

---

# Protocols for Belief Merge: Reaching Agreement via Communication

ALEXANDRU BALTAG, *ILLC, University of Amsterdam, the Netherlands. E-mail: A.Baltag@uva.nl*

SONJA SMETS, *ILLC, University of Amsterdam, the Netherlands. E-mail: S.J.L.Smets@uva.nl*

## Abstract

We investigate the issue of reaching doxastic agreement among the agents of a group by “sharing” information via successive acts of sincere, persuasive and public communication within the group. The topic relates to “preference aggregation” in Social Choice theory, where the problem is to find a natural and fair merge operation for aggregating the agents’ preferences into a single group preference. In this paper we study this topic within the setting of Dynamic Epistemic Logic (DEL) and its recent extensions to belief revision theory. First we interpret the agents’ preference relations as “doxastic preferences” or “doxastic plausibility orders”. Using these plausibility structures, we can express several types of relevant doxastic attitudes: ranging from an agent’s simple beliefs and conditional beliefs to her defeasible knowledge and irrevocable knowledge. Next, we model communication via a series of doxastic/epistemic actions that affect the agent’s plausibility structures. Because these structures can change in different ways, a group of agents has to adopt a specific communication protocol in order to aggregate their doxastic attitudes. We study the type of merges that are realizable via specific communication protocols and we highlight important factors which may influence the result, such as “the order of the speakers”, or the listeners’ *doxastic attitudes* towards the speakers (i.e their opinions concerning the reliability of the incoming information).

*Keywords:* dynamic epistemic logic, belief revision theory, belief merge, preference aggregation, doxastic logic, epistemic protocols

## 1 Introduction

As usually considered in Social Choice theory, the problem of *preference aggregation* is to find a natural and fair “*merge*” operation (subject to various naturalness or fairness conditions), for aggregating the agents’ preferences into a single “group preference”. Depending on the stringency of the required fairness conditions, one can obtain either an Impossibility theorem (e.g Arrow’s theorem [3]) or a classification of the possible types of “reasonable” merge operations [2]. In this paper we propose a more “*dynamic*” approach to this issue. Dynamically speaking, “merging” preference relations means finding an action or a sequence of actions (a protocol) that, when applied to any arbitrary multi-agent preference model, produces a new model in which all the agents’ preference relations are the same. When the new relations are the result of a specific merge operation, we say that we have “realized” this operation via the given (sequence of) action(s). One would like to know what types of merges are realizable

## 2 Protocols for Belief Merge: Reaching Agreement via Communication

by using only specific types of preference-changing actions.

In a doxastic/epistemic setting, the agents' preference relations are interpreted as “doxastic preferences” or “doxastic plausibility” orders. These encode the agents' beliefs, but in fact they capture *all* their doxastic-epistemic attitudes: their “knowledge” (in the sense of absolutely certain, un-revisable, irrevocable knowledge, i.e. the epistemic concept mostly used in Logic, Computer Science and Economics), their “strong beliefs”, their “safe beliefs” (also known as “defeasible knowledge”, i.e. the epistemic concept used mostly by philosophers and researchers in Belief Revision theory), as well as their “conditional beliefs” (encoding their “belief-revision strategy”, i.e. their contingency plans for belief change). In other words, an agent's doxastic preference structure captures *all her “information”*: both her “hard” (absolutely certain, infallible) information and her “soft” (potentially fallible) information. In this context, a preference merge operation corresponds to a way of *combining the agents' information into a single “group information”*.

Similarly, preference-changing actions can be interpreted in a doxastic setting as acts of *communication* or *persuasion*. But not every preference-changing action can be understood in this way: there has to be a specific relation between one agent's (the speaker's) prior preferences before the action and the whole group's posterior preferences. Actions in which this relation holds will be instances of what we call in this paper *sincere, persuasive and public communication*.

An announcement of some information  $P$  is said to be “public” when it is common knowledge that this particular message  $P$  is announced and that all the agents are adopting the same attitude towards the (plausibility of the) announcement: they all adopt the same opinion about the reliability of this information. Depending on the specific common attitude, there are three main possibilities that have been discussed in the literature: 1) the information  $P$  is *certainly* true: it is *common knowledge* that the message is necessarily truthful; (2) the announcement is *strongly believed* by all agents to be true: it is *common knowledge that everybody strongly believes* that the speaker tells the truth; (3) the announcement is (*simply*) *believed*: it is *common knowledge that everybody believes (in the simple, “weak” sense)* that the speaker tells the truth. These three alternatives correspond to three forms of “learning” a public announcement, forms discussed in [15, 17] in a Dynamic Epistemic Logic context: “update” <sup>1</sup>  $!P$ , “radical upgrade”  $\uparrow P$  and “conservative upgrade”  $\uparrow P$ . Under various names, they have been previously proposed in the literature on Belief Revision, e.g. by Rott [29] and Boutilier [13], and in the literature on dynamic semantics for natural language by Veltman [34]. The first operation (update) models “*truthful public announcements*” of “*hard*” information; the other two are models of “*soft*” public announcements.

“Sincerity” of a communication act can be defined as sharing of information that *was already “accepted” by the speaker* (before the act). The meaning of “acceptance” depends on the form of communication: as we'll see, for updates with “hard” information, acceptance means “knowledge” (in the irrevocable sense), while for upgrades with “soft” information, acceptance just means some type of “belief” or “strong belief”

---

<sup>1</sup>Note that in Belief Revision, the term “belief update” is used for a totally different operation (the Katzung-Mendelzon update [27]), while what we call “update” is known as “conditioning”. We choose to follow here the terminology used in Dynamic Epistemic Logic, but we want to warn the reader against any possible confusions with the KM update.

(depending on whether the upgrade is “conservative” or “radical”). But, as a general concept, sincerity requires that the speaker’s prior doxastic attitude (knowledge, belief, strong belief) about the announced sentence *matches* the attitude induced in the listener (towards the same sentence) by her announcement

“Persuasiveness” requires that the communicated information *becomes commonly “accepted” by all the agents* (in the same sense of “acceptance” that the speaker has adopted): this means that, after the act, everybody *commonly* exhibits *the same doxastic attitude as the speaker* (knowledge, belief or strong belief) towards the communicated information. So, after a persuasive communication, all agents reach a *partial agreement*, namely with respect to the specific information that has been communicated.

In a cooperative setting, the *goal* of “sharing” doxastic information is *reaching “agreement”* with respect to *all* the (relevant) issues. Indeed, the natural stopping point of iterated sharing is when *nothing is left for further sharing or persuading*; i.e. *complete* agreement. Any further sincere persuasive communication is *redundant* at that stage: it can no longer change any agent’s doxastic structure. This happens exactly when all the agents’ relevant doxastic attitudes towards all issues are *exactly the same*. (Which attitude is relevant depends again of the type of communication: “knowledge” for updates, “belief” for conservative upgrades, “strong belief” for radical upgrades). This means that the agents’ (relevant) accessibility relations (i.e. respectively, the knowledge relations, the belief structure or the strong belief structure) *became identical*: we say that these structures *have “merged” into one*.

So we arrive in a natural way at the main issue addressed in this paper: *the “dynamic merge” of doxastic structures by sincere and persuasive public communication*. In this context, it is natural to investigate the realizability of merge operations via (1) updates, (2) radical upgrades and (3) conservative upgrades. We show that, in the first case, only the epistemic structures (given by the “hard” knowledge relations) can be merged; and moreover, the only form of realizable merge is in this case the so-called “parallel merge” [2], given by the intersection of all preference relations. Epistemically, this corresponds to the familiar concept of “distributed knowledge”. The realizability result is constructive, it comes with a specific announcement-based protocol for realizing this merge. This is essentially the algorithm in van Benthem’s paper “One is a Lonely Number” [14]: the agents announce “all they know”, in no particular order. In the second case (radical or lexicographic upgrade), the “defeasible knowledge” structures are merged, but in fact this implies that all the other doxastic attitudes become the same: *the agents’ whole “doxastic preference” structures are merged*. The natural analogue of the above-mentioned protocol for radical (or lexicographic) upgrades, in the absence of hard information, realizes now a different type of merge which is called the “lexicographic merge”. In the presence of both hard and soft information, a new protocol which is a mixture of the previous two protocols, is needed to merge the whole plausibility structure of the agents. With this complex protocol we can realize the “priority merge”. Finally, in the case of conservative upgrades, only the (simple) belief structures (given by the doxastic relations) can be merged.

This *surface similarity* between the different cases is pleasing, but in fact it hides *deeper dissimilarities*. As we mentioned, the realizable merge is *unique* only in the

#### 4 Protocols for Belief Merge: Reaching Agreement via Communication

first case. This is *not* true in the other cases: a whole class of merge operations can be realized by radical or conservative upgrades. Moreover, in the first case, the order in which the announcements are made is irrelevant, while in the other cases *the order matters*: if the upgrades are performed in a different order than the one prescribed in the protocol, then different merge operations may be realized! Finally, in the first case, the merge may be realized by allowing *only one announcement by each agent* (of “all she knows”). But this is *not true* in the other cases: the agents may have to make many soft announcements, including announcing facts that may already be entailed by their previous announcements!

Some of the questions we address in this paper came to our attention after hearing a presentation by J. van Benthem on “The Social Choice Behind Belief Revision” at the workshop “Dynamic Logic Montreal” in 2007 [16]. Van Benthem’s view was that belief dynamics in itself can be captured as a form of preference merge (between the prior doxastic preferences and the on-going doxastic preferences about the new information). One can see that our approach here is actually *the dual* of the perspective adopted in [16]: implementing preference merge dynamically by successive belief revisions, instead of understanding belief revision in terms of preference merge.

In the next section we introduce the necessary background on different notions of knowledge, belief and other doxastic attitudes. The main focus will be on the semantics, which is given via preference models. In section 3, we introduce the main concepts of belief dynamics, following the work in [4, 5, 6, 7, 15] on joint upgrades, as models for “sincere, persuasive public announcements”. In section 4 we present three natural merge operations: parallel merge, lexicographic merge and (relative) priority merge. In section 5 we present the protocols for dynamic realizations of parallel merge, lexicographic merge and priority merge, giving counterexamples that point out the differences between them. We end with a short note and an open question in our Conclusions section.

This paper was first presented at the FAMAS-MALLOW workshop in Italy in 2009 and is a slight extended version of the proceedings-paper in [10].

## 2 Plausibility Structures and Doxastic Attitudes

This paper is set within the formal framework of Dynamic Epistemic Logic (DEL) and its recent extensions on dealing with mechanisms for belief revision. We use this section to review some basic notions and results from [8]. In particular we work with finite “plausibility” frames, in the sense of our papers [4, 5, 6, 7, 8, 9]. These kind of semantic structures are the *natural multi-agent generalizations of structures that are standard*, in one form or another, in Belief Revision Theory: see e.g. Halpern’s “preferential models” [24], Spohn’s ordinal-ranked models [30], Board’s “belief-revision structures” [19], Grove’s “sphere” models [23]. Similar as in most of our previous work in belief revision theory, we adopt in this paper a pure qualitative logical setting.<sup>2</sup> Hence our approach retains the qualitative flavor of the original AGM theory [1] in belief revision. But note that unlike the settings in [8, 9], we restrict here to the *finite* case, for reasons of simplicity.

**Plausibility Frames** For a given finite set  $\mathcal{A}$  of labels called “agents”, a (*finite*,

---

<sup>2</sup>For a detailed discussion of our motivation for this choice, see [8].

*multi-agent*) *plausibility frame* is just a finite, multi-agent Kripke frame  $(S, R_a)_{a \in \mathcal{A}}$  consisting of a set  $S$  of states or so-called possible worlds and accessibility relations  $R_a \subseteq S \times S$  for each agent in  $\mathcal{A}$ . The accessibility relations  $R_a \subseteq S \times S$  in a plausibility frame are usually denoted by  $\leq_a$  and are called “plausibility orders” or “doxastic preference” relations. These accessibility relations are assumed to be *locally connected preorders*. Note that a “locally connected preorder”  $\leq \subseteq S \times S$  is a reflexive and transitive relation such that: if  $s \leq t$  and  $s \leq w$  then either  $t \leq w$  or  $w \leq t$ ; and if  $t \leq s$  and  $w \leq s$  then either  $t \leq w$  or  $w \leq t$ . We refer to [8] for a justification and motivation for these conditions.<sup>3</sup> We use the notation  $s \sim_a t$  for the *comparability relation* with respect to  $\leq_a$ :  $s \sim_a t$  iff either  $s \leq_a t$  or  $t \leq_a s$ . For simplicity, throughout this paper we make the additional assumption that *the transitive closure of the union of all comparability relations*  $(\bigcup_{a \in \mathcal{A}} \sim_a)^*$  is “globally connected”<sup>4</sup>: this means that every state  $w$  can be reached from every other state  $s$  by a finite chain  $w = w_0 \sim_{a_0} w_1 \sim_{a_1} \dots w_n \sim_{a_n} s$ .

We write  $s <_a t$  for the corresponding *strict order relation*:

$$s <_a t \text{ iff } s \leq_a t \text{ but } t \not\leq_a s$$

And we write  $s \cong_a t$  for the corresponding *indifference relation*:

$$s \cong_a t \text{ iff both } s \leq_a t \text{ and } t \leq_a s$$

When using the  $R_a$  notation for the preference relations  $\leq_a$ , we also use the notations  $R_a^<$ ,  $R_a^\sim$  and  $R_a^\cong$  to denote the corresponding strict order, comparability and indifference relations  $<_a$ ,  $\sim_a$  and  $\cong_a$ .

In a plausibility frame, the comparability relations  $\sim_a$  are *equivalence relations*, hence they induce *partitions*. We denote by  $s(a) := \{t \in S : s \sim_a t\}$  the  $\sim_a$ -*partition cell* of  $s$ , comprising all  $a$ ’s *epistemic alternatives* for  $s$ . Finally, we use  $\rightarrow_a$  to denote the “best alternative” or “most preferred” relation  $\rightarrow_a$ , given by:

$$s \rightarrow_a t \text{ iff } t \geq_a t' \text{ for all } t' \in s(a).$$

**Plausibility Models** A (*finite, multi-agent, pointed*) *plausibility model* is a structure  $\mathbf{S} = (S, \leq_a, \|\cdot\|, s_0)_{a \in \mathcal{A}}$ , consisting of a plausibility frame  $(S, \leq_a)_{a \in \mathcal{A}}$  together with a *valuation* and a *designated state*  $s_0 \in S$ , which is called the “actual state”. The valuation  $\|\cdot\| : \Phi \rightarrow \mathcal{P}(S)$  maps every element  $p$  of some given set  $\Phi$  of “atomic sentences” into a set of states  $\|p\| \subseteq S$ .

**(Common) Knowledge and (Conditional) Belief** Given a plausibility model  $\mathbf{S}$ , the sets  $P, Q \subseteq S$  of states, an agent  $a \in \mathcal{A}$  and some group  $G \subseteq \mathcal{A}$ , we define the set of most plausible states satisfying a given proposition  $P$  as:

$$\text{best}_a P = \text{Max}_{\leq_a} P := \{s \in P : t \leq_a s \text{ for all } t \in P\}$$

The irrevocable type of knowledge of an agent is written as standard in epistemic logic:

$$K_a P := \{s \in S : s(a) \subseteq P\}$$

<sup>3</sup>In the infinite case, one has to add a well-foundedness condition, obtaining “locally well-preordered” relations (see [8]).

<sup>4</sup>This simplifying assumption allows us to avoid the need for the technical modal notion of bisimilarity.

## 6 Protocols for Belief Merge: Reaching Agreement via Communication

Simple beliefs are defined as follows :

$$B_a P := \{s \in S : best_a s(a) \subseteq P\}.$$

Conditional beliefs are defined via binary operators labeled by agents:

$$B_a^Q P := \{s \in S : best_a(s(a) \cap Q) \subseteq P\}.$$

Operators for group knowledge and group beliefs are set as follows:

The operator for “everybody knows” is:

$$EkP := \bigcap_{a \in \mathcal{A}} K_a P$$

The operator for “everybody believes” is:

$$EbP := \bigcap_{a \in \mathcal{A}} B_a P$$

The operator for common knowledge among the agents is:

$$CkP := \bigcap_{n \in \mathbb{N}} Ek^n P$$

(where  $Ek^1 P := EkP$  and  $Ek^{n+1} := Ek(Ek^n P)$ ). The operator for common belief  $CbP$  is similarly defined (by replacing  $Ek$  with  $Eb$  in the definition of common knowledge).

**Interpretation.** The elements of  $S$  represent the “possible worlds” or possible states of a system and we view them as the *possible descriptions of the real world*. The *correct* description of the real world, which might be unknown to the agents, is given by the “actual state”  $s_0$ . The atomic sentences  $p \in \Phi$  represent “*ontic*” (*non-doxastic*) facts, that might hold or not in a given state. The valuation tells us which facts hold at which worlds. For each agent  $a$ , the equivalence relation  $\sim_a$  represents the agent  $a$ ’s *epistemic indistinguishability relation*, inducing  $a$ ’s *information partition*;  $s(a)$  is the state  $s$ ’s information cell with respect to  $a$ ’s partition: if  $s$  were the real state, then agent  $a$  would consider all the states  $t \in s(a)$  as “epistemically possible”.  $K_a P$  is the proposition “agent  $a$  knows  $P$ ”: observe that this is indeed the same as Aumann’s partition-based definition of knowledge. The plausibility relation  $\leq_a$  is agent  $a$ ’s “*doxastic preference*” relation: her *plausibility order* between her “epistemically possible” states. So we read  $s \leq_a t$  as “agent  $a$  considers  $t$  at least as plausible as  $s$  (though the two are *epistemically indistinguishable*)”. This is meant to capture the agent’s (*conditional*) beliefs about the state of the system. Note that  $s \leq_a t$  implies  $s \sim_a t$ , so that the agent only compares the plausibility of states that are epistemically indistinguishable: so we are not concerned here with counterfactual beliefs (going against the agent’s knowledge), but only with conditional beliefs (if given new evidence that must be compatible with prior knowledge). So  $B_a^Q P$  is read “agent  $a$  believes  $P$  conditional on  $Q$ ” and means that, if  $a$  would receive some further (certain) information  $Q$  (to be added to what she already knows) then she would believe that  $P$  was the case. So conditional beliefs  $B_a^Q$  give descriptions of the agent’s *plan* (or

commitments) about what would she believe (about the current state) if she would learn some new information  $Q$ . To quote J. van Benthem in [15], conditional beliefs are “static pre-encodings” of the agent’s potential belief changes in the face of new information. The above definition says that  $B_a^Q P$  holds iff  $P$  holds in all the “best” (i.e. the most plausible)  $Q$ -states (that are consistent with  $a$ ’s knowledge). In particular, a *simple (non-conditional) belief*  $B_a P$  holds iff  $P$  holds in all the best states that are epistemically possible for  $a$ .

**Kripke Modalities** For any binary accessibility relation  $R \subseteq S \times S$  and set  $P \subseteq S$  in a given model  $\mathbf{S}$ , the corresponding *Kripke modality* is given by:

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

We think of sets  $P \subseteq S$  as *propositions*, we write  $s \models_{\mathbf{S}} P$  instead of  $s \in P$  and  $\mathbf{S} \models P$  instead of  $s_0 \in P$ .<sup>5</sup>

It is easy to see that *belief is the Kripke modality*  $B_a = [\rightarrow_a]$  for the “best alternative” relation  $\rightarrow_a$  defined above. Similarly, knowledge is the Kripke modality for the epistemic relation  $K_a = [\sim]$ .

**Safe belief as “defeasible knowledge”** The Kripke modality for the plausibility relation  $\Box_a := [\leq_a]$  was called “safe belief” in [8], and “the preference modality” in [18]. It was also considered by Stalnaker in [31], as a formalization of Lehrer’s notion of “defeasible knowledge”. According to this so-called *defeasibility theory of knowledge*, a belief counts as “knowledge” if it is *stable under belief revision with any true information*. Indeed, the safe belief modality has the property that *it is conditionally believed under any true condition*:

$$s \models \Box_a Q \text{ iff: } s \models B_a^P Q \text{ for all } P \text{ such that } s \models P.$$

For this reason, we’ll refer to  $\Box$  using either of the terms “safe belief” and “defeasible knowledge”. In contrast, we called the knowledge concept captured by the  $K$  modality “irrevocable knowledge”, since it is a belief that is stable under revision with *any* information (including false ones):

$$s \models K_a Q \text{ iff: } s \models B_a^P Q \text{ for all } P.$$

There are other differences: irrevocable knowledge  $K$  satisfies the axioms of the modal system  $S5$ , so it is fully introspective; in contrast, defeasible knowledge  $\Box$  is only positively introspective, but not necessarily negative introspective. In fact, the complete logic of  $\Box$  is the modal logic  $S4.3$  (see [31]). *An agent’s belief can be safe without him necessarily “knowing” this* (in the “strong” sense of the irrevocable knowledge  $K$ ): “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. Indeed, *the only way for an agent to know a belief to be safe is to actually know it to be truthful*. This is captured by the valid identity:

$$K_a \Box_a P = K_a P.$$

In other words: *knowing that something is safe to believe is the same as just knowing it to be true*. In fact, *all beliefs held by an agent “appear safe” to him* : in order

---

<sup>5</sup>If no confusion is possible, we drop the subscript  $\mathbf{S}$  in our notation and simply write  $s \models P$

## 8 Protocols for Belief Merge: Reaching Agreement via Communication

to believe them, he has to believe that they are safe. This is expressed by the valid identity:

$$B_a \Box_a P = B_a P$$

saying that: *believing that something is safe to believe is the same as just believing it*<sup>6</sup>. Contrast this with the situation concerning “knowledge”: in our logic (as in most standard doxastic-epistemic logics), we have the identity:

$$B_a K_a P = K_a P.$$

So *believing that something is known is the same as knowing it*!

The difference between  $K$  and  $\Box$  and their different properties, expressed by the above identities, are enough to solve the so-called “Paradox of the Perfect Believer” in [22, 36, 33, 28, 37, 21]: when we say that somebody “only believes that she knows something (without really knowing it)”, we’re using the word “knowledge” in a different sense than the fully introspective  $K$  modality. A natural reading is to interpret it as the defeasible knowledge  $\Box$ , in which case “believing that you know” is the same as “believing”, by the identity  $B_a \Box_a P = B_a P$ .

**“Strong Belief”** Another important doxastic attitude, called *strong belief*, is given by:

$$Sb_a P = \{s \in S : s(a) \cap P \neq \emptyset \text{ and } w >_a t \text{ for all } t \in s(a) \cap P \text{ and all } w \in s(a) \setminus P\}.$$

So  $P$  is a strong belief at a state  $s$  iff  $P$  is *epistemically possible* and moreover *all epistemically possible  $P$ -states at  $s$  are more plausible than all epistemically possible non- $P$  states*. This notion was called “strong belief” by Battigalli and Siniscalchi [12], while Stalnaker [32] calls it “robust belief”. It is easy to see that we have the following equivalence:

$$s \models Sb_a P \text{ iff } (s \models B_a P \text{ and } s \models B_a^Q P \text{ for every } Q \text{ such that } s \models \neg K_a(Q \rightarrow \neg P)).$$

In other words: *something is strongly believed if and only if it is believed and if this belief can only be defeated by evidence (truthful or not) that is known to contradict it*. An example is the “presumption of innocence” in a trial: requiring the members of the jury to hold the accused as “innocent until proven guilty” means asking them to start the trial with a “strong belief” in innocence.

The operators  $ESbP$  (“everybody strongly believes  $P$ ”) and  $CSbP$  (“ $P$  is common strong belief”) are defined in a similar manner as  $EbP$  and  $CbP$  (simply by replacing belief with strong belief in their definitions).

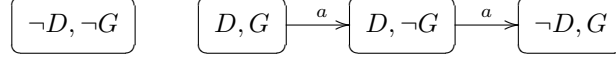
**Example 1:** Consider the situation of Professor Albert Winestein. Albert feels that he is a genius. He knows that there are only two possible explanations for this feeling: either he is a genius or he is drunk. He doesn’t feel drunk, so he believes that he is a sober genius. However, *if* he realized that he’s drunk, he’d think that his genius feeling was just the effect of the drink; i.e. *after learning he is drunk* he’d come to *believe that he was just a drunk non-genius*. In reality though, Albert is *both drunk and a genius*.

---

<sup>6</sup>The proof is an easy semantic exercise, which can be rendered in English as: saying that “the best worlds have the property that all the worlds at least as good as them are  $P$ -worlds” is equivalent to simply saying that “the best worlds are  $P$ -worlds”.

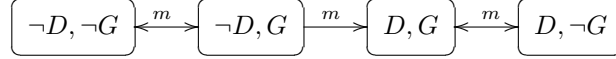


We can represent Albert's information and (conditional beliefs) by the following plausibility model:



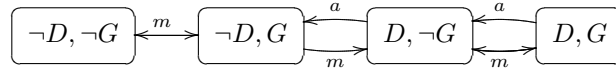
Here, as in all other drawings, we use *labeled arrows for plausibility relations*  $\leq_a$  (not for the “best alternative” relations  $\rightarrow_a$ !), going from less plausible to more plausible worlds, but *we skip loops and composed arrows* (since  $\leq_a$  are reflexive and transitive). The real world is  $(D, G)$ . Albert considers  $(D, \neg G)$  as being more plausible than  $(D, G)$ , and  $(\neg D, G)$  as more plausible than  $(D, \neg G)$ . Albert can distinguish all these worlds from  $(\neg D, \neg G)$ , since (in the real world) he knows (“ $K_a$ ”) that either  $D$  or  $G$  holds.

Consider another agent in our story, Professor Mary Curry. She is pretty sure that Albert is drunk: she can see this with her very own eyes. But Mary is *completely indifferent with respect to Albert's genius*: so she considers the possibility of genius and the one of non-genius as equally plausible. However, having a philosophical mind, Mary is aware of the possibility that the testimony of her eyes may in principle be wrong: it is in principle possible that Albert is not drunk, despite the presence of the usual symptoms. Mary's beliefs are captured by her plausibility order:



We can see from the drawing that Mary strongly believes  $D$ , and in fact in the real world  $(D, G)$  her belief in  $D$  is also safe: so she “knows” that Albert is drunk, in the sense of defeasible knowledge (although she doesn't know it, in the sense of  $K$ ). But she is completely indifferent with respect to  $G$ : hence she considers the possibility of  $G$  and  $\neg G$  as equally plausible.

To put together the agents' plausibility orders, we need to be told what do they know about each other. Suppose all their opinions as described above (i.e. all their conditional beliefs) are *common knowledge*: essentially, this means their doxastic preferences are common knowledge. We thus obtain the following multi-agent plausibility model:

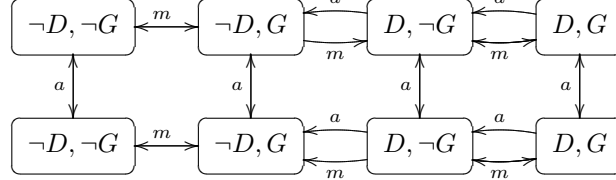


At the real world  $(D, G)$ , one can check that  $B_a G$  is true. Further, Albert does not know  $G$  in the sense of  $K$ , hence we have  $(D, G) \models \neg K_a G$  while  $(D, G) \models K_a (D \vee G)$ . Moreover, he doesn't “know”  $G$  in the defeasible sense either:  $(D, G) \models \neg \Box_a G$ . Indeed, his belief in  $G$  is *not safe*, since  $B_a^D \neg G$  holds in the real world: so if Albert would learn (correctly) that he was drunk, then he would lose his (true) belief in being a genius.

**Example 2** Let us now relax our assumptions about the agents' mutual knowledge: suppose that only Albert's opinions are common knowledge. In addition, we suppose that it is common knowledge that Mary has no opinion on Albert's genius (so she considers genius and non-genius as equi-plausible), but we assume that she has a

strong opinion about his drunkenness: she can see him, so judging by his behavior she either strongly believes that he is drunk or she strongly believes he is not drunk. However, her actual opinion about this is unknown to Albert, who thus considers both opinions as equally plausible.

The resulting model immediately becomes more complex:



The real world is represented by *the upper*  $(D, G)$  state. One can check that, in the real world, Mary still strongly believes that Albert is drunk; but he does not know this: Mary's plausibility relation between  $D$  and  $\neg D$  is unknown to Albert. However, he knows that either she strongly believes  $D$  or she strongly believes  $\neg D$ .

We can go on and modify the example further, by allowing that Albert's plausibility is not commonly known either etc. But, for simplicity of drawing, we stop here: when less common knowledge is assumed, more worlds are possible, and hence the drawings get more and more complex.

### 3 Belief Dynamics: Sincere, Persuasive Public Communication

In this section we focus on *belief dynamics* and are mainly concerned with the following question, namely what happens when some proposition  $P$  is *publicly announced*? According to Dynamic Epistemic Logic, this induces, not only a revision of beliefs, but a *change of the model*: a “revision” of the whole relational structure, changing the agents' plausibility orders. However, the specific change depends on *the agents' attitudes to the plausibility* of the announcement: *how certain is the new information*? Three main possibilities have been discussed in the literature:

- (1) the announcement  $P$  is *certainly* true: it is *common knowledge* that the speaker tells the truth;
- (2) the announcement is *strongly believed* to be true by everybody: it is *common knowledge that everybody strongly believes* that the speaker tells the truth;
- (3) the announcement is (*simply*) *believed*: it is *common knowledge that everybody believes (in the simple, “weak” sense)* that the speaker tells the truth.

These three alternatives correspond to three forms of “joint learning”, forms discussed in [15, 17] in a Dynamic Epistemic Logic context: “update” <sup>7</sup>  $!P$ , “*radical upgrade*”  $\uparrow P$  and “*conservative upgrade*”  $\uparrow P$ . Under various names, the single-agent versions of these doxastic transformers have been previously proposed by e.g. Rott [29], Boutilier [13] and Veltman [34].

<sup>7</sup>Note that in Belief Revision, the term “belief update” is used for a totally different operation (the Katzung-Mendelzon update[27]), while what we call “update” is known as “conditioning”. We choose to follow here the terminology used in Dynamic Epistemic Logic, but we want to warn the reader against any possible confusions with the KM update.

We will use “joint upgrades” as a general term for all these three model transformers, and denote them in general by  $\uparrow P$ , where  $\uparrow \in \{!, \uparrow, \uparrow\}$ . Formally, each of our joint upgrades is a (possibly partial) function taking as inputs pointed models  $\mathbf{S} = (S, \leq_a, \|\cdot\|, s_0)$  and returning new (“upgraded”) pointed models  $\uparrow P(\mathbf{S}) = (S', \leq'_a, \|\cdot\|', s'_0)$ , with  $S' \subseteq S$ . Since upgrades are purely doxastic, they *won't affect the real world or the “ontic facts” of each world*: i.e. they all satisfy  $s'_0 = s_0$  and  $\|p\|' = \|p\| \cap S'$ , for atomic  $p$ . So, in order to completely describe a given upgrade, we only have to specify (a) *its possible inputs*  $\mathbf{S}$ , (b) *the new set of states*  $S'$ ; (c) *the new accessibility relations*  $\leq'_a$ .

**(1) Learning Certain information: Joint “Update”.** The update  $!P$  is an operation on pointed models which is *executable* (on a pointed model  $\mathbf{S}$ ) *iff*  $P$  is true (at  $\mathbf{S}$ ) and which *deletes all the non- $P$ -worlds from the pointed model, leaving everything else the same*. Formally, an update  $!P$  is an upgrade such that: (a) it takes as inputs only pointed models  $\mathbf{S}$ , such that  $\mathbf{S} \models P$ ; (b) the new set of states is  $S' = \{s \in S : s \models P\}$ ; (c)  $s \leq'_a t$  iff  $s \leq_a t$  and  $s, t \in S'$ .

**(2) Learning from a Strongly Trusted Source: (Joint) “Radical” Upgrade.** The “radical upgrade” (or “lexicographic upgrade”)  $\uparrow P$ , as an operation on pointed plausibility models, can be described as “promoting” all the  $P$ -worlds within each information cell so that they become “better” (more plausible) than all  $\neg P$ -worlds in the same information cell, while keeping everything else the same: the valuation, the actual world and the relative ordering between worlds within either of the two zones ( $P$  and  $\neg P$ ) stay the same. Formally, a radical upgrade  $\uparrow P$  is (a) a *total* upgrade (taking as input *any* model  $\mathbf{S}$ ), such that (b)  $S' = S$ , and (c):  $s \leq'_a t$  iff either ( $s, t \notin P$  and  $s \leq_a t$ ) or ( $s, t \in P$  and  $s \leq_a t$ ) or ( $t \in s(a)$  and  $s \notin P$  and  $t \in P$ ).

**(3) “Barely believing” what you hear: (Joint) “Conservative” Upgrade.** The so-called “conservative upgrade”  $\uparrow P$  (called “minimal conditional revision” by Boutilier [13]) performs in a sense the minimal possible revision of a model that is forced by believing the new information  $P$ . As an operation on pointed models, it can be described as “promoting” only the “best” (most plausible)  $P$ -worlds, so that they become the most plausible in their information cell, while keeping everything else the same. Formally,  $\uparrow P$  is (a) a *total* upgrade, such that (b)  $S' = S$ , and (c):  $s \leq'_a t$  iff either ( $t \in \text{best}_a(s(a) \cap P)$ ) or ( $s \notin \text{best}_a(s(a) \cap P)$  and  $s \leq_a t$ ).

**Sincerity** Intuitively, an announcement is “sincere” when it reflects the speaker’s prior epistemic state: *it induces in the listener the same attitude towards his statement that he himself had before the announcement*. More precisely:

**Definition:** A public announcement  $!P$  made by an agent  $a$  is *sincere* iff  $a$  knows  $P$ , i.e. if  $K_a P$  holds in the original model (before the announcement). An announcement  $\uparrow P$  made by an agent  $a$  is *sincere* iff  $a$  strongly believes  $P$  (before the announcement). An announcement  $\uparrow P$  made by an agent  $a$  is *sincere* iff  $a$  believes  $P$  (before the announcement).

**Redundancy versus Informativity** A joint upgrade  $\uparrow P$  is *redundant on a model*  $\mathbf{S}$  if the upgraded model is *the same*<sup>8</sup> as the original one:  $\uparrow P(\mathbf{S}) = \mathbf{S}$ . An upgrade  $\uparrow P$  is *informative (on  $\mathbf{S}$ )* if it is not redundant.

<sup>8</sup>Usually, in modal logic two models are “essentially the same” when they are bisimilar. Our “global connectedness” assumption allows us to avoid the technical notion of bisimilarity, and simply use *identity* of models instead.

**Proposition 1**

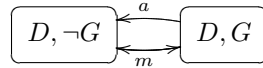
1. In a pointed model  $\mathbf{S}$ ,  $!P$  is *redundant* iff  $P$  is *common knowledge* in  $\mathbf{S}$ .
2.  $\uparrow P$  is *redundant* iff it is *common knowledge that  $P$  is common strong belief*. Equivalently, iff it is *common knowledge that everybody strongly believes  $P$* .
3.  $\uparrow P$  is *redundant* iff it is *common knowledge that  $P$  is common belief*. Equivalently, iff it is *common knowledge that everybody believes  $P$* .

**Invariance under communication:** For a given upgrade type  $\dagger \in \{!, \uparrow, \uparrow\}$ , we say that a pointed model  $\mathbf{S}$  is *invariant under  $\dagger$ -communication* iff, for all propositions  $P$ , any sincere announcement of the form  $\dagger P$  made by any agent in  $\mathcal{A}$  is redundant in  $\mathbf{S}$ .

**Proposition 2**

1. A pointed model  $\mathbf{S}$  is *invariant under  $!$ -communication* iff all (irrevocable) knowledge is common knowledge, i.e. for all propositions  $P \subseteq S$  and all agents  $a, b \in \mathcal{A}$ ,  $K_a P$  holds in  $\mathbf{S}$  iff  $K_b P$  holds in  $\mathbf{S}$ ; equivalently: iff all the agents' epistemic relations coincide:  $\sim_a = \sim_b$  for all  $a, b \in \mathcal{A}$ .
2. A pointed model  $\mathbf{S}$  is *invariant under  $\uparrow$ -communication* iff all “defeasible knowledge” is common defeasible knowledge, i.e. for all propositions  $P \subseteq S$  and all agents  $a, b \in \mathcal{A}$ ,  $\Box_a P$  holds in  $\mathbf{S}$  iff  $\Box_b P$  holds in  $\mathbf{S}$ ; equivalently: iff all strong beliefs (conditional beliefs) are common strong beliefs (common conditional beliefs); equivalently: iff all agents' plausibility relations coincide:  $\leq_a = \leq_b$  for all  $a, b \in \mathcal{A}$ .
3. A pointed model  $\mathbf{S}$  is *invariant under  $\uparrow$ -communication* iff all (simple) beliefs are common beliefs, i.e. for all propositions  $P \subseteq S$  and all agents  $a, b \in \mathcal{A}$ ,  $B_a P$  holds in  $\mathbf{S}$  iff  $B_b P$  holds in  $\mathbf{S}$ ; equivalently: iff all agents' “best alternative” relations coincide:  $\rightarrow_a = \rightarrow_b$  for all  $a, b \in \mathcal{A}$ .

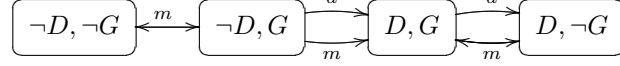
**Example 3** Suppose that in the situation in Example 1 above, a *trusted, infallible source publicly announces* that Albert is drunk: this is “hard”, incontrovertible information, corresponding to a *joint update*  $!D$ . The updated model is



After the update, Albert starts to wrongly believe that  $\neg G$  is the case! This is an example of *true but un-safe belief*: it can be lost after acquiring (new) true information.

**Example 4** Consider again the situation in example 3, but instead of Albert of receiving the information from an infallible source, he receives the information from Mary. Mary announces publicly (to Albert) that  $D$  is the case and we assume that Mary's announcement is both *sincere* and *persuasive*: she tells what she thinks and she convinces Albert. Since Mary is a fallible agent (and not an infallible source), this announcement is *soft*: in principle, she could be wrong, or she could lie, or she could simply guess and be right only by chance. So we cannot interpret Mary's announcement as a “hard” update  $!D$ , since such an announcement wouldn't be sincere: the update  $!D$  would automatically change Mary's order (making her irrevocably know

$D$ , when she didn't know it before!). But we *can model it as a “soft” announcement*  $\uparrow D$ ; i.e. after hearing it, all agents upgrade with  $D$ : they *start to prefer any  $D$ -world to any  $\neg D$ -world*. The upgraded model is



Note that *Mary's order is left unchanged*, so the announcement *was indeed sincere*.

**Example 5** What if instead Mary announces that she “knows” that Albert is drunk? If we take this in the sense of irrevocable knowledge  $K$ , then such an announcement would *not* be sincere: indeed, in the original situation of Example 1,  $K_m D$  was false. However, she *did “know” it* in the sense of *defeasible knowledge*  $\Box_m D$ : she correctly believed  $D$ , and this belief was safe. This “knowledge” was fallible, and she was aware of this: she didn't believe that she knows irrevocably ( $\neg B_m K_m D$ ), but she believed that she “knows” defeasibly ( $B_m \Box_m D$ ). Hence, she is sincere if she announces that she “knows” in this sense. Assuming that Albert is also aware of the fallibility of her knowledge, but that he still highly trusts her to be right, we can interpret this as a sincere and persuasive announcement of the form  $\uparrow (\Box D)$ . Its effect is the same as in Example 4: the upgraded model is the same.

**Counterexample 6** Note that *simply announcing that Mary believes  $D$ , or even that she strongly believes  $D$ , won't do*: this *will not be persuasive*, since it will not change Albert's beliefs about the facts of the matter ( $D$  or  $\neg D$ ), although it may change his beliefs about her beliefs. *Being informed of another's beliefs is not enough to convince you of their truth*. Indeed, Mary's beliefs are already common knowledge in the initial model of Example 1: so an upgrade  $\uparrow (B_m D)$  would be superfluous!

**Persuasiveness** So what is needed for persuasive communication is that the speaker (Mary) “converts” the others to her own beliefs. For this, she should *not* simply announce that she believes them. Instead, she can either announce that something *is* the case (when in fact she *just strongly believes* that it is the case), or else announce that she *defeasibly “knows” it* (when she *only believes that she “knows” it*, and in fact this implies that she strongly believes that she “knows”).

## 4 Merge Operations

A *merge operation*, or “aggregation procedure”, is an operator taking any sequence  $\{R_i\}_{1 \leq i \leq n}$  of preference relations into a “group preference” relation:

$$\bigodot_i R_i = R_1 \bigodot R_2 \bigodot \cdots R_n$$

In [2] the authors give a *general classification* of types of preference merge, in a very general context, subject to some minimal “*fairness*” and *rationality conditions*. They show that all the merge operations satisfying these conditions can be represented as compositions of *only two* basic merge operators: “parallel merge” and “lexicographic merge”.

**Parallel Merge** The merge operation we consider first can be thought of as the *most “democratic” form of aggregation*: everybody has a veto, so that group preferences

are unanimous preferences. Following [2], we call it *parallel merge*. It simply takes the merged relation to be the *intersection*  $\bigodot_a R_{a \in \mathcal{A}} := \bigcap_{a \in \mathcal{A}} R_a$  of all the preference relations of the agents in the given group  $\mathcal{A}$ .<sup>9</sup>

Parallel merge is particularly well suited for aggregating the agents' "hard information" (irrevocable knowledge)  $K$ , i.e. for merging the epistemic relations  $\{\sim_a\}_{a \in \mathcal{A}}$ . Since if we consider absolutely certain and fully introspective knowledge, there is no danger of introducing an inconsistency. The agents can pool their information in a *completely symmetric manner, accepting the other's bits without reservations*. In fact, parallel merge of the agents' irrevocable knowledge gives us the standard concept of "*distributed knowledge*":

$$DKP = [\bigcap_{a \in \mathcal{A}} R_a]P.$$

This matches to the notion of distributed knowledge that was first introduced in the work of Halpern and Moses [25]. In [20] the authors say that a group has distributed knowledge of  $P$  if roughly speaking the agents' combined knowledge implies  $P$ . Or "distributed knowledge can be viewed as what a wise man - one who has complete knowledge of what every member of the group knows - would know" [20, p3].

**Lexicographic Merge** When the group is hierarchically structured according to some total order (on agents), called a "*priority order*", then the agents with higher priority are thought of as having a higher "epistemic expertise" than the agents with lower priority. For a group  $\mathcal{A} = \{a, b\}$  of two agents, in which  $a$  has higher priority, we can think of  $a$  as the "expert" (or the professor) and of  $b$  as the "layman" (or the student). In this context, the natural doxastic merge operation is the so-called *lexicographic merge*. For two agents  $a, b$ , the "lexicographic merge"  $R_{a/b}$  gives priority to agent  $a$ 's strong (i.e. strict) preferences over  $b$ 's: first, *the strict preference order of  $a$  is adopted by the group*; and *when  $a$  is indifferent between two options, then  $b$ 's preference is adopted*; finally,  $a$ -incomparability gives group incomparability. Formally:

$$R_{a/b} := R_a^< \cup (R_a^= \cap R_b) = R_a^< \cup (R_a \cap R_b) = R_a \cap (R_a^< \cup R_b).$$

The lexicographic merge is *particularly suited for aggregating "soft information"* (strong beliefs, safe beliefs, conditional beliefs) *in the absence of any hard information*: since soft information is not fully reliable (because of lack of negative introspection for  $\Box$ , and because of potential falsity for belief, conditional belief and strong belief), some "screening" must be applied to some agents' information (and so some hierarchy must be enforced), in order to ensure consistency of the merge.

**(Relative) Priority Merge** Note that, *in lexicographic merge, the first agent's priority is "absolute"*. But *in the presence of "hard" information*, the lexicographic merge of soft information must be modified as follows: one first needs to pool together all the hard information and then use it to restrict the lexicographic merge of soft information. This leads us to a "more democratic" combination of both Parallel Merge and Lexicographic Merge, called the "*(relative) priority merge*"  $R_{a \otimes b}$ :

---

<sup>9</sup>From a purely formal perspective, parallel merge resembles the so-called "non-prioritized belief revision" (see e.g. [26]). But note that "merge" is *not* "revision"!

$$R_{a \otimes b} := (R_a^< \cap R_b^<) \cup (R_a^{\cong} \cap R_b^{\cong}) = R_a \cap R_b^< \cap (R_a^< \cup R_b^<).$$

In a Relative Priority Merge, both agents have a “veto” with respect to group incomparability. Here the group can only compare options that *both* agents can compare; and whenever the group can compare two options, everything goes on as in the lexicographic merge. Agent  $a$ ’s order gets priority, while  $b$ ’s order is adopted only when  $a$  is indifferent.

Since our plausibility structures they encode both the “hard” and the “soft” information possessed by the agent, it seems that Priority Merge is best suited for aggregating the agents’ plausibility relations.

**Example 7:** If in Example 1, we give priority to Mary, the relative priority merge  $R_{m \otimes a}$  of Mary’s and Albert’s original plausibility orders amounts to:



If instead we give priority to Albert, we simply obtain Albert’s order as our “merge”:

$$R_{a \otimes m} = R_a.$$

It is important to note that in *both* cases of Example 7, some of the resulting *joint beliefs are wrong*: when giving priority to Mary, both agents end up believing  $\neg G$ ; while if we give priority to Albert, they both end up believing  $\neg D$ . In fact, *no type of hierarchic belief merge is a warranty of veracity*.

## 5 “Realizing” Merge Dynamically

Intuitively, the *purpose of sharing* hard knowledge, defeasible knowledge or beliefs is to achieve a state in which *there is nothing else to share*, i.e. one in which any further sharing is redundant: all hard knowledge, or defeasible knowledge, or beliefs, are already shared in common. For sharing via a specific type of public communication  $\dagger \in \{!, \uparrow, \uparrow\}$ , this happens precisely when the model-changing process induced by  $\dagger$ -type sharing *reaches a fixed point of  $\dagger$ -communication*: a model that is invariant under that particular type of announcements.

For every specific type of public communication  $\dagger \in \{!, \uparrow, \uparrow\}$ , agent  $a$ ’s “*relevant structure*” in a model  $\mathbf{S}$  is given by:  $a$ ’s *epistemic* relation  $\sim \subseteq S \times S$  in the case of *updates*  $!$ ;  $a$ ’s *plausibility* relation  $\leq_a$  in the case of *radical upgrade*  $\uparrow$ ; and  $a$ ’s *doxastic* “best alternative” relations  $\rightarrow_a$  in the case of conservative upgrade.

**Upgrade sequences** We define a (finite)  $\dagger$ -*upgrade sequence* to be a finite sequence  $\dagger \vec{P} = (\dagger P^1, \dots, \dagger P^n)$  of upgrades  $\dagger P^i$  of the given type  $\dagger \in \{!, \uparrow, \uparrow\}$ . Any  $\dagger$ -upgrade sequence induces a (partial) function, mapping every pointed model  $\mathbf{S}$  into a finite sequence  $\dagger \vec{P}(\mathbf{S}) = (\mathbf{S}^i)_i$  of *pointed models*, defined inductively by:  $\mathbf{S}^0 := \mathbf{S}$ ; and  $\mathbf{S}^{i+1} := \dagger P^i(\mathbf{S}^i)$ , if this is defined (and undefined otherwise). A  $\dagger$ -upgrade sequence  $\dagger \vec{P}$  is a *sincere  $\dagger$ -communication sequence* if all its upgrades are sincere for at least

one agent at the moment of speaking: i.e. for every  $i \leq n$  there exists  $a_i \in \mathcal{A}$  such that  $\dagger P^i$  is sincere for  $a_i$  on  $\mathbf{S}^i$ .

A sincere  $\dagger$ -communication sequence  $\dagger \vec{P}$  is said to be *exhaustive* on a model  $\mathbf{S}$  if the last model  $\mathbf{S}^n$  of the induced sequence  $\dagger \vec{P}(\mathbf{S})$  is invariant under (sincere)  $\dagger$ -communication; equivalently, iff it is *maximal*: it cannot be extended to any longer  $\dagger$ -communication sequence. By Proposition 2, the last model  $\mathbf{S}^n$  generated by an exhaustive  $\dagger$ -communication sequence is one in which *all the agents' "relevant structures"  $R_a^n$  coincide*.

**Realization of a Merge Operation** A sincere and exhaustive  $\dagger$ -communication sequence *realizes a given preference merge operation*  $\otimes$  on a given pointed model  $\mathbf{S}$  if, for every agent  $b \in \mathcal{A}$ , the relevant structure  $R_b^n$  in the last generated model is the  $\otimes$ -merge of the initial relevant structures  $\{R_a^0\}_{a \in \mathcal{A}}$ : i.e.  $R_b^n = \otimes_{a \in \mathcal{A}} R_a^0$ , for all  $b \in \mathcal{A}$ . A merge operation  $\otimes$  is *realizable by  $\dagger$ -communication* if there exists some sincere exhaustive  $\dagger$ -communication sequence that realizes  $\otimes$ . The merge operation is said to be *constructively realizable by  $\dagger$ -communication* if *there exists a protocol* such that every  $\dagger$ -communication sequence that complies with the protocol is exhaustive and realizes  $\otimes$ .

For each of the above types of public communication ( $!$ ,  $\uparrow$ ,  $\dagger$ ), we can ask which merge operations are realizable, or constructively realizable. The answer depends on the constraints (*transitivity, connectedness* etc.) satisfied by the agents' relevant structures (epistemic, doxastic or plausibility relations).

**Proposition 3** *Parallel merge is the only merge operation that is realizable by updates (i.e. by  $!$ -communication). Moreover, parallel merge is constructively realizable by updates. The protocol is as follows: in no particular order, the agents have to publicly announce "all that they know" (in the sense of irrevocable knowledge  $K$ ). More precisely, for each set of states  $P \subseteq S$  such that  $P$  is known to a given agent  $a$  at the moment when it's her turn to speak, a public announcement  $!P$  is made.*

This essentially is the protocol in van Benthem's paper "One is a Lonely Number" [14]. Formally, the protocol consists of  $n$  steps (where  $n$  is the number of agents in  $\mathcal{A}$ ), where each step is itself a sequence of announcements by the same agent: in the first step, one of the agents, say  $a$ , announces "all he knows". This is the sequence of announcements:

$$\sigma_a := \prod \{!P : P \subseteq S \text{ such that } s \models K_a P\}$$

(where  $\prod$  is sequential composition of actions). Then, another agent  $b$  performs a similar step  $\sigma_b$

$$\sigma_b := \prod \{!P : P \subseteq S \text{ such that } s \models [\sigma_a] K_b P\}$$

(announcing all she knows *after the first step*); etc.

**Important Observations:** (1) *The order in which the agents make the announcements doesn't actually matter. The agents may even "interrupt" each other: any exhaustive  $!$ -communication sequence produces the same result.* (2) *The protocol can be simplified by restricting it only to knowledge announcements, i.e. of the form  $!(K_a P)$  (for each  $P$  such that  $K_a P$  holds): instead of announcing all they know, the agents announce that they know all that they know.* (3) *The protocol can be simplified by allowing each agent to make only one announcement: instead of successively announcing everything he knows, he can just announce the conjunction  $!(\bigwedge \{P : \mathbf{S} \models K_a P\})$*



of all the things he knows; or, equivalently and more efficiently he can announce the proposition  $!s(a)$  in the current state  $s$ .

**Proposition 4** *Assume given a model in which all knowledge is public knowledge: there is no private “hard” information (i.e. the agents’ epistemic relations  $\sim_a$  are the same). Then, for every given priority order  $(a_1, \dots, a_n)$  on agents, the corresponding Lexicographic Merge (of plausibility relations) is constructively realizable by radical upgrades (i.e. by  $\uparrow$ -communication) on this mode. The protocol is a natural modification of the previous one: following the priority order, the agents have to publicly announce “all that they strongly believe”. More precisely, a joint radical upgrade  $\uparrow P$  is performed, for each set  $P \subseteq S$  such that  $P$  is strongly believed by the given agent  $a$  at the moment when it’s her turn to speak.*

Formally, the protocol consists again of  $n$  steps, each step being a sequence of announcements by the same agent: first, the first agent (according to the priority order), say  $a$ , announces all that he strongly believes. This is the sequence of radical upgrades:

$$\rho_a := \prod \{\uparrow P : P \subseteq S \text{ such that } \mathbf{S} \models Sb_a P\}.$$

Then, the next agent in the hierarchy, say  $b$ , performs a similar step

$$\rho_b := \prod \{\uparrow P : P \subseteq S \text{ such that } \mathbf{S} \models [\rho_a]Sb_b P\}$$

(announcing all she strongly believes *after* the first step); etc.

**Important Observations:** (0) This protocol is operational only on a restricted class of models in which the agents hold no different “hard information” (1) *In this protocol the order of the announcements matters:* the agents have to respect the priority order. Moreover, *no interruptions are allowed:* agents with lower priority can speak *only after the agents with higher priority finished* announcing all their strong beliefs. Any interruptions may lead to the realization of *complete different merge operations* (see the Counterexample below)! So lexicographic merge is *not* the only merge operation realizable by radical upgrades. (2) This protocol can also be simplified by restricting it only to “*defeasible knowledge*” announcements, i.e. announcements of the form  $\uparrow (\Box_a P)$ . But recall that, unlike irrevocable knowledge, defeasible knowledge is not negatively introspective: so the agents don’t know for sure what things they “know” and what not, and hence the best they can do is to announce all the things they *believe they “know”*. But, since believing to (indefeasibly) “know” is the same as believing, they have to *announce that they “know”  $P$ , for each proposition  $P$  which they believe*. So the simplified protocol replaces e.g. the first step by the following sequence of radical upgrades

$$\rho'_a := \prod \{\uparrow (\Box_a P) : P \subseteq S \text{ such that } \mathbf{S} \models B_a P\}.$$

(3) Unlike the case of upgrades and parallel merge, in general the above protocol *actually requires multiple announcements by the same agents, including announcing facts that may already be entailed by their previous announcements!* A sequence of radical upgrades is in general *not* equivalent to a radical upgrade, so there is no way to compress the sequences  $\rho_a$  or  $\rho'_a$  into a single upgrade!

**Proposition 5** *For every given priority order  $(a_1, \dots, a_n)$  on agents, the corresponding (relative) Priority Merge (of plausibility relations) is constructively realizable by a*

*series of updates and radical upgrades*. The protocol is a concatenation of the previous two protocols (for realizing parallel merge and lexicographic merge). More precisely, first the agents have to publicly announce *all that they know*: for each set  $P \subseteq S$  for which  $KP$  is the case, a joint *update*  $!P$  is performed; and next, *following the priority order, the agents have to publicly announce “all that they strongly believe”*: i.e. after the updates, the agents perform joint *radical upgrades*  $\uparrow P$ , for each set  $P \subseteq S$  such that  $P$  is *strongly believed by the given agent at that time*.

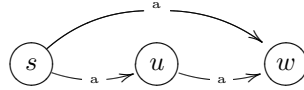
**Important Observations:** (1) *As in the case of lexicographic merge, the order of the announcements matters in the second part of the protocol to realize the priority merge:* the agents have to respect the priority order. (2) This protocol can be *simplified* in two ways: first, by replacing (as in the case of lexicographic merge) the “strong belief” announcements by “*defeasible knowledge*” announcements, i.e. performing steps  $\rho'_a$  consisting of a series of announcements of the form  $\uparrow (\Box_a P)$ ; second, by concatenating first, for each agent, the corresponding steps  $\sigma_a; \rho'_a$  of the two protocols. In this way, the combined protocol becomes much simpler: following the priority order  $a, b, \dots$ , the first agent  $a$  performs  $\sigma_a$  followed by  $\rho'_a$ , then the second agent performs  $\sigma_b$  followed by  $\rho'_b$  etc, until there is no more information to be shared in this way. An illustration of this way to realize the priority merge is given in Example 8. (3) When priority merge has been realized there is no private information left behind, hence all agents’ knowledge has become common knowledge and all their beliefs, strong beliefs, safe beliefs, conditional beliefs etc are common.

**Example 8** Recall the initial order of Mary and Albert in Example 1. Consider the protocol:

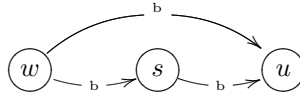
$$\uparrow \Box_m D; (!K_a(D \vee G)); \uparrow \Box_a \neg G$$

The first is a sincere announcement by Mary, the rest are sincere announcements by Albert. This communication sequence yields the model presented in Example 7, as the result of the priority merge  $R_{m \otimes a}$  of the two plausibility orders.

**Counterexample 9** To show the *non-uniqueness* of priority merge among  $\uparrow$ -realizable merge operations and the *order-dependency* of the above protocol, note first that the priority merge of the ordering



with the ordering

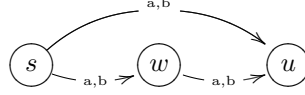


is equal to either of the two orders (depending on which agent has priority). But consider now the following public dialogue

$$\uparrow \Box_b u \cdot \uparrow \Box_a (u \vee w)$$

This first is a *sincere announcement* by  $b$ , the second is *sincere announcement* by  $a$ . This is an exhaustive  $\uparrow$ -communication sequence, but note that the strict priority order required by the above protocol is *not* respected here: the first speaker  $b$  is

“interrupted” by the second speaker  $a$  *before* she finished announcing all his strong beliefs. (Indeed,  $s \vee u$  is also a strong belief of agent  $b$ , though one that is entailed by the first announcement; nevertheless,  $b$  should have first announced this second strong belief before  $a$  would have been allowed to speak!) And, indeed, the resulting model, though a fixed point of  $\uparrow$ -communication (since all the plausibility relations come to coincide), *realizes a different merge operation than either of the two priority merges*:



**The Power of Agendas** This order-dependence illustrates a phenomenon well-known in Social Choice Theory: the *important role of the person who “sets the agenda”*: the “Judge” who assigns *priorities to witnesses’ stands*; the chairman or moderator who determines the *order of the speakers* in a meeting, as well as the *issues* to be discussed and the *relative priority of each issue*.

**Proposition 6** *For every given priority order  $(a_1, \dots, a_n)$  on agents, the corresponding priority merge of both (hard) knowledge and doxastic “best alternatives” relations  $\{\rightarrow_a\}_a$  is constructively realizable by both updates and conservative upgrades (i.e. by a combination of ! and  $\uparrow$ -communication). The protocol is the natural modification of the previous one: First all the agents have to publicly announce first all they know (in the sense of  $K$ ); and then following the priority order, they announce “all that they (simply) believe”. More precisely after the updates, for each set of states  $P \subseteq S$  such that  $P$  is believed by the given agent  $a$ , a joint conservative upgrade  $\uparrow P$  is performed.*

Similar observations as the ones following Proposition 5 apply to the case of doxastic conservative upgrades: the priority merge is *not* the only realizable merge operation; the order of announcements *does* matter; in general, the protocol *may* require *multiple announcements by the same agents*.

## 6 Conclusion

In this paper, we focused on *dynamically realizing by public communication three specific forms of merge of (hard and soft) information*: parallel merge, lexicographic merge and priority merge. But, as we saw, depending on the “agenda”, soft announcements *can realize a whole plethora of merge operations*. Nevertheless, *not everything goes*: the requirements imposed on the plausibility relations generally pose restrictions to which kinds of merge are realizable. This raises an important open question: *characterize the class of merge operations realizable by radical upgrades (or by combinations of radical upgrades and updates, or by conservative upgrades etc).*

In on-going work, first presented at the workshop on “Formal theories of communication” held in February 2010 at the Lorentz center in the Netherlands, we show that the setting presented in this paper can be extended by making *explicit* the agents’ *doxastic attitudes towards the reliability of the incoming information* [11]. In that context, one can formally define the “honesty” of a communication act by requiring that the speaker has *the same attitude* towards the announced information that (s)he expects to induce in the listeners. For example, assume that it is common knowledge that a given speaker is considered to be infallible and that all agents consider the pub-

lic message coming from this source to be necessarily true. In this case the speaker is “honest” if it is indeed the case that he himself knows (in the irrevocable sense) that the information is true. If on the other hand the information coming from this source is strongly (resp. simply) believed by the listeners, but *not known* to be true, then the speaker is “honest” if he himself strongly (resp. simply) believes the information. The details of this more sophisticated formalization are left for future work.

## 7 Acknowledgements

The authors thank the organizers of FAMAS-MALLOW 2009 for creating the platform at which this work was first presented. The second author acknowledges that this work is part of the VIDI research programme number 639.072.904 of the Netherlands Organization for Scientific Research. In the final writing stage this paper benefited from the second author’s received funding from the European Research Council under the European Community’s Seventh Framework Programme (FP7/2007-2013) with ERC Grant agreement nr 283963.

## References

- [1] C.E. Alchourron, P. Gardenfors and D. Makinson. “On the logic of theory change: partial meet contraction and revision functions”, *Journal of Symbolic Logic*, 50(2):510530, 1985.
- [2] H. Andreka, M. Ryan and P-Y. Schobbens “Operators and Laws for Combining Preference Relations”, *Journal of Logic and Computation*, 12(1), 13–53, 2002.
- [3] K.J. Arrow, “A Difficulty in the Concept of Social Welfare”, *Journal of Political Economy*, 58(4), 328–346, 1950.
- [4] A. Baltag and S. Smets, “Conditional doxastic models: a qualitative approach to dynamic belief revision”, *Electronic Notes in Theoretical Computer Science*, 165, 5–21, 2006.
- [5] A. Baltag and S. Smets, “The Logic of Conditional Doxastic Actions: A Theory of dynamic multi-agent belief revision”, in S. Artemov and R. Parikh (eds.), *Proceedings of the Workshop on Rationality and Knowledge*, 13–30, ESSLLI 2006.
- [6] A. Baltag and S. Smets, “Dynamic Belief Revision over Multi-Agent Plausibility Models”, in G. Bonanno, W. van der Hoek, M. Wooldridge (eds.), *Proceedings of the 7th Conference on Logic and the Foundations of Game and Decision (LOFT 2006)*, 11–24, University of Liverpool, 2006.
- [7] A. Baltag and S. Smets, *Probabilistic Dynamic Belief Revision*, in J. van Benthem and S. Ju and F. Veltman (eds.), *Proceedings of LORI’07*, College Publications London, 21–39, 2007.
- [8] A. Baltag and S. Smets, “A Qualitative Theory of Dynamic Interactive Belief Revision”, in G. Bonanno, W. van der Hoek, M. Wooldridge (eds.), *Logic and the Foundations of Game and Decision Theory, Texts in Logic and Games*, 3, 9–58, Amsterdam University Press, 2008.
- [9] A. Baltag and S. Smets, “The Logic of Conditional Doxastic Actions”, in R. van Rooij and K. Apt (eds.), *New Perspectives on Games and Interaction, Texts in Logic and Games*, 4, 9–31, Amsterdam University Press, 2008.
- [10] A. Baltag and S. Smets, “Talking Your Way into Agreement: Belief Merge by Persuasive Communication”. CEUR Workshop Proceedings, Vol. 494, Proceedings of the Second Multi- Agent Logics, Languages, and Organisations Federated Workshops Turin, Italy, September 7-10, pp. 129 - 141. 2009
- [11] A. Baltag and S. Smets, “Doxastic Attitudes and Group Communication” and “Merging beliefs by honest persuasive communication”, Presentations held at the workshop on “Formal theories of communication” at the Lorentz center in the Netherlands, February 2010. <http://www.lorentzcenter.nl/lc/web/2010/381/participants.php3?wsid=381>
- [12] P. Battigalli and M. Siniscalchi, “Strong Belief and Forward Induction Reasoning”, *Journal of Economic Theory*, 105, 356–391, 2002.

- [13] C. Boutilier, “Iterated Revision and Minimal Change of Conditional Beliefs”, *JPL*, 25(3), 262–305, 1996.
- [14] J.F.A.K. van Benthem, “One is a lonely number”. In P. Koepke Z. Chatzidakis and W. Pohlers, (eds.) *Logic Colloquium 2002*, 96–129, ASL and A.K. Peters, Wellesley MA, 2006.
- [15] J.F.A.K. van Benthem, “Dynamic logic of belief revision”, *JANCL*, 17 (2), 129–155, 2007.
- [16] J.F.A.K. van Benthem, “Priority Product Update as Social Choice” (Expanded version), Unpublished Manuscript, November 2007
- [17] J.F.A.K. van Benthem, *Logical Dynamics of Information and Interaction*, Cambridge Univ. Press, 2010.
- [18] J.F.A.K. van Benthem and F. Liu, “Dynamic logic of preference upgrade”, *Journal of Applied Non-Classical Logics*, University of Amsterdam, 17 (2), 157–182, 2007.
- [19] O. Board, “Dynamic interactive epistemology”, *Games and Economic Behaviour*, 49, 49–80, 2002.
- [20] R. Fagin and J. Halpern, Y. Moses, M. Vardi, *Reasoning about Knowledge*, MIT Press, 1995.
- [21] N. Friedmann and J.Y. Halpern, “Conditional logics of belief revision”, *Proc. of 12th National Conference in Artificial Intelligence*, AAAI Press, Menlo Park, CA, 915–921, 1994.
- [22] P. Gochet and P. Gribomont, “Epistemic Logic”, D.M. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, Elsevier, 7, 99–195, 2006.
- [23] A. Grove, “Two modellings for theory change”, *Journal of Philosophical Logic*, v17, 157–170, 1988.
- [24] J.Y. Halpern, *Reasoning about Uncertainty*, MIT Press, Cambridge MA, 2003.
- [25] J.Y. Halpern and Y. Moses. “Knowledge and common knowledge in a distributed environment”, *Journal of the ACM*, 37(3):549–587, 1990.
- [26] S.O. Hansson. “A Survey of Non-Prioritized Belief Revision”, *Erkenntnis*, 50: 413–47, 1999.
- [27] H. Katsuno and A. Mendelzon, “On the difference between updating a knowledge base and revising it”, *Cambridge Tracts in Theoretical Computer Science*, 183–203, 1992.
- [28] J.-J.Ch. Meyer and W. van der Hoek, *Epistemic Logic for AI and Computer Science*, *Cambridge Tracts in Theoretical Computer Science*, 41, Cambridge University Press, Cambridge, 1995.
- [29] H. Rott, “Conditionals and theory change: revisions, expansions, and additions” in *Synthese*, 81(1), 91–113, 1989.
- [30] W. Spohn, “Ordinal conditional functions: a dynamic theory of epistemic states”, in W.L. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, vol. II, 105–134, 1988.
- [31] R. Stalnaker, “On Logics of Knowledge and Belief”, *Philosophical Studies*, vol. 128, 169–199, 2006.
- [32] R. Stalnaker, “Knowledge, Belief and Counterfactual Reasoning in Games”, *Economics and Philosophy*, vol. 12, 133–163, 1996.
- [33] W. van der Hoek, “Systems for knowledge and beliefs”, *Journal of Logic and Computation*, 3, nr. 2, 173–195, 1993.
- [34] F. Veltman, “Defaults in Update Semantics”, *Journal of Philosophical Logic*, 25, 221–261, 1996.
- [35] P. Blackburn, M. de Rijke and Y. Venema, *Modal Logic*, Cambridge University Press, 2002.
- [36] F.P.J.M. Voorbraak, *As Far as I Know*, Utrecht University, Utrecht, NL, *Questiones Inertiae* volume VII, 1993.
- [37] T. Williamson, “Some philosophical aspects of reasoning about knowledge”, *Proceedings of TARK’01*, J. van Benthem (ed.), 97–97, Morgan Kaufmann Publishers, San Francisco, 2001.

Received 22 January 2011.