Conditional Doxastic Models:
A Qualitative Approach to Dynamic Belief Revision

Alexandru Baltag [1]

Computing Laboratory
Oxford University
Oxford, UK.

Sonja Smets [2,3]

Center for Logic and Philosophy of Science
Vrije Universiteit Brussel
Brussels, Belgium

Abstract

In this paper, we present a semantical approach to multi-agent belief revision and belief update. For this, we introduce relational structures called *conditional doxastic models* (CDM's, for short). We show this setting to be equivalent to an epistemic version of the classical AGM Belief Revision theory. We present a *logic of conditional beliefs* that is complete w.r.t. CDM's. Moving then to *belief updates* (sometimes called "dynamic" belief revision) induced by epistemic actions, we consider two particular cases: *public announcements* and *private announcements to subgroups* of agents. We show how the standard semantics for these types of updates can be appropriately modified in order to apply it to CDM's, thus incorporating belief revision into our notion of update. We provide a complete axiomatization of the corresponding dynamic doxastic logics. As an application, we solve a "cheating version" of the Muddy Children Puzzle.

> *Key words:* belief revision, belief update, conditional belief, dynamic epistemic
> logic, public announcement, modal logic, multi-agent system

1 Introduction

Once upon a time there were three very wise children, playing in a garden, under the tall trees. Despite their father's warning, naughty Adam and Eve got mud on their foreheads, but obedient Mary stayed clean. Then the father came to them and said:"Behold, at least one of you is dirty".

The story might have easily gone the usual way, with the father repeatedly asking'em if they (knew, or justifiably believed, that they) were dirty or not, until the children had arrived

---

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: `www.elsevier.nl/locate/entcs`

to the correct answer by the sheer power of pure logic. But ... pretty Eve was an impatient girl: before answering any questions, she quickly took a glance into her pocket-mirror, without anybody even suspecting this. So she immediately answered "yes, dear father, I know: I'm dirty, and I am sorry", while the others could only confess their ignorance. Dirty girl Eve, indeed!

But what would the other two answer if the compassionate father repeated the Question? Adam's answer can be correctly predicted using a special case (*private announcements to subgroups*) of the *logic of epistemic actions*, introduced in [5,6,4] and which we will hereby refer to as "the Rightful Logic". Indeed, Eve's peek in the mirror can be thought of as a fully private announcement (that "Eve is dirty") having only herself as the recipient. Using Rightful Logic, one can prove that Adam will come to the incorrect (but logically justified) conclusion that he's clean. This agrees with our intuitions: not suspecting any cheating, Adam will reason that Eve could have known she was muddy only if she was in fact the *only* muddy one. Moreover, Adam will never be able to retract his wrong answer: Rightful Logic simply cannot allow him to change his mind. Poor naughty Adam: the dirty boy is condemned to be forever wrong; but this surely serves him right!?

Sadly, the Rightful Logic predicts an even more unfortunate ending to our story: after hearing Eve's answer, innocent Mary will simply go mad! She will simultaneously believe that she's dirty and that she's clean, so her second answer will only be an inconsistent mumble. Indeed, according to the Rightful semantics of private and public announcements, the set of "possible worlds" that she considers as possible (after Eve's answer) is empty. Moreover, Mary is condemned to perpetual madness: no future communication can heal her inconsistencies.

This is in total contrast to our intuitions: a wise Mary should just conclude that Eve has somehow cheated, obtaining the desired information by some other process than pure reasoning (e.g. by looking in a mirror or by some other equivalent secret action). Mary should thus answer "I don't know" to fathers' second repetition of the question, but then in the third round of questioning (after hearing Adam's wrong answer), she should finally say "Now I know, dear father: I'm clean". Correct answer, instead of inconsistent mumble: what a happy ending for the immaculate Mary!

The purpose of this paper is twofold: *first*, to develop a *Kripke-model based, qualitative, multi-agent version of the classical Belief Revision theory*, which we call *the logic of conditional*[4] *beliefs*; *second*, we use this to propose a modified semantics for private and public announcements, and to axiomatize the corresponding dynamic doxastic logic, which one may call "the Merciful Logic" (of public/private announcements). By incorporating the main ideas of classical Belief Revision theory into our basic semantic structures, the Merciful Logic will save Mary from madness, will lead her to Truth, and could even give another chance to Adam to redeem himself, if the father asked the Question once again.

The first goal is met by replacing the usual doxastic/epistemic Kripke models with semantic structures called "conditional doxastic models" (CDM's). It is important to note that our approach differs from the recent semantical literature on the topic of (dynamic or static) belief revision (e.g. [3,10,16,22,23]) in the following sense. Most Kripke-style models proposed for multi-agent belief revision are based on *specific mechanisms that rely on quantitative notions*, such as "degrees of belief", plausibility functions, graded models or probabilistic measures of belief.[5] However, classical (AGM) belief revision theory is a qualitative theory, based on

---

[4] or "hypothetical"

[5] One could argue that the degrees of belief can be given by a plausibility order relation, so by a qualitative,

simple postulates concerning a basic operation (revision), of great generality and simplicity. Our approach retains this qualitative flavor of the classical AGM theory.

It is true that we also give a Representation Theorem, showing that any CDM can be represented as arising from a *(multi-agent) epistemic plausibility model* (based on a family of "well-preorderings"). [6] Such models are closer to the ones encountered in the standard literature on belief revision, being a simple variation on a theme pursued first by Gardenfors (total preorders as plausibility relations) and later by Spohn [24] (ordinal-valued plausibility functions). However, *the correspondence between CDM's and plausibility models is not one-to-one*: the same CDM corresponds to many different plausibility models. This means that, if we take the conditional doxastic structure as *fundamental*, we can easily see that all the other above-mentioned descriptions are *somewhat redundant by comparison*: they include irrelevant features, such as specific ordinal assignments, or plausibility comparisons between states that are epistemically distinguishable. For this and other reasons [7], we strongly prefer the qualitative description in terms of conditional doxastic maps, which can be seen as a natural extension of the standard definition of doxastic Kripke models, and which gives rise in a natural way to *conditional belief operators*, and thus to a *conditional doxastic logic CDL*. Indeed, the semantic structure of our CDM's matches perfectly the structure of our logic *CDL*, so that a complete axiomatization can be easily obtained by a simple modal translation of our semantic clauses.

In this sense, our approach is close to the one in Johan van Benthem's recent (unpublished) paper [27], of which we became aware only at a late stage of writing this paper. Though based on ("quantitative") models involving degrees of plausibility, the approach in [27] abstracts away from the details of modeling when considering the associated modal *logic*, which (is *not* based on any "graded belief" operator, as in e.g. [3,10], but) is a simple language of *conditional beliefs and update modalities*, virtually identical to ours (for public announcements). As a result, the main "reduction axiom" in [27], which computes (in the style of the Action-Knowledge Axiom in [6,5]) the conditional beliefs after a public announcement in terms of initial beliefs, is identical to our corresponding axiom. In a sense, our approach here is simply to go one step further, and abstract away (from the specific details of a particular quantitative implementation of belief revision operators) *on the semantic side as well*. This leads to a perfect match between the syntax (based on conditional beliefs) and the semantics (in terms of conditional doxastic models), giving our logic a broader, more general scope of application and a greater transparency. In its turn, this greatly facilitates the move to more general contexts: one can easily produce in this way appropriate analogues of the reduction axioms for *private announcements*, and in fact (in unpublished work [8]) we obtain natural generalizations to the case of *arbitrary epistemic/doxastic actions*.

Our concepts of conditional belief and of CDM can also be seen in the context of the wide logical-philosophical literature on notions of *conditional*, see e.g. [1,25,19,18,9]. One can of course look at our conditional belief operators as non-classical (and non-monotonic!) implications. Indeed, there have been various attempts and discussions concerning using conditionals to deal with belief revision (see e.g. [11,15,20]). We will show that our operators avoid the

---

[6] This result can be seen as an analogue in our semantic context of Gardenfors' representation theorem in [12], representing the AGM revision operator in terms of the minimal valuations for some total preorder on valuations.

order-theoretic notion, but in fact the way belief revision or update are defined makes an essential use of the "arithmetic" of these (finite or transfinite) degrees, e.g. in [24] and [3]; hence, the quantitative flavor.

[7] The notion of *equivalence* between models is sensitive to the choice of definition. We think that the "right" such notion for our logic is the natural concept of *bisimilarity* between CDM's.

known paradoxes arising from such mixtures of conditional and belief revision, by *failing to satisfy the so-called Ramsey test* (except in absolute, unconditional contexts). Indeed, as argued in [27], the usual statement of the Ramsey test is based on a confusion between *knowledge of a conditional* with premise ϕ (or rather, between the "static" belief revision with ϕ, as captured by our *hypothetical beliefs*) and the *knowledge/belief held after learning ϕ* (i.e. the "dynamic" belief revision). The approach in [21] seems also to be closely related to ours: the "models" considered there for belief revision and belief update are *of the same type* (except for being single-agent) as our CDM's. They consider some natural semantic conditions, in correspondence with modal axioms, but they do not focus on the same set of postulates as us. [8]

The plan of this paper is the following. In the next section we briefly review some basic notions about knowledge-belief (KB) models and doxastic-epistemic logic. In section **3**, we "revise" the standard (syntactic) AGM revision theory, to make it applicable to a (multi-agent) epistemic/doxastic language, by considering revision of beliefs *against a knowledge base*; this imposes a weakening of the standard "Success" postulate. Then we convert the (revised) belief revision postulates into *semantic* clauses on KB models, obtaining a *semantic counterpart of the (revised) AGM theory.* In section **4**, we define our central semantic notion, *conditional doxastic models* (CDM's), and we prove this setting to be actually equivalent (modulo the usual KB conditions) with the above-mentioned "semantic AGM" postulates. We also show this to be equivalent to a definition in terms of "well-preordered" plausibility relations. In section **5**, we move to *belief updates*, by changing the usual semantics of *public announcements* to make them act on CDM's in the natural way, thus allowing beliefs to be "dynamically revised" when learning new information. In section **6**, we extend this setting to *private announcements to subgroups*, we give a complete axiomatization (using "reduction axioms" in the style of [6,5,26,27]), and then we apply this logic to the task of "saving Mary" from the cheaters, in the above Muddy-Children-type scenario.

2 Preliminaries: KB-Models and Belief-Knowledge Logic

A *knowledge-belief frame* (KB-frame for short, see e.g. [17], pg. 89) is a Kripke frame of the form $(S, \rightarrow_a, \sim_a)_{a \in \mathcal{A}}$, with a given set of states S and two binary relations for each agent; the first relation $\sim_a$ is meant to capture the *knowledge* of agent a, while the second $\rightarrow_a$ captures his *beliefs*. A KB frame is required to satisfy the following natural conditions: (1) each $\sim_a$ is reflexive: $s \sim_a s$; (2) if $s \sim_a t$ then we have: $s \rightarrow_a w$ iff $t \rightarrow_a w$, and also $s \sim_a w$ iff $t \sim_a w$; (3) if $s \rightarrow_a t$ then $s \sim_a t$; (4) for every $s \in S$ there exists some $t \in S$ such that $s \rightarrow_a t$.

The first clause expresses the *truthfulness* of knowledge, the second expresses *full introspection* (an agent knows what he believes/knows and what not), the third says that *agents believe everything they know*, and the last (seriality) says that *beliefs are consistent.* A *knowledge-belief model* (KB-model) is a Kripke model having an underlying KB-frame.

By replacing the accessibility relations with their image-maps [9], we obtain *an equivalent definition* of a more "coalgebraic" flavor: a KB-frame is a structure $(S, \bullet_a, \bullet(a))_{a \in \mathcal{A}}$, where S is a set of states and $\bullet_a, \bullet(a) : S \rightarrow \mathcal{P}(S)$ are maps satisfying the following conditions:

(1) $s \in s(a)$;   (3) $s_a \subseteq s(a)$

(2) if $t \in s(a)$, then $s_a = t_a, s(a) = t(a)$;   (4) $s_a \neq \emptyset$.

---

[8] The notion of update considered in [21] is also completely different from our corresponding notion.

[9] The image-map of a relation $R \subseteq S \times S$ is the map $\widehat{R} : S \rightarrow \mathcal{P}(S)$, $\widehat{R}(s) := \{t \in S : sRt\}$.

The maps $\bullet_a$ and $\bullet(a)$ are called *appearance maps*: s_a is the *doxastic appearance* of s to a (or the *theory of a about s*), and $s(a)$ is the *epistemic appearance* of s to a (or the *knowledge of a about s*). The equivalence between the two definitions of knowledge-belief models is easily verified [10].

Given a knowledge-belief model **S**, an *S-proposition* (or *S-theory*) is simply any set $P \subseteq S$ of states in S. This is of course a purely extensional and semantical notion of proposition/theory, to be distinguished from the syntactical and intensional notions of "sentence" and "theory". For any *S*-proposition P and agent $a \in \mathcal{A}$, we can define as usually the *S*-propositions $B_a P$ ("agent a believes P") and $K_a P$ ("agent a knows P") by the *standard Kripke definitions of modalities* (for the accessibility relations $\rightarrow_a$ and $\sim_a$). In terms of appearance maps, these definitions can be given in the form of *Galois dualities* (between appearance and knowledge/belief):

$$s \in B_a P \text{ iff } s_a \subseteq P \qquad\qquad s \in K_a P \text{ iff } s(a) \subseteq P$$

We can define operations on *S*-propositions: *negation* $\neg P := S \setminus P$, *conjunction* $P \wedge Q := P \cap Q$, *general true belief* $EbP := \bigcap_{a \in \mathcal{A}} B_a P$ ("everybody believes P") and *general knowledge* $EkP := \bigcap_{a \in \mathcal{A}} K_a P$. Finally, we define *common true belief* $CbP := \bigcap_{n \geq 0} (Eb)^n P = P \cap EbP \cap Eb(EbP) \cap \cdots$ and *common knowledge* $CkP := \bigcap_{n \geq 0} (Ek)^n P = P \cap EkP \cap Ek(EkP) \cap \cdots$.

The *Belief-Knowledge Logic* (BKL) is a logic whose syntax is given by:
$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_a\varphi \mid K_a\varphi \mid Cb\varphi \mid Ck\varphi$$
The semantics is given by the obvious compositional clauses: p is given by the valuation, $\|\neg\varphi\|_S := \neg\|\varphi\|_S$ etc. As standard, we also use the notation $s \models_S \varphi$ for $s \in \|\varphi\|_S$. Observe that general belief and general knowledge are *definable* in BKL, by putting: $Eb\varphi := \bigwedge_{a \in \mathcal{A}} B_a\varphi$, $Ek\varphi := \bigwedge_{a \in \mathcal{A}} K_a\varphi$. Under various names, BKL is a well-known logic and its *complete proof system*, which we will also denote by BKL, is given by familiar axioms and rules, see e.g. [17] (pg. 94, where this proof system is called KL).

3   A semantic, multi-agent, epistemic AGM theory

Classical AGM theory. Classical belief revision takes a *syntactic* view of theories: we are given a family $\mathcal{T}$ of all "theories", whose members are assumed to be *deductively closed sets of sentences* (over some given language). Let $\bot$ be the inconsistent theory (containing all sentences). The *expansion* $T + \varphi$ of a theory $T \in \mathcal{T}$ with a sentence φ is defined as $T + \varphi := \{\psi : T \cup \{\varphi\} \vdash \psi\}$. Now the belief revision operator $*$ can be introduced by means of the standard AGM postulates:

(*1) $T * \varphi$ is a theory;   (*5) $T * \varphi = \bot$ iff $\vdash \neg\varphi$;

(*2) $\varphi \in T * \varphi$;   (*6)   if $\vdash \varphi \leftrightarrow \psi$ then $T * \varphi = T * \psi$;

(*3-4)   if $\vdash \varphi$ then $T * \varphi = T$;   (*7-8)   if $\neg\psi \notin T * \varphi$ then $T * (\varphi \wedge \psi) = (T * \varphi) + \psi$

Revising the Revision Theory: epistemic AGM. In order to apply belief revision to theories in a *doxastic-epistemic language*, we need to revise the "Success" postulate (*5) in an obvious way, since *agents' beliefs about their own beliefs or knowledge are certain*, and thus they *should not be revised*. More generally, if something is "known", than it should not be subject to revision; or, in other words, *any attempt to "revise" with a sentence whose negation is "known" should lead to a contradiction*. This leads us to a "revision of this belief revision postulate", by replacing (*5) with its "epistemic version":

(*5e) $T * \varphi = \bot$ iff $T \vdash K\neg\varphi$ (i.e. iff $(K\neg\varphi) \in T$).

This revised system, composed of postulates (*1), (*2), (*3-4), (*5e), (*6), (*7-8), is called *epistemic AGM*. If, as usually, we assume that knowledge satisfies the *Necessitation rule* (from $\vdash \varphi$ infer $\vdash K\varphi$), then from this and (*5e) we obtain as a consequence the desirable half of (*5): if $\vdash \neg\varphi$ then $T * \varphi = \bot$.

---

[10] One way by putting $\bullet_a = \widehat{\rightarrow_a}$ and $\bullet(a) = \widehat{\sim_a}$ (where $\widehat{R}$ is the image-map of R), and the opposite way by putting: $s \rightarrow_a t$ iff $t \in s_a$, and $s \sim_a t$ iff $t \in s(a)$.

Multi-agent AGM. To apply the postulates to theories written in the logic BKL, we need a multi-agent version of Epistemic AGM. So we need to *restate postulate (*5e) using the labelled operator K_a, for all agents a*. But in addition, observe that *the notion of "theory" and the "revision" operation become relative to agents*: a set of sentences might well be a possible theory for agent a, but not for an agent b. "Theories" in AGM are supposed to be complete descriptions of (the agent's) beliefs about the world. So, for example, a theory that leaves open the question whether $K_b p$ holds or not (for some given fact p) *cannot ever be the (complete) theory describing agent b's beliefs* (though it can perfectly well describe completely agent a's beliefs): due to introspection, b cannot be uncertain about his own knowledge.

So we need to assume as given, for each agent a, a family $\mathcal{T}_a$ of "a-theories". We assume these to be deductively closed sets of sentences in the logic BKL; as pointed above, we also need to require a minimal notion of *introspectiveness*: an a-theory should settle all the questions concerning a's beliefs and knowledge. In addition, we want each revision operator $*_a$ to act on a-theories, and so to state the postulate (*5e), we need to require the inconsistent theory to be an a-theory.

So we formulate our revised *multi-agent (epistemic) AGM* postulates, by giving, for every agent $a \in \mathcal{A}$: a family $\mathcal{T}_a \subseteq \mathcal{P}(BKL)$ of sets of sentences in the language BKL, called *a-theories*, and a belief revision operator $*_a : \mathcal{T}_a \times BKL \to \mathcal{T}_a$, taking pairs of a-theories and BKL-sentences into new a-theories; and requiring them to satisfy the following conditions: **(T1)** $\perp \in \mathcal{T}_a$ (where $\perp := BKL$ is the inconsistent theory, containing all the sentences in BKL); **(T2)** every $T \in \mathcal{T}_a$ is deductively closed, w.r.t. the complete proof system of BKL; **(T3)** for every $\varphi \in BKL$ and every $T \in \mathcal{T}_a$, we have either $K_a \varphi \in T$ or $(\neg K_a \varphi) \in T$; **(T4)** all the above postulates of epistemic AGM, in which we label with agent names both the knowledge K_a and the revision $*_a$ operators. Observe that it is not necessary to require an introspective condition corresponding to (T3) for *belief*, since this follows from the above conditions, given the axioms of BKL. Indeed, one can easily prove that for every $\varphi \in BKL$ and every $T \in \mathcal{T}_a$, we have either $B_a \varphi \in T$ or $(\neg B_a \varphi) \in T$.

Semantic Belief Revision. To develop a *semantical counterpart of multi-agent (epistemic) AGM*, we assume as given a KB-model $\mathbf{S}$. We need to replace in the above postulates the *syntactic* notion of a "theory" as set of sentences with the semantic notion of $\mathbf{S}$-theory (i.e. set of states in S); similarly, we replace sentences by S-propositions (also set of states). Observe that each $\mathbf{S}$-theory $T \subseteq S$ gives rise to a syntactic theory $th(T) = \{\phi \in BKL : t \models_{\mathbf{S}} \phi \text{ for all } t \in T\}$. In addition to the above postulates, we have to make our *belief revision theory consistent with our theory of beliefs (given by the model $\mathbf{S}$)*: namely, we have to add a postulate **(T0)** requiring that, for each agent a, *the agent's current beliefs form an a-theory*. Finally, we need to replace the operation $T + \phi$ and the "deductive closure" of a theory with their semantic counterparts. To do this, observe first that the partial order on theories is inverted for semantic theories: for $\mathbf{S}$-theories $T, T' \subseteq S$ we have $T \subseteq T'$ iff $th(T') \subseteq th(T)$. The inconsistent theory $\perp$ is now represented by the empty set of states $\emptyset \subseteq S$. The deductive closure of the union of two syntactic theories corresponds to the *intersection* of the corresponding semantic theories (sets of states). Hence, *expansion $T + P$ of a semantic theory $T \subseteq S$ with a semantic proposition $P \subseteq S$ is simply given by the *intersection* $T \cap P$. As a result, we obtain the following definition:

Semantic Version of Epistemic AGM Postulates. Given a KB-model $\mathbf{S}$, an *AGM belief revision theory for $\mathbf{S}$* is defined by giving, for each agent a, a family of S-theories $\mathcal{T}_a \subseteq \mathcal{P}(S)$, called *$a$-theories over $\mathbf{S}$*, and an operation $*_a : \mathcal{T}_a \times \mathcal{P}(S) \to \mathcal{T}_a$, such that for all $T \in \mathcal{T}_a$, $P \subseteq S$, we have:

(T0) $s_a \in \mathcal{T}_a$, for all $s \in S$;

(T1) $\emptyset \in \mathcal{T}_a$;

(T2) if $T \in \mathcal{T}_a$, then for all $s, t \in T$ we have $s_a = t_a$ and $s(a) = t(a)$.

(*1) $T *_a P \in \mathcal{T}_a$;

(*2) $T *_a P \subseteq P$;

(*3-4) $T *_a S = T$;

(*5e) $T *_a P = \emptyset$ iff $T \subseteq K_a \neg P$ (iff $T(a) \cap P = \emptyset$) ;

(*6) if $P = Q$ then $T *_a P = T *_a Q$;

(*7-8) if $T *_a P \cap Q \neq \emptyset$ then $T *_a (P \cap Q) = T *_a P \cap Q$,

where we used the notation $T(a) := \{t(a) : t \in T\}$, to indicate the "knowledge of a in T". Observe that, in fact, the above semantic version of the AGM postulate (*6) is *superfluous*: it is always trivially satisfied, due to the extensionality of S-theories.

4   Conditional Doxastic Models

We give now a setting that is *equivalent to semantic (multi-agent epistemic) AGM*, though it is much simpler in formulation. Namely, we enrich our knowledge-belief models to capture a notion of *conditional belief*. A *conditional doxastic frame (CD-frame, for short)* $(S, \{\bullet_a^P\}_{a \in \mathcal{A}, P \subseteq S})$ consists of a set of states S, together with a family of *conditional (doxastic) appearance* maps, one for each agent a and each possible condition $P \subseteq S$. These are required to satisfy the following conditions:

(i)   if $s \in P$ then $s_a^P \neq \emptyset$;   (ii)   if $P \cap s_a^Q \neq \emptyset$ then $s_a^P \neq \emptyset$;

(iii)   if $t \in s_a^P$ then $s_a^Q = t_a^Q$;   (iv) $s_a^P \subseteq P$;

(v) $s_a^{P \cap Q} = s_a^P \cap Q$,   if $s_a^P \cap Q \neq \emptyset$.

A *conditional doxastic model* (CDM, for short) is a Kripke model whose underlying frame is a CD-frame. The conditional appearance s_a^P captures *the way a state s appears to an agent a, given some additional (plausible, but not neccesarily truthful) information P*. More precisely: whenever s is the current state of the world, then after receiving new information P, agent a will come to believe that any of the states $s' \in s_a^P$ might have been the current state of the world (as it was before receiving information P).

Using conditional doxastic appearance, the *knowledge $s(a)$ possessed by agent a about state s* (i.e. the *epistemic appearance* of s) can be defined as the *union of all conditional doxastic appearances*. In other words, *something is known iff it is believed in any conditions*: $s(a) := \bigcup_{Q \subseteq S} s_a^Q$. Using this, we can see that the first condition above in the definition of conditional doxastic frames captures the *truthfulness of knowledge*. The second condition states the *success of belief revision, when consistent with knowledge*: *if something is not known to be false, then it can be consistently entertained as a hypothesis*. The third condition expresses *full introspection of (conditional) beliefs*: agents *know* their own conditional beliefs, so they cannot revise their beliefs about them. The fourth condition says *hypotheses are hypothetically believed*: when making a hypothesis, that hypothesis is taken to be true. The last condition describes *minimality of revision*: when faced with new information Q, agents keep as much as possible of their previous (conditional) beliefs s_a^P.

These requirements can be seen as strengthenings of the clauses defining a KB-frame: indeed, *every CD-frame is a KB-frame*. To see this, it is enough to define $s_a := s_a^S$, and check this satisfies all the KB assumptions. In other words: *we can recover the unconditional ("default") beliefs as conditional beliefs with respect to some trivially true condition*.

Alternatively, we can define a conditional doxastic frame *relationally* as a tuple $(S, \{\xrightarrow{P}_a\}_{a \in \mathcal{A}, P \subseteq S})$, where $\xrightarrow{P}_a$ are binary relations, satisfying the clauses: (1.) if $s \in P$ then there exists some state t such that $s \xrightarrow{P}_a t$; (2.) if $s \xrightarrow{Q}_a t \in P$, then there exists a state $w \in S$ such that $s \xrightarrow{P}_a w$; (3.) if $s \xrightarrow{P}_a t$ then for every state $w \in S$ we have: $s \xrightarrow{Q}_a w$ iff $t \xrightarrow{Q}_a w$; (4.) if $s \xrightarrow{P}_a t$ then $t \in P$; (5.) if there exists $s \xrightarrow{P}_a t \in Q$ then, for all every $w \in S$, we have: $s \xrightarrow{P \cap Q}_a w$ iff $s \xrightarrow{P}_a w \in Q$. It is easy to see that the this

definition of conditional doxastic frames is equivalent to the above one [11] .

Applying the standard Kripke relational definition of modalities to the conditional doxastic relations $s \xrightarrow{P}_a t$, we obtain a new operator B_a^P on S-propositions, expressing *conditional beliefs*; in terms of appearance maps, the definition says that *conditional belief is the Galois dual of conditional appearance*:

$$B_a^P Q := \{s \in S : s_a^P \subseteq Q\}$$

We read this as saying that *agent a believes Q conditional of P*. More precisely, this says that: *if the agent would learn P, then (after learning) he would come to believe that Q was the case in the current state (before the learning)*. Notice that beliefs conditional to the trivially true proposition S coincide with the usual, unconditional beliefs: $B_a^S Q = B_a Q$.

As a consequence of the above postulates, the *knowledge operator*, defined (as in the previous section) as the Galois dual of epistemic appearance $K_a P := \{s \in S : s(a) \subseteq P\}$, has the following property:

$$K_a P = \bigcap_{Q \subseteq S} B_a^Q P = B_a^{\neg P} \emptyset = B_a^{\neg P} P$$

We can also define *conditional versions of general belief, common true belief, knowledge, general knowledge, common knowledge*, by putting: $Eb^P Q := \bigcap_{a \in \mathcal{A}} B_a^P Q$, $Cb^P Q := \bigcap_{n \geq 0} (Eb^P)^n Q = Q \cap Eb^P Q \cap Eb^P(Eb^P)Q \cap ...$, $K_a^P Q := K_a(P \to Q)$, $Ek^Q := \bigcap_{a \in \mathcal{A}} K_a^P Q$, $Ck^P Q := \bigcap_{n \geq 0} (Ek^P)^n Q$

Theorem. A CDM is equivalent to a semantic AGM theory over a KB-model.

Proof. Given an AGM theory over a KB-model, we define $s_a^P := s_a *_a P$, and check this satisfies the clauses of a CDM. For the converse, start with a CDM, and put $\mathcal{T}_a := \{s_a^P : s \in S, P \subseteq S\}$. Define a revision operator $T *_a Q$ for our theories $T = s_a^P \in \mathcal{T}_a$, by cases: if we have $P \cap s(a) = \emptyset$ (i.e. if $s_a^P = \emptyset$), then we put $s_a^P *_a Q := \emptyset$; if we have $P \cap s(a) \neq \emptyset$, but $P \cap Q \cap s(a) = \emptyset$, then put $s_a^P *_a Q := s_a^Q$; else, put $s_a^P *_a Q := s_a^{P \cap Q} = s_a^P \cap Q = s_a^Q \cap P$. It is easy to check the KB conditions. ⊣

Examples of CDM's: *Any KB-model is a CDM*; indeed, we can trivially convert a KB-model into a CDM, by putting: $s_a^P = s_a \cap P$, whenever $s_a \cap P \neq \emptyset$, and $s_a^P = s(a) \cap P$ otherwise. Of course, this is *only one* way to organize a KB-model as a CDM, a very special case corresponding to *the most trivial belief revision policy*, encoded in the principle: "when your *beliefs* are contradicted by new facts, *give them all up* and stick with what you *know*". A *more general example* is given by:

Plausibility Models: An *epistemic plausibility frame* is a structure $(S, \sim_a, \leq_a)_{a \in \mathcal{A}}$, consisting of a set S endowed with a family of equivalence relations $\sim_a$ and a family of "well-preorders" $\leq_a$, one for each agent a. Here, a "well-preorder" is just a preorder $\leq$ on S such that every subset has minimal elements; i.e. for every set $T \subseteq S$ there exists $t \in T$ such that $t \leq t'$ for all $t' \in T$. An epistemic plausibility frame together with a valuation gives an epistemic plausibility model. Plausibility frames for only one agent and without the epistemic relations have been used as models for AGM belief revision in [12,22] etc. A more *concrete example* of plausibility frames was given by W. Spohn in [24], in terms of ordinal preference maps assigning ordinals $d(s)$ ("the degree of plausibility" of s) to each state $s \in S$. In our epistemic multi-agent context, this would give us structures consisting of a multi-agent knowledge frame $(S, \sim_a)_{a \in \mathcal{A}}$, together with an ordinal plausibility map $d_a : S \to Ord$ (where Ord is the family of all ordinals).

Any epistemic plausibility model gives rise to a CDM, in a canonical way, by putting

$$s_a^P := Min_{\leq_a} \{t \in P : t \sim_a s\}$$

where $Min_{\leq_a} T = \{t \in T : t \leq_a t' \text{ for all } t' \in T\}$ is the set of minimal elements in T. We call this *the canonical CDM associated to the given plausibility model*. The converse is given by the following:

---

[11] In one way they are equivalent by putting $\bullet_a^P = \widehat{\xrightarrow{P}_a}$ (where $\widehat{R}$ is the image-map of R), and in the opposite way by putting: $s \xrightarrow{P}_a t$ iff $t \in s_a^P$.

Representation Theorem. Every CDM is the canonical CDM of some epistemic plausibility model.

Proof. Given a CDM $\mathbf{S} = (S, \{\bullet_a^P\}_{a \in A, P \subseteq S})$, take for each a some arbitrary well-ordering $\leq^a$ of the family $\{s(a) : s \in S\}$ of all epistemic appearances. Define $s \sim_a t$ iff $s(a) = t(a)$. Define $s \leq_a t$ by: either $s(a) \leq^a t(a)$, or $s(a) = t(a)$, $s \in t_a^{\{s,t\}}$. It is easy to check that this is an epistemic plausibility model, whose canonical CDM is $\mathbf{S}$ itself. $\dashv$

So our setting in terms of CDM's is equivalent to a more standard one in terms of plausibility models. Nevertheless, the proof of the above theorem shows the correspondence is not one-to-one[12]: the same CDM corresponds canonically to many plausibility models. In its turn, the same plausibility model corresponds to many Spohn-type models (in terms of plausibility degrees). In this paper, *we take the conditional doxastic structure as fundamental*, since we are interested in a logic of conditional beliefs. This means that, for our purposes, not only the actual assignment d_a of ordinal degrees of plausibility, but even much of the induced structure of the plausibility relations $\leq_a$, is *irrelevant: they contain superfluous features*. The important thing are the corresponding conditional doxastic maps.

Conditional Doxastic Logic (CDL). We now change BKL to a version in which belief operators are conditionalized. The syntax of CDL is given by:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_a^\varphi \varphi \mid Cb^\varphi \varphi \mid Ck^\varphi \varphi \ ,$$

while the semantics is given by the obvious compositional clauses for the interpretation map $\| \bullet \|_\mathbf{S} : CDL \to \mathcal{P}(S)$ in a CDM $\mathbf{S}$. In this logic, the *knowledge modality* can be defined as an abbreviation, putting $K_a\phi := B_a^{\neg\phi}\bot$ (where $\bot = p \wedge \neg p$ is an inconsistent sentence), or equivalently $K_a\phi := B_a^{\neg\phi}\phi$.[13] It is easy to see that this agrees semantically with the previous definition of the semantic knowledge operator (as the Galois dual of epistemic appearance): $\|K_a\phi\|_\mathbf{S} = K_a\|\phi\|_\mathbf{S}$. We also define $K_a^\theta\varphi := K_a(\theta \to \varphi)$, $Eb^\theta\varphi := \bigwedge_{a \in \mathcal{A}} B_a^\theta\varphi$, $Ek^\theta\varphi := \bigwedge_{a \in \mathcal{A}} K_a^\theta\varphi$.

Theorem. A sound and complete proof system for CDL is obtained as follows: first, include *all the axioms and rules of classical propositional logic*; second, include *Necessitation Rules* for all modalities: *from $\vdash \varphi$ infer $\vdash B_a^\psi\varphi$, $\vdash Ck\varphi$ and $\vdash Cb^\psi\varphi$* ; third, include the following axioms:

Normality:
$$\vdash B_a^\theta(\varphi \to \psi) \to (B_a^\theta\varphi \to B_a^\theta\psi)$$
$$\vdash Cb^\theta(\varphi \to \psi) \to (Cb^\theta\varphi \to Cb^\theta\psi)$$
$$\vdash Ck^\theta(\varphi \to \psi) \to (Ck^\theta\varphi \to Ck^\theta\psi)$$

Truthfulness of Knowledge:
$$\vdash K_a\varphi \to \varphi$$

Persistence of Knowledge:
$$\vdash K_a\varphi \to B_a^\psi\varphi$$

Full Introspection:
$$\vdash B_a^\psi\varphi \to K_a B_a^\psi\varphi$$
$$\vdash \neg B_a^\psi\varphi \to K_a\neg B_a^\psi\varphi$$

Hypotheses are (hypothetically) accepted:
$$\vdash B_a^\varphi\varphi$$

Minimality of revision:
$$\vdash \neg B_a^\varphi\neg\psi \to (B_a^{\varphi \wedge \psi}\theta \leftrightarrow B_a^\varphi(\psi \to \theta))$$

Fixed-Point Axioms:
$$\vdash Cb^\theta\varphi \to \varphi \wedge Eb^\theta Cb^\theta\varphi$$
$$\vdash Ck^\theta\varphi \to \varphi \wedge Ek^\theta Ck^\theta\varphi$$

Induction Axioms:
$$\vdash Cb^\theta(\varphi \to Eb^\theta\varphi) \to (\varphi \to Cb^\theta\varphi)$$
$$\vdash Ck^\theta(\varphi \to Ek^\theta\varphi) \to (\varphi \to Ck^\theta\varphi)$$

Related to our topic are the standard philosophical problems of using conditionals in belief revision

---

[12] Indeed, this is shown by the arbitrary choice of the well-orders $\leq^a$.

[13] This way of defining knowledge in terms of doxastic conditionals can be traced back to [25].

models (see for instance [11,15,20]). In this context, it is interesting to see how our conditional belief operators, understood as conditionals, can avoid *Gardenfors' triviality result* [11], which has been used to argue that standard AGM theory is incompatible with a conditional-based view of belief revision. This result was based on the assumption that any such conditional should satisfy the so-called *Ramsey test* [19]. Following [20], the Ramsey test can be stated in syntactic terms as saying that

$$(R) \quad \text{"if } P \text{ then } Q\text{"} \in T \text{ iff } Q \in T * P$$

If we interpret the conditional "if P then Q" as our conditional belief $B_a^P Q$, interpret the revision operator as the operator $*_a$ defined above (in the proof of the theorem on the equivalence between CDM's and AGM theories over KB models), and interpret "theories" T to mean elements of $\mathcal{T}_a$ in a CDM (as defined in the above-mentioned proof, i.e. theories of the form $T = s_a^R$, for some proposition R), then we obtain the following semantic version of the Ramsey test:

$$(R*?) \quad \text{for every } R \subseteq S : \ s_a^R \subseteq B_a^P Q \text{ iff } s_a^R *_a P \subseteq Q.$$

It is easy to check this is *false*: given our CDM postulates, the $(R*)$-test fails under this interpretation. To see this, observe that the left-hand side of $(R*?)$ is equivalent to $\forall t \in s_a^R : t_a^P \subseteq Q$. But, by our postulates, $t \in s_a^R$ implies $t_a^P = s_a^P = s_a *_a P$. So, whenever $s_a^R \neq \emptyset$, the left-hand side of $(R*?)$ is simply equivalent to $s_a *_a P \subseteq Q$, which is in general *not* equivalent to the right-hand side $s_a^R *_a P \subseteq Q$. So we see that the Ramsey test could only succeed if we had $s_a = s_a^R$ in general, i.e. *if conditional beliefs would collapse to unconditional ones*: this is in a way our own *semantic version of Gardenfors' triviality result*. On the other hand, observe that $(R*?)$ *does hold in unconditional contexts*, that is for $R = S$ (i.e. for theories of the form $T = s_a$): $s_a \subseteq B_a^P Q$ iff $s_a *_a P = s_a^P \subseteq Q$.

The deep flaw underlying the Ramsey test is that it treats *hypothetical beliefs about beliefs* in the same way as *hypothetical beliefs about facts*; the test would succeed only if, when making a hypothesis, agents would revise their beliefs about their own beliefs in the same way they revise their factual beliefs. But this is inconsistent with the *restrictions posed by introspective knowledge to belief revision*: introspective agents *know* their own beliefs, and so cannot accept hypotheses that go against this knowledge. A hypothetical belief system (e.g. the theory s_a^R in the above counterexample) may include different ontic statements than the unconditional belief system (s_a); but it includes *precisely the same doxastic/epistemic statements $B_a^P Q$* as this unconditional belief system. Due to introspection, beliefs about beliefs *cannot be revised*, not in the sense of ("static") belief revision that we have here. [14] Only a *"dynamic" kind of belief revision* (that aims to represent the revised beliefs of the agent *about the situation after the revision*) would satisfy some (suitably modified) Ramsey test.

5 Dynamic Belief Revision: Public Announcements

The belief revision encoded in the conditional doxastic models above is of a *static*, purely *hypothetical*, nature. Indeed, the revision operators cannot alter models in any way: all the possibilities are already there, so both the unconditional and the revised, conditional beliefs *refer to the same world and the same moment in time*. [15] In contrast, a *belief update* is a dynamic form of belief revision, meant to capture the actual change of beliefs induced by learning (or by other forms of epistemic/doxastic

---

[14] Remember that $B_a^P Q$ means *"if a would learn P, then he would come to believe that Q had been the case (before the learning)"*. Suppose you happen to believe $\neg P$, and somebody asks you: "If I was to tell you that P was the case, would that change your mind about the fact that you currently believe $\neg P$?". Clearly, the correct answer is: "No, it wouldn't. It would indeed change my belief about P, but not my belief about the fact that now I believe $\neg P$".

[15] Indeed, the postulate (*2) (and the corresponding clause (4) in the definition of CD-frames) can only hold if a revision with so-called Moore sentences (e.g. $\varphi \wedge \neg K_a \varphi$ is understood to be *only hypothetically possible*. The agent a's actual beliefs after learning such a sentence *cannot* possibly include the sentence itself.

actions). As already noticed before [13,6,5], the original model does not usually include enough states to capture all the epistemic possibilities that arise in this way. So, contrary to the previous section, we now allow for belief revisions that change the original CDM. In this section we focus on *public announcements*, which change epistemic (and conditional doxastic) models in a *minimal* way: they can only *shrink* the model by "relativization" to a given sentence.

Given a model (CDM) **S**, denote by $s_a^Q, \|.\|$ the appearance maps and valuation **in S**. For any S-proposition $P \subseteq S$, we define the *relativized CDM* **P!(S)** by taking:

(i) the set of states of **P!(S)** is the set P,

(ii) $(s_a^Q)_{\mathbf{P!(S)}} := s_a^Q$, for every $s \in P$ and $Q \subseteq P$,

(iii) $\|p\|_{\mathbf{P!(S)}} := \|p\| \cap P$.

As an immediate consequence, we get that *unconditional beliefs after the update come from prior conditional beliefs*: $(s_a)_{\mathbf{P!(S)}} = (s_a^P)_{\mathbf{P!(S)}} = s_a^P$.

We interpret the action $P!$ as a transition relation from any current state $s \in \mathbf{S}$ satisfying P to the state $s \in \mathbf{P!(S)}$. The *syntax* of our public announcement logic is obtained by simply adding constructs involving dynamic modalities $< \varphi! > \varphi$ to the syntax of CDL. For the *semantics* we include the following extra clause: $\| < \varphi! > \psi \|_{\mathbf{S}} = \|\psi\|_{\|\varphi\|_{\mathbf{S}}!(S)}$.

To obtain a *sound and complete proof system*, we add reduction axioms for public announcements to the axioms of CDL:

$$< \varphi! > p \quad \leftrightarrow \quad \varphi \wedge p \qquad\qquad < \varphi! > \neg\psi \quad \leftrightarrow \quad \varphi \wedge \neg < \varphi! > \psi$$

$$< \varphi! > (\psi \wedge \theta) \quad \leftrightarrow \quad < \varphi! > \psi \vee < \varphi! > \theta \qquad < \varphi! > Ck_a^\theta \psi \quad \leftrightarrow \quad \varphi \wedge Ck_a^{<\varphi!>\theta} < \varphi! > \psi$$

$$< \varphi! > B_a^\theta \psi \quad \leftrightarrow \quad \varphi \wedge B_a^{<\varphi!>\theta} < \varphi! > \psi \qquad < \varphi! > Cb_a^\theta \psi \quad \leftrightarrow \quad \varphi \wedge Cb_a^{<\varphi!>\theta} < \varphi! > \psi$$

where p's denote atomic sentences.

6 Private Announcements to subgroups

In this section we deal with the specific action of "privately learning a fact" or, more generally, a "private announcement $P!_A$ to a subgroup of agents". The intuition is that the announcement is broadcasted to the agents of a group A, while the outsiders $B \not\in A$ do not suspect this is happening. For simplicity, we consider here the case in which *it is common knowledge that nothing else can happen: this particular announcement (of this particular sentence P to the group A) is the only message that may be broadcasted at this time*; the only alternative is *no message being sent*, i.e. the silent action $\tau_A P$ in which *"nothing happens"* (but in which the outsiders *don't know this*, so *they think it is possible* that the message P was in fact broadcasted to group A).

Given a CDM **S** and an S-proposition $P \subseteq S$, we define the *a new, updated CDM* under private announcements **P!$_A$(S)** as follows: for each of the old states $s \in S$ we take two *distinct new copies* $P!_A(s)$ (meant to denote the state after P was announced to the group A) and $\tau_A P(s)$ (meant to denote the corresponding state in which *nothing really happened*, but the outsiders $b \not\in A$ *consider possible* that P was announced to group A). Then the new model $\mathbf{S}'$ is obtained by putting:

(i) the new set of states of **P!$_A$(S)** is the set $S' = P!_A(P) \cup \tau_A P(S)$

(ii) for all $a \in A$: $P!_A(s)_a^Q := P!_A(s_a^{P!_A^{-1}(Q)})$ and $\tau_A P(s)_a^Q := \tau_A P(s_a^{\tau_A P^{-1}(Q)})$

(iii) for all $b \not\in A$: $\tau_A P(s)_b^Q = P!_A(s)_b^Q := \tau_A P(s_b^{\tau_A P^{-1}(Q)})$, if $s(b) \cap \tau_A P^{-1}(Q) \neq \emptyset$;
and $\tau_A P(s)_b^Q = P!_A(s)_b^Q := P!_A(s_b^{P!_A^{-1}(Q)})$, otherwise

(iv) $\|p\|_{\mathbf{S}'} := P!_A(\|p\|_{\mathbf{S}}) \cup \tau_A P(\|p\|_{\mathbf{S}})$,

where we used the notations $\sigma(Q) := \{\sigma(s) : s \in Q\}$ and $\sigma^{-1}(Q') := \{s \in S : \sigma(s) \in Q'\}$ for any of the two "actions" $\sigma \in \{P!_A, \tau_A P\}$, and for all sets $Q \subseteq S$, $Q' \subseteq S'$. We explain the clauses for conditional belief update. For clause 2: the insiders know in any case which action σ happened (be it $P!_A$ or $\tau_A P$), so if after that action they are given some new information Q they apply the following algorithm. They first reconsider their beliefs about the past states, in the view of the new information: they might have to revise these beliefs with the fact that, after this specific action happens, Q becomes true; so they revise their beliefs about the past state with $\sigma^{-1}(Q)$; then they run back to present, by applying action σ to the states allowed by these past beliefs: this gives their current belief about the state of the world after the action σ. In clause 3, the outsiders apply essentially the same algorithm, but (not knowing which action really happens) they keep the default belief that what they see, that is action $\tau_A P$ (i.e. "nothing"), is what is happening; so they apply the above algorithm only to action $\sigma = \tau_A P$; *unless* this is contradicted by the new information Q, i.e. unless it is already known beforehand that Q cannot become true after $\tau_A P$; in which case, they "revise their belief about the current action": they realize that $P!_A$ is happening, so they apply the above algorithm to $\sigma = P!_A$.

For our *syntax*, we replace the public announcement modalities above $<\varphi!>$ with dynamic modalities $<\varphi!_A>$ and $<\tau_A\varphi>$ corresponding to the two types of action above. [16] The *semantics* is given by the standard PDL clause: $\| <\sigma> \psi\|_{\mathbf{S}} := \{s \in S : \sigma(s) \text{ exists and } \sigma(s) \in \|\psi\|_{\mathbf{S}'}\}$. To get a *complete proof system*, we replace the reduction axiom for conditional beliefs with the axioms:

$$< \varphi!_A > B_a^\theta \psi \leftrightarrow \varphi \wedge B_a^{<\varphi!_A>\theta} < \varphi!_A > \psi$$
$$< \tau_A\varphi > B_a^\theta \psi \leftrightarrow B_a^{<\tau_A\varphi>\theta} < \tau_A\varphi > \psi$$
$$< \varphi!_A > B_b^\theta \psi \leftrightarrow \varphi \wedge B_b^{<\tau_A\varphi>\theta} < \tau_A\varphi > \psi \wedge (K_b[\tau_A\varphi]\neg\theta \rightarrow B_b^{<\varphi!_A>\theta} < \varphi!_A > \psi)$$
$$< \tau_A\varphi > B_b^\theta \psi \leftrightarrow B_b^{<\tau_A\varphi>\theta} < \tau_A\varphi > \psi \wedge (K_b[\tau_A\varphi]\neg\theta \rightarrow B_b^{<\varphi!_A>\theta} < \varphi!_A > \psi),$$

for all *insiders* $a \in A$ and all *outsiders* $b \notin B$. With these modifications, and by eliminating the axioms and rules referring to common knowledge and common belief, we obtain a *sound and complete proof system for the logic of private announcements (without common knowledge/belief)*. [17] And finally, here is the promised *"dynamic analogue" of the Ramsey test* (which is *valid*, unlike its static counterpart):

$$R!_a(s)_a \subseteq B_a[P!_a]Q \text{ iff } P!_a(R!_a(s)_a) \subseteq Q$$

Back to Mary. Having introduced these models, we return to the example presented in the introduction. So given Eve (e), Adam (a) and Mary (m), we denote the states in the initial model $\mathbf{S}$ by $x = (x^e, x^a, x^m)$, with $x^e, x^a, x^m \in \{0, 1\}$ where 0 =clean and 1 =dirty. The epistemic uncertainty relation is clear: agents see each other but not themselves, so $x(i) = \{y \in S : y^j = x^j \text{ for all } j \neq i\}$. We can convert this into a KB-model by e.g. assuming that agents start by being "cautious", i.e. *believing only what they know*. This sets $x_i = x(i)$. For conditional beliefs, we can use e.g. the "most trivial" belief revision policy (introduced in section 4): $s_i^P = s_i \cap P$, whenever $s_i \cap P \neq \emptyset$, and $s_i^P = s(i) \cap P$ otherwise. [18]

Now, the real state of the world is $w = (1, 1, 0)$. Mary's initial belief and knowledge is $w_m = w(m) = \{(1, 1, 1), (1, 1, 0)\}$. After father's announcement, the state $(0, 0, 0)$ is eliminated. Eve's peek in the mirror can be modeled as a private announcement $\gamma = 1^e!_e$ to herself (where the atomic sentence 1^e

---

[16] For reasons of simplicity, we eliminate common knowledge and common belief operators.

[17] We considered only this restricted logic for simplicity. As in the simpler case of purely epistemic updates with a private announcement (without belief revision) in [5], one can also obtain a complete axiomatization of the logic with common knowledge and common belief, by adding some generalized "Dynamic-Epistemic Induction" proof rules.

[18] But note that this particular choice of a trivial belief revision policy is irrelevant for the rest of this argument: the same analysis as below can be applied to *any* CDM based on the above KB model.

indicates that Eve's forehead is dirty), and the alternative (no peeking) is denoted by $\tau := \tau_e 1^e$. After that, in the resulting model $\mathbf{S}'$, Mary's knowledge is $\gamma(w)(m) = \{\tau(1,1,1), \tau(1,1,0), \gamma(1,1,1), \gamma(1,1,0)\}$, while her belief is $\gamma(w)_m = \{\tau(1,1,1), \tau(1,1,0)\}$. But after Eve's public announcement $(K_e 1^e)!$, the model shrinks to the set $S'' := \|K_e 1^e\|_{\mathbf{S}'} = \{\gamma(1, x^a, x^m) : x^a, x^m \in \{0,1\}\}$. Mary's unconditional beliefs after the public announcement are obtained using her prior *conditional* beliefs $(\gamma(w)_m)_{\mathbf{S}''} = \gamma(w)_m^{S''}$. To evaluate the last term we need to use the second case of clause 3 in the definition of private announcements (since we have $w_m \cap \tau^{-1}(S'') = \emptyset$): $(\gamma(w)_m)_{\mathbf{S}''} = \gamma_m^{S''} = \gamma(w_m^{\gamma^{-1}(S'')}) = \gamma(w(m) \cap \gamma^{-1}(S'')) = \gamma(w(m)) = \{\gamma(1,1,1), \gamma(1,1,0)\}$. This is a non-empty set of possible states: so Mary is still *sane*! Moreover, all her possible states are outputs of the action γ: she *knows* that γ has happened. In other words: Mary discovers that *cheating (γ) has taken place*!

7 Conclusion

We have presented here a new, qualitative semantic implementation of the *AGM* belief revision theory, in terms of conditional doxastic models. Based on this, we proposed a revised semantics for public and private announcements, incorporating belief revision into the notion of update. This "Merciful Logic" solves problems such as the ones posed by the above cheating version of the Muddy Children Puzzle, preventing agents from going mad when their beliefs are invalidated.

The semantical structures used in this paper have an algebraic counterpart. In [7], a first attempt has been made to work out an algebraic setting for multi-agent belief revision. In unpublished work [8], we generalize the present setting to allow other types of actions. More precisely, all epistemic action models in [6,4,5] can be conditionalized. The updated CDM's are actually the result of taking the "update product" (in a sense that refines the concept in [5]) of the initial conditional doxastic *state model* with a *conditional doxastic action model*. But the definition of the general update is rather complex and technical, and would require a lot of preparation and justification. To build a case for it, we chose for simplicity (and due to lack of space) to concentrate here on two very special cases, of great intuitive appeal. But the general picture can already be glimpsed from our example: looking back at Mary, Adam and Eve, it is obvious that when Mary revises her beliefs as part of the update with action $(K_e 1^e)!$, she actually deduces that the cheating action $\gamma = 1^e!_e$ has happened (instead of τ); so she revises not only her static beliefs about propositions, but also her *beliefs about actions*.

References

[1] E. W. Adams. A logic of conditionals. *Inquiry* 8: 166-197. 1965.

[2] C.E. Alchourron, P. Gardenfors, D. Makinson. On the Logic of Theory Change: Partial Meet Contraction and Revision Functions. *The Journal of Symbolic Logic*, **50**, No 2, 510-530. 1985.

[3] G. Aucher. *A Combined System for Update Logic and Belief Revision*. Master's thesis, ILLC, University of Amsterdam, the Netherlands, 2003.

[4] A. Baltag. A Logic for Suspicious Players: Epistemic Actions and Belief Updates in Games. *Bulletin of Economic Research*, **54**(1), 1-46. 2002.

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

[6] 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 7th Conference on Theoretical Aspects of Rationality and Knowledge*, (TARK'98), 43-56. 1998.

[7] A. Baltag and M. Sadrzadeh. The Algebra of Multi-Agent Dynamic Belief Revision. To appear in *Electronic Notes in Theoretical Computer Science*, proceedings of IJCAI. 2005.

[8] 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*.

[9] J. Bennett. *A philosophical guide to conditionals*. Oxford Univ. Press. 2003.

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

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

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

[13] J. Gerbrandy. Dynamic Epistemic logic. In L.S. Moss et al. (eds), *Logic, Language and Information*, vol. 2, CSLI Publications, Stanford University, 1999.

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

[15] A. Fuhrmann and I. Levi. Undercutting and the Ramsey Test for Conditionals. *Synthese*,**101**, 157-169. 1994.

[16] A. Grove. Two Modellings for Theory Change. *Journal of Philosophical Logic* , **17**:2, 157-170. 1988.

[17] J.-J. Ch. Meyer, W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge Univ. Press, 1995.

[18] D. Lewis. *Counterfactuals*. Blackwell Publishing, Oxford,1973.

[19] F. Ramsey. *The Foundations of Mathematics and Other Essays*. Kegan Paul, London, 1931.

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

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

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

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

[24] 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.

[25] R.C. Stalnaker. A Theory of Conditionals. In N. Rescher (ed.), *Studies in Logical Theory*, Oxford, Blackwell, APQ Monograph No2, 1968.

[26] J. van Benthem, J. van Eijck and B. Kooi. Logics of Communication and Change. Presented at TARK2006 Singapore, available at http://staff.science.uva.nl/ johan/publications.html.

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