Dynamic Belief Revision over Multi-Agent Plausibility Models

Alexandru Baltag* and Sonja Smets†

1 Introduction

In this paper, we develop a notion of *doxastic actions*, general enough to cover all examples of communication actions and most other belief-changing actions encountered in the literature, but also flexible enough to deal with the issue of *(static and dynamic) revision of beliefs*. This can be seen as a natural extension of the work in [3, 4] on "epistemic actions", incorporating ideas from the semantics of belief revision and of conditional belief, along the lines pioneered in [2] and [11], but using the conditional belief approach adopted in [22, 10, 9] and adapted in [25] to the context of dynamic belief revision.

In [6] and [7], we introduced two equivalent semantic settings for conditional beliefs in a multi-agent epistemic context (*conditional doxastic models* and *epistemic plausibility models*), taking the first setting as the basic one. Here, we adopt the second setting, which is closer to the standard notions used in belief revision literature (being in fact just a multi-agent, "epistemically friendly" version of Grove's relational setting in [14]). We use this setting to define notions of *knowledge, belief* and *conditional belief*, along the standard lines. We also introduce an apparently very useful, though not so well-studied, notion of "weak (non-introspective) knowledge", going back to Stalnaker's *defeasibility analysis of knowledge* in [22]. We call this notion *safe belief*, to distinguished from our (standard, fully introspective) concept of knowledge. We extend the logic of conditional belief operators from [6] with modalities for safe belief, and study some of the logical laws governing these notions.

Moving to the dynamic setting, we introduce *plausibility pre-orders on actions*, developing a notion of "action plausibility models", that extends the "epistemic action models" from [3], along similar lines to the work in [2, 11]. We introduce a natural operation of *anti-lexicographic update product* of such models, which extends the corresponding notion from [3]. From a formal point of view, this can be seen as just one of the options already explored in [11] (generalizing [2]), among many other possible formulas for combining the "degrees of belief" of actions and states. But here we justify our option, arguing that it is the only one that is consistent with our interpretation of action models. The main idea is that *beliefs about changes induce (and "encode") changes of beliefs*.

Our approach differs in flavor from some of the closely related semantical literature on the topic of (dynamic or static) belief revision (e.g. [14, 19, 20, 2, 11, 8]) in the following sense. Most Kripke-style models proposed for multi-agent belief revision are based on *quantitative measures of belief*, such as "degrees of belief", ordinal plausibility functions, graded models or probabilistic measures of belief. Indeed, the update mechanisms proposed in [2, 11, 8] are essentially quantitative, appealing to various arithmetical formulas to compute the degree of belief (or probability) of the output-states in terms of the degrees/probabilities of the input-states and the degrees/probabilities of the actions. This leads to an increase in complexity, both in the computation of updates and in the corresponding logical systems. Moreover, there seems to be no canonical choice for the arithmetical formula for updates, various authors proposing various formulas; we see no transparent reason to prefer one to the others, the resulting complexity and plurality of systems leading the reader to a failure of intuition, and to a feeling of arbitrariness. In contrast, classical (AGM) belief revision theory is a qualitative theory, based on simple, natural, intuitive postulates concerning a basic operation (revision), of great generality and simplicity.

*Computing Laboratory, Oxford University, Oxford UK. Email: Alexandru.Baltag@comlab.ox.ac.uk

†Center for Logic and Philosophy of Science, Vrije Universiteit Brussel and CPNSS, London School of Economics. Post-Doctoral Researcher at the Flemish Fund for Scientific Research. Email: sonsmets@vub.ac.be

Our approach retains this qualitative flavor [1] of the AGM theory, and aims at building a theory of "dynamic" belief revision of equal simplicity and naturality as the classical "static" account.

In this sense, our approach may be subsumed under the general program of viewing belief revision as *conditional* or counterfactual reasoning. As such, our formulation of "static" belief revision is somewhat close to the one in [22], [9], [10], [17]: as in [9], the preference relation is assumed to be well-(pre)ordered, an assumption forced by the conditional view of belief revision; as a result, our modal axioms for conditional beliefs in [6] are very close to the ones in [9]. Our approach is also closely related to the one in [25] and [18], which abstract away from the quantitative details of the plausibility maps when considering the associated *logic*: instead of using "graded belief" operators, as in e.g. [2, 11], or probabilistic modal logic, as in [8], both our account and the ones in [25] and [18] concentrate on the simple, qualitative language of *conditional beliefs, knowledge and action modalities* (to which we add here the *safe belief* operator). As a consequence, one obtains simple, general, elegant *logical laws of dynamic belief revision*. Given any complete axiomatization of the "static" logic of conditional beliefs and safe belief, our reduction laws give a *complete axiomatization of the logic of belief updates induced by doxastic actions*. Compared both to our older axiomatization in [7] and to the system in [2], one can easily see that the introduction of the safe belief operator leads to a major simplification of the reduction laws.

2 Epistemic Plausibility Models

A *(multi-agent) epistemic plausibility frame* (EPF, for short) is a structure $(S, \sim_a, \leq_a)_{a\in\mathcal{A}}$, consisting of a set S of "states", endowed with a family of equivalence relations $\sim_a$, called *epistemic indistinguishability relations*, and a family of "well-preorders" $\leq_a$, called *(a priori) plausibility relations*. Here, a "well-preorder" is just a preorder[2] $\leq$ on S such that *every non-empty subset has minimal elements*. Using the notation

$$Min_{\leq} T := \{t \in T : t \leq t' \text{ for all } t' \in T\}$$

for the set of $\leq$-minimal elements of T, the last condition says that: for every set $T \subseteq S$, if $T \neq \emptyset$ then $Min_{\leq} T \neq \emptyset$. As usually, we write $s <_a t$ iff $s \leq_a t$ but $t \not\leq_a s$ (the *"strict" plausibility relation*), and $s \simeq_a$ iff both $s \leq_a t$ and $t \leq_a s$ (the "equal plausibility" relation, or "doxastic indistinguishability").

Observe that the conditions on the preorder $\leq_a$ are a *semantical analogue* of Grove's conditions for the (relational version of) his models in [14]. The standard formulation of Grove models is in terms of a "system of spheres" (weakening Lewis' similar notion), but it is equivalent (as proved in [14]) to a relational formulation. Grove's postulates are still *syntax-dependent*, e.g. existence of minimal elements is required only for subsets that are *definable* in his language. We prefer a purely semantic condition, independent of the choice of a language, both for reasons of elegance and simplicity and because we want to be able to consider more than one language for the same structure.[3] So we adopt the natural semantic analogue of Grove's condition, simply requiring that *every* subset has minimal elements: this will allow our conditional operators to be well-defined on sentences of *any* extension of our logical language. Also, observe that the minimality condition implies, by itself, that the relation $\leq_a$ is both *reflexive* (i.e. $s \leq_a s$ for all $s \in S$) and *connected* (i.e. either $s \leq_a t$ or $t \leq_a s$, for all $s, t \in S$). Note that, when the set S is *finite*, a well-preorder is nothing but a connected preorder. This shows that our notion of frame subsumes, not only Grove's setting, but some of the other settings proposed for conditionalization.

Given an epistemic plausibility frame S, an *S-proposition*, or *S-theory*, is any subset $T \subseteq S$. Intuitively, we say that *the state s satisfies the proposition/theory T* (and write $s \models T$) if $s \in T$. Observe that an EPF is just a special case of a *relational frame* (or *Kripke frame*). So, as it is standard for Kripke frames in general, we can define an *epistemic plausibility model* (EPM, for short) to be an (epistemic)

[1] One could argue that our plausibility pre-order relation can be equivalently encoded in a quantitative notion (of ordinal degrees of plausibility, such as [21]), but in fact the way belief update is defined in our account does not use in any way the ordinal "arithmetic" of these degrees, but only their qualitative, relational features.

[2] i.e. a reflexive and transitive relation. In fact, reflexivity doesn't have to be required here, since it follows from the next condition on the existence of minimal elements.

[3] Imposing syntactic-dependent conditions in the very definition of a class of structures makes the definition meaningful only for one language; or else, the meaning of what, say, a plausibility model is won't be *robust*: it will change whenever one wants to extend the logic, by adding a few more operator. This is very undesirable, since then one cannot compare the expressivity of different logics on the same class of models.

plausibility frame S together with a valuation map $\| \bullet \| : \Phi \to \mathcal{P}(S)$, mapping every element of a given set Φ of "atomic sentences" into S-propositions.

Interpretation. The elements of S will be interpreted as the *possible states* of a system (or "possible worlds"). The atomic sentences $p \in \Phi$ represent *"ontic" (non-doxastic) facts* about the world, that might hold or not in a given state, while the valuation tells us which facts hold at which worlds. The equivalence relations $\sim_a$ capture *agent a's knowledge about the actual state of the system* (intuitively based on the agent's *(partial) observations of this state*): two states s, t are *indistinguishable for agent a* if $s \sim_a t$. In other words, when the actual state of the system is s, then agent a knows only the state's equivalence class $s(a) := \{t \in S : s \sim_a t\}$. Finally, the plausibility relations $\leq_a$ capture *agent a's a priori (conditional) beliefs about the (possible) states* of the system : if the agent is given *some information about the history* (past and current state) of some virtual (not yet actualized, hence not directly observable) system S, and if this information is enough to determine that the virtual state is either s or t, but not to determine which of the two, agent a will believe the state to be s iff $s <_a t$; will believe the state to be t iff $t <_a s$; otherwise (if $s \simeq_a t$), the agent will be indifferent between the two alternatives, i.e. will not be able to decide to believe any one alternative rather than the other.

The words "a priori" and "virtual" in the lines above are meant to express the fact that our plausibility relations refer to *prior beliefs*, in the sense of being prior to obtaining any direct knowledge about the actual state; another way to express this is that they do not refer to the actual state at all, but to some (unknown) virtual state (about which the agent has no direct information, but is told that it is either s or t). This *a priori* belief has to be distinguished from the agent's *actual belief about the current state* of the system: the agent will typically have *some direct knowledge* (based on his observations) about the current state; for instance, the state might be such that she can *see* that some facts hold. The agent's actual beliefs (including his conditional beliefs) will typically differ from his prior conditional beliefs. As explained below (in defining the "local plausibility" relation $\trianglelefteq_a$), we will assume these actual beliefs to be computable in a rather simple manner, by summing up the agent's *knowledge* of the current state with her prior conditional beliefs about the same state (seen as a virtual state).

Observe that this is a *qualitative* interpretation of the plausibility relations, in terms of *conditional beliefs* rather than "degrees of belief": there is no scale of beliefs allowing for "intermediary" stages between believing and not believing. Instead, all beliefs are equally "firm" (though conditional): given the condition, something is either believed or not. To repeat, writing $s <_a t$ is for us just a way to say that: whenever agent a is informed that some unknown (not directly observable) possible state of a virtual system is either s or t (without being given enough information about the history of this state to determine which of the two), she believes that state to be s. There is no need for us to refer in any way to the "strength" of this agent's belief: though she might have beliefs of unequal strengths, we are not interested in modeling this quantitative aspect. Instead, we give the agent some information about a state of a virtual system (that it is either s or t) and we ask her a *yes-or-no question* ("Do you believe that virtual state to be s ?"); we write $s <_a t$ iff the agent's answer is "yes". This is a firm answer, so it expresses a firm belief. "Firm" does not imply "un-revisable" though: if later we reveal to the agent that the state in question was in fact t, she should be able to accept this new information; after all, the agent should be introspective enough to realize that her belief, however firm, was just a belief.

One possible objection against this qualitative interpretation is that our postulate that $\leq_a$ is a well-preorder (and so in particular a connected pre-order) introduces a hidden "quantitative" feature; indeed, any such preorder can be equivalently described using a plausibility map as in e.g. [21], assigning ordinals to states. Our answer is that, first, the specific ordinals will not play any role in our definition of a dynamic belief update; and second, all our postulates can be given a justification in purely qualitative terms, using conditional beliefs. The transitivity condition for $\leq_a$ is just a *consistency* requirement imposed on a rational agent's conditional beliefs. And the existence of minimal elements in any non-empty subset is simply the natural extension of the above setting to *general* conditional beliefs, not only conditions involving two states: more specifically, for any possible condition $P \subseteq S$ about a system S, the S-proposition $Min_{\leq_a} P$ is simply a way to encode everything that agent a believes (*a priori*) about a virtual state of the system, when given only the information that the state satisfies the condition P.

Another possible argument against our insistence for a qualitative interpretation is that it is *irrelevant*: the underlying mathematical structure (given by plausibility frames) is exactly the same as in a degrees-of-belief interpretation; as a consequence, the choice of interpretation will not make any difference for the "static" belief revision theory (i.e. for the rest of this section). Indeed, observe that the conditional beliefs expressed here are in some sense "static": although dealing with a virtual state (thus including future states), they express the agent's *prior* view (belief) about them, given some hypothetical new information $\{s, t\}$. Even after we will factore in the actual knowledge about the current state (by defining the "local plausibility" relations $\trianglelefteq_a$ below), we will still be able only to compute the agent's *current* belief about the (actual) state. But there is still no real belief *change*, no belief updating, and this is captured by the fact that all states considered live in the same model S, endowed with the same plausibility relation $\leq_a$. So our answer to the above objection is that the choice of interpretation *will* make a difference when we'll have to decide on a specific mechanism for *dynamic* belief revision in section 3: as we shall argue, *only one* such mechanism seems consistent with our qualitative interpretation.

Example 0. Alice is offered a free trip to Las Vegas, provided that she accepts to play the lottery there. Observe that the outcome of playing the lottery is a *virtual* state of a system S, whose possible states correspond to all the possible amounts of money she might win, or lose, in total. But these are just virtual states, which might never be actualized: she might as well turn down the offer and stay home, or go somewhere else etc. Not being able to decide whether to accept or not, Alice goes to a fortune-teller called Jasmine, who gazes in her palm and sees that her fortune line is splitting in two in the near future: this means, says Jasmine, that when playing the lottery she will either lose all her money (state s) or win the Jackpot (state t). Provided that Alice believes Jasmine (-conditionalization!), what will she do next? If she immediately decides to go to Las Vegas, we write $t <_a s$; if she immediately decides to turn down the offer, we write $s <_a t$; if she's still undecided and goes looking for a second opinion (say, by asking a Game Theory expert), we write $s \simeq_a t$. Observe that we do not care about how strong is Alice's belief that she will win the Jackpot (or that she'll lose her money); we don't even care about what decision she'll take after looking for a second opinion; and, of course, we don't care about what would Alice believe in the actual event of her going to Las Vegas and, say, losing only a moderate amount of money: that would be an *a posteriori* belief about the actual outcome. The only thing we care about, when writing down Alice's plausibility relation, is what decision she takes based only on her belief in Jasmine's words.

Notation: local plausibility, doxastic indistinguishability, full indistinguishability. In addition to the "a priori" plausibility relations above, we introduce *local plausibility relations* $\trianglelefteq_a :=\sim_a \cap \leq_a$, or in other words:

$$s \trianglelefteq_a t \text{ iff } s \sim_a t \text{ and } s \leq_a t.$$

Intuitively, this relation combines plausibility with knowledge, to capture the agent's plausibility relation *at a given state*: hence, the name "local plausibility". A detailed interpretation of this notion will be given below. Note that this relation is also a *preorder*, but not necessarily a well-preorder. It does satisfy a weaker condition though, namely that (seen as a set of pairs) *it is a union of mutually disjoint well-preorders*. A relation with this property will be called a *locally well-preordered* relation.

Similar to strict plausibility, we also have a "strict" (i.e. asymmetric) version of local plausibility: $s \triangleleft_a t$ iff $s \trianglelefteq_a t$ but $t \not\trianglelefteq_a s$. As in the case of the plausibility relations, the local plausibility, being a preorder, induces an equivalence relation $\cong_a$ ("full indistinguishability"), canonically defined by putting: $s \cong_a t$ iff both $s \trianglelefteq_a t$ and $t \trianglelefteq_a s$. Observe that two states are fully indistinguishable $s \cong_a t$ iff they are both epistemically indistinguishable $s \sim_a t$ and doxastically indistinguishable $s \simeq_a t$.

Interpretation. Recall that $\leq_a$ captures the agent's (conditional) beliefs *about some virtual state*. In contrast, the local plausibility relation $\trianglelefteq$ captures the agent's *(conditional) beliefs about the actual, current state* of the system: $s \triangleleft_a t$ means that, when given the (correct) information that the actual state of the system is either s or t, agent a *won't know* which of the two (since they are epistemically indistinguishable) but he'll *believe* (based on his a priori plausibility relation $\leq_a$) that the state was in fact s. This explains our definition of $\trianglelefteq_a$ as $\sim_a \cap \leq_a$. Note that, when dealing with the actual state, the plausibility relation is only relevant when having to choose between epistemically indistinguishable states: whenever the agent can use his observations (knowledge) to distinguish the two states, his *a priori* beliefs (expressed by the plausibility relation $\leq_a$) will be overridden by the actual knowledge (about the actual state). Observe that this is still a *static* situation, and this is even more important when talking about the "current state": the conditional information $\{s, t\}$ is still *conditional, i.e. hypothetical.* If the agent would *actually receive* such new information $\{s, t\}$, this would typically *change the current state* of the system (to a new state, different from both s and t, since agent a will have *more* information in this new state). The agent himself should be aware of this change (through introspection), and so his actual belief after receiving the new information should be that the (new, actual) state of the system is different from both s and t. To capture this change we will have to move to *dynamic* belief revision (or belief "update"), to be treated in the next section. But to get back to our static situation, the correct reading of $s \triangleleft_a t$ is thus that: agent a *doesn't know* whether the actual state of the system is s or t, but *after learning that it is either* s or t he *will come to believe* that this actual state (*before the learning*) *was* in fact s.

To better express the difference between $\leq_a$ and $\trianglelefteq_a$, we could differentiate between the agent's beliefs *about* a (given, virtual) state and his beliefs *at* a (given, actual) state. The first are *a priori beliefs* about possible states in general, the second are the actual beliefs that are held by the agent at a given state of the system. The two are of course related: the actual beliefs at a state are a result of combining (via $\trianglelefteq_a = \leq_a \cap \sim_a$) the a priori beliefs with the actual knowledge possessed by the agent at the current state.

Example 0**, continued**. When a state of the system described in the previous example (Alice playing the Las Vegas lottery) is actualized, say by Alice playing the lottery and in the end winning (or losing) some money, Alice's beliefs at this state are given by her local plausibility relation $\trianglelefteq_a$ (not by $\leq_a$). Assuming that she actually *knows* how much money she won or lost, it follows that $\trianglelefteq_a$ is simply the identity relation.

In the rest of this paper, we will only be concerned with the agents' beliefs *at* a state, and for this purpose the *local plausibility* relation $\trianglelefteq_a$ contains all the relevant information. Indeed, observe that we can recover the *epistemic indistinguishability* relation via the identity

$$\sim_a = \trianglelefteq_a \cup \trianglerighteq_a ,$$

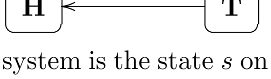
(where $\trianglerighteq_a = (\trianglelefteq_a)^{-1}$ is the converse relation), identity which easily follows from the definition of plausibility frames. It is true that we *cannot recover* the full plausibility relation $\leq_a$, since $\trianglelefteq_a$ can only compare states that are epistemically indistinguishable. But, as argued above, we do not need to compare the others when dealing only with beliefs *at* a state. In other words, if we are only interested in localized beliefs (and knowledge), then our epistemic plausibility models contain *redundant information.* To capture the relevant core of our models we introduce a new notion: local plausibility models.

A *local plausibility frame* (LPF, for short) is a structure $(S, \trianglelefteq_a)_{a\in\mathcal{A}}$, consisting of a set of states S together with a family of locally well-preordered relations $\trianglelefteq_a$ (in the sense defined above: disjoint unions of well-preorders), one for each agent $a \in \mathcal{A}$. A *local plausibility model* (LPM, for short) is an LPF together with a valuation map. Given an LPF, one can *define epistemic indistinguishability* relations by putting $\sim_a := \trianglelefteq_a \cup \trianglerighteq_a$ (where $\trianglerighteq_a = (\trianglelefteq_a)^{-1}$ is the converse), and *full indistinguishability* relations by putting $\cong_a := \trianglelefteq_a \cap \trianglerighteq_a$. Strong local plausibility $\triangleleft_a$ is defined as above.

Proposition. Every EPF $(S, \sim_a, \leq_a)$ generates a LPF, in the obvious (thus canonical) way (by putting $\trianglelefteq_a := \sim_a \cap \leq_a$). Conversely, every LPF $(S, \trianglelefteq_a)$ is the canonical LPF generated by some EPF, having the same epistemic indistinguishability relation $\sim_a = \trianglelefteq_a \cup \trianglerighteq_a$.

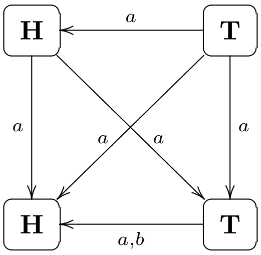
When drawing an LPM in the following examples, we use *labeled arrows to represent the strict converse local plausibility relations* $\triangleright_a$, going from less plausible states to more plausible ones:

Example 1. The referee (Charles) informs Alice and Bob that there is a coin lying face up on the table in front of them. The face is covered (so Alice and Bob cannot see it), but Charles knows which face is up (since he put the coin there). Based on previous experience, (it is common knowledge that) Alice and Bob believe that the upper face is Heads (-they know that Charles has a strong preference for Heads). And in fact, they're right: the coin lies Heads up.



In this picture, the actual state of the system is the state s on the left. We will refer to this plausibility model as $\mathbf{S}$. Alice and Bob don't know which is the real state, but they believe it is s. The strict (local) plausibility relation for Charles is not drawn, since it is trivial (the empty set): in each case, Charles *knows* the state, so the local plausibility relation $\trianglelefteq_c$ is the identity, and thus the strict relation $\triangleleft_c$ is empty.

Example 2. Alice has to get out of the room for a minute, which creates an opportunity for Bob to quickly raise the cover in her absence and take a peek at the coin. He does that (in the presence of Charles) and so he sees that the coin is Heads up. After Alice returns, she obviously doesn't know whether or not Bob took a peek at the coin, but she believes he didn't do it: taking a peek is against the rules of the game, and so she trusts Bob not to do that (and trusts Charles to enforce the rules). The situation now can be depicted in the following way:



Let us call this model $\mathbf{S}'$. The actual state s_1' is the one in the upper left corner, in which Bob took a peek and saw the coin Heads up, while the state t_1' in the upper right corner represents the other possibility, in which Bob saw the coin lying Tails up. The two lower states s_2' and t_2' represent the case in which Bob *didn't take a peek*. Observe that the above drawing includes the (natural) assumption that Alice keeps her previous belief that the coin lies Heads up (since there is no reason for her to change her mind). Moreover, we assumed that she will keep this belief even if she'd be told that Bob took a peek: this is captured by the a-arrow from t_1' to s_1'. This seems natural: Bob's taking a peek doesn't change the upper face of the coin, so it shouldn't affect Alice's prior belief about the coin.

Epistemic Appearance and (Conditional) Doxastic Appearance of a State. Given an *LPM* $\mathbf{S}$, the *epistemic appearance* of a state s to an agent a is simply the equivalence class $s(a) := \{t \in S : s \sim_a t\}$ of s with respect to the equivalence relation $\sim_a$, i.e. the set of all states that are epistemically indistinguishable from the actual state s. This precisely captures *all the knowledge (about the current state) possessed by the agent* at state s, and it correctly describes *the way state s appears to agent a (for all he knows).* To similarly capture belief, we define the *doxastic appearance* of state s to agent a to be the set $s_a := Min_{\leq_a} s(a) = Min_{\trianglelefteq_a} s(a)$ of all $\trianglelefteq_a$-minimal states of $s(a)$: these are the "most plausible" states that are consistent with the agent's knowledge at state s. We can put this in a relational form, by defining a *doxastic accessibility relation* $\to_a$ by putting

$$s \to_a t \quad \text{iff} \quad t \in s_a \quad \text{iff} \quad t \sim_a s \text{ and } t \trianglelefteq_a t' \text{ for all } t' \sim_a s$$

We read this as saying that: when the actual state is s, agent a *believes* that any of the states t with $s \to_a t$ *may be* the actual state. We can extend this to capture *conditional beliefs* (in full generality), by associating to each S-proposition $P \subseteq S$ and each state $s \in S$ the *conditional doxastic appearance* s_a^P *of state s to agent a, given (information) P.* This can be defined as the set

$$s_a^P := Min_{\leq_a} s(a) \cap P = Min_{\trianglelefteq_a} s(a) \cap P$$

of all $\trianglelefteq_a$-minimal states of $s(a) \cap P$: these are the "most plausible" states satisfying P that are consistent with the agent's knowledge at state s. To put this relationally, we can define: $s \to_a^P t$ iff $t \in s_a^P$; or, equivalently:

$$s \to_a^P t \text{ iff } t \in P, t \sim_a s \text{ and } t \trianglelefteq_a t' \text{ for all } t' \in P \text{ such that } t' \sim_a s.$$

The conditional belief operator B_a^P (as defined below) will simply by the Kripke modality associated to the accessibility relation to_a^P. We mention here that the local plausibility frames can be alternatively described only in terms of the conditional doxastic maps (or relations) $s \mapsto s_a^P$: indeed, in [6, 7] we axiomatized a notion of *conditional doxastic frames* as structures of type $(S, \to_a^P)_{a \in \mathcal{A}, \mathrm{P} \subseteq \mathrm{S}}$, and proved them to be equivalent to *LPF*'s.

One can easily see how this approach relates to a more widely adopted definition for conditional beliefs; in [9], [11], [25], this definition involves the assumption of a "truly local" *plausibility relation at a given state* $s \leq_a^w t$, to be read as: "at state w, agent a considers state s at least as plausible as state t ".

Given such a relation, the conditional belief operator is usually defined in terms that are equivalent to putting $s_a^P := Min_{\leq_a^s} P$, and then taking the Kripke modality corresponding to the associated relation $\rightarrow_a^P$ (defined as above by $s \rightarrow_a Pt$ iff $t \in s_a^P$). One could easily restate our above definition in this form, by putting:

$$s \rightarrow_a^w t \;\text{ iff }\; \text{ either } t \not\sim_a w \;\text{ or }\; s \sim_a w \sim_a t, s \leq_a t.$$

The converse problem is studied in [9], which gives the precise conditions that need to be imposed on the "truly local" relations $\leq_a^w$ in order to recover a global plausibility relation $\leq_a$.[4]

Epistemic and (Conditional) Doxastic Modalities. Recall that, given an LPF (and thus an EPF) S, an S-proposition is just a subset $Q \subseteq S$. We have thus the usual Boolean operations with propositions $P \wedge Q := P \cap Q$, $P \vee Q : P \cup Q$, $\neg P := S \setminus P$, $P \rightarrow Q := \neg P \vee Q$ etc., as well as Boolean constants $\top_S := S$ and $\bot := \emptyset$ for the "always true" and "always false" S-propositions. But, in addition, we can introduce *modalities* (i.e. unary operations on propositions). In particular, any binary relation $R \subseteq S \times S$ on $\mathbf{S}$ gives rise to a *Kripke modality* $[R] : \mathcal{P}(S) \rightarrow \mathcal{P}(S)$, defined by

$$[R]Q := \{s \in S : \forall t\,(sRt \Rightarrow t \in Q)\}.$$

As instances of this, we define *epistemic and (conditional) doxastic modalities* as the Kripke modalities for the epistemic and (conditional) doxastic accessibility relations: $K_a Q := [\sim_a]Q$, $B_a Q := [\rightarrow_a]Q$, $B_a^P Q := [\rightarrow_a^P]Q$. We also define the dual (Diamond) of the K-modality as $\tilde{K}_a P := \neg K_a \neg Q$. We read $K_a Q$ as saying that agent *a knows Q*, and read $B_a Q$ ($B_a^P Q$) as saying that agent *a believes Q* (*given condition P*, or *the hypothesis P*). We interpret the conditional belief statement $s \in B_a^P Q$ in the following way: if the actual state is s, then after coming to believe that P is the case (at this actual state), agent a will believe that Q *was* the case (at the same actual state, before his change of belief). In other words, conditional beliefs B_a^P give descriptions of the agent's *plan* (or *commitments*) about what he will believe about the current state after receiving new (believable) information. In Johan van Benthem's words in [25]: conditional beliefs, though *static* (i.e. referring to the same state of the world, not to the states generated by belief changes), "pre-encode" the agent's potential belief changes in the face of (potential) new information. Finally, observe that conditional beliefs generalize the local plausibility relations, which are nothing but special cases of conditional belief:

$$s \trianglelefteq_a t \text{ iff } s \in B_a^{\{s,t\}}\{s\} \text{ iff } t \in B_a^{\{s,t\}}\{s\}.$$

Relations between knowledge, belief and conditional belief. Observe that, for all propositions $Q \subseteq S$, we have $B_a Q = B_a^S Q$, i.e. *(unconditional) belief is the same as belief conditionalized by the trivially true proposition* $\top_S = S$. Similarly, observe that $K_a Q = \bigcap_{P \subseteq S} B_a^P Q$, or equivalently:

$$s \in K_a P \;\text{ iff }\; s \in B_a^P Q \;\text{ for all } P \subseteq S. \tag{1}$$

This gives a characterization of *knowledge as "absolute" belief, invariant under any belief revision*: a given belief is "known" iff it cannot be revised, i.e. it would be still believed in any condition.[5] Observe that this resembles Stalnaker's *defeasibility analysis* of knowledge in [23], based on the idea that "if a person has knowledge, than that person's justification must be sufficiently strong that it is not capable of being defeated by evidence that he does not possess" (Pappas and Swain [16]). However, Stalnaker interprets "evidence" as "true information", and thus the concept of knowledge in [23] differs from ours (corresponding in fact to what we will call "safe belief"). Unlike the one in [23], our notion of knowledge *is negatively introspective* and corresponds to interpreting "evidence" in the above quote as meaning "any information, be it truthful or not". Thus, our "knowledge" is more robust than Stalnaker's: it resists any belief revision, not capable of being defeated by *any* evidence (including false evidence). This is a very "strong" notion of knowledge (implying "absolute certainty" and full introspection), which seems to us to

[4] The analysis in [9] applies to a system without knowledge operators, context in which the global relation $\leq_a$ coincides with our local relation $\trianglelefteq_a$. But the same conditions, when applied to our knowledge-enriched setting, ensure the recovery of the relation $\trianglelefteq_a$.

[5] This of course assumes agents to be "rational" in a sense that excludes "fundamentalist" or "dogmatic" beliefs, i.e. beliefs in unknown propositions but refusing any revision, even when contradicted by facts. But this "rationality" assumption is already built in our plausibility models, which satisfy an epistemically friendly version of the standard AGM postulates of rational belief revision. See [6] for details.

fit better with the standard usage of the term in Computer Science literature. Nevertheless, we consider the weaker concept to be equally important, and introduce below under the name of "safe belief".

Another identity that can be easily checked is:

$$K_a P = B_a^{\neg Q} Q = B_a^{\neg Q} \bot \tag{2}$$

(where $\bot := \emptyset$). In other words: something is "known" if conditionalizing our belief with its negation is impossible (i.e. it would lead to an inconsistent belief). This corresponds to yet another of Stalnaker's notions of knowledge, defined in [22] in terms of doxastic conditionals, using the above identity (2).

Safe Beliefs. One last modality will play an important role: the Kripke modality $\Box_a$ associated to the converse $\trianglerighteq_a$ of the local plausibility relation, i.e. given by

$$\Box_a Q := [\trianglerighteq_a] Q$$

for all **S**-propositions $Q \subseteq S$. We read $s \in \Box_a Q$ as saying that: *at state s, agent a's belief of Q (being the case) is safe*; or at *state s, a safely believes that Q*. We will explain this reading below, but first observe that: $\Box_a$ is an $S4$-modality (since $\trianglerighteq_a$ is reflexive and transitive), but not necessarily $S5$; i.e. *safe beliefs are truthful* ($\Box_a Q \subseteq Q$) *and positively introspective* ($\Box_a Q = \Box_a \Box_a Q$), but not necessarily negatively introspective; *knowledge implies safe belief*: $K_a Q \subseteq \Box_a Q$ (i.e. known statements are safely believed); and *safe belief implies belief*: $\Box_a Q \subseteq B_a Q$ (i.e. if something is safely believed then it is believed).

The last observation can be strengthened to characterize safe belief in a similar way to the above characterization (1) of knowledge (as belief invariant under any revision): *safe beliefs are precisely the beliefs which are persistent under revision with any true information.* Formally, this says that:

$$s \in \Box_a Q \quad \text{iff} \quad s \in B_a^P Q \;\text{ for all } P \subseteq S \text{ such that } s \in P \tag{3}$$

We can thus see that "safe belief" coincides with Stalnaker's non-standard notion of "knowledge" described above, and formally defined in [23] as: "an agent knows that φ if and only if φ is true, she believes that φ, and she continues to believe φ if any *true* information is received". As mentioned above, we prefer to keep the name "knowledge" for the strong notion (which gives absolute certainty), and call this weaker notion "safe belief": indeed, these are beliefs that are "safe" to hold, in the sense that no future learning of truthful information will force us to revise them. Comparing the last identity with the above characterization of knowledge, we can see that safe beliefs can indeed be understood as a form of *"weak (non-negatively-introspective) knowledge"*. But observe that *an agent's belief can be safe without him necessarily knowing this*: "safety" (similarly to "truth") is an *external* property of the agent's beliefs, that can be ascertained only by comparing his belief-revision system with reality. In fact, *all* beliefs held by an agent "appear safe" to him: in order to believe them, he has to believe this belief to be safe. This is expressed by the valid identity

$$B_a Q = B_a \Box_a Q \,,$$

saying that: *believing something is the same as believing that it is safe to believe it.* Moreover, the only way for an agent to *know* that one of his beliefs is safe is to actually *know it to be truthful*, i.e. to have actual *knowledge* (not just a belief) of its truth. This is captured by the valid identity

$$K_a Q = K_a \Box_a Q$$

In other words: *knowing something is the same as knowing that it is safe to believe it.*

Another important observation is that *one can characterize belief and conditional belief in terms of knowledge and safe belief.* We only give here the characterization[6] of (unconditional) belief:

$$B_a Q = \tilde{K}_a \Box_a Q \tag{4}$$

Examples 1 and 2 continued. Observe that in both Examples 1 and 2 above, Alice holds a *true belief* (at the real state) that the coin lies Heads up: the actual state satisfies $B_a\mathbf{H}$. In both cases, this true belief is *not knowledge* (since Alice doesn't know the upper face), but nevertheless in Example 1, this

[6] In which recall that $\tilde{K}_a P = \neg K_a \neg Q$.

belief is *safe* (altough it is *not known by the agent to be safe*): no additional truthful information (about the real state s) can force her to revise this belief. (To see this, observe that any *new* truthful information would reveal to Alice the real state s, thus confirming her belief that Heads is up, which in this way would become knowledge.) So in the first model $\mathbf{S}$ we have $s \models \Box_a \mathbf{H}$ (where s is the actual state). In contrast, in Example 2, Alice's belief (that the coin is Heads up), though true, is *not safe*. There is some piece of correct information (about the real state s'_1) which, if learned by Alice, would make her change this belief: we can represent this piece of correct information as the doxastic proposition $\mathbf{H} \to \mathbf{K_b H}$. It is easy to see that the actual state s'_1 of the model $\mathbf{S'}$ satisfies the proposition $B_a^{\mathbf{H}\to\mathbf{K_b H}}\mathbf{T}$ (since $(\mathbf{H} \to \mathbf{K_b H})_{\mathbf{S'}}$ $= \{s'_1, t'_1, t'_2\}$ and the minimal state in the set $s'_1(a) \cap \{s'_1, s'_1, t'_2\} = \{s'_1, t'_1, t'_2\}$ is t'_2, which satisfies $\mathbf{T}$.) So, if given this information, Alice would come to wrongly believe that the coin is Tails up!

3 Modeling Belief Change: Plausibility Action Models

Doxastic Propositions. A *doxastic proposition* is a map $\mathbf{P}$ assigning to each plausibility model $\mathbf{S}$ an $\mathbf{S}$-proposition, i.e. a set of states $\mathbf{P_S} \subseteq S$. We denote by *Prop* the family of all doxastic propositions. Note that all the operations (Boolean operations, doxastic and epistemic modalities) defined above on $\mathbf{S}$-propositions, for any given *EPM* $\mathbf{S}$, induce corresponding operations on doxastic propositions: this is done point-wise (or if we like, "model-wise"), since doxastic propositions are just maps defined on the family of all models; for instance, for any doxastic proposition $\mathbf{P}$ and any agent a, we define the proposition $K_a \mathbf{P}$ by putting: $(K_a \mathbf{P})_{\mathbf{S}} := K_a \mathbf{P_S}$, for all plausibility models $\mathbf{S}$.

In [3], it was argued that *epistemic actions should be modeled in essentially the same way as epistemic states*, and this common setting was taken to be given by *epistemic Kripke models*. Since in this paper we enriched our state models with doxastic plausibility relations to deal with (conditional) beliefs, it is natural to follow [3] into extending the similarity between actions and states to this setting, thus obtaining *(epistemic) action plausibility models*. The idea of such an extension was first developed in [2] (for a different notion of plausibility model and a different notion of update product), then generalized in [11], where many types of action plausibility models and notions of update product are explored, in the context of the "degrees of belief" interpretation. As we shall see, our notion of anti-lexicographic product update can be recovered in this context as one of the possible choices enumerated in [11]. But the advantage of our qualitative interpretation is that it seems to justify only *one* such choice: the one we have chosen here.

In our sense, an *action plausibility model* (*APM*, for short) is just an epistemic plausibility frame $(\Sigma, \sim_a, \leq_a)_{a\in\mathcal{A}}$, together with a *precondition map* $pre : \Sigma \to Prop$ associating to each element of Σ some doxastic proposition $pre(\sigma)$. As in [3], we call the elements of Σ *(basic) epistemic actions*, and we call pre_σ *the precondition of action* σ. The basic actions $\sigma \in \Sigma$ are taken to represent some *deterministic* actions of a particularly simple nature. Intuitively, the precondition defines the *domain of applicability* of σ: this action can be executed on a state s iff s satisfies its precondition. The plausibility pre-orderings $\leq_a$ give the agent's beliefs about which actions are more plausible than others. As for state models, one can consider the associated *local plausibility relation* $\trianglelefteq_a$ on actions, and thus reduce the action models to *local action plausibility models* $(\Sigma, \trianglelefteq_a)_{a\in\mathcal{A}}$, which contain all the information that we will need.

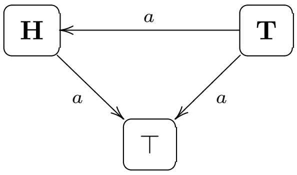
Interpretation: Beliefs about Changes encode Changes of Beliefs. The name "doxastic actions" might be a bit misleading; the elements of a plausibility model are not intended to represent "real" actions in all their complexity, but only the *doxastic changes* induced by these actions: each of the nodes of the graph represents a *specific kind of change of beliefs (of all the agents)*. As in [3], we only deal here with pure "belief changes", i.e. actions that do not change the "ontic" facts of the world, but only the agents' beliefs[7]. Moreover, we think of these as *deterministic* changes: there is at most one output of applying an action to a state.[8] Intuitively, the precondition defines the *domain of applicability* of σ: this action can be executed on a state s iff s satisfies its precondition. The plausibility pre-orderings $\leq_a$ give the agent's prior conditional beliefs about (virtual) actions. But this should be interpreted as *beliefs about*

[7] We stress this is a minor restriction, and it is very easy to extend this setting to "ontic" actions. The only reason we stick with this restriction is that it simplifies the definitions, and that it is general enough to apply to all the actions we are interested here, and in particular to all *communication actions*.

[8] As in [3], we will be able to represent non-deterministic actions as sums (unions) of deterministic ones.

changes, that *encode changes of beliefs*. In this sense, we use such "beliefs about actions" as a way to represent doxastic changes: the information about how the agent changes her beliefs is captured by our action plausibility relations. So we read $\sigma <_a \sigma'$ as saying that: if, prior to any action happening, agent a is given some information about some possible history of a virtual system until and including some action happening at some moment , and if this information is enough to determine that the action is *either* σ *or* σ', but not enough to determine which of the two, then she believes the action to be σ . As before, the *local plausibility relation* $\trianglelefteq_a$ adds to the agent's prior beliefs her direct knowledge about the current action in an actual system, and thus computes the *actual (conditional) beliefs about the current action.*

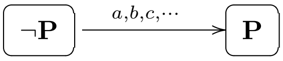
Example 3: Private (Group) Announcements. Let us consider the *action* that produced the situation represented in Example 2 above. This was the action of Bob taking a peek at the coin, while Alice is away (but in the presence of Charles). Recall that we assumed that Alice *believes that nothing is really happening* in her absence (since she assumes the others play by the rules), though obviously she *doesn't know* this (that nothing is happening). In the literature on dynamic-epistemic logic, this action is usually called *a private announcement to a subgroup*: the group consisting of Bob and Charles gains common knowledge of which face is up, while the outsider Alice believes nothing is happening. To represent this, we consider an action model $\mathbf{\Sigma}$ consisting of three "actions": the actual action σ in which Bob takes a peek and sees the coin lying Heads up; the alternative possible action ρ is the one in which Bob sees the coin lying Tails up; finally, the action τ is the one in which "nothing is really happening" (as Alice believes). As before, we draw arrows for the (converse) strict local plausibility relations $\rhd_a$:



Here, the action σ is the one in the upper left corner, having precondition $\mathbf{H}$: indeed, this can happen iff the coin is really lying Heads up; similarly, the action ρ in the upper right corner has precondition $\mathbf{T}$, since it can only happen iff the coin is Tails up; finally, the action τ is the lower one, having as precondition the "universally true" doxastic proposition $\top$ (defined by $\top_{\mathbf{S}} := S$, for all state models $\mathbf{S}$): indeed, this action can always happen (since in it, nothing is really happening!). The plausibility relations reflect what each agent believes: in each case, both Bob and Charles know exactly what is happening, so their local plausibility relations are the identity (and so their strict relations are empty). Alice believes nothing is happening, so τ is the most plausible action for her (to which all her arrows are pointing); but she also keeps her belief that $\mathbf{H}$ is the case, so she considers σ as more plausible than τ.

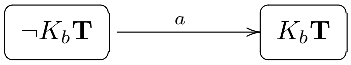
Example 3 (b): Public Group Announcements of "Hard Facts". A special case of the private announcement to a subgroup is when the subgroup consists of all the agents: *truthful public announcement* $\mathbf{P}!$ of some epistemic proposition $\mathbf{P}$. This is the action described in [25] as (public) "belief change under hard facts". The action model consists of only one node, labeled with $\mathbf{P}$.

Example 3 (c): Public Announcements of "Soft Facts": The Lexicographic Update. To allow for "soft" belief revision, an action $\Uparrow \mathbf{P}$ was introduced in [25], essentially adapting to public announcements the lexicographic policy for belief revision in [17]. This action changes the current preorder on the state model, such that *all* $\mathbf{P}$*-worlds become better than all* $\neg\mathbf{P}$*-worlds, and within the two zones, the old ordering remains.* In our setting, this can be represented by the *right node* of the following model:



Our anti-lexicographic update product, defined below, will make this action model "act" on a state exactly like the lexicographic update $\Uparrow \mathbf{P}$ in [25].

Example 4: Successful Lying. Suppose now that, after the previous action (of Bob taking a peek in Alice's absence, followed by the return of an unsuspicious Alice), Charles loudly announces: "Bob has cheated, he took a peek and saw the coin was lying Tails up". For our purposes, we can formalize the content of this announcement as $K_b\mathbf{T}$, i.e. saying that "Bob knows the coin is lying Tails up". (If we want to incorporate the information that this is uttered by Charles, who thus claims to know this, we may instead formalize the message as $K_c K_b \mathbf{T}$; but this is a minor change, with no relevance to our modeling.) This is a *public announcement*, but not a truthful one: it is a *lie*! However, let us suppose that this is a *successful lie*: Alice trusts Charles, so she believes him. The action is given by the *left node* in the following model Σ':



This model has two actions: α, the one on the left, is the real action (successful lying) that is taking place, has precondition $\neg K_b \mathbf{T}$, since it takes place when in fact Bob *doesn't know* the coin is Tails up (and moreover he knows it is Heads up); β, the one on the right, is the action that Alice believes to be happening, namely the one in which Charles is telling the truth: so β is just a truthful public announcement, having $K_b \mathbf{T}$ as precondition (since it can happen only if Bob really knows $\mathbf{T}$).

The Product Update of Two Plausibility Models. We are ready now to define the *updated (state) plausibility model*, representing the way an action lying in an *APM* $\mathbf{\Sigma} = (\Sigma, \sim_a, \leq_a, pre)_{a \in \mathcal{A}}$ will act on an input-state lying in an initially given (state) *EPM* $\mathbf{S} = (S, \sim_a, \leq_a, \| \bullet \|)_{a \in \mathcal{A}}$. We denote this updated model by $\mathbf{S} \otimes \mathbf{\Sigma}$, and we call it the *update product* of the two models. Its states are elements (s, σ) of the Cartesian product $S \times \Sigma$. More specifically, the set of states of $\mathbf{S} \otimes \mathbf{\Sigma}$ is

$$S \otimes \Sigma := \{(s, \sigma) : s \in pre(\sigma)_{\mathbf{S}}\}$$

The valuation is given by the original input-state model: for all $(s, \sigma) \in S \otimes \Sigma$, we put $(s, \sigma) \models p$ iff $s \models p$. As epistemic uncertainty relations, we take the *product* of the two epistemic uncertainty relations[9]: for $(s, \sigma), (s', \sigma') \in S \otimes \Sigma$,

$$(s, \sigma) \sim_a (s', \sigma') \text{ iff } \sigma \sim_a \sigma', s \sim_a s'.$$

Finally, we define the plausibility relation as the *anti-lexicographic preorder relation on pairs* (s, σ), i.e.:

$$(s, \sigma) \leq_a (s', \sigma') \text{ iff either } \sigma <_a \sigma' \text{ or } \sigma \simeq_a \sigma', s \leq_a s'.$$

The above definition roughly corresponds to one of the update product constructions encountered in the literature (essentially incorporating the so-called "*maximal-Spohn revision*" into plausibility action models). But, in the context of our qualitative (conditional) interpretation of plausibility models, we will argue below that this is essentially the only coherent option. In any case, we regard this construction as the most natural analogue in a belief-revision context of the similar notion in [3, 4].

Note that our operation $\otimes$ on epistemic plausibility models induces a corresponding operation (also denoted by $\otimes$) on the corresponding *local* plausibility models. This can be defined directly: given a "state" model $\mathbf{S} = (S, \trianglelefteq_a)_{a \in \mathcal{A}}$ and an "action" model $\mathbf{\Sigma} = (\Sigma, \trianglelefteq_a)_{a \in \mathcal{A}}$, we define the update product $\mathbf{S} \otimes \mathbf{\Sigma}$ by taking the same states $S \otimes \Sigma$ as above, taking the same valuation and putting:

$$(s, \sigma) \trianglelefteq_a (s', \sigma') \text{ iff either } \sigma \lhd_a \sigma', s \sim_a s' \text{ or } \sigma \cong_a \sigma', s \leq_a s' .$$

Since in our examples we used only the local plausibility relations, this product operation on local plausibility models will be the one actually used in practice.

Interpretation. To explain the definition of the anti-lexicographic product update, recall first that we only deal with *pure "belief changes"*, not affecting the "facts": this explains our "conservative" valuation. Second, the product construction on the epistemic indistinguishability relation $\sim_a$ is the same as in [3]: if two indistinguishable actions are successfully applied to two indistinguishable input-states, then their output-states are indistinguishable. Third, the anti-lexicographic preorder gives "priority" to the *action* plausibility relation; this is not an arbitrary choice, but is motivated by our above-mentioned interpretation of "actions" as specific types of *belief changes*. The action plausibility relation captures what agents *really believe is going on at the moment*; while the input-state plausibility relations only capture *past beliefs*. The doxastic action is the one that "changes" the initial doxastic state, and not vice-versa. If the "believed action" α requires the agent to revise some past beliefs, then so be it: this is

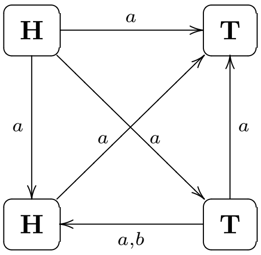
[9] Observe that this is precisely the uncertainty relation of the epistemic update product, as defined in [3].

the whole point of believing α, namely to use it to revise or update one's past beliefs. For example, in a successful lying, the action plausibility relation makes the hearer believe that the speaker is telling the truth; so she'll accept this message (unless contradicted by her knowledge), and change her past beliefs appropriately: this is what makes the lying "successful". Giving priority to action plausibility does not in any way mean that the agent's belief in actions is "stronger" than her belief in states; it just captures the fact that, at the time of updating with a given action, *the belief about the action is what is actual, is the current belief about what is going on, while the beliefs about the input-states are in the past.*[10] The belief update *induced by a given action* is nothing but *an update with the (presently) believed action.*

In other words, the anti-lexicographic product update reflects our Motto above: *beliefs about changes* (as formalized in the action plausibility relations) *are nothing but ways to encode changes of belief* (i.e. ways to change the original plausibility order on states). This simply expresses our *particular interpretation* of the (strong) plausibility ordering on actions, and is thus a matter of *convention*: we decided to introduce the order on actions to encode corresponding *changes of order* on states. *The product update is a consequence of this convention*: it just says that a strong plausibility order $\alpha <_a \beta$ on actions corresponds indeed to a change of ordering, (from whatever the ordering was) between the original input-states s, t, to the order $(s, \alpha) <_a (t, \beta)$ between output-states; while equally plausible actions $\alpha \simeq_a \beta$ will leave the initial ordering unchanged: $(s, \alpha) \leq_a (t, \beta)$ iff $s \leq_a t$. This becomes obvious if we recall the *meaning* of our plausibility relations, for both actions and states: if all the available information about the history of the system is enough to reduce the uncertainty to a binary alternative (between two current states, or two current actions), but it is still compatible with *both* options, then the agent believes that the current state or action is the one that is minimal in the strict order $<_a$; or, if none is strictly smaller, the agent remains undecided. Applying this to the current action, we notice that the "current" state is part of the past history of the current action. Hence, when choosing what to believe about the action, the agent will by definition choose the minimal action in the case that both are compatible with the given histories (i.e. with the corresponding states); the other action is simply eliminated from her current beliefs. But once the agent chooses to believe an action, this will obviously affect her beliefs about its past history: the states leading to any action that is *not believed* to be happening are themselves *no longer believed* to have occurred. Only in the case that none of the two actions is excluded from the current belief (i.e. when the agent is undecided between the actions), will the previous beliefs about past history be maintained essentially unchanged: so when $\sigma \simeq_a \sigma'$, the current belief about the (current) state is simply decided by the past beliefs. In other words: the product update is just a formalization of *our qualitative interpretation* of action plausibility models, and thus it doesn't impose any further limitation to our setting.

Example 3, continued. To check the correctness of our update operation, take the update product of the (local plausibility) state model $\mathbf{S}$ from Example 1 with the (local plausibility) action model $\mathbf{\Sigma}$ in Example 3. As predicted, the resulting state model is isomorphic to the model $\mathbf{S}'$ in Example 2.

Example 4, continued. Applying the action model $\mathbf{\Sigma}'$ in Example 4 (for "successful lying") to the state model $\mathbf{S}'$ from Example 2, we obtain the following:



Dynamic Modalities: The Weakest Precondition of a Doxastic Action. Given a doxastic action α (living in some APM $\mathbf{\Sigma}$, we can define a corresponding dynamic modality, capturing the *weakest precondition* of α: for every proposition $\mathbf{P}$, the proposition $[\alpha]\mathbf{P}$ is given by $([\alpha]\mathbf{P})_{\mathbf{S}} := \{s \in S : \text{ if } (s, \alpha) \text{ is defined } \in \mathbf{S} \otimes \mathbf{\Sigma} \text{ then } (s, \alpha) \in \mathbf{P}_{\mathbf{S} \otimes \mathbf{\Sigma}}\}$.

[10] Of course, *at a later moment*, the above-mentioned belief about action (*now* belonging to the past) might be itself revised. But this is another, *future update*.

The Laws of Dynamic Belief Revision. The following “dynamic reduction” laws hold:

$$\begin{array}{rcl}
[\alpha]p & = & pre_\alpha \rightarrow p \\
[\alpha]\neg\mathbf{P} & = & pre_\alpha \rightarrow \neg[\alpha]\mathbf{P} \\
[\alpha](\mathbf{P}\wedge\mathbf{Q}) & = & pre_\alpha \rightarrow [\alpha]\mathbf{P}\wedge[\alpha]\mathbf{Q} \\
[\alpha]K_a\mathbf{P} & = & pre_\alpha \rightarrow \bigwedge_{\beta\sim_a\alpha} K_a[\beta]\mathbf{P} \\
[\alpha]\Box_a\mathbf{P} & = & pre_\alpha \rightarrow \bigwedge_{\beta\lhd_a\alpha} K_a[\beta]\mathbf{P} \wedge \bigwedge_{\gamma\cong_a\alpha} \Box_a[\gamma]\mathbf{P}\ ,
\end{array}$$

for all propositions $\mathbf{P}, \mathbf{Q}$, actions α and atomic sentences p. Using above-mentioned characterizations of (conditional) belief, we can deduce reduction laws for B_a and $B_a^{\mathbf{P}}$. We give here the one for simple belief:

$$\begin{array}{rcl}
[\alpha]B_a\mathbf{P} & = & pre_\alpha \rightarrow \bigwedge_{\beta\sim_a\alpha} \tilde{K}_a < \beta > \Box_a\mathbf{P} \\
 & = & pre_\alpha \rightarrow \bigwedge_{\beta\sim_a\alpha} \left(\tilde{K}_a pre_\alpha \wedge \bigwedge_{\gamma\lhd_a\beta} K_a[\gamma]\mathbf{P} \wedge \bigwedge_{\delta\cong_a\beta} B_a^{pre_\alpha}[\delta]\mathbf{P} \right)
\end{array}$$

Theorem (**Relative Completeness**). Given any complete axiomatization of the logic of conditional beliefs and safe belief, the above reduction laws give a *complete axiomatization of the dynamic logic of doxastic actions*, for the semantics given by plausibility models and the anti-lexicographic product update.

Acknowledgments. We thank Hans van Ditmarsch and an anonymous referee for their extremely useful critical comments on this paper, as well as for helping us to get a much better view of the history of the subject. We also thank Johan van Benthem for his illuminating discussion of the topic. This paper is essentially based (as were our previous contributions to dynamic belief revision, and many other related papers by various authors in this field) on insights arising from his pioneering work.

References

[1] C.E. Alchourron, P. Gardenfors, D. Makinson. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *Journal of Symbolic Logic*, **50** (2), 510-530. 1985.

[2] G. Aucher. *A Combined System for Update Logic and Belief Revision.* Master’s thesis, Univ. of Amsterdam. 2003.

[3] A. Baltag and L.S. Moss. Logics for Epistemic Programs. *Synthese*, **139**, 165-224. 2004.

[4] A. Baltag, L.S. Moss and S. Solecki. The Logic of Common Knowledge, Public Announcements, and Private Suspicions. In I. Gilboa (ed.), *Proceedings of the TARK’98*, 43-56. 1998.

[5] A. Baltag and M. Sadrzadeh. The Algebra of Multi-Agent Dynamic Belief Revision. 2005. To appear in *ENTCS*.

[6] A. Baltag and S. Smets. Conditional Doxastic Models: A Qualitative Approach to Dynamic Belief Revision. 2006. Submitted to WOLLIC ’06. Available at http://www.vub.ac.be/CLWF/SS/WOLLIC.pdf

[7] A. Baltag and S. Smets. The Logic of Conditional Doxastic Actions: A Theory of Dynamic Multi-Agent Belief Revision. 2006. Submitted to the ESSLLI’06 Workshop on Rationality and Knowledge. Available at http://www.vub.ac.be/CLWF/SS/RAK.pdf.

[8] B. P. Kooi. Probabilistic Dynamic Epistemic Logic. *Journal of Logic, Language and Information* 12, 381-408. 2003.

[9] O. Board. Dynamic interactive epistemology. *Games and Economic Behaviour* **49**, 49-80. 2002.

[10] G. Bonanno. A Simple Modal Logic for Belief Revision. *Synthese* **147: 2**, 193-228. 2005.

[11] H. van Ditmarsch. Prolegomena to Dynamic Logic for Belief Revision. *Synthese*, **147**, 229-275. 2005.

[12] P. Gardenfors. Belief Revisions and the Ramsey Test for Conditionals. *Philosophical Review*, **95**, 81-93. 1986.

[13] P. Gardenfors. *Knowledge in Flux: Modelling the Dynamics of Epistemic States.* MIT Press, Cambridge MA. 1988.

[14] A. Grove. Two Modellings for Theory Change. In *Journal of Philosophical Logic*, **17**, 157-170. 1988.

[15] J. Y. Halpern. *Reasoning about Uncertainty.* MIT Press, 2003.

[16] G. Pappas and M. Swain (eds). *Essays on Knowledge and Justification.* Cornell Univ. Press, Ithaca, NY. 1978.

[17] H. Rott. Conditionals and theory change: revisions, expansions, and additions. *Synthese*, **81**, 91-113. 1989.

[18] M. Ryan, P.Y. Schobbens. Counterfactuals and updates as inverse modalities. *Journal of Logic, Language and Information.* 1997.

[19] K. Segerberg. Irrevocable Belief Revision in Dynamic Doxastic Logic. *Notre Dame Journal of Formal Logic*, **39**, No 3, 287-306. 1998.

[20] K. Segerberg. Default Logic as Dynamic Doxastic Logic. *Erkenntnis*, **50**, 333-352. 1999.

[21] W. Spohn. Ordinal conditional functions: A dynamic theory of epistemic states. In W.L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change and Statistics*, vol. 2, 105-134. Reidel, Dordrecht, 1988.

[22] R.C. Stalnaker. A Theory of Conditionals. *Studies in Logical Theory*, Oxford, Blackwell, APQ Monograph No2, 1968.

[23] R.C. Stalnaker. Knowledge, Belief and Counterfactual Reasoning in Games. *Economics and Philosophy* **12**, 133-163. 1996.

[24] J. van Benthem, J. van Eijck and B. Kooi. Logics of Communication and Change. Available at http://staff.science.uva.nl/ johan/publications.html.

[25] J. van Benthem. Dynamic Logic for Belief Change. Working-paper (version 30 november) 2005.