Agreeing to Agree: Reaching Unanimity via Preference Dynamics Based on Reliable Agents

Sujata Ghosh
Indian Statistical Institute
Chennai, India
sujata@isichennai.res.in

Fernando R. Velázquez-Quesada
Grupo de Lógica, Lenguaje e Información
Universidad de Sevilla, Spain
FRVelazquezQuesada@us.es

ABSTRACT

Situations akin to public deliberation leading to preference changes are modelled. A set of agents is considered, each endowed with a preference relation over a set of objects and a reliability relation over the involved agents. Different ways in which the public announcement of the current individual preferences can influence the agents' future preferences are studied. Special emphasis is given to ways in which the repetitive public announcement of the individual preferences lead to a unanimity on preferences.

Categories and Subject Descriptors

I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal logic

Keywords

preference, preference change, lexicographic, unanimity

1. INTRODUCTION

There are two important aspects of any democratic decision: aggregation and deliberation. Till now, there has been a lot of work on the 'aggregation' aspect: *'when a group needs to make a decision, we are faced with the problem of aggregating the views of the individual members of that group into a single collective view that adequately reflects the "will of the people"'* [6]. These kinds of situations are studied not only in social choice theory [1], but also on preference/belief change/merge/aggregation (e.g., [13, 18, 9]).

On the other hand, some authors (e.g., [5, 14] and others) have questioned the epistemic values of the aggregation process, typically achieved by a voting procedure which considers the individual preferences but not their origins. They point to the merits of the deliberative process, which makes people reflect on their preferences and thus influences possible changes. Indeed, while deliberation can be thought of as a combination of evidences and reasons, an aggregation process as voting merely considers the opinions they result in. In fact, [5] goes so far as to say that, *'there would not be any need for an aggregation mechanism, since a rational decision would tend to produce unanimous preferences'*. The process of deliberation is thus an important aspect of group decision making: when it leads to unanimity, there is no need to consider some (possibly artificial) aggregation process. Only if unanimity cannot be reached via deliberation, one can then resort to the process of aggregation.

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*
Copyright © 2015, International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

The present work contributes to the formal study of this deliberative perspective. Rather than studying methods through which the agents' individual views can be aggregated to obtain the group's views, it studies the way these individual preferences may change towards reaching unanimity, thus recreating a form of public deliberation. Individual preferences can change for different reasons, and this work focuses on the changes that result from the (public) announcement of such preferences and the *reliability* each agent has on one another (e.g., an agent adopts the preferences of the people she relies on the most). This notion of agent reliability can be understood as some form of *trust* among agents, a concept that has been studied in, e.g., [19, 16] (see the discussion later on reliability). Of course, one can consider somebody else extremely reliable, yet not share any of his/her preferences at all. But, under certain circumstances, agents do choose to revise their preferences based on the preferences of those on whom they rely the most. For example, if one visits a new country, one would rely on local friends about restaurant preferences, even though one might already have some preferences based on `http://www.tripadvisor.co.uk`, say. The following example, to which we will come back while setting up the formal framework, shows a situation in which it benefits to reach an agreement through deliberation.

Example Consider three colleagues Alan (a), Barbara (b) and Chiara (c) who want to order food for an office party in one of the three restaurants of the neighbourhood. Alan is not a foodie and feels that Chiara knows better about it than Barbara, whereas Chiara feels that she has ample knowledge of the local restaurants, and does not rely on Alan's opinion at all. Barbara once had a good experience based on Alan's suggestion, and hence relies on him the most. In this situation, if they come to know about one another's preferences, can they reach an agreement?

2. BASIC DEFINITIONS

The focus of this work is public deliberation, so let Ag be a *finite non-empty* set of agents with $|Ag| = n \geq 2$ (if $n = 1$, there is no scope for joint discussion). Below we present the most important definitions of this framework.

Definition 1 (*PR* frame). *A* preference and reliability (*PR*) frame F *is a tuple* $\langle W, \{\leq_i, \preccurlyeq_i\}_{i \in Ag}\rangle$ *where*

- W *is a finite non-empty set of worlds;*
- $\leq_i \subseteq (W \times W)$ *is a total preorder (a total, reflexive and transitive relation), agent* i*'s* preference relation *over worlds in* W *(*$u \leq_i v$ *is read as "world* v *is at least as preferable as world* u *for agent* i*");*
- $\preccurlyeq_i \subseteq (Ag \times Ag)$ *is a total order (a total, reflexive, transitive and antisymmetric relation), agent* i*'s* reliability relation *over agents in* Ag *(*$j \preccurlyeq_i j'$ *is read as "agent* j' *is at least as reliable as agent* j *for agent* i*").*

The assumptions about the preference and the reliability relations will be discussed in Subsection 2.1. For now, here are further useful definitions.

Definition 2. *Let* $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i \in Ag}\rangle$ *be a PR frame.*

- $u <_i v$ *("*u *is less preferred than* v *for agent* i*") iff*$_{def}$ $u \leq_i v$ *and* $v \not\leq_i u$.
- $u \simeq_i v$ *("*u *and* v *are equally preferred for agent* i*") iff*$_{def}$ $u \leq_i v$ *and* $v \leq_i u$.
- $j \prec_i j'$ *("*j *is less reliable than* j' *for agent* i*") iff*$_{def}$ $j \preccurlyeq_i j'$ *and* $j' \not\preccurlyeq_i j$.
- $\mathrm{mr}(i) = j$ *iff*$_{def}$ $j' \preccurlyeq_i j$ *for every* $j' \in Ag$.

By denoting with w_i the world where 'restaurant i is the best' ($i = 1, 2, 3$), the example's situation can be represented by a *PR* frame $F_{exp} = \langle \{w_1, w_2, w_3\}, \{\leq_i, \preccurlyeq_i\}_{i \in Ag}\rangle$ in which $\leq_a$: $w_1 <_a w_2 <_a w_3$, $\leq_b$: $w_3 <_b w_2 <_b w_1$ and $\leq_c$: $w_2 <_c w_1 <_c w_3$, and also $\preccurlyeq_a$: $a \prec_a b \prec_a c$, $\preccurlyeq_b$: $b \prec_b c \prec_b a$ and $\preccurlyeq_c$: $a \prec_c b \prec_c c$.

2.1 A formal language

Throughout this paper, let At be a countable set of atomic propositions.

Definition 3 (Language). *Formulas* φ, ψ *and relational expressions* π, σ *of the language* $\mathcal{L}$ *are given by*

$$\varphi, \psi ::= p \mid j \sqsubseteq_i j' \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\pi\rangle\,\varphi$$
$$\pi, \sigma ::= 1 \mid \leq_i \mid \geq_i \mid -\pi \mid \pi \cup \sigma \mid \pi \cap \sigma$$

with $p \in At$ *and* $i, j, j' \in Ag$. *Propositional constants* ($\top, \bot$), *other Boolean connectives* ($\wedge, \rightarrow, \leftrightarrow$) *and the dual modal universal operators* $[\pi]$ *are defined as usual* ($[\pi]\,\varphi := \neg\langle\pi\rangle\,\neg\varphi$ *for the latter). Define also, for any relational expression* π, *the operator* $\boxed{\pi}$ *as* $\boxed{\pi}\,\varphi := [-\pi]\,\neg\varphi$. *All these abbreviations will facilitate both the writing of formulas as well as the presentation of the axiom system.*

The set of formulas of $\mathcal{L}$ contains atomic propositions (p) and formulas describing the agents' reliability relations ($j \sqsubseteq_i j'$), and it is closed under negation ($\neg$), disjunction ($\vee$) and modal operators of the form $\langle\pi\rangle$ with π a relational expression. The set of relational expressions contains the constant 1, the preference relations ($\leq_i$) and their respective converse ($\geq_i$; [3]), and it is closed under Boolean operations over relations (the so called *boolean modal logic*; [10]).

The following two definitions establish what a model is and how formulas of $\mathcal{L}$ are interpreted over them.

Definition 4 (*PR* model). *A PR* model M *is a tuple* $\langle F, V\rangle$ *where* F *is a PR frame and* $V : At \to \wp(W)$ *is a* valuation function. *A pair* (M, w) *with* M *a PR model and* w *a world in it is called a* pointed *PR* model.

Definition 5 (Semantic interpretation). *The truth definition of formulas in* $\mathcal{L}$ *at pointed PR models and the relation* R_π *for each relational expressions* π *are given by*

$$
\begin{array}{lll}
(M,w) \Vdash p & \textit{iff} & w \in V(p) \\
(M,w) \Vdash j \sqsubseteq_i j' & \textit{iff} & j \preccurlyeq_i j' \\
(M,w) \Vdash \neg\varphi & \textit{iff} & (M,w) \nVdash \varphi \\
(M,w) \Vdash \varphi \vee \psi & \textit{iff} & (M,w) \Vdash \varphi \ \textit{or}\ (M,w) \Vdash \psi \\
(M,w) \Vdash \langle\pi\rangle\,\varphi & \textit{iff} & \exists u \in W \textit{ s.t. } R_\pi wu \textit{ and } (M,u) \Vdash \varphi
\end{array}
$$

and

$$
\begin{array}{llll}
R_1 := W \times W & & R_{-\pi} := (W \times W) \setminus R_\pi \\
R_{\leq_i} := \leq_i & & R_{\pi\cup\sigma} := R_\pi \cup R_\sigma \\
R_{\geq_i} := \{(v,u) \mid u \leq_i v\} & & R_{\pi\cap\sigma} := R_\pi \cap R_\sigma
\end{array}
$$

As a consequence of the previous definition,

$$
\begin{array}{lll}
(M,w) \Vdash \langle 1\rangle\,\varphi & \text{iff} & \exists u \in W \text{ s.t. } (M,u) \Vdash \varphi \\
(M,w) \Vdash [\pi]\,\varphi & \text{iff} & \forall u \in W,\ R_\pi wu \text{ implies } (M,u) \Vdash \varphi \\
(M,w) \Vdash \boxed{\pi}\,\varphi & \text{iff} & \forall u \in W,\ (M,u) \Vdash \varphi \text{ implies } R_\pi wu
\end{array}
$$

Thus, while $\langle 1\rangle$ is the global existential modality, $\boxed{\pi}$ is the window operator [11]. A formula φ is true in a model M ($M \Vdash \varphi$) if $(M, w) \Vdash \varphi$ for all worlds w in M.

On the language's operators Relational expressions allow us to define modalities for relations that can be defined from $\leq_i$ (see Definition 2). For example,

$$
\simeq_i := \leq_i \cap \geq_i \qquad \not\geq_i := -\geq_i \qquad <_i := \leq_i \cap \not\geq_i
$$
$$
\not\leq_i := -\leq_i \qquad >_i := \not\leq_i \cap \geq_i
$$

and hence

$$
\langle\simeq_i\rangle\,\varphi := \langle\leq_i \cap \geq_i\rangle\,\varphi,\quad \langle\not\geq_i\rangle\,\varphi := \langle-\geq_i\rangle\,\varphi,\quad \langle<_i\rangle\,\varphi := \langle\leq_i \cap \not\geq_i\rangle\,\varphi,
$$
$$
\langle\not\leq_i\rangle\,\varphi := \langle-\leq_i\rangle\,\varphi,\quad \langle>_i\rangle\,\varphi := \langle\not\leq_i \cap \geq_i\rangle\,\varphi.
$$

These modalities will be useful not only for expressing the agents' preferences, but also for providing axioms for the preference upgrade operations of Section 3.

Consider propositions p_i ($i = 1, 2, 3$), read as 'restaurant i is the best'. Consider the *PR model* M_{exp}: $\langle F_{exp}, V\rangle$ where F_{exp} is as before and V is given by $V(p_i) = \{w_i\}$ for each i. This describes the introduction example aptly. We will see how the dynamics work, once we introduce the preference upgrade operations.

For an axiom system, we have the following.

Theorem 1. *Table 1 provides a sound and complete axiom system (with* i *any agent and* π *any relational expression) for* $\mathcal{L}$ *with respect to PR models.*

Proof. *The first five blocks of the axiom system are known to be sound and complete for the fragment of the language they take care of:* **(1)** *axioms and rules in the first block take care of propositional validities;* **(2)** *those in the second establish that every modality is normal;* **(3)** *axioms in the third state that* $\leq_i$ *is a reflexive, transitive and total relation (recall that* $\langle 1\rangle$ *is the global existential modality);* **(4)** *those in the fourth establish that* $\geq_i$ *is the converse of* $\leq_i$ *[3];* **(5)** *axioms and rules of the fifth block characterise validities involving relational expressions, in particular, those involving Boolean relational operations [10]. Showing that axioms in the sixth block characterise the total, reflexive, transitive and antisymmetric relation* $\preccurlyeq_i$ *is straightforward.* $\square$

$\vdash \varphi$ for every propositional tautology φ		From $\vdash \varphi$ and $\vdash \varphi \to \psi$ infer $\vdash \psi$	
$\vdash [\pi]\,(\varphi \to \psi) \to ([\pi]\,\varphi \to [\pi]\,\psi)$	*(K_π)*	From $\vdash \varphi$ infer $\vdash [\pi]\,\varphi$	*(N_π)*
$\vdash \varphi \to \langle \leq_i \rangle\,\varphi$	*($T_\leq$)*	$\vdash \langle \leq_i \rangle \langle \leq_i \rangle\,\varphi \to \langle \leq_i \rangle\,\varphi$	*($4_\leq$)*
$\vdash (\langle 1 \rangle\,\varphi \wedge \langle 1 \rangle\,\psi) \to (\langle 1 \rangle\,(\varphi \wedge \langle \leq_i \rangle\,\psi) \vee \langle 1 \rangle\,(\psi \wedge \langle \leq_i \rangle\,\varphi))$			*($totality_\leq$)*
$\vdash \varphi \to [\leq_i]\,\langle \geq_i \rangle\,\varphi$	*($Con1_\leq$)*	$\vdash \varphi \to [\geq_i]\,\langle \leq_i \rangle\,\varphi$	*($Con2_\leq$)*
$\vdash \varphi \to \langle 1 \rangle\,\varphi$	*(T_E)*	$\vdash \langle 1 \rangle \langle 1 \rangle\,\varphi \to \langle 1 \rangle\,\varphi$	*(4_E)*
$\vdash \varphi \to [1]\,\langle 1 \rangle\,\varphi$	*(5_E)*		
$\vdash [1]\,\varphi \leftrightarrow ([\pi]\,\varphi \wedge \boxed{\pi}\,\neg\varphi)$	*(1_1)*	$\vdash \boxed{1}\,\top$	*(1_2)*
$\vdash [-\pi]\,\varphi \leftrightarrow \boxed{\pi}\,\neg\varphi$	*($-_1$)*	$\vdash [\pi]\,\neg\varphi \leftrightarrow \boxed{-\pi}\,\varphi$	*($-_2$)*
$\vdash \langle \pi \cup \sigma \rangle\,\varphi \leftrightarrow (\langle \pi \rangle\,\varphi \vee \langle \sigma \rangle\,\varphi)$	*($\cup$)*	$\vdash \boxed{\pi \cap \sigma}\,\varphi \leftrightarrow (\boxed{\pi}\,\varphi \wedge \boxed{\sigma}\,\varphi)$	*($\cap$)*
From $\vdash [\pi]\,\varphi \to ([\sigma]\,\varphi \to [\rho]\,\varphi)$ infer $\vdash [\pi]\,\varphi \to (\boxed{\rho}\,\neg\varphi \to \boxed{\sigma}\,\neg\varphi)$			*(BR)*
$\vdash j \sqsubseteq_i j' \;\vee\; j' \sqsubseteq_i j$ for all agents j, j'			*(totality)*
$\vdash j \sqsubseteq_i j$ for every agent j			*(reflexivity)*
$\vdash (j \sqsubseteq_i j' \wedge j' \sqsubseteq_i j'') \to j \sqsubseteq_i j''$ for all agents j, j', j''			*(transitivity)*
$\vdash j \sqsubseteq_i j' \to \neg(j' \sqsubseteq_i j)$ for all agents j, j' with $j \neq j'$			*(antisymmetry)*

Table 1: Axiom system for $\mathcal{L}$. w.r.t. *PR* models.

On preference and reliability As mentioned earlier, this work models situations akin to public deliberation. Individual preferences are announced, and each agent changes her preferences upon getting information about the others', influenced by her reliability over agents (including herself, so she might consider herself as more reliable than some agents but also as less reliable than some others).

In the introduced framework, the agents' preferences are represented by a binary relation, a strategy used by several works in the formal study of preferences (see, e.g., [1, 13] and further references therein). Such relation is typically assumed to be at least reflexive and transitive. Our extra totality assumption simply forbids incomparable worlds. This is not a matter of principle but rather of convenience.

The notion of *reliability* requires a deeper discussion. It is related to that of *trust*, a concept that has been important within artificial societies (e.g., [7]), and for which there are several proposals for its formal representation. It is worthwhile to discuss, albeit briefly, how *reliability* as discussed here relates to *trust*.

Though there are proposals that consider trust to be an attitude of an agent who believes that another agent has a given property [8], a common understanding of this concept is as "agent i trusts agent j's judgement about φ" (called "trust on credibility" in [4]). And, while there are approaches that define trust in terms of other attitudes, as knowledge, beliefs, intentions and goals (e.g., [4, 16]), others define it as a semantic primitive, typically by means of a *neighbourhood function* $N_{i,j} : W \to \wp(\wp(W))$ that assigns, to every pair of agents i, j in every world w, a set of sets of worlds $N_{i,j}(w)$; then it is said that agent i trusts agent j's judgement about φ at world w if and only if the set of worlds in W where φ holds is in $N_{i,j}(w)$ [19].[1] The notion of trust is not represented with normal modal semantics in order to avoid closure under logical consequence: that agent i trusts agent j's judgement about some formula does not imply that i also trust j on the formula's logical consequences.

In contrast, reliability as discussed here is closer to the notion of trust of [17], where it is understood as an ordering among sets of sources of information (cf. the discussion in [12]). Observe, then, how while one difference between the more standard representation of trust (on credibility) and what is called reliability here is that the former parametrises trust with a formula (which can be understood as a topic or area of expertise), the key distinction is that the latter does not yield any *absolute* judgements ("i relies on j's judgement [about φ]"), but only *comparative* ones ("for i, agent j' is at least as reliable as agent j"). For the purposes of this work, such comparative judgements will suffice.

In particular, our reliability relation is asked to be a total, reflexive, transitive and antisymmetric relation. Reflexivity and transitivity are natural requirements for an ordering, and totality simply forbids incomparability, as before. Antisymmetry prevents different agents from being equally-reliable, thus forcing every agent to always select, among any set of agents, a most reliable one. As Section 5 discusses, these assumptions can be dropped, leading to a setting that is more technically involved but also more realistic.

It is also worthwhile to emphasize that the approaches mentioned above as well as the present one consider reliability/trust as a notion that can be extracted from a single snapshot of the agent's attitudes. This is because the aim is to understand how reliability/trust in a source affects the way the information the source provides is assimilated. Other approaches (e.g., [21]), follow the opposite direction, using the way the information is assimilated to define the degree of trust in the source, thus defining reliability/trust not in terms of a single snapshot but rather in terms of at least two (the one before receiving the information, and the one afterwards), and hence understanding this notion not statically but rather dynamically.

3. PREFERENCE DYNAMICS

Intuitively, a public announcement of the agents' individual preferences might induce an agent i to adjust her own preferences according to what has been announced and the reliability she assigns to the set of agents.[2] Thus, agent i's preference ordering *after* such announcement, $\leq'_i$, can be defined in terms of the just announced preferences (the agents' preferences *before* the announcement, $\leq_1, \ldots, \leq_n$) and how much i relied on each agent (i's reliability *before* the announcement, $\preccurlyeq_i$): $\leq'_i := f(\leq_1, \ldots, \leq_n, \preccurlyeq_i)$ for some function f. Here are some such functions.

Definition 6 (Drastic upgrade). *Agent i takes 'as is' the preference ordering of her most reliable agent. More precisely,*

$$u \leq'_i v \quad \mathit{iff}_{def} \quad u \leq_{\mathrm{mr}(i)} v$$

Definition 7 (Radical upgrade). *Agent i takes the preference ordering of her most reliable agent, and in the zones of equally-preferable worlds she uses her old ordering. More precisely,*

$$u \leq'_i v \quad \mathit{iff}_{def} \quad u <_{\mathrm{mr}(i)} v \;\vee\; (u \simeq_{\mathrm{mr}(i)} v \;\wedge\; u \leq_i v)$$

Definition 8 (Tie-breaker upgrade). *Agent i keeps her old ordering, using that of her most reliable agent to 'break ties' in equally-preferable zones. More precisely,*

$$u \leq'_i v \quad \mathit{iff}_{def} \quad u <_i v \;\vee\; (u \simeq_i v \;\wedge\; u \leq_{\mathrm{mr}(i)} v)$$

[1] Some variants (e.g., [20]) deal with *graded trust*.

[2] Note that we do not study the formal representation of such announcement, but rather the representation of its effects.

The following definitions of the new preference ordering are more elaborated: they use more than just the current preference orderings of i and her most reliable agent.

Definition 9 (Lexicographic upgrade). *Agent i takes the preference ordering of her most reliable agent; within the zones of equally-preferable worlds she uses the ordering of her second most reliable agent; within the zones of equally-preferable worlds she uses the ordering her third most reliable agent, and so on. More precisely, if agent i's reliability ordering is given by $a_1 \preccurlyeq_i \cdots \preccurlyeq_i a_n$, then*

$$u \leq'_i v \text{ iff}_{def} \; (u <_{a_n} v) \;\vee\; (u \simeq_{a_n} v \wedge u <_{a_{n-1}} v) \;\vee\; (u \simeq_{a_n} v \wedge u \simeq_{a_{n-1}} v \wedge u <_{a_{n-2}} v) \;\vee\; \cdots \;\vee\; (u \simeq_{a_n} v \wedge \cdots \wedge u \simeq_{a_2} v \wedge u \leq_{a_1} v)$$

The next one differs from the previous in that it uses the agent's original preference ordering as the most important.

Definition 10 (Lexicographic tie-breaker). *Agent i keeps her preference ordering; within the zones of equally-preferable worlds she uses the ordering of her most reliable agent; within the zones of equally-preferable worlds she uses the ordering of her second most reliable agent, and so on. More precisely, if agent i's reliability ordering* without including herself *is given by $a_1 \preccurlyeq_i \cdots \preccurlyeq_i a_{n-1}$, then*

$$u \leq'_i v \text{ iff}_{def} \; (u <_i v) \;\vee\; (u \simeq_i v \wedge u <_{a_{n-1}} v) \;\vee\; (u \simeq_i v \wedge u \simeq_{a_{n-1}} v \wedge u <_{a_{n-2}} v) \;\vee\; \cdots \;\vee\; (u \simeq_i v \wedge u \simeq_{a_{n-1}} v \wedge \cdots \wedge u \simeq_{a_2} v \wedge u \leq_{a_1} v)$$

A general lexicographic upgrade operation The upgrades defined so far can be seen as particular instances of a general case in which an ordered list indicates the priority of the preference orderings that are involved in the upgrade.

Definition 11 (General lexicographic upgrade). *A* lexicographic list *$\mathcal{R}$ over W is a finite non-empty list whose elements are indices of preference orderings over W, with $|\mathcal{R}|$ the list's length and $\mathcal{R}[k]$ its kth element ($1 \leq k \leq |\mathcal{R}|$). Intuitively, $\mathcal{R}$ is a priority list of preference orderings, with $\leq_{\mathcal{R}[1]}$ the one with the highest priority. Given $\mathcal{R}$, the preference ordering $\leq_{\mathcal{R}} \subseteq (W \times W)$ is defined as*

$$u \leq_{\mathcal{R}} v \quad \text{iff}_{def} \quad \underbrace{\Big(u \leq_{\mathcal{R}[\,|\mathcal{R}|\,]} v \;\wedge\; \bigwedge_{k=1}^{|\mathcal{R}|-1} u \simeq_{\mathcal{R}[k]} v\Big)}_{1} \;\vee\; \underbrace{\bigvee_{k=1}^{|\mathcal{R}|-1} \Big(u <_{\mathcal{R}[k]} v \;\wedge\; \bigwedge_{l=1}^{k-1} u \simeq_{\mathcal{R}[l]} v\Big)}_{2}$$

Thus, $u \leq_{\mathcal{R}} v$ holds if this agrees with the least prioritised ordering ($\leq_{\mathcal{R}[|\mathcal{R}|]}$) and for the rest of them u and v are equally preferred (part 1), or if there is an ordering $\leq_{\mathcal{R}[k]}$ with a strict preference for v over u and all orderings with higher priority see u and v as equally preferred (part 2).

The upgrades defined before are all special cases of this general lexicographic upgrade. In fact, the reader might wonder why to use the list $\mathcal{R}$ when the reliability ordering gives us already an ordering among preference relations. The key is that $\mathcal{R}$ includes only the preference relations (strictly speaking, the *indices* of the preference relations) that are actually used when building up the new preference ordering. In this sense, the lexicographic upgrade (Definition 9) is the most natural upgrade since it uses the reliability relation at its fullest (its list $\mathcal{R}$ is exactly i's reliability ordering), but the general lexicographic upgrade allows to work also with other natural upgrades, as the radical or the tie-breaker one.

Proposition 1. *Let $\mathcal{R}$ be a lexicographic list over W. If* every *ordering $\mathcal{R}[k]$ ($1 \leq k \leq |\mathcal{R}|$) is reflexive (transitive, total, respectively), then so is $\leq_{\mathcal{R}}$.*

Proof. *Reflexivity. Take any $u \in W$. Every $\leq_{\mathcal{R}[k]}$ is reflexive ($1 \leq k \leq |\mathcal{R}|$), so $u \simeq_{\mathcal{R}[k]} u$ for all such k. Hence, by part 1 of $\leq_{\mathcal{R}}$'s definition, $u \leq_{\mathcal{R}} u$.*

Transitivity. Suppose (I) $u \leq_{\mathcal{R}} v$ and (II) $v \leq_{\mathcal{R}} w$. From $\leq_{\mathcal{R}}$'s definition, (I) $u \leq_{\mathcal{R}[|\mathcal{R}|]} v$ and, for the rest of the orderings, $u \simeq v$, or there is an ordering $k_1 < |\mathcal{R}|$ such that $u <_{\mathcal{R}[k_1]} v$ and, for all orderings with higher priority, $u \simeq v$, and (II) $v \leq_{\mathcal{R}[|\mathcal{R}|]} w$ and, for the rest of the orderings, $v \simeq w$, or there is an ordering $k_2 < |\mathcal{R}|$ such that $v <_{\mathcal{R}[k_2]} w$ and, for all orderings with higher priority, $v \simeq w$. Each item is a disjunction, so there are four cases; here are the second and the fourth.

- *Suppose $u \leq_{\mathcal{R}[|\mathcal{R}|]} v$ and, for the rest of the orderings, $u \simeq v$, and there is $k_2 < |\mathcal{R}|$ such that $v <_{\mathcal{R}[k_2]} w$ and, for all orderings with higher priority, $v \simeq w$. First, for all orderings with higher priority than that of k_2, $u \simeq w$. Second, focus on k_2. On the one hand, since $u \leq v$ for all orderings, $u \leq_{\mathcal{R}[k_2]} v$. On the other hand, $v <_{\mathcal{R}[k_2]} w$ implies $v \leq_{\mathcal{R}[k_2]} w$. Hence, by $\leq_{\mathcal{R}[k_2]}$'s transitivity, $u \leq_{\mathcal{R}[k_2]} w$. Now observe how $w \not\leq_{\mathcal{R}[k_2]} u$; otherwise from it, $u \leq_{\mathcal{R}[k_2]} v$ and transitivity we would have $w \leq_{\mathcal{R}[k_2]} v$, a contradiction. Then $u <_{\mathcal{R}[k_2]} w$. Thus, by putting the two pieces together, part 2 of $\leq_{\mathcal{R}}$'s definition yields $u \leq_{\mathcal{R}} w$.*
- *Suppose there is $k_1 < |\mathcal{R}|$ such that $u <_{\mathcal{R}[k_1]} v$ (so $u \leq_{\mathcal{R}[k_1]} v$) and, for all orderings with higher priority, $u \simeq v$, and there is $k_2 < |\mathcal{R}|$ such that $v <_{\mathcal{R}[k_2]} w$ (so $v \leq_{\mathcal{R}[k_2]} w$) and, for all orderings with higher priority, $v \simeq w$.*

  Case $k_1 = k_2$. First observe how, for all orderings with higher priority than that of k_1, $u \simeq w$. Second, from $<_{\mathcal{R}[k_1]}$'s transitivity, $u <_{\mathcal{R}[k_2]} w$. Hence, by part 2 of the definition, $u \leq_{\mathcal{R}} w$.

  Case $k_2 > k_1$, so k_1's priority is higher than k_2's priority. First observe how, for all orderings with higher priority than that of k_1, $u \simeq w$. Second, focus on k_1. On the one hand, from $u \leq_{\mathcal{R}[k_1]} v$ and $v \leq_{\mathcal{R}[k_1]} w$ (by $k_2 > k_1$) we get $u \leq_{\mathcal{R}[k_1]} w$. Moreover, $w \not\leq_{\mathcal{R}[k_1]} u$; otherwise from it, $v \leq_{\mathcal{R}[k_1]} w$ and transitivity we would have $v \leq_{\mathcal{R}[k_1]} u$, a contradiction. Then, $u <_{\mathcal{R}[k_1]} w$. Hence, by part 2 of the definition, $u \leq_{\mathcal{R}} w$.

  Case, $k_1 > k_2$: as the previous one.

Totality. Take any u and v; since all orderings in $\mathcal{R}$ are total, $u < v$, $v < u$ or else $u \simeq v$ for each one of them. Now, if $u \simeq v$ for all orderings, part 1 of $\leq_{\mathcal{R}}$'s definition give us $u \leq_{\mathcal{R}} v$. Otherwise there is at least one ordering for which either $u < v$ or else $v < u$; among them, denote by k the one with the highest priority, so u and v are equally preferable for every ordering with higher priority. If k is the ordering with the lowest priority (i.e., $k = |\mathcal{R}|$), part 1 of $\leq_{\mathcal{R}}$'s definition implies $u \leq_{\mathcal{R}} v$ or $v \leq_{\mathcal{R}} u$, according to k's opinion. If not, part 2 of $\leq_{\mathcal{R}}$'s definition implies $u \leq_{\mathcal{R}} v$ or $v \leq_{\mathcal{R}} u$, again according to k's opinion. □

As a consequence of this proposition, the general lexicographic upgrade preserves total preorders (and thus our class of semantic models) *when every preference ordering in* $\mathcal{R}$ *satisfies the requirements.*

The formal language In order to describe the changes the general upgrade operation brings about, the language is extended in the following way.

DEFINITION 12. *The language* $\mathcal{L}_{\{\mathrm{fx}\}}$ *extends* $\mathcal{L}$ *with a modality* $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle$ *for every agent* $i \in Ag$ *and every lexicographic list* $\mathcal{R}$*. Given a PR pointed model* (M, w)*, define*

$$(M, w) \Vdash \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi \quad \textit{iff} \quad \big(\mathrm{fx}^i_{\mathcal{R}}(M), w\big) \Vdash \varphi$$

where the PR model $\mathrm{fx}^i_{\mathcal{R}}(M)$ *is exactly as* M *except in* $\leq_i$*, which is now given by* $\leq_{\mathcal{R}}$ *(Definition 11). Observe how, since the general lexicographic upgrade is a total function, the semantic interpretation of* $[\mathrm{fx}^i_{\mathcal{R}}]\, \varphi := \neg\langle \mathrm{fx}^i_{\mathcal{R}} \rangle \neg\varphi$ *is*

$$(M, w) \Vdash [\mathrm{fx}^i_{\mathcal{R}}]\, \varphi \quad \textit{iff} \quad \big(\mathrm{fx}^i_{\mathcal{R}}(M), w\big) \Vdash \varphi$$

that is, $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi \leftrightarrow [\mathrm{fx}^i_{\mathcal{R}}]\, \varphi$*.*

Thus, the modality $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle$ allows us to express the effects of upgrading the preference relation $\leq_i$ via the general lexicographic upgrade with a lexicographic list $\mathcal{R}$ while keeping the remaining preference relations as before.

For an axiom system for the modality $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle$, we will provide recursion axioms: valid formulas and validity-preserving rules indicating how to translate a formula with the new modality into a provably equivalent one without them. Then, while soundness follows from the validity of these new axioms, completeness follows from the completeness of the basic system. The reader is referred to Chapter 7 of [24] for an extensive explanation of this technique.

In our particular case, the modalities can take the form of any relational expression. Hence, given any relational expression in the model $\mathrm{fx}^i_{\mathcal{R}}(M)$, we must provide a 'matching' relational expression in the original model M. The relational transformer defined below, similar in spirit to the program transformers of [23] for providing recursion axioms for regular *PDL*-expressions [15] (in their case, after action-model updates; [2]), captures this.

DEFINITION 13 (RELATIONAL TRANSFORMER). *Let* $\mathcal{R}$ *be a lexicographic list and* i *be an agent. A* relational transformer $Tx^i_{\mathcal{R}}$ *is a function from relational expressions to relational expressions defined as follows.*

$$Tx^i_{\mathcal{R}}(\leq_i) := \underbrace{\Big(\leq_{\mathcal{R}[\,|\mathcal{R}|\,]} \;\cap\; \bigcap_{k=1}^{|\mathcal{R}|-1} \simeq_{\mathcal{R}[k]}\Big)}_{1} \;\cup\; \underbrace{\bigcup_{k=1}^{|\mathcal{R}|-1} \Big(<_{\mathcal{R}[k]} \;\cap\; \bigcap_{l=1}^{k-1} \simeq_{\mathcal{R}[l]}\Big)}_{2}$$

$$Tx^i_{\mathcal{R}}(\geq_i) := \underbrace{\Big(\geq_{\mathcal{R}[\,|\mathcal{R}|\,]} \;\cap\; \bigcap_{k=1}^{|\mathcal{R}|-1} \simeq_{\mathcal{R}[k]}\Big)}_{1} \;\cup\; \underbrace{\bigcup_{k=1}^{|\mathcal{R}|-1} \Big(>_{\mathcal{R}[k]} \;\cap\; \bigcap_{l=1}^{k-1} \simeq_{\mathcal{R}[l]}\Big)}_{2}$$

$Tx^i_{\mathcal{R}}(1) := 1$	$Tx^i_{\mathcal{R}}(-\pi) := -\,Tx^i_{\mathcal{R}}(\pi)$
$Tx^i_{\mathcal{R}}(\leq_j) := \leq_j$ *for* $i \neq j$	$Tx^i_{\mathcal{R}}(\pi \cup \sigma) := Tx^i_{\mathcal{R}}(\pi) \cup Tx^i_{\mathcal{R}}(\sigma)$
$Tx^i_{\mathcal{R}}(\geq_j) := \geq_j$ *for* $i \neq j$	$Tx^i_{\mathcal{R}}(\pi \cap \sigma) := Tx^i_{\mathcal{R}}(\pi) \cap Tx^i_{\mathcal{R}}(\sigma)$

A relational transformer $Tx^i_{\mathcal{R}}$ takes a relational expression representing a relation in the model $\mathrm{fx}^i_{\mathcal{R}}(M)$ and returns a matching relational expression representing a relation in the original model M. The cases for the basic relational expressions, $\leq_i$ and $\geq_i$, are the important ones; thanks to the expressivity of the 'static' language, they can use Definition 11 directly in order to indicate, first, that $\leq_i$ in $\mathrm{fx}^i_{\mathcal{R}}(M)$ corresponds to $\leq_{\mathcal{R}}$, the result of applying the general lexicographic upgrade with lexicographic list $\mathcal{R}$ in M, and second, that the relation $\geq_i$ in $\mathrm{fx}^i_{\mathcal{R}}(M)$ is simply the converse of $\leq_i$ in the same model, and hence the converse of $\leq_{\mathcal{R}}$ in M. The remaining cases take care of the constant 1, the basic relational expressions for agents other than i and the complement, union and intersection of relations. With $Tx^i_{\mathcal{R}}$ defined, it is possible now to provide the promised recursion axioms.

$\vdash \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, p \;\leftrightarrow\; p$	$\vdash \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, (\varphi \vee \psi) \;\leftrightarrow\; \big(\langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi \vee \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \psi\big)$
$\vdash \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, j \sqsubseteq_i j' \;\leftrightarrow\; j \sqsubseteq_i j'$	$\vdash \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \langle \pi \rangle\, \varphi \;\leftrightarrow\; \langle Tx^i_{\mathcal{R}}(\pi) \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi$
$\vdash \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \neg\varphi \;\leftrightarrow\; \neg\langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi$	From $\vdash \varphi$ infer $\vdash [\mathrm{fx}^i_{\mathcal{R}}]\, \varphi$

Table 2: Recursion axioms for $\mathcal{L}$ plus $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle$ w.r.t. *PR* **models.**

THEOREM 2. *The axioms and rules on Table 1 together with those on Table 2 provide a sound and complete axiom system (with* i *any agent and* π *any relational expression) for* $\mathcal{L}$ *plus* $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle$ *with respect to PR models.*

PROOF (SKETCH). *The rule and the axioms for atomic propositions, reliability formulas, negation and disjunction are standard for an operation without precondition that does not affect atomic propositions (and, in our case, neither reliability). The crucial axiom, the one for modalities with relational expressions, makes key use of the lexicographic relational transformer of Definition 13. As an example, when* π *is* $\leq_i$*, the right-hand side of the axiom becomes*

$$\langle \underbrace{\Big(\leq_{\mathcal{R}[\,|\mathcal{R}|\,]} \;\cap\; \bigcap_{k=1}^{|\mathcal{R}|-1} \simeq_{\mathcal{R}[k]}\Big)}_{1} \;\cup\; \underbrace{\bigcup_{k=1}^{|\mathcal{R}|-1} \Big(<_{\mathcal{R}[k]} \;\cap\; \bigcap_{l=1}^{k-1} \simeq_{\mathcal{R}[l]}\Big)}_{2} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi$$

which, by the axiom for $\cup$ *(Table 1) and some commutation, turns into*

$$\underbrace{\overbrace{\langle <_{\mathcal{R}[1]} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi}^{k=1} \vee \cdots \vee \overbrace{\langle \simeq_{\mathcal{R}[1]} \cap \cdots \cap \simeq_{\mathcal{R}[|\mathcal{R}|-2]} \cap <_{\mathcal{R}[|\mathcal{R}|-1]} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi}^{k=|\mathcal{R}|-1}}_{2}$$
$$\vee\; \underbrace{\langle \simeq_{\mathcal{R}[1]} \cap \cdots \cap \simeq_{\mathcal{R}[\,|\mathcal{R}|-1\,]} \cap \leq_{\mathcal{R}[\,|\mathcal{R}|\,]} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle\, \varphi}_{1}$$

It is now clear what the axiom states: after a general lexicographic upgrade for i *with* $\mathcal{R}$ *there will be a* $\leq_i$*-reachable* φ*-world,* $\langle \mathrm{fx}^i_{\mathcal{R}} \rangle \langle \leq_i \rangle\, \varphi$*, if and only if, currently, there is a* $<_{\mathcal{R}[1]}$*-reachable world that will satisfy* φ *after the operation (part 2 with* $k = 1$*), or else ..., or else there is a* $(\simeq_{\mathcal{R}[1]} \cap \cdots \cap \simeq_{\mathcal{R}[|\mathcal{R}|-2]} \cap <_{\mathcal{R}[|\mathcal{R}|-1]})$*-reachable world that will satisfy* φ *after the operation (part 2 with* $k = |\mathcal{R}| - 1$*), or else there is a* $\simeq_{\mathcal{R}[1]} \cap \cdots \cap \simeq_{\mathcal{R}[\,|\mathcal{R}|-1\,]} \cap \leq_{\mathcal{R}[\,|\mathcal{R}|\,]}$*-world that will satisfy* φ *after the operation (part 1). This is simply the unfolding of the definition of* $\leq_{\mathcal{R}}$ *(Definition 11), and the axiom's validity is straightforward.* □

Example 1. *As examples, consider the drastic and the radical upgrades (Definitions 6 and 7, respectively). Both are instances of the general lexicographic upgrade, with their respective lexicographic lists being* $\langle\!\langle\, \mathrm{mr}(i) \rangle\!\rangle$ *(thus,* $|\mathcal{R}| = 1$*) and* $\langle\!\langle i\ ;\ \mathrm{mr}(i) \rangle\!\rangle$ *(thus,* $|\mathcal{R}| = 2$*).*[3] *Then, their respective recursion axioms for* $\leq_i$ *are*

$$\langle \mathrm{fx}^i_{\mathcal{R}} \rangle \langle \leq_i \rangle \varphi \leftrightarrow \underbrace{\langle \leq_{\mathcal{R}[1]} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \varphi}_{1}$$

$$\langle \mathrm{fx}^i_{\mathcal{R}} \rangle \langle \leq_i \rangle \varphi \leftrightarrow \Big(\underbrace{\langle <_{\mathcal{R}[1]} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \varphi}_{2} \vee \underbrace{\langle \simeq_{\mathcal{R}[1]} \cap \leq_{\mathcal{R}[2]} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \varphi}_{1} \Big)$$

that is,

$$\langle \mathrm{fx}^i_{\mathcal{R}} \rangle \langle \leq_i \rangle \varphi \leftrightarrow \underbrace{\langle \leq_{\mathrm{mr}(i)} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \varphi}_{1}$$

$$\langle \mathrm{fx}^i_{\mathcal{R}} \rangle \langle \leq_i \rangle \varphi \leftrightarrow \Big(\underbrace{\langle <_{\mathrm{mr}(i)} \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \varphi}_{2} \vee \underbrace{\langle \simeq_{\mathrm{mr}(i)} \cap \leq_i \rangle \langle \mathrm{fx}^i_{\mathcal{R}} \rangle \varphi}_{1} \Big)$$

The reader might have noticed that, while this paper's main interest are simultaneous preference upgrades, the model operations and modality of this section deal with single agent upgrades. This presentation style has been chosen in order to simplify notation and (more importantly) readability, but the provided definitions can be easily extended in order to match our goals. In particular, the model operation of Definition 12 can be extended to simultaneous upgrades by asking for a list $\mathbf{R}$ of lexicographic lists (with $\mathbf{R}_i$ the list for agent i). Then its correspondent modality, $\mathrm{fx}_{\mathbf{R}}$, is still axiomatised by the presented system as long as the relational transformer is changed by making the cases for each agent i relative to i's lexicographic list $\mathbf{R}_i$ (thus removing the cases "for agents different from i").

Going back to the example from the introduction and its modelling in Section 2 in terms of M_{exp}, note how

$$M_{exp} \Vdash \langle \leq_b \rangle p_1 \wedge \langle \mathrm{fx}_{\mathbf{R}} \rangle \langle \leq_b \rangle p_3$$

that is, Barbara preferred restaurant 1 but after hearing the others' preferences she prefers restaurant 3. At each world of the model, there exists some preferred world for agent b where p_1 holds. If the model gets updated by an announcement of preferences from all the agents, then there exists some preferred world for agent b in the new model where p_3 holds.

Limits of the general lexicographic upgrade Lexicographic upgrades are not the only 'reasonable' upgrades an agent can perform. One could also think of changing the old preference orderings in a more constrained way, focusing on a specific part of the ordering of the reliable agent(s). Here is a possibility.

Example 2. *Agent i can upgrade her preferences by placing her most reliable agent's most preferred worlds above the rest, then using her old ordering within each zone.*[4] *Thus, for example, if agent a is agent b's most reliable agent and the individual preferences are as below*

$$\leq_a\colon w_2 <_a w_1 <_a w_3 \simeq_a w_4, \qquad \leq_b\colon w_3 <_b w_4 <_b w_1 <_b w_2$$

then such upgrade on b's preferences will create two zones, the upper one with a's most preferred worlds (w_3 *and* w_4)*, and the lower one with the remaining worlds* (w_1 *and* w_2)*. Within each zone, b's old preferences will apply, thus producing* $w_3 <'_b w_4$ *and* $w_1 <'_b w_2$*. The final result is then*

$$\leq'_b\colon w_1 <'_b w_2 <'_b w_3 <'_b w_4$$

Now, observe how no lexicographic list can produce this outcome. First, no singleton list does the job, as $\leq'_b$ *is different from both* $\leq_a$ *and* $\leq_b$*. The list* $\langle\!\langle b\ ;\ a \rangle\!\rangle$ *also fails, as it would give* $\leq_a$ *the highest priority, thus producing an ordering with* w_2 *strictly below* w_1*, different from what* $\leq'_b$ *states. Finally,* $\langle\!\langle a\ ;\ b \rangle\!\rangle$ *fails too, as it will give priority to* $\leq_b$*, thus putting* w_4 *strictly below* w_1*, again different from what* $\leq'_b$ *establishes.*

The upgrade defined in the previous example is not lexicographic: it does not create a preference ordering following a priority list of orderings. Instead, it uses a partition of the domain to create an ordered set of layers, using then a 'default' ordering to sort the worlds within each layer. A more general preference upgrade operation including lexicographic upgrades as well as those of the form just discussed is left for future work.

4. REACHING UNANIMITY

The following are the crucial concepts of this section.

Definition 14 (Unanimity and stability). *Let* $Ag = \{a_1, \ldots, a_n\}$ *be the set of agents; let* $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i \in Ag} \rangle$ *be a PR frame.*

- *There is* unanimity *at F whenever* $\leq_{a_1} = \cdots = \leq_{a_n}$*.*
- *There is* stability *at F* under a given preference upgrade policy f *whenever* $F_\gamma = F_{\gamma+1}$ *for every* $\gamma \geq 1$*, with* $F_1 := F$ *and* $F_{\gamma+1}$ *the result of applying f to all the individual preference relations in* F_γ*.*

Note how our defined notion of unanimity is very restrictive: it asks for the agents' preferences to be completely identical. Broader definitions (e.g., full agreement up to a subset of the domain), though out of the scope of the present work, would allow to cover further scenarios.

Here is a result relating unanimity and stability under the general lexicographic upgrade.

Proposition 2. *Under the general lexicographic upgrade, unanimity implies stability.*

Proof. *Under such upgrade with* $\mathbf{R}_i$ *being the list for each* $i \in Ag$*, the new preference of each agent i becomes the old preference of* $\mathbf{R}_i$*'s highest prioritised ordering (i.e.,* $\leq'_i = \leq_{\mathbf{R}_i[1]}$*) with the possible exception of the cases where* $\leq_{\mathbf{R}_i[1]}$ *establishes equal preferability, which might be broken by some* $\leq_{\mathbf{R}_i[k]}$ $(1 < k \leq |\mathbf{R}_i|)$*. But there is unanimity: worlds equally preferable for* $\leq_{\mathbf{R}_i[1]}$ *are also equally preferred for every other ordering so no tie will be broken. Hence,* $\leq'_i = \leq_{\mathbf{R}_i[1]}$*. But, again by unanimity,* $\leq_{\mathbf{R}_i[1]} = \leq_i$*, and thus* $\leq'_i = \leq_i$*: individual preferences will be the same after one and hence any number of further iterations.* □

Thus, under the general lexicographic upgrade, once everybody agrees on something, further 'discussion' will not change the agents' preferences.

Drastic upgrade The drastic upgrade (Definition 6) is the simplest one among those that have been presented: each

[3] In the lists, the ordering with the highest priority is the rightmost one.

[4] This upgrade is called *lexicographic* in [22].

agent simply takes 'as is' the preference ordering of the agent on whom she relies the most. Observe that when this upgrade is repeatedly applied to every agent's preferences, any ordering any agent will eventually have should be the preference ordering of some agent at the initial stage. The following definition indicates whose original preference ordering each agent has at each stage of the iteration.

DEFINITION 15 (AGENT RELIABILITY STREAM). *Let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be a PR frame; let i be an agent. An i* reliability stream *from F is a function $\alpha_i : \mathbb{N} \to Ag$ such that* **(1)** *$\alpha_i[0] := i$ and,* **(2)** *for every $\ell \geq 0$, $\alpha_i[\ell{+}1] := \mathrm{mr}(\alpha_i[\ell])$.*

The following proposition characterises the cases where drastic upgrade leads to unanimity.

PROPOSITION 3. *Let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be a PR frame in which the $\leq_i$ are all different.*[5] *The iterative application of drastic upgrade over the agents' individual preferences starting from F reaches unanimity (and, by Proposition 2, stability) if and only if there is an $\ell \in \mathbb{N}$ such that $\alpha_{a_1}[\ell] = \cdots = \alpha_{a_n}[\ell]$, with α_a agent a's reliability stream built from her reliability ordering at F.*

PROOF. *Immediate, as an agent reliability stream indicates whose original preference ordering each agent i will have at each stage of the iteration.* □

Here is a propositional dynamic logic [15] based decision procedure that uses the previous proposition's characterisation to determine whether simultaneous drastic upgrades will lead to unanimity in a given frame. It builds a relational framework (model and language) in which states represent agents and transitions between them represent the way each agent's individual preference will change.

DEFINITION 16. *Let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be a PR frame in which the individual preference orderings are all different. Define the set of atomic propositions At' as $\{p_i \mid i \in Ag\}$, and then the relational model M_F based on At' as*

$$\langle S := \{s_i \mid i \in Ag\}, R := \{(s_i, s_j) \mid j = \mathrm{mr}(i)\}, s_i \overset{V}{\longmapsto} \{p_i\}\rangle$$

Thus, S has a state for each agent i, R takes each s_i to the state representing i's most reliable agent, and V makes each p_i the unique atom true at each s_i. (Observe how R is total and deterministic. Define $R^0 := \{(s_i, s_i) \mid s_i \in S\}$, $R^1 := R$, and $R^{\ell+1} := R^\ell \circ R^1$. Denote by $R^\ell[s_i]$ the unique s_j such that $R^\ell s_i s_j$.) The language $\mathcal{L}_{At'}$ is given by $\varphi, \psi ::= p_i \mid \neg\varphi \mid \varphi \vee \psi \mid \langle\cdot\rangle\,\varphi \mid \langle *\rangle\,\varphi$, with these formulas interpreted in multi-pointed models: tuples $(M_F, s_1\cdots s_n)$ where each s_k is a state in M_F. The semantic interpretation for $\neg$ and $\vee$ are standard; for the rest,

$$\begin{array}{lll} (M_F, s_1\cdots s_n) \Vdash p_i & \textit{iff} & p_i \in \bigcap_{k=1}^n V(s_k) \\ (M_F, s_1\cdots s_n) \Vdash \langle\cdot\rangle\,\varphi & \textit{iff} & (M_F, R[s_1]\cdots R[s_n]) \Vdash \varphi \\ (M_F, s_1\cdots s_n) \Vdash \langle *\rangle\,\varphi & \textit{iff} & \exists \ell \in \mathbb{N}\ s.t. (M_F, R^\ell[s_1]\cdots R^\ell[s_n]) \Vdash \varphi \end{array}$$

Observe how while $\langle\cdot\rangle\,\varphi$ indicates that a φ-situation is reachable in one 'multi-transition', $\langle *\rangle\,\varphi$ indicates that a φ-situation is reachable after some finite (possibly zero) number of them. Note how an atom is true at $(M_F, s_1\cdots s_n)$ if and only if it holds in every state in $\{s_1, \ldots, s_n\}$. Since each atom is true at only one state, an atom holds at $(M_F, s_1\cdots s_n)$ if and only if all states in $\{s_1, \ldots, s_n\}$ are the same.

[5]In case some individual preferences are identical, the proposition holds by using the same 'agent-name' for them.

PROPOSITION 4. *Let $Ag = \{a_1, \ldots, a_n\}$ be the set of agents; let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be a PR frame in which the individual preference orderings are all different.*[6] *The iterative application of drastic upgrade over the agents' individual preference starting from F reaches unanimity (and, by Proposition 2, stability) if and only if $(M_F, s_{a_1}\cdots s_{a_n}) \Vdash \langle *\rangle\,(p_{a_1} \vee \cdots \vee p_{a_n})$.*

PROOF. *Observe how $(M_F, s_{a_1}\cdots s_{a_n}) \Vdash \langle *\rangle\,(p_{a_1} \vee \cdots \vee p_{a_n})$ holds exactly when, starting from $s_{a_1}\cdots s_{a_n}$, all states converge to the same state (any of them, hence the disjunction of atoms) after a finite number of steps. This is exactly Proposition 3's characterisation for reaching unanimity under iterative drastic upgrades.* □

Lexicographic upgrade The drastic upgrade reassigns the agents' initial preferences following the reliability ordering. In contrast, the more general lexicographic upgrade (Definition 9) might 'create' new preference orderings by breaking equal-preferability between worlds. However, the following proposition shows that this creation of new preference orderings cannot go on forever: it stops after one step.

PROPOSITION 5. *Let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be a PR frame; let $F' = \langle W, \{\leq'_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be the result of lexicographic upgrades at F. If $u \simeq'_j v$ for some agent $j \in Ag$, then such 'tie' will not be broken by further applications of such upgrade.*

PROOF. *If $u \simeq'_j v$, then $u \simeq_{j'} v$ for all agents j'. Hence, further iterations of the upgrade will not break such tie.* □

Thus, after one step, the lexicographic upgrade behaves exactly as the drastic one: at each step, each agent just adopts the 'old' preference ordering of her most reliable agent as her 'new' one. Then Proposition 3 can be used to characterise the cases in which unanimity will be reached.

General lexicographic upgrade Observe the reason why no ties can be broken after one step of the lexicographic upgrade: such upgrade uses the preferences of all agents to break possible ties. Then, the agent whose preference can break a tie, if any, is always 'reachable' within one step. Thus, Proposition 5 can be generalised to those cases of the general lexicographic upgrade (Definition 11) in which the lexicographic lists used by each agent contains all agents.

PROPOSITION 6. *Let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be a PR frame; let $F' = \langle W, \{\leq'_i, \preccurlyeq_i\}_{i\in Ag}\rangle$ be the result of general lexicographic upgrades at F, with each $i \in Ag$ using some lexicographic list $\mathbf{R}_i$ containing all agents. If $u \simeq'_j v$ for some agent $j \in Ag$, then such 'tie' will not be broken by further applications of such upgrade with the same lists $\mathbf{R}_i$.*

If the lexicographic lists do not include all agents, the agent whose preference can break a tie might not be 'reachable' in one step, and thus such ties will be broken only later (if ever). For example, consider agents a, b, c with preferences $w_1 \simeq_a w_2 \simeq_a w_3$, $w_1 \simeq_b w_2 <_b w_3$ and $w_1 <_c w_2 \simeq_c w_3$, and with reliability such that $\mathrm{mr}(a) = b$ and $\mathrm{mr}(b) = c$. By using radical upgrade (Definition 7; an instance of the general lexicographic upgrade whose lexicographic list contains only two agents), agent a's preference tie between w_1 and w_2 will be broken only at the second round.

[6]If some individual preferences are identical, proceed as indicated in Proposition 3.

Nevertheless, even if the lexicographic lists do not include every agent, there is a finite number of rounds after which all ties that can be broken will be broken.

PROPOSITION 7. *Let $F = \langle W, \{\leq_i, \preccurlyeq_i\}_{i \in Ag} \rangle$ be a PR frame; let $F' = \langle W, \{\leq'_i, \preccurlyeq_i\}_{i \in Ag} \rangle$ be the result of simultaneous general lexicographic upgrades at F, with each $i \in Ag$ using some lexicographic list $\mathbf{R}_i$. For each agent i there is a number m_i such that if $u \simeq'_i v$ after m_i steps, then such 'tie' will not be broken by further applications of such upgrade with the same lists $\mathbf{R}_i$.*

PROOF. *Define the relation $\mathbb{R}$ on Ag as $i\mathbb{R}j$ iff $j \in \mathbf{R}_i$; let $\mathbb{R}^+$ denote the transitive closure of $\mathbb{R}$. For any $i, j \in Ag$, define r_i^j, the degree of reachability of agent j from agent i with respect to the list $\mathbf{R}_i$ as follows:*

$$r_i^j := \begin{cases} 0 & \text{if } j \notin \mathbb{R}^+(i); \\ k & \text{otherwise, with } k \text{ the smallest number s.t. } j \in \mathbb{R}^k(i). \end{cases}$$

Thus, intuitively, r_i^j is the minimal number of steps agent i needs in order to be able to use agent j's preferences, and hence $m_i := \max\{r_i^j \mid j \in Ag\}$ is the number of steps agent i needs to be able to use the preferences of all the agents she can eventually use. Now, we claim that if there is a tie for agent i between worlds u and v after m_i rounds of general lexicographic upgrade with the lexicographic lists in $\mathbf{R}$, then such tie will not be broken afterwards. The reason is clear: after m_i steps agent i will have used all the preference orderings she can use to try to break such tie. If after such a number of rounds the tie prevails, then it will not be affected by further iterations. □

Thus, after a finite number of steps ($\max\{m_i \mid i \in Ag\}$), the general lexicographic upgrade behaves exactly as the drastic one: at each step, each agent just adopts the 'old' preference ordering of her most reliable agent as her 'new' one. Then, again, Proposition 3 can be used to characterise the cases in which unanimity will be reached.

Example Finally, going back to the running example from the introduction, the following deliberation process uses the general lexicographic upgrade with the full reliability ordering as each agent's lexicographic list (thus behaving as the lexicographic upgrade of Definition 9).

EXAMPLE 3. *Consider the initial preference orderings below, with reliability given by $\preccurlyeq_a$: $a \prec_a b \prec_a c$, $\preccurlyeq_b$: $b \prec_b c \prec_b a$ and $\preccurlyeq_c$: $a \prec_c b \prec_c c$.*

$$\leq_a: w_1 <_a w_2 <_a w_3, \ \leq_b: w_3 <_b w_2 <_b w_1, \ \leq_c: w_2 <_c w_1 <_c w_3$$

As a result of an announcement of preferences, b follows a, a follows c and c remains the same:

$$\leq'_a: w_2 <'_a w_1 <'_a w_3, \ \leq'_b: w_1 <'_b w_2 <'_b w_3, \ \leq'_c: w_2 <'_c w_1 <'_c w_3$$

After a second announcement, a's preference ordering does not change (c's did not change) and then b, who has followed a, will coincide with c: unanimity has been reached after two steps (cf. Proposition 3).

$$\leq''_a: w_2 <''_a w_1 <''_a w_3, \leq''_b: w_2 <''_b w_1 <''_b w_3, \leq''_c: w_2 <''_c w_1 <''_c w_3$$

In case Alan had a different reliability ordering, viz. $\preccurlyeq_a$: $c \prec_a a \prec_a b$, unanimity would not have been reached.

Observe how, when the preference orderings are total orders (i.e., no two worlds are equally preferred), the general lexicographic upgrade behaves exactly as the drastic upgrade.

5. CONCLUSIONS AND FURTHER WORK

This work approaches collective decision scenarios not by looking for preference aggregation procedures, but by representing a process of deliberation through which agents share and then change their individual preferences. It introduces a framework for representing both the agents' preferences about objects and their reliability ordering among agents, together with a formal language for describing such structures as well as a sound and complete axiom system. Then it explores some individual preference upgrade policies, introducing a general operation of which the previous ones are particular cases, providing an axiom system for a modality describing the operation's effect. Finally, with the formal framework established, this work provides characterisations of the situations in which the iterative application of the general operation leads to a state where unanimity of preferences has been achieved.

The presented results are the initial steps towards a formal study of the deliberation process; here are some further questions to be studied. **(1)** As discussed, there are 'reasonable' upgrade policies that go beyond the general lexicographic upgrade; how to define a general upgrade operation that covers such cases? **(2)** Which other preference upgrade policies arise when the requirements on the preference relation are weakened? And when the reliability relation is generalised to an ordering among sets of agents? **(3)** Maybe more interesting, the announcement of the individual preferences triggers not only a change in preferences but also a change in reliability (e.g., an agent relies more on the people with whom she shares the same preferences). It is worthwhile to explore policies for such change. **(4)** The introduced framework represents an agent's preferences and reliability, but it leaves out epistemic notions. A further extension allowing to represent the knowledge/beliefs agents have about one another's preferences would raise the issue of strategic behaviour, and therefore the topic of manipulation. **(5)** Finally, a reasonable combination of the deliberation and the aggregation perspectives will provide a more realistic model for group decision making.

Acknowledgements
The authors thank the three anonymous reviewers of AAMAS 2015 for their insightful comments which helped to improve the paper.

REFERENCES

[1] K. J. Arrow, A. K. Sen, and K. Suzumura, editors. *Handbook of Social Choice and Welfare*. Elsevier, 2002. Two volumes.

[2] A. Baltag, L. S. Moss, and S. Solecki. The logic of public announcements and common knowledge and private suspicions. In I. Gilboa, editor, *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-98), Evanston, IL, USA, July 22-24, 1998*, pages 43–56. Morgan Kaufmann, 1998.

[3] J. P. Burgess. Basic tense logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume II, chapter 2, pages 89–133. Reidel, 1984.

[4] R. Demolombe. To trust information sources: A proposal for a modal logic framework. In C. Castelfranchi and Y.-H. Tan, editors, *Trust and*

Deception in Virtual Societies. Kluwer Academic, Dordrecht, 2001.

[5] J. Elster. The market and the forum. In J. Elster and A. Hylland, editors, *Foundations of Social Choice Theory*. Cambridge University Press, Cambridge, 1986.

[6] U. Endriss. Logic and social choice theory. In A. Gupta and J. van Benthem, editors, *Logic and Philosophy Today*, volume 2, pages 333–377. College Publications, 2011.

[7] R. Falcone, K. S. Barber, J. Sabater-Mir, and M. P. Singh, editors. *Trust in Agent Societies, 11th International Workshop, TRUST 2008, Estoril, Portugal, May 12-13, 2008. Revised Selected and Invited Papers*, volume 5396 of *Lecture Notes in Computer Science*. Springer, 2008.

[8] R. Falcone and C. Castelfranchi. Social trust: A cognitive approach. In C. Castelfranchi and Y.-H. Tan, editors, *Trust and Deception in Virtual Societies*, pages 55–90. Kluwer Academic, Dordrecht, 2001.

[9] D. M. Gabbay, O. Rodrigues, and G. Pigozzi. Connections between belief revision, belief merging and social choice. *Journal of Logic and Computation*, 19(3):445–446, 2009.

[10] G. Gargov and S. Passy. A note on boolean modal logic. In P. P. Petkov, editor, *Mathematical Logic*, pages 299–309. Springer US, 1990.

[11] R. I. Goldblatt. Semantic analysis of orthologic. *Journal of Philosophical Logic*, 3(1):19–35, Mar. 1974.

[12] A. I. Goldman. Experts: Which ones should you trust? *Philosophy and Phenomenological Research*, 63(1):85–110, 2001.

[13] T. Grüne-Yanoff and S. O. Hansson, editors. *Preference Change*, volume 42 of *Theory and Decision Library*. Springer, 2009.

[14] J. Habermas. *Between Facts and Norms: Contributions to a Discursive Theory of Law and Democracy*. MIT Press, Cambridge, 1996.

[15] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, Cambridge, MA, 2000.

[16] A. Herzig, E. Lorini, J. F. Hübner, and L. Vercouter. A logic of trust and reputation. *Logic Journal of the IGPL*, 18(1):214–244, 2010.

[17] W. H. Holliday. Trust and the dynamics of testimony. In L. Kurzen, D. Grossi, and F. R. Velázquez-Quesada, editors, *Logic and Interactive RAtionality. Seminar's yearbook 2009*, pages 118–142. Institute for Logic, Language and Computation, Universiteit van Amsterdam, Amsterdam, The Netherlands, 2010.

[18] S. Konieczny and R. P. Pérez. Logic based merging. *Journal of Philosophical Logic*, 40(2):239–270, 2011.

[19] C.-J. Liau. Belief, information acquisition, and trust in multi-agent systems – a modal logic formulation. *Artificial Intelligence*, 149(1):31–60, 2003.

[20] E. Lorini, G. Jiang, and L. Perrussel. Trust-based belief change. In T. Schaub, G. Friedrich, and B. O'Sullivan, editors, *ECAI 2014 – 21st European Conference on Artificial Intelligence, 18-22 August 2014, Prague, Czech Republic – Including Prestigious Applications of Intelligent Systems (PAIS 2014)*, volume 263 of *Frontiers in Artificial Intelligence and Applications*, pages 549–554. IOS Press, 2014.

[21] B. Rodenhäuser. *A Matter of Trust: Dynamic Attitudes in Epistemic Logic*. PhD thesis, Institute for Logic, Language and Computation (ILLC), Universiteit van Amsterdam (UvA), Amsterdam, The Netherlands, 2014. ILLC Dissertation Series DS-2014-04.

[22] J. van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17(2):129–155, 2007.

[23] J. van Benthem, J. van Eijck, and B. Kooi. Logics of communication and change. *Information and Computation*, 204(11):1620–1662, 2006.

[24] H. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic*. Number 337 in Synthese Library. Springer, 2008.