive. Moreover, the analogy with modal logic might also be useful for the further *conceptual* development of belief revision. For example, once a correspondence between $\Box$ and $*$ in terms of the sentential operator $[*\_]$ is established, it is clear that there must be an equally interesting "dual revision operator" $\langle *\_\rangle$ that stands to $[*\_]$ as $\Diamond$ stands to $\Box$ and which can be defined in terms of $[*\_]$, namely, as $\neg[*\_]\neg$.

Summarizing, we end up with the following list of points that establish DDL as an interesting alternative to traditional belief revision:

Dynamic doxastic logic is useful or even necessary for

- studying beliefs about revision and studying nested revision
- proof- and complexity theoretic control over linguistic expressiveness
- indexical treatment of doxastic actions
- developing belief revision in analogy to modal logic

This is not to say that DDL is preferable to standard belief revision in each and every context. For some purposes the study of nested revision is not needed, and in such cases the higher degree of expressiveness of DDL with respect to nesting does not show up as an advantage. For example, compare the theoretically and practically successful formalization of default information by computer scientists in terms of meta-linguistic non-monotonic consequence relations $\mid\!\sim$, the properties of which are inter-translatable with properties of the standard belief revision operator (as was shown by Gärdenfors and Makinson, 1994). In that context, standard belief revision and DDL would simply fare equally well. The general upshot of the discussion in this section is rather that, while DDL does not do worse than AGM style revision in most contexts, in some contexts the further development of belief revision within DDL has a clear payoff.

The areas in which DDL goes beyond traditional belief revision are more or less new logical territory. While there are various suggestions of AGM type accounts of *iterated* belief revision as in

$$\theta \in (K * \phi) * \psi,$$

there is no established standard of iterated belief revision, let alone a DDL counterpart of such a standard (though DDL can be seen to supply a *minimal* logic of iterated revision). The situation is even worse with respect to the logic of *doxastic attitudes towards* belief revision (as for example (1) above) or the logic of *nested* belief revision (as for example (2)). Accordingly, we will limit ourselves to restricted fragments of the full and unrestricted language of DDL when we turn to the development of DDL in terms of axiomatic and semantic systems in the next section.

3 How: axiomatic and semantic systems of DDL

3.1 The Ramsey test: a division of paradigms

As explained in the previous section, the driving idea of DDL is that formulæ such as $[*\phi]\theta$ are used to express doxastic actions on the same linguistic level on which also the arguments and the outcomes of these doxastic actions are expressed. Since DDL is furthermore motivated by the strong analogy of the operators $[*\phi]$ with standard modal operators, the intended semantics of DDL will be a possible worlds semantics, according to which a set of possible worlds or points is assigned to each such formula $[*\phi]\theta$.[4] Thus, not just factual formulæ such as ϕ and θ will be regarded as expressing sets of possible worlds but also doxastic formulæ as $[*\phi]\theta$. For this reason DDL is bound to face a serious logical challenge: the danger of getting entangled in the potentially paradoxical consequences of combining belief revision for an object language F with a representation of the revision operator in terms of formulæ in F.

The possibly devastating effects of such a combination first showed up when Gärdenfors considered a doxastic interpretation of conditionals in terms of the so-called Ramsey test for conditionals:

Ramsey test $\phi \Rightarrow \theta \in K$ iff $\theta \in K * \phi$

The problem is that if the AGM axioms are formulated for a belief revision operator that applies to a propositional language with a new conditional sign $\Rightarrow$, and if these axioms are subsequently extended by the Ramsey test for $\Rightarrow$, a contradiction can be derived in the resulting system:

Theorem 1 (Gärdenfors, 1986) *The AGM axioms of $*$ are inconsistent with the Ramsey test for conditionals in any modelling allowing for at least three propositions which are pairwise consistent but jointly inconsistent.*

[4] Actually, each of these "possible worlds" will be regarded as consisting of two independent components, namely a doxastic component and an environmental one; compare Sects 3.3 and 3.4.

(Cf. Segerberg, 1989 for a logical reconstruction, and Lindström and Rabinowicz, 1995 for an analysis and overview of possible responses to Gärdenfors's result.)

The relevance of Gärdenfors's theorem for DDL lies in the fact that, in order to maintain consistency, either

(i) the logical axioms and rules for **B** and $[*\_]$ must not allow the derivation, for all ϕ and θ, of a formula of the form

$$\mathbf{B}(\chi[\phi,\theta]) \leftrightarrow [*\phi]\mathbf{B}\theta$$

where $\chi[\phi,\theta]$ is some formula that is built syntactically from ϕ and θ,

or

(ii) the logical axioms and rules for $[*\_]$ must not conform to the AGM postulates (assuming that the doxastic logic of **B** is fixed).

In both cases, by the guiding idea of DDL, all formulæ of the form $[*\phi]\mathbf{B}\theta$ should still lie on the same linguistic level as ϕ and θ. But while in case (i) not all formulæ of the form $[*\phi]\mathbf{B}\theta$ express propositions that can be expressed by static belief formulæ, in case (ii) not all formulæ of the form $[*\phi]\mathbf{B}\theta$ express belief revision in the sense of AGM.

The alternative (i) can be seen to be a consequence of the DDL account of AGM that we introduce in Sect. 3.3. In contrast, the DDL version of KGM, which we focus on in Sect. 3.4, is an instance of the alternative (ii) (and consequently its dynamic operator will be denoted by '$[\star\_]$' rather than by '$[*\_]$'). In this sense, the acceptance or denial of Ramsey test-like postulates corresponds to the division of DDL into the two paradigmatic systems that we are going to explain and compare in the next sections. We will return to the Ramsey test and its plausibility from the viewpoint of AGM and KGM at the end of Sect. 3.5. Before we turn to the DDLs underlying AGM and KGM, we will compare Gärdenfors's theorem on the Ramsey test with another, even more famous impossibility result: Arrow's theorem on social rankings.

3.2 Gärdenfors vs. Arrow

The comparison between, on the one hand, Gärdenfors's limitative result on belief revision and the Ramsey test, and on the other hand, Arrow's limitative result on social choice and the Non-Dictatorship condition, would of course merit a much more detailed discussion than we are able to give in view of the overall aims of this paper [but see Leitgeb (2005, unpublished manuscript), for a more thorough development of their logical relationship and for the proof of theorem 2 below]. We will restrict ourselves merely to a motivation of the claim that the two results seem to reflect a common underlying formal pattern.

Let us start with standard belief revision. According to Grove's (1988) well-known representation theorem, revision operators $*$ can be put into one-to-one correspondence to spheres models or to ranked models of possible worlds, that is, semantic objects of the following kind (Fig. 1).

Instead of proper spheres, we use a graphical representation in terms of layers or ranks: the lowest layer corresponds to the innermost sphere, taking the union of the lowest layer with the second layer from below corresponds to the next larger sphere, and so forth. K is the set of formulas that are true in all worlds which are members

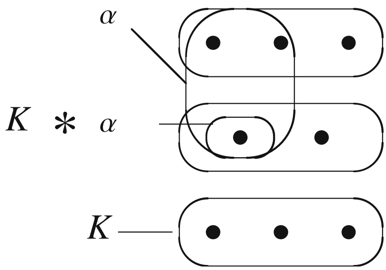


Fig. 1 Spheres models for belief revision

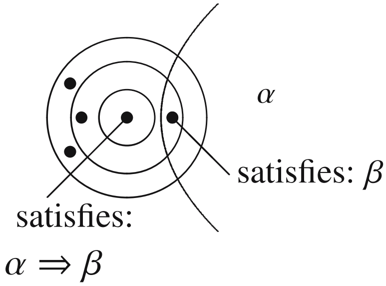


Fig. 2 Spheres models for conditionals

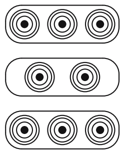


Fig. 3 Spheres models for conditionals and belief revision

of the lowest layer; $K * \alpha$ is the set of formulas which are satisfied by all those worlds that have minimal rank among the worlds that satisfy α.

Accordingly, the logic of subjunctive conditionals was proved valid by Lewis (1973) with respect to a sphere semantics that is similar to the semantics for $*$. Such sphere models can be depicted as shown in Fig. 2.

Hence, the counterfactual conditional $\alpha \Rightarrow \beta$ is true in the world w the spheres system of which we are looking at, as β is satisfied by all those worlds that are most similar to w among the worlds that make α true.

Now let us assume that we intend to apply the belief revision operator to formulæ in a language with non-material conditionals where these conditionals obey a Lewis-style semantics. In such a case, the ranked models for $*$ actually look more like this (Fig. 3).

Every world is "surrounded" by a sequence of Lewis' spheres by which conditionals can be evaluated in these worlds. (In order to keep the diagram concise we did not actually draw all the worlds that populate these spheres.)

While the left-hand side of the Ramsey test expresses a constraint on the "small" rankings of worlds given by the spheres around the worlds in the lowest layer, the right-hand side of the Ramsey test tell us something about the "big" ranking that defines the semantics of the revision operator. Gärdenfors's impossibility theorem may thus be seen to express the fact that the "small" rankings cannot correspond to the "big" ranking along the lines suggested by the Ramsey test. But this sounds familiar in view of Arrow's theorem (Arrow, 1963), which says that there is no function that extends any given set of individual rankings $\leqslant_i \subseteq A \times A$ (for fixed individuals $i \in N = \{1, \ldots, n\}$ and a set A of alternatives) to a social ranking $\leqslant \subseteq A \times A$, such that certain axioms are satisfied (e.g., the Pareto condition: if $x \leqslant_i y$ for all i, then $x \leqslant y$; the non-existence of a "dictator" i; and so forth). In fact one can show that this is more than just an analogy.

In what follows, let $\mathcal{L}$ be a classic propositional language with at least two and at most finitely many propositional variables. Let $\mathcal{L}_{\Rightarrow}$ be the language $\mathcal{L}$ extended by (and closed under) a new conditional sign $\Rightarrow$. The expansion operator $+$ takes subsets of $\mathcal{L}_{\Rightarrow}$ and formulæ in $\mathcal{L}_{\Rightarrow}$ as arguments, i.e., for all $K \subseteq \mathcal{L}_{\Rightarrow}$, for all $\alpha \in \mathcal{L}_{\Rightarrow}$: $K + \alpha := Cn(K \cup \{\alpha\})$, where Cn is logical closure in conditional logic (which we also refer to as: $\mathcal{L}_{\Rightarrow}$-deductive closure). The conditional logic in question can be much weaker than Lewis's standard system; for example, the axiomatic counterpart of the "centeredness" condition on spheres is not needed. (That is why we did not use Lewis's symbol '$\Box\!\!\rightarrow$' but rather '$\Rightarrow$' to denote our non-material conditional sign.) We also speak of $\mathcal{L}$-consistency and $\mathcal{L}_{\Rightarrow}$-consistency: the former is consistency in classic propositional logic, the latter consistency in the chosen system of conditional logic. Accordingly, an $\mathcal{L}$-world ω is a maximally $\mathcal{L}$-consistent set of formulæ in $\mathcal{L}$, whereas an $\mathcal{L}_{\Rightarrow}$-world w is a maximally $\mathcal{L}_{\Rightarrow}$-consistent set of formulæ in $\mathcal{L}_{\Rightarrow}$. Hence, an $\mathcal{L}_{\Rightarrow}$-world w can be an extension of an $\mathcal{L}$-world ω, in the sense that $w \cap \mathcal{L} = \omega$. Worlds are assumed to satisfy formulas if and only if they contain these formulas as members.

We will now reconstruct Arrow's theorem in a system of belief revision for a language with the conditional operator $\Rightarrow$. The rôles of the set A of alternatives and the set N of individuals in Arrow's social choice account will be played by sets of $\mathcal{L}$-worlds; for simplicity, we choose these sets to be both identical to the set of *all* $\mathcal{L}$-worlds. Since $\mathcal{L}$ was assumed to contain at least two and at most finitely many propositional variables, both A and N are finite sets with at least four members. The ordering of the set of "alternatives" by an "individual" ω can be represented by a spheres system around ω, or, equivalently: by extending ω to an $\mathcal{L}_{\Rightarrow}$-world w, as the set of conditionals satisfied by w encodes a spheres system in light of Lewis's completeness theorem. Finally, the social ranking of "alternatives" corresponds to the ranking of $\mathcal{L}_{\Rightarrow}$-worlds as being given by the belief revision operator, where belief sets have to be understood as $\mathcal{L}_{\Rightarrow}$-deductively closed subset of $\mathcal{L}_{\Rightarrow}$ accordingly. Since N is intended to be the set of all $\mathcal{L}$-worlds, no belief set is allowed to contain a non-tautological formula of $\mathcal{L}$—therefore, distinct belief sets differ only in terms of the conditionals they contain. The ranking of alternatives by an individual ought to be determined uniquely, so there should not be a belief set K and an $\mathcal{L}$-world ω, such that there are two distinct $\mathcal{L}_{\Rightarrow}$-worlds w and w' with $K \cup \omega \subseteq w, w'$. Furthermore, as we want A to be the set of all $\mathcal{L}$-worlds as well, if $\gamma_\omega \in \mathcal{L}$ is the state description of an $\mathcal{L}$-world ω, then no belief set shall include an $\mathcal{L}_{\Rightarrow}$-formula of the form $\gamma_\omega \rightarrow (\alpha \Rightarrow \bot)$ with $\alpha \in \mathcal{L}$ being $\mathcal{L}_{\Rightarrow}$-consistent and $\bot$ being $\mathcal{L}_{\Rightarrow}$-inconsistent. These are the constraints on belief sets that we are going to presuppose.

Keeping in mind the intended identifications between items of the social choice setting and those in the belief revision framework for conditionals, one can show that Arrow's assumptions in his Dictator theorem have the following revision counterparts:

- (Independence of Irrelevant Alternatives; IIA)
  For all $\mathcal{L}_{\Rightarrow}$-consistent belief sets K, K', for all $\alpha \in \mathcal{L}$:
  If

– for all $\mathcal{L}_{\Rightarrow}$-worlds w, w', such that (i) w is $\mathcal{L}_{\Rightarrow}$-consistent with K, (ii) w' is $\mathcal{L}_{\Rightarrow}$-consistent with K', and (iii) w and w' satisfy precisely the same formulæ in $\mathcal{L}$, it holds that
for all $\mathcal{L}$-worlds ω, ω' which are $\mathcal{L}$-consistent with α,
$w \models (\omega \vee \omega') \Rightarrow \omega$ iff $w' \models (\omega \vee \omega') \Rightarrow \omega$,
then for all $\beta \in \mathcal{L}$: $\beta \in K * \alpha$ iff $\beta \in K' * \alpha$.

- (Pareto; P)
For all $\mathcal{L}_{\Rightarrow}$-consistent belief sets K, for all $\alpha, \beta \in \mathcal{L}$:
if $\alpha \Rightarrow \beta \in K$, then $\beta \in K * \alpha$.
- (Non-Dictatorship; ND)
There is no $\mathcal{L}$-world ω, such that:
for all consistent belief sets K, for all $\mathcal{L}_{\Rightarrow}$-worlds w which are $\mathcal{L}_{\Rightarrow}$-consistent with K and which satisfy ω, for all $\alpha, \beta \in \mathcal{L}$:
if $w \models \alpha \Rightarrow \beta$, then $\beta \in K * \alpha$.

It follows that there is a belief revision counterpart of Arrow's theorem:

Theorem 2 *$K*1$–$K*8$ (see Gärdenfors, 1988), IIA, P, ND are jointly inconsistent* (*given our background assumptions on belief sets and on the number of possible worlds involved*).

The Pareto condition P that was introduced above is just the left-to-right direction of the Ramsey test for conditionals. Independence of Irrelevant Alternatives and Non-Dictatorship are harder to interpret from the revision-theoretic point of view, but the result by itself should be sufficient to confirm that Gärdenfors's and Arrow's Impossibility theorems are related logically, which should mutually strengthen the importance that these two theorems had in the fields from which they originated, and also beyond.

3.3 Basic iterative belief change I: the AGM paradigm

3.3.1 Basic language

In DDL we encounter several object languages, depending on what abilities the agent or the agents are supposed to have. In the case studied in this section two assumptions are made: (i) the agent is able to hold beliefs only about the environment in which he moves, an environment that is not supposed to include his beliefs, and (ii) he will change his beliefs only with regard to information about the environment. These assumptions explain the cumbersome definitions that follow.

The set pB of *pure Boolean formul*æ are those formulæ that are built exclusively from propositional letters and truth-functional connectives (which are also called Boolean operators). The set F of all formulæ is the smallest set to satisfy the following constraints:

$\mathsf{pB} \subseteq \mathsf{F}$,
F is closed under Boolean operators,
if $\phi \in \mathsf{pB}$ then $\mathbf{B}\phi \in \mathsf{F}$ and $\mathbf{K}\phi \in \mathsf{F}$,
if $\phi \in \mathsf{pB}$ and $\theta \in \mathsf{F}$ then $[*\phi]\theta \in \mathsf{F}$.

Thus in addition to truth-functional operators we have *modal operators*: on the one hand the *static doxastic operators* $\mathbf{B}$ and $\mathbf{K}$, on the other hand for every Boolean formula ϕ the *dynamic operator* $[*\phi]$. The static operators **B** and **K** are intended to formalize revisable and nonrevisable belief, respectively. To coin a slogan: **K** stands for "knowledge", not for 'knowledge'. (What we call knowledge in ordinary life situations is often just that: beliefs that we are not prepared to revise, at least not then and there.) Note that there are no nestings of doxastic operators in basic DDL. (In so-called one-shot DDL there are no iterations of dynamic operators either; cf. Segerberg, 1999.) We write **b** and **k** for the duals of **B** and **K**, and $\langle *\phi \rangle$ for the dual of $[*\phi]$. Thus **b** and **k** may be seen as short for $\neg\mathbf{B}\neg$ and $\neg\mathbf{K}\neg$, respectively, while $\langle *\phi \rangle$ is short for $\neg[*\phi]\neg$.

Convention: Throughout this paper, whenever we display formulæ or formula schemata we assume, without necessarily saying so explicitly, that the expressions displayed represent well-formed formulæ. For example, if we discuss an expression $[*\phi](\psi \to \mathbf{B}\theta)$, we take it for granted that ϕ and θ are pure Boolean.

3.3.2 Basic axiom system

Our axiom system consists of four blocks: classical, modal, AGM proper, and extra. Each block contains certain postulates, that is, axioms (described by means of axiom schemata), and rules of inference.

The axioms of the classical block are the tautologies of classical propositional logic; the single rule is modus ponens (MP). The postulates of the modal block should yield, for each modal operator $\triangle \in \{\mathbf{B}, \mathbf{K}, [*\phi] \colon \phi \in \mathsf{pB}\}$, the smallest normal modal logic K. For example, it is a standard fact about normal modal logics that the following postulates would suffice:

$\triangle(\phi \wedge \psi) \leftrightarrow (\triangle\phi \wedge \triangle\psi)$,
$\triangle\top$,
if $\phi \leftrightarrow \psi$ is a theorem, then $\triangle\phi \leftrightarrow \triangle\psi$ is also a theorem.

The block AGM proper corresponds, clause for clause, to the original AGM-postulates into DDL language (our numbering follows that of the original list of postulates):

(∗2) $[*\phi]\mathbf{B}\phi$,

(∗3) $[*\top]\mathbf{B}\phi \to \mathbf{B}\phi$,

(∗4) $\mathbf{b}\top \to (\mathbf{B}\phi \to [*\top]\mathbf{B}\phi)$,

(∗5) $[*\phi]\mathbf{B}\bot \to \mathbf{K}\neg\phi$,

(∗6) $\mathbf{K}(\phi \leftrightarrow \psi) \to ([*\phi]\mathbf{B}\theta \leftrightarrow [*\psi]\mathbf{B}\theta)$,

(∗7) $[*(\phi \wedge \psi)]\mathbf{B}\theta \to [*\phi]\mathbf{B}(\psi \to \theta)$,

(∗8) $\langle *\phi \rangle \mathbf{b}\psi \to ([*\phi]\mathbf{B}(\psi \to \theta) \to [*(\phi \wedge \psi)]\mathbf{B}\theta)$.

The final extra block consists of postulates that seem to be (more or less) implicit in the original AGM theory:

(∗0) $\theta \leftrightarrow [*\phi]\theta$, if θ is pure Boolean,

(∗DF) $\langle *\phi\rangle\theta \leftrightarrow [*\phi]\theta$,

(∗KB) $\mathbf{K}\phi \rightarrow \mathbf{B}\phi$,

(∗KK) $\mathbf{K}\theta \leftrightarrow [*\phi]\mathbf{K}\theta$.

Revision by tautology, the doxastic action denoted by $*\top$ (called "consolidation" by Sven Ove Hansson), makes a difference only if the agent's belief set is inconsistent. The postulate (∗DF) reflects the fact that AGM-style belief revision commits itself to a total-functional notion of revision according to which every input leads to unique revision output; as Rabinowicz and Lindström (1994) have shown, it is possible to develop belief revision without adopting this constraint. The postulate (∗KK) reflects another assumption that may or may not have been held by the founding fathers of AGM: that what they called the background theory (our nonrevisable beliefs) is not modified by ordinary belief change.

A formula ϕ is *provable* in this system if there is a sequence $\psi_0, \ldots, \psi_n$ of formulæ such that, for all $i \leqslant n$, either (i) ψ_i is a postulate, or (ii) there are $j, k < i$ such that $\psi_k = (\psi_j \rightarrow \psi_i)$, or (iii) there are θ and θ' such that $\psi_i = (\triangle\theta \leftrightarrow \triangle\theta')$ and, for some $j < i, \psi_j = (\theta \leftrightarrow \theta')$, where $\triangle$ is $\mathbf{B}$ or $\mathbf{K}$ or $[*\tau]$, for some $\tau \in \mathsf{pB}$. Furthermore, ϕ is *derivable* in this system from a set Σ of formulæ if there are formulæ $\sigma_0, \ldots, \sigma_{n-1} \in \Sigma$ such that $(\sigma_0 \wedge \cdots \wedge \sigma_{n-1}) \rightarrow \phi$ is provable.[5] Finally, a set Σ of formulæ is *consistent* in the system if $\perp$ is not derivable from Σ.

3.3.3 Semantics

In his extraordinary book (Lewis, 1973), David Lewis suggested several types of modellings—among them sphere systems, selection functions and entrenchment orderings—that can be used for the semantical analysis of important kinds of conditional logic. Given the connexion between conditionals and belief change, it is not surprising that Lewis's modellings can be adapted to the analysis of belief change logic. Here we choose the sphere systems modelling as being (in our subjective view) the most visual.

We employ the rudiments of topology. A *topological space* (U, T) is a set U together with a topology T, that is, a set of subsets of U that is closed under finite intersection and arbitrary union. (Since $\bigcap \emptyset = U$ and $\bigcup \emptyset = \emptyset$, it follows that both U and $\emptyset$ are elements of T.) One may refer to U by itself as a topological space if it is clear which topology T one has in mind. A so-called *Stone topology* (U, T) satisfies two further conditions:

If $u \in U$ and $v \in U$ and $u \neq v$, then there is set $X \subseteq U$ such that both X and $U - X$ are elements of T and furthermore $u \in X$ and $v \in U - X$. (TOTAL SEPARATION)
If $S \subseteq T$ and $U = \bigcup S$, then there is a finite subset $S_0 \subseteq S$ such that $U = \bigcup S_0$. (COMPACTNESS)

The elements of a topology T are said to be *open*. A *closed* set is subset of U whose complement in U is open. If X is any subset of U, then we write $\mathsf{C}X$ for the *closure* of X, that is, the smallest closed set that includes X. A subset of U can be open and closed at the same time, and if it is, it is said to be *clopen*. It can be shown that every

[5] If $n = 0$ the conjunction $\sigma_0 \wedge \cdots \wedge \sigma_{n-1}$ is empty and is identified with $\top$.

open set in a Stone space is the union of a set of clopen sets, and hence that every closed set is the intersection of a set of clopen sets.

To bring out the connexion with finitary logic, in this paper we will refer to the clopen sets as the *propositions* of (U, T) and the closed sets as the (*semantic*) *theories* of (U, T). The reference to (U, T) may be omitted if it can be done without causing confusion.

A *sphere system* in a Stone space in (U, T), here called an *onion*, is a nonempty set O of theories in (U, T) that is linearly ordered by set inclusion and is also closed under arbitrary nonempty intersection:

$X \subseteq Y$ or $Y \subseteq X$, for all $X, Y \in O$,
if C is a nonempty subset of O, then $\bigcap C \in O$.

A *revision frame* is a structure (U, T, H, R) such that (U, T) is a Stone space, H is a set of onions, R is a function from the set of propositions to $H \times H$, and the conditions (o1)–(o4) below are satisfied. We say, after Lewis, that a proposition P is *entertainable in O* if P intersects some element of O; that is, if $P \cap \bigcup O \neq \emptyset$. Write $O \bullet P$ for the family of elements of O that intersect with P; that is, $O \bullet P = \{X \in O : P \cap X \neq \emptyset\}$. Thanks to compactness, if P is entertainable in O, then $O \bullet P$ contains a smallest element.

(o1) For every proposition P, if $(O, O') \in R^P$, then either P is entertainable in O and $\bigcap O' = P \cap Z$, where Z is the smallest element of $O \bullet P$, or else $\bigcap O' = \{\emptyset\}$. (ONION REVISION)
(o2) $\complement \bigcup O = \complement \bigcup O'$, for all $O, O' \in H$. (ONION COMMITMENT)
(o3) For every $O \in H$ there is some $O' \in H$ such that $(O, O') \in R^P$. (ONION SERIALITY)
(o4) If $(O, O') \in R^P$ and $(O, O'') \in R^P$, then $O' = O''$. (ONION FUNCTIONALITY)

The singleton set $\{\emptyset\}$ is an onion, and it called the *trivial onion*. Notice that because of condition (o2) ONION COMMITMENT, if the onion set of a revision frame contains the trivial onion, then it has no other elements; in this degenerate case, the frame is itself trivial: the *trivial frame*. Note that the trivial onion always results when an onion is revised by a nonentertainable proposition.

A *valuation* in a Stone space is a function from the set of propositional letters to the set of clopen subsets of the space. A valuation V can always be lifted by the obvious conditions to a function $\overline{V}$ defined on the set of all pure Boolean formulæ (with the range still included in the set of clopen sets):

$\overline{V}(\mathbb{P}) = V(\mathbb{P})$, for all propositional letters $\mathbb{P}$,
$\overline{V}(\phi \wedge \psi) = \overline{V}(\phi) \cap \overline{V}(\psi)$,
$\overline{V}(\phi \vee \psi) = \overline{V}(\phi) \cup \overline{V}(\psi)$,
$\overline{V}(\neg\phi) = U - \overline{V}(\phi)$, etc.

When it is clear what valuation V is understood, we will write $[\![\phi]\!]$ for $\overline{V}(\phi)$. Note that this notation (in which the reference to V is tacit) is meaningful only if ϕ is a pure Boolean formula.

A *revision model* is a revision frame together with a valuation. In other words, (U, T, H, R, V) is a revision model if (U, T, H, R) is a revision frame and V is a valuation in (U, T). The notion of *truth* of a formula in a revision model, symbolized by the symbol $\vDash$, is defined relative to a pair (O, u), where O is an onion and u is a point (that is, $O \in H$ and $u \in U$). (Intuitively, O is, or represents, the belief state of the agent; u is the state of the environment.) Truth-conditions:

$(O, u) \vDash \mathbb{P}$ iff $u \in V(\mathbb{P})$, for every propositional letter $\mathbb{P}$,
[conditions for the Boolean operators]
$(O, u) \vDash \mathbf{B}\phi$ iff $\bigcap O \subseteq [\![\phi]\!]$,
$(O, u) \vDash \mathbf{K}\phi$ iff $\bigcup O \subseteq [\![\phi]\!]$,
$(O, u) \vDash [*\phi]\theta$ iff, for all O', if $(O, O') \in R^{[\![\phi]\!]}$ then $(O', u) \vDash \theta$.

Because of conditions (o3) ONION SERIALITY and (o4) ONION FUNCTIONALITY there is, for every onion O and pure Boolean formula ϕ, a unique onion $O^{[\![\phi]\!]}$ such that $(O, O^{[\![\phi]\!]}) \in H$. In the AGM setting it would therefore be possible to replace the last truth-condition by one that is more specific:

$$(O, u) \vDash [*\phi]\theta \text{ iff } (O^{[\![\phi]\!]}, u) \vDash \theta.$$

A formula is *valid* in a revision frame if it is true relative to all pairs of onions and points. A set Σ of formulæ is *satisfiable* in a revision frame (U, T, H, R) if there is an onion $O \in H$ and a point $u \in U$ such that, for all $\phi \in \Sigma$, $(O, u) \vDash \phi$, where the turnstile refers to some model on (U, T, H, R).

One can show that our axiom system is complete in the following sense:

Theorem 3 *A formula is provable in our axiom system if and only if it is valid in all revision frames.*

In fact, our system is even strongly complete in the following sense:

Theorem 4 *A set of formulæ is consistent in our axiom system if and only if it is satisfiable in some revision frame.*

For proofs of these results, see Segerberg (2005, unpublished manuscript).

3.3.4 Extensions

Basic AGM can be extended in various directions. For example, it should be possible to take higher-order belief into account, to give a substantial theory of iterated belief change (the modelling presented here allows iteration but has nothing interesting to say about it beyond what is contained implicitly in basic AGM), to include several agents, and perhaps to combine belief revision with action (on the last topic, however, see the following section). But in spite of much discussion, particularly of the problem of iteration, progress has been slow. For some recent efforts, see Segerberg (2003) on higher order belief, Rott (2006) and Segerberg (2005, unpublished manuscript), on iteration. See also van Ditmarsch (2005) for revision in multi-agent settings and with action operators.

3.4 Basic iterative belief change II: the KGM paradigm

3.4.1 Basic language

One notable difference between AGM and KGM is apparent already in their underlying languages. While the language of AGM is based on that of the classical propositional calculus, KGM extends the classical, truth-functional connectives by a new conditional operator for which we will use David Lewis's symbol $\Box\!\!\rightarrow$. Let pc be the set of all *pure conditional formulæ*, that is, the set of formulæ generated from the set of propositional letters by $\Box\!\!\rightarrow$ and the truth-functional operators. (The truth-functional connective $\rightarrow$ is of course also a conditional operator: the material conditional. Nevertheless, when we speak of the conditional operator in the sequel, it is always the nonmaterial $\Box\!\!\rightarrow$ we have in mind.)

The set of all *update formulæ* is defined as the smallest set F satisfying the following constraints:

$\mathsf{pc} \subseteq \mathsf{F}$,
F is closed under Boolean operators,
if $\phi \in \mathsf{pc}$ then $\mathbf{B}\phi \in \mathsf{F}$ and $\mathbf{K}\phi \in \mathsf{F}$,
if $\phi \in \mathsf{pc}$ and $\theta \in \mathsf{F}$ then $[\star\phi]\theta \in \mathsf{F}$.

As before, $\mathbf{B}$ and $\mathbf{K}$ are static doxastic operators, while $[\star\phi]$ is a dynamic operator for every pure Boolean formula ϕ. We use the following abbreviations: $\phi \Diamond\!\!\rightarrow \psi$ for $\neg(\phi\Box\!\!\rightarrow \neg\psi)$, $\Box\phi$ for $\neg\phi\Box\!\!\rightarrow \bot$, and $\Diamond\phi$ for $\neg(\phi\Box\!\!\rightarrow \bot)$.

3.4.2 Basic axiom system

Our axiom system consists of four blocks: classical, conditional, modal, and KGM. The classical block is as before. The conditional block consists of postulates that together are sufficient for David Lewis's logic VCU: for example, the inference rule

(RC′) if $\phi \leftrightarrow \phi'$ and $\psi \leftrightarrow \psi'$ as well as $\phi\Box\!\!\rightarrow \psi$ are theorems, then $\phi'\Box\!\!\rightarrow \psi'$ is also a theorem

and the axiom schemata

(NN′) $(\phi\Box\!\!\rightarrow (\psi \wedge \theta)) \leftrightarrow ((\phi\Box\!\!\rightarrow \psi) \wedge (\phi\Box\!\!\rightarrow \theta))$,

(N′) $\phi\Box\!\!\rightarrow \top$,

($\star$1) $\phi\Box\!\!\rightarrow \phi$,

($\star$2) $(\phi\Diamond\!\!\rightarrow \psi) \rightarrow \Diamond\psi$,

($\star$3) $\phi \rightarrow (\top\Box\!\!\rightarrow \phi)$,

($\star$4) $\phi \rightarrow (\top\Diamond\!\!\rightarrow \phi)$,

($\star$5) $((\phi \wedge \psi)\Box\!\!\rightarrow \theta) \rightarrow (\phi\Box\!\!\rightarrow (\psi \rightarrow \theta))$,

($\star$6) $(\phi\Diamond\!\!\rightarrow \psi) \rightarrow ((\phi\Box\!\!\rightarrow (\psi \rightarrow \theta)) \rightarrow ((\phi \wedge \psi)\Box\!\!\rightarrow \theta))$,

($\star$7) $\Box\phi \rightarrow \Box\Box\phi$,

($\star$8) $\Diamond\phi \rightarrow \Box\Diamond\phi$.

The modal block is as before. Finally the KGM block: the rule

(RC$''$) if $\phi \leftrightarrow \psi$ is a theorem, then $[\star\phi]\theta \leftrightarrow [\star\psi]\theta$ is also a theorem, for all pure conditional formulæ ϕ and ψ,

and the axiom schemata

($\star$0) $\theta \leftrightarrow [\star\phi]\theta$, if θ is a pure conditional formula,

($\star$DF) $\langle\star\phi\rangle\theta \leftrightarrow [\star\phi]\theta$,

($\star$RR) $\mathbf{B}(\phi \Box\!\!\rightarrow \psi) \leftrightarrow [\star\phi]\mathbf{B}\psi$,

($\star$KB) $\mathbf{K}\phi \leftrightarrow \mathbf{B}\Box\phi$.

The schema ($\star$RR) is a counterpart of what Grahne (1991) called Ramsey's Rules, that is, the belief update version of the Ramsey test for conditionals.

We did not introduce any non-material conditional sign into the language of AGM-style DDL, let alone a Ramsey-test axiom schema for such new conditionals and the sentential belief revision operator. As mentioned at the beginning of this section, if we had done so, then by Gärdenfors's theorem this would have led to an inconsistent system (given some very mild assumptions).

3.4.3 Semantics

For the semantics of our DDL system of KGM, an approach in terms of selection functions is convenient. Suppose that (U, T) is a Stone space and f is a function from the set of propositions to the set of theories. Then f is a *selection function* if the following conditions are satisfied: for all propositions P and Q,

(i) $fP \subseteq P$,

(ii) if $P \subseteq Q$ and $fP \neq \emptyset$, then $fQ \neq \emptyset$,

(iii) if $P \subseteq Q$ and $P \cap fQ \neq \emptyset$, then $fP = P \cap fQ$.

An *update frame* is a triple (U, T, F) such that (U, T) is a Stone space and F is a function assigning to each element $u \in U$ a selection F_u and, furthermore, T is closed under the binary operations $\Box$ and $\rhd$, where for all propositions P and Q,

$$P \Box Q =_{\mathrm{df}} \{u \in U : F_u P \subseteq Q\},$$

$$P \rhd Q =_{\mathrm{df}} \{u \in U : F_u P \cap Q \neq \emptyset\}.$$

It is clear that if (U, T, F) is a given update model and V is a valuation in (U, T), then V can be extended to a function $\overline{V}$ by the usual conditions for the truth-functional connectives plus the new condition

$$\overline{V}(\phi \Box\!\!\rightarrow \psi) = \{u \in U : \overline{V}(\phi)\Box\overline{V}(\psi)\} = \{u \in U : F_u(\overline{V}(\phi)) \subseteq \overline{V}(\psi)\}.$$

It follows that

$$\overline{V}(\phi \Diamond\!\!\rightarrow \psi) = \{u \in U : \overline{V}(\phi) \rhd \overline{V}(\psi)\} = \{u \in U : F_u(\overline{V}(\phi)) \cap \overline{V}(\psi) \neq \emptyset\}.$$

As in the preceding section we will employ the notation $[\![\phi]\!]$ for the set $\overline{V}(\phi)$. Note, however, that this time the notation covers not just pure Boolean formulæ but all pure conditional formulæ.

An *update model* is an update frame together with a valuation. The notion of *truth* of a formula ϕ in an update model (U, T, F, V) relative to a pair (B, u), where B is a theory and u is a point—in symbols, $(B, u) \vDash \phi$—is defined as follows:

$(B, u) \vDash \mathbb{P}$ iff $u \in V(\mathbb{P})$, for every propositional letter $\mathbb{P}$,

[conditions for the Boolean operators]

$(B, u) \vDash \mathbf{B}\phi$ iff $B \subseteq [\![\phi]\!]$,

$(B, u) \vDash \mathbf{K}\phi$ iff $K \subseteq [\![\phi]\!]$, where $K = \bigcup_{v \in B}\{F_v P : P$ is a proposition$\}$,

$(B, u) \vDash \phi \Box\!\!\rightarrow \psi$ iff, for all $v \in F_u[\![\phi]\!]$, $(B, v) \vDash \psi$,

$(B, u) \vDash [\star\phi]\theta$ iff $(B', u) \vDash \theta$, where $B' = \mathsf{C} \bigcup_{v \in B} F_v[\![\phi]\!]$.

Concepts of validity and satisfaction are defined in the KGM context in the same way as in the AGM context above. There are also similar completeness theorems:

Theorem 5 *A formula is provable in our axiom system if and only if it is valid in all update frames.*

Theorem 6 *A set of formulæ is consistent in our axiom system if and only if it is satisfiable in some update frame.*

Proofs of these results, which generalize those of Grahne (1991), are given in Segerberg (2005, unpublished manuscript).

3.4.4 Doxastic actions and real events

David Lewis's official reading of $\phi \Box\!\!\rightarrow \psi$ was "if it were the case that ϕ, then it would be the case that ψ". For $[\star\phi]\psi$ a natural reading is "after the agent has updated his beliefs by (the information that it is now the case that) ϕ, it is the case that ψ" or more briefly "after update by ϕ, ψ". The term $\star\phi$ was not given an independent meaning in the semantics above, but it would be possible to do so by defining (with respect to a given model)

$$[\![\star\phi]\!] = \{(X, Y) \colon X, Y \text{ are theories in } (U, T) \ \&\ Y = \mathsf{C} \bigcup_{x \in X} F_x[\![\phi]\!]\}.$$

Notice that with this definition $[\star\phi]$ behaves like a normal modal box operator. Semantically $[\![\star\phi]\!]$ represents the updating by the proposition $[\![\phi]\!]$, what may be called a *doxastic action* (or a *doxastic event*—in this paper we make no distinction between actions and events).

The validity of the Ramsey schema ($\star$RR) suggests a correlation between beliefs, conditionals and doxastic actions. In order to examine this correlation more closely, let us extend our object language by introducing a notation for a limited class of *real events*, namely, "resultative" events: events describable in terms of their results. Let ∂

be a new operator that operates on pure conditional formulæ to produce terms with which we can form modal box operators of the type $[\partial\phi]$. The intended meaning of a term $\partial\phi$ would be "the event resulting in (its being the case that) ϕ". Accordingly, the intended reading of a formula $[\partial\phi]\psi$ would be "after the event resulting in (its being the case that) ϕ, it is the case that ψ" or more briefly "after ϕ has just been realized, ψ". There is a connexion between conditionals and real actions and a related connexion between beliefs and doxastic actions. It will be instructive to explain this claim in some detail.

Consider a certain update model. Define, for pure Boolean ϕ,

$[\![\partial\phi]\!] = \{(u,v)\colon v \in F_u[\![\phi]\!]\}$,
$(B,u) \vDash [\partial\phi]\psi$ iff, for all v, if $(u,v) \in [\![\partial\phi]\!]$ then $(B,v) \vDash \psi$.

A notable consequence of these definitions is the validity of the schema

$$(\phi \Box\!\!\rightarrow \psi) \leftrightarrow [\partial\phi]\psi. \tag{3}$$

In typical applications of belief change theory, we assume some regularity in the way the environment may change—in a physical context there are laws of nature, in games there are rules. In general there have to be describable limits to how things can change, or theorizing would be futile. Let us introduce some (perhaps euphemistic) terminology: we say that the agent is *knowledgeable* (with respect to the way the environment can change) if the schema

$$(\phi \Box\!\!\rightarrow \psi) \rightarrow \mathbf{B}(\phi \Box\!\!\rightarrow \psi)$$

is valid, and *correct* if the schema

$$\mathbf{B}(\phi \Box\!\!\rightarrow \psi) \rightarrow (\phi \Box\!\!\rightarrow \psi)$$

is valid. Furthermore, let us say that the agent is *well-educated* if both knowledgeable and correct. Thus for well-educated agents the schema

$$(\phi \Box\!\!\rightarrow \psi) \leftrightarrow \mathbf{B}(\phi \Box\!\!\rightarrow \psi)$$

is valid, making the distinction between truth of and belief in conditional statements invisible. And just as the schemata (3) and ($\star$RR) are valid, so is the schema

$$\mathbf{B}[\partial\phi]\psi \leftrightarrow [\star\phi]\mathbf{B}\psi.$$

This is a precise sense in which beliefs, conditionals, and real and doxastic actions correlate in KGM.

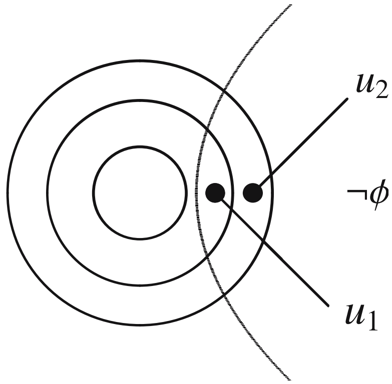
3.5 Comparing AGM and KGM

We collect our remarks under four headings.

3.5.1 Revision and update

In the literature on belief change the distinction between static and dynamic environment has become important. Ignoring the fact that belief change can take other forms than the relatively simple form of adding new data that we have been dealing with here, it seems right to say that belief change due to new information in an unchanging environment has come to be called *belief revision* (the static case, in the sense that

Fig. 4 Doxastic preference



the "world" remains unchanged), while it is fairly generally accepted to use the term *belief update* for belief change that is due to reported changes in the environment itself (the dynamic case, in the sense that the "world" changes; compare our analysis in the last subsection). It has been held for some time that these cases support different logics, with AGM recognized as the logic of revision and KGM as the logic of update. This tradition is reflected in the terminology of this paper where revision frames and update frames play the leading semantic rôles in the sections on AGM and KGM, respectively.

The established tradition notwithstanding, it would be interesting to see a really convincing argument for tying AGM revision to static environments. We saw in the previous section how easily real change mixes with doxastic change in the KGM modelling, and it is certainly not clear today how to make AGM onions cope with real change. But it is also not clear that belief date update has to be interpreted as reflecting a proper change in the environment. We think the actual difference between the intended interpretation of revision and update is given by the fact that the former belief change follows a *doxastic* order of "fallback positions" (Lindström & Rabinowicz, 1990) while the latter conforms to a *worldy* similarity order of states of affairs—the one "rides" on a subjective structure, the other on an objective one, but neither is necessarily tied to either invariances or changes in the environment. Let us argue for this in more detail.

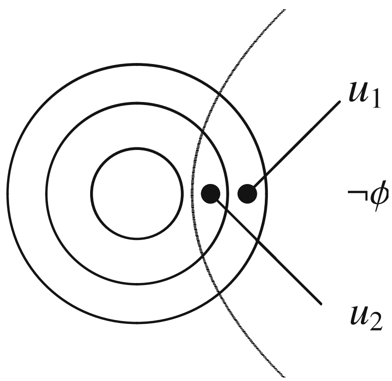
According to the intended interpretation of belief revision, belief states are analyzed in terms of onions: for example, we might consider an onion that determines a belief (simpliciter) in ϕ, as well as two fallback positions: one that includes a point u_1 with $u_1 \models \neg\phi$ such that there is no smaller fallback position that includes a point that satisfies $\neg\phi$, and a larger fallback position which both includes u_1 and a different point u_2 with $u_2 \models \neg\phi$. By the intended reading of the subset relation, the former fallback position is doxastically preferable to the latter. If the belief in ϕ is to be revised by $\neg\phi$ with respect to this onion, then this preference shows up in the way that the agent does not believe u_2 to be possible after the revision but he believes u_1 to be possible (see Fig. 4).

The intended interpretation of the semantics for belief update depends crucially on the manner in which selection functions f are interpreted. The standard interpretation is in terms of environmental change; but there is another plausible way of interpreting selection functions, one that enables us to demonstrate that update does not necessarily correspond to environmental changes. Lewis famously considered objective similarity relations between possible worlds to be determinable from the objective spheres systems that in turn can be defined in terms of given objective selection functions: for example, the point u_2 with $u_2 \vDash \neg\phi$ could be determined to be most similar to the actual world among the set of points that satisfy $\neg\phi$, while the point u_1 with $u_1 \vDash \neg\phi$ would be less similar to the actual world than u_2. Consequently, with respect to such a selection function, if the actual world satisfies a Lewis-type conditional $\neg\phi \Box\!\!\rightarrow \theta$ then θ must be satisfied by u_2 though not necessarily by u_1; in fact, given weak additional assumptions, there will be such a formula θ such that θ is *not* satisfied by u_1. Since belief update is explained semantically in terms of the same selection function that determines the truth condition of $\Box\!\!\rightarrow$, if the agent's belief in ϕ is to be updated by $\neg\phi$ with respect to this selection function then the objective similarity relation that is determined by it shows up in the way that the agent does not believe u_1 to be possible after the update but he believes u_2 to be possible (see Fig. 5).

Thus, given new evidence, we find that in the case of belief revision the agent tries to change his beliefs in a manner such that the worlds that he subsequently believes to be in comprise the *subjectively most plausible deviation* from the worlds he originally believed to inhabit. However, when confronted with the same evidence in belief update, the agent tries to change his beliefs in a way such that the worlds that he subsequently believes to be in are *as objectively similar as possible* to the worlds he originally believed to be the most plausible candidates for being the actual world.

It is tempting to relate these different views on belief change to the traditional distinction of indicative and subjective conditionals. Using the stock example: everyone considers the indicative 'If Oswald did not kill Kennedy somebody else did' as acceptable, but many regard the subjunctive 'If Oswald had not killed Kennedy somebody else would have' as false. The latter seems to imply that there is a world u_2 in which Kennedy was not killed at all, such that u_2 is regarded by many to be maximally similar to the actual world among those worlds in which Oswald did not kill Kennedy. At the same time, the most plausible worlds in which Oswald did not kill Kennedy are still considered to be worlds in which Kennedy was killed by someone—the latter belief would be preserved if the new information 'Oswald did not kill Kennedy' were added to one's current belief set. Let u_1 be one of the "maximally plausible" worlds just mentioned: then we could actually have a situation of precisely the form described above, that is, while u_2 is regarded objectively more similar to the

Fig. 5 Objective similarity



actual world than u_1, the latter is nevertheless considered subjectively more plausible than the former. According to this interpretation, belief revision would correspond to belief change "in the indicative mood", whereas belief update would correspond to belief change in "the subjunctive mood", where in both cases the corresponding "changes" that apply to the actual world are merely hypothetical. (Our discussion of the indicative vs. subjunctive distinction and the Oswald-Kennedy conditionals is of course oversimplified and not much more than schematic; in particular, temporal aspects have been disregarded completely.)

3.5.2 Types of beliefs

While we are waiting for a clear theoretical account of these issues, let us reflect on the different types of belief that are involved in the two paradigms. Let us presuppose the "environmental change" interpretation of update: then the KGM case seems comparatively straightforward. The immediate concern here is beliefs about the current state of the environment; let us call this *particular belief*. But the agent is also able to hold beliefs about the ways in which the environment may change; let us call such belief about environmental regularity *systematic belief*. It is striking that, given a body of systematic beliefs and an initial set of particular beliefs, in KGM all future particular beliefs are determined by reports about what happens. So KGM, unlike basic AGM, is a theory of iterated belief change. For example,

$$[\star\phi][\star\psi]\mathbf{B}\theta \leftrightarrow \mathbf{B}(\phi \Box\!\!\rightarrow (\psi \Box\!\!\rightarrow \theta))$$

is a theorem schema of KGM. So if you understand conditionals (of this kind), you understand iterated belief change.

In AGM the situation with regard to types of belief is different. Again we have particular beliefs. But how the agent reacts in the face of new data is different. His behaviour is not determined by his beliefs about any standard operating procedures that regulate the environment but in some completely different way. In Lindström and Rabinowicz's apt terminology, onion elements represent possible "fallback" theories. Their terminology spells out the point of the AGM approach: that the belief state of an agent must not be identified with his belief set, that is, set of particular beliefs—one must also take into account what may be called the agent's doxastic dispositions. If the agent is challenged to give up his current belief set, then in the normal case he has already prepared a collection of backup or default theories: "if I cannot believe A, at least I will believe B; and if that, too, is not possible, at the very least I will believe C; …" (and so on). But the agent's choice of fallbacks need not have anything to do with the environment. For example, in an economic or political context, the structure of a belief state (onion) may be rooted in value judgements or tactical considerations. (This is particularly common in the case of complex agents consisting of lower-level agents.) The beliefs behind such doxastic dispositions (the layers of the onion) may perhaps be glossed as *default beliefs*.

3.5.3 Minimality

The notion of minimality is common to Stalnaker/Lewis type conditional logic and belief revision; it was found already in Ramsey (1978), the inspiration of both traditions. But exactly what is minimized in a theory is sometimes difficult to say. Intuitively, in a theory of belief change it is usually loss of information that should be minimized. Technically, in a modelling of the sort considered in this paper, this is achieved by minimizing “distance”, going for points that are as “close to” a point of reference as possible. Lewis achieves minimality by maximizing what he calls similarity between worlds. But what counts as similarity in a given application remains to be explained.

In the KGM paradigm the notion of minimality may perhaps be said to be implicit, but it does not show up in an explicit way. To say, with respect to a certain update frame and the “environmental change” interpretation of KGM, that there is a sense in which a point v is close to a point u if $(u, v) \in [\![\partial\phi]\!]$ would be misleading: v is simply one of those points at which the system may land if the event $[\![\partial\phi]\!]$ is initiated at u. Furthermore, what B' minimizes, if B and B' are belief sets and $B' = \bigcup_{x\in B} F_x[\![\phi]\!]$, is too obvious to deserve a special name: B' is simply the smallest set containing all possible end-points if the event $[\![\partial\phi]\!]$ is initiated at a point in B. It is easier to explain the “subjunctive mood” interpretation of KGM in terms of minimality, since according to such an account the task of belief update is rather to minimize what is believed to be the “objective distance” between the current belief set and the belief set after the update.

The question of minimality makes clear sense in the case of the AGM paradigm. The fallbacks in an onion represent theories that the agent is willing, under certain circumstances, to endorse at least tentatively. Those theories are all “close” to the theory he currently holds, the theory corresponding to the innermost element (the belief set). Confronted with unassailable evidence adverse to his current theory, the agent will fall back on his strongest alternative; that is, on the “closest” theory available to him.

3.5.4 The Ramsey test again

Finally let us return to the Ramsey test which we used at the beginning of Sect. 3 to distinguish the two paradigms of AGM and KGM. Given what was said above about the logical and philosophical properties of the two paradigms, it is perhaps now clearer why the Ramsey test *ought not* to be expected to hold for belief revision while a Ramsey-test-like logical law *should* be expected to hold for update. The latter reflects, on the level of belief, an objective change of environmental states or an objectively given similarity relation for such states; therefore, every result of a belief update corresponds logically to a previously held belief in some sort of factual law. In contrast, belief revision conforms to a ranking of environmental states in terms of their subjective plausibility as fallback positions: there is no reason to believe that this doxastic structure could be “read into the environment” in the way expressed by the Ramsey test for conditionals. Gärdenfors’s Theorem (given a weak auxiliary hypothesis) even proves the impossibility of such a reading.

4 Where to: an outlook

We end with a tentative list of aims that we think might inspire the future development of DDL.

- Axiomatization of specific methods of iterated revision: Every semantically formulated suggestion for an iterated revision scheme ought to be characterized logically in DDL on the basis of a corresponding soundness and completeness theorem (much as various specific assumptions on modal accessibility relations can be represented adequately by an extension of the minimal modal logic K). What do these systems look like? In addition to special older suggestions in the literature, some very preliminary but more systematic answers to these questions can be found in Rott (2006), Segerberg (2005, unpublished manuscript), and van Benthem (Forthcoming).

- Axiomatization and semantics of doxastic attitudes towards revision and update: What are the proper logical systems that govern formulæ such as (1) on p. 3
- Axiomatization and semantics of nested revision: What are the proper logical systems that govern formulæ such as (2) on p. 3
- Axiomatization and semantics of revision and update in a framework with more than one agent: What are the proper logical systems for formulæ that express belief changes of an agent x (or a group g) induced by an agent y (or a group g')? (For some progress on this front, see van Ditmarsch, 2005.)
- Philosophical interpretation of revision and update: In what ways can and should the axiomatic and semantic differences between revision and update be interpreted? How do these differences relate to the differences between axiomatic and semantic systems for iterative and subjunctive conditionals?
- Comparison of belief update and probabilistic imaging: Results by Adams (1998) and Gärdenfors (1988) support the view that AGM type belief revision can be understood as the qualitative counterpart of probabilistic conditionalization. The question remains how to solve the equation

$$\text{KGM} : \text{AGM} = x : \text{conditionalization}$$

  for the variable x. We hypothesize the answer to be Lewis's (1976) method of *probabilistic imaging* which was generalized later by Gärdenfors (see Sect. 5.3. of Gärdenfors, 1988).
- Combinations of logical systems for qualitative belief change (such as DDL) with systems of probabilistic logic: See Kooi (2003) for recent progress in this direction.
- Comparison of limitative results for DDL with limitative results in other areas: As demonstrated above (see Leitgeb, 2005, unpublished manuscript for details), there is a close logical relationship between Gärdenfors's impossibility result for belief revision with conditionals and Arrow's classic impossibility result for social choice. Indeed, both the assumptions and the conclusion of the latter can be expressed in a system of belief revision for conditionals and Arrow's proof can be carried out within that system. Most remarkably, the *Pareto* assumption of Arrow's result translates into the left-to-right direction of the Ramsey test for conditionals. This raises the question of whether responses to Arrow's theorem in the theory of social choice can be related to responses to Gärdenfors's theorem in the theory of belief revision and if so, whether this yields additional information on the interpretation of AGM vs. KGM and their formal treatment in DDL. There is also a recent rise of interest in limitative results on judgment aggregation and their representation in logical systems (see e.g. Pauly and van Hees, 2006) which is likely to lead to equally interesting results in DDL with operators for both individual and social belief change.
- Unification of DDL and DEL: Simultaneous to the introduction of DDL, systems of so-called *Dynamic Epistemic Logic* (DEL) emerged in which belief change is studied in the same way as in DDL, that is, in terms of axiomatic and semantic system for dynamic sentential operators (Baltag et al., 1998; van Benthem, 1996; Gerbrandy, 1999; cf. Plaza, 1989; for an overview, see: van Ditmarsch et al.,

2006). For example, Baltag et al. (1998) showed how a dynamic logic of public announcements operators could be developed; public announcement is regarded as one among many possible types of belief update, where 'update' is not used as a name for $\star$ from above but rather as a general term for all sorts of belief changes (so-called "Amsterdam update"). While DDL focused on the logic of belief *revision* operators in a *single agent* framework, DEL concentrated on the logical study of belief *expansion* in a *multi-agent* framework (e.g., applications of the public announcement operator $[\phi!]$ correspond semantically to shrinking a possible worlds model to its set of ϕ-worlds, where the logical outcomes of this process can be investigated in terms of a static common knowledge operator for a group of agents). We predict that the two research programmes of DDL and DEL will merge in the long run into the single logical endeavour of *DBC*: *dynamic logics of belief change*.

Acknowledgements The idea of writing this paper was conceived during the academic year 2004–2005 which the authors were privileged to spend in the Philosophy Department at Stanford University. Most of the material in the paper was presented at the ESF Exploratory Workshop "Understanding the Dynamics of Knowledge" in Siena in November 2005. We want to thank the organizers of the workshop as well as Johan van Benthem, Wiebe van der Hoek, Hans Rott, Frank Veltman and other participants for many constructive comments. We are also grateful to no less than three anonymous referees for their extensive and very helpful remarks. Finally, we want to thank Fabio Paglieri for his assistance and generous support.

References

Adams, E. W. (1998). *A primer of probability logic*. Stanford: CSLI Lecture Notes.

Alchourrón, C., Gärdenfors, P., & Makinson, D. (1985). On the logic of theory change. *The Journal of Symbolic Logic, 50*, 510–530.

Arrow, K. A. (1963). *Social choice and individual values* (2nd ed.). New Haven: Yale University Press.

Baltag, A., Moss, L. S., & Solecki, S. (1998). The logic of public announcements, common knowledge and private suspicions. In I. Gilboa (Ed.), *Proceedings TARK 1998* (pp. 43–56). Los Altos: Morgan Kaufmann Publishers.

van Benthem, J. (1996). *Exploring logical dynamics*. Stanford: CSLI Publications.

van Benthem, J. (Forthcoming). Dynamic logic for belief revision. *Journal of Applied Non-Classical logics*.

van Ditmarsch, H. (2005). Prolegomena to dynamic logic for belief revision. *Synthese: Knowledge, Rationality and Action, 147*, 229–275 (KRA pages 41–87).

Gärdenfors, P. (1986). Belief revisions and the Ramsey test for conditionals. *Philosophical Review, 95*, 81–93.

Gärdenfors, P. (1988). *Knowledge in flux. Modeling the dynamics of epistemic states*. Cambridge, Mass: MIT Press.

Gärdenfors, P., & Makinson, D. (1994). Nonmonotonic inference based on expectations. *Artificial Intelligence, 65*, 197–245.

Gerbrandy, J. (1999). Dynamic epistemic logic. In L. S. Moss et al. (Eds.), *Logic, language, and information* (vol. 2, pp. 269–276). Stanford: CSLI Publications.

Grahne, G. (1991). Updates and counterfactuals. In J. A. Allen, R. Fikes, & E. Sandewall (Eds.), *Principles of knowledge representation and reasoning* (pp. 269–276). San Mateo: Morgan Kaufmann.

Grove, A. (1988). Two modellings for theory change. *Journal of Philosophical Logic, 17*, 157–170.

Katsuno, H., & Mendelzon, A. (1992). On the difference between updating a knowledge base and revising it. In P. Gärdenfors (Ed.), *Belief revision* (pp. 183–203). Cambridge: Cambridge University Press.

Kooi, B. P. (2003). Probabilistic dynamic epistemic logic. *Journal of Logic, Language and Information, 12*(4), 381–408.

Leitgeb, H. (Forthcoming). Towards a logic of type-free modality and truth. In C. Dimitracopoulos et al. (Eds.), *Logic Colloquium 05,* lecture Notes in Logic, ASL.

Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
Lewis, D. (1976). Probabilities of conditionals and conditional probabilities. *The Philosophical Review, 85*, 297–315.
Van Linder, B., van der Hoek, W., & Meyer, J.-J. (1995). Actions that make you change your mind. In A. Laux, & H. Wansing (Eds.), *Knowledge and belief in philosophy and artificial intelligence* (pp. 103–146). Berlin: Akademie Verlag.
Lindström, S., & Rabinowicz, W. (1990). Epistemic entrenchment with incomparabilities and relational belief revision. In A. Fuhrmann, & M. Morreau (Eds.), *The logic of theory change*, Lecture Notes in Artificial Intelligence 465 (pp. 93–126). Berlin: Springer Verlag.
Lindström, S., & Rabinowicz, W. (1995). The Ramsey test revisited. In G. Crocco, L. Fariñas del Cerro, & A. Herzig (Eds), *Conditionals from philosophy to computer science* (pp. 157–202). Oxford: Oxford University Press.
Lindström, S., & Segerberg, K. (2006). Model logic and philosophy. In P. Blackburn, J. van Benthem, & F. Wolter (Eds.), *Handbook of model logic*. Elsevier.
Pauly, M., & van Hees, M. (2006). Logical constraints on judgement aggregation. *Journal of Philosophical logic 36*(6), 569–585.
Plaza, J. A. (1989). Logic of public communications. In M. L. Emrich et al. (Eds.), *Proceedings of the fourth international symposium on methodologies for intelligent systems* (pp. 201–216). Oak Ridge National Laboratory, ORNL/DSRD-24.
Rabinowicz, W., & Lindström, S. (1994). How to model relational belief revision. In D. Prawitz & D. Westerståhl (Eds.), *Logic and philosophy of science in Uppsala* (pp. 69–84). Dordrecht: Kluwer.
Ramsey, F. P. (1978). *Foundations: Essays in philosophy, logic, mathematics and economics*, D. H. Mellor (Ed.). London: Routledge and Kegan Paul.
Rott, H. (2006). Shifting priorities: Simple representations of twenty-seven iterated theory change operators. In H. Lagerlund, S. Lindström, & R. Sliwinski (Eds.), *Modality matters* (pp. 359–385). Uppsala, Uppsala Philosophical Studies 53.
Segerberg, K. (1989). A note on an impossibility theorem of Gärdenfors. *Noûs, 23*, 351–354.
Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic, 39*, 287–306.
Segerberg, K. (1999). Two traditions in the logic of belief: Bringing them together. In H. J. Ohlbach, & U. Reyle (Eds), *Logic, language and reasoning: Essays in honour of Dov Gabbay* (pp. 135–147). Dordrecht: Kluwer.
Segerberg, K. (2001). The basic dynamic doxastic logic of AGM. In M.-A. Williams, & H. Rott (Eds.), *Frontiers in belief revision* (pp. 57–84). Dordrecht: Kluwer.
Segerberg, K. (2003). Moore problems in full dynamic doxastic logic. In J. Malinowski, & A. Pietruszczak (Eds.), *Essays in logic and ontology dedicated to Jerzy Perzanowski* (pp. 11–25). Poznań Studies in the Sciences and the Humanities 50, Amsterdam: Rodopi.