Valuing others' opinions: Preference, belief and reliability dynamics

Sujata Ghosh[1] and Katsuhiko Sano[2]
[1] *Indian Statistical Institute, Chennai, India*
[2] *Department of Philosophy, Graduate School of Letters, Hokkaido University, Sapporo, Japan*
sujata@isichennai.res.in, katsuhiko.sano@gmail.com

Keywords: Preference, Belief, Reliability, Hybrid Logic, Public Announcement Logic, Propositional Dynamic Logic

Abstract: Deliberation often leads to changes in preferences and beliefs of an agent, influenced by the opinions of others, depending on how reliable these agents are according to the agent under consideration. Sometimes, it also leads to changes in the opposite direction, that is, reliability over agents gets updated depending on their preferences and/or beliefs. There are various formal studies of preference and belief change based on reliability and/or trust, but not the other way around – this work contributes to the formal study of the latter aspect, that is, on reliability change based on agent preferences. In process, some policies of preference change based on agent reliabilities are also discussed. A two-dimensional hybrid language is proposed to describe such processes, and axiomatisations and decidability are discussed.

1 INTRODUCTION

Deliberation forms an important component in any decision-making process. It is basically a conversation through which individuals provide their opinions regarding certain issues, give preferences among possible choices, justify these preferences. This process may lead to changes in their opinions, because they are influenced by one another. A factor that sometimes plays a big role in enforcing such changes, is the amount of reliability the agents have on one another's opinions. Such reliabilities may change as well through this process of deliberation, e.g. on hearing someone else's preferences about a certain issue, one can start or stop relying on that person's opinion. One may tend to unfriend certain friends hearing about their preferences regarding certain issues (e.g. Helen De Cruz's recent remarks in her article 'Being Friends with a Brexiter?' in the Philosophers On series of the Daily Nous blog[1]).

Formal studies on preferences (cf. (Arrow et al., 2002; Endriss, 2011)) and trust (cf. (Liau, 2003; Demolombe, 2004; Herzig et al., 2010)) abound in the literature on logic in artificial intelligence. Recently, there has been work on relating the notions of belief and trust, e.g. about agents changing their beliefs based on another agent's announcement depending on how trustworthy that agent is about the issue in question (e.g. see (Lorini et al., 2014)). And, also on relating preference and reliability, e.g. about agents changing their preferences based on another agent's preferences, on whom he or she relies the most (Ghosh and Velázquez-Quesada, 2015a; Ghosh and Velázquez-Quesada, 2015b). A pertinent issue that arises in this context is: an agent's assessment of another individual's reliability might change as well. How would one model that? This work precisely provides a way to answer this question. We focus on reliability changes based on (public) announcement of individual preferences and we provide formal frameworks to describe such changes. In process, we also provide some policies of preference change as well. Note that the notion of reliability considered here is not topic-based (in contrast to the notion of trust described in (Lorini et al., 2014)) but deals with only comparative judgements about agents (cf. Section 2 for details). The following provides an apt example of the situations we would like to model:

> *Our Running Example*: Consider three flatmates Isabella, John and Ken discussing about redecorating their house and they were wondering whether to put a print of Monet's picture on the left wall or on the right wall of the living room. Isabella and Ken prefer to put it on the right wall, while John wants to put it on the left. Isabella has more faith in John's taste than on hers and Ken's, and John has more faith in Isabella's taste than on his' and Ken's.

[1] `http://dailynous.com/2016/06/28/philosophers-on-brexit/#DeCruz`

Ken has full faith in his own taste. As long as Isabella's and John's preferences are different and each think that the other's taste is better (by taking their preferences into consideration), the three flatmates would never reach an agreement. But it so happens that on hearing about John's and Ken's choices, Isabella starts relying more on Ken, whereas even after hearing about Isabella's and Ken's choices John's reliability attribution to Isabella does not change.

To model such situations we introduce a two-dimensional hybrid logic framework extending the basic logic proposed in (Ghosh and Velázquez-Quesada, 2015a; Ghosh and Velázquez-Quesada, 2015b) in the line of those developed in (Gargov et al., 1987; Sano, 2010; Seligman et al., 2013). We add dynamic operators to model preference and reliability changes. The main novelty of this work is that reliability changing policies based on agent preferences are introduced and studied formally, which has not been dealt with before. In addition, reliabilities are modelled (more naturally) as total pre-orders (instead of total orders (Ghosh and Velázquez-Quesada, 2015a; Ghosh and Velázquez-Quesada, 2015b)), and preference changing policies are modified accordingly. The proposed logic is expressive enough to deal with both these kinds of changes.

2 TWO-DIMENSIONAL HYBRID LOGIC

Let us first motivate our assumptions on preference and reliability orders that we make below in the lines of (Ghosh and Velázquez-Quesada, 2015a). As mentioned earlier, we are modelling situations akin to joint deliberation where agents announce their preferences. Each agent can change her preferences upon getting information about the other agents' preferences, 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). Agents can also change their opinions regarding how reliable they think the other agents are in comparison to themselves, influenced by the announced preferences of those agents.

The agents' preferences are represented by binary relations (as in (Arrow et al., 2002; Grüne-Yanoff and Hansson, 2009) and further references therein), which is typically assumed to be reflexive and transitive. This paper also follows this ordinary assumption, and so, we note that this assumption do allow the possibility of incomparable worlds.

The notion of *reliability* is related to that of *trust*, a well-studied concept (e.g., (Falcone et al., 2008)), with several proposals for its formal representation, e.g. an attitude of an agent who believes that another agent has a given property (Falcone and Castelfranchi, 2001). One also says that "an agent i trusts agent j's judgement about φ" (called "trust on credibility" in (Demolombe, 2001)). Trust can also be defined in terms of other attitudes, such as knowledge, beliefs, intentions and goals (e.g., (Demolombe, 2001; Herzig et al., 2010)), or as a semantic primitive, typically by means of a *neighbourhood function* (Liau, 2003). Some others (e.g., (Lorini et al., 2014)) deal with *graded trust*.

Reliability as discussed here is closer to the notion of trust in (Holliday, 2010), where it is understood as an ordering among sets of sources of information (cf. the discussion in (Goldman, 2001)). Such a notion of reliability 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, similar to (Ghosh and Velázquez-Quesada, 2015a), such comparative judgements suffice.

In contrast to (Ghosh and Velázquez-Quesada, 2015a), our reliability relation is assumed to be a reflexive, transitive and total relation. Reflexivity and transitivity are, more often than not, natural requirements for an ordering and totality disallows incomparability, as before. The changes in reliability for an agent depend on the information assimilated (similar to approaches like (Rodenhäuser, 2014)), in particular, about the other agents' preferences.

The focus of this work is joint deliberation, so let A be a *finite non-empty* set of agents ($|A| = n \geq 2$).

Definition 1. A *PR* (*preference/reliability*) *frame* F is a tuple $(W, \{\leq_i, \preccurlyeq_i\}_{i\in A})$ where **(1)** W is a finite non-empty set of worlds; **(2)** $\leq_i \subseteq W \times W$ is a preorder (i.e., a reflexive and transitive relation), agent i's *preference relation* among worlds in W ($u \leq_i v$ is read as "world v is at least as preferable as world u for agent i"); **(3)** $\preccurlyeq_i \subseteq A \times A$ is a total pre-order (i.e., a connected pre-order), agent i's *reliability relation* among agents in A ($j \preccurlyeq_i k$ is read as "agent k is at least as reliable as agent j for agent i"). Let $mr(i)$ denote the set of all maximally reliable agents for i.

We define $u <_i v$ ("u is less preferred than v for agent i") as $u \leq_i v$ and $v \not\leq_i u$, and $u \simeq_i v$ ("u and v are equally preferred for agent i") as $u \leq_i v$ and $v \leq_i u$. Moreover, $j \prec_i k$ ("j is less reliable than k for agent i") is defined as $j \preccurlyeq_i k$ and $k \not\preccurlyeq_i j$, and $j \approx_i k$ ("j and k are equally reliable for agent i") as $j \preccurlyeq_i k$ and $k \preccurlyeq_i j$.

Example 1. Recall the example in Section 1. Put $A = \{i, j, k\}$, where i, j, and k represent Isabella, John,

and Ken, respectively. By denoting with w_x the world where 'Monet's picture is at wall x' ($x = l, r$), the example's situation can be represented by a *PR* frame $F_{exp} = (\{w_l, w_r\}, \{\leq_y, \preccurlyeq_y\}_{y \in A})$ in which the preference orders are given by: $w_l <_i w_r$, $w_r <_j w_l$ and $w_l <_k w_r$, and the reliability orders are given by: $i \approx_i k \prec_i j$, $j \approx_j k \prec_j i$ and $j \approx_k i \prec_k k$.

In (Ghosh and Velázquez-Quesada, 2015a), Ghosh and Velázquez-Quesada propose a language to talk about the preference changes and their effects. Following the semantic idea of (Seligman et al., 2013), we extend their syntax for the static language into a two-dimensional syntax with the help of dependent product of two hybrid logics (Sano, 2010). Let P be a countable infinite set of propositional variables, $\mathsf{N}_1 = \{\mathbf{a}, \mathbf{b}, \mathbf{c}, \dots\}$ be a countable infinite set of world-nominals (syntactic names for worlds) and let $\mathsf{N}_2 = \{\mathbf{i}, \mathbf{j}, \mathbf{k}, \dots\}$ be a countable infinite set of agent-nominals (syntactic names for agents).

Definition 2 (Language $\mathcal{HL}$)**.** Formulas $\varphi, \psi, \dots$ (read φ as "the current agent satisfies the property φ in the current state" or indexically as "I am φ in the current state") and relational expressions (or program terms) $\pi, \rho, \dots$ of the language $\mathcal{HL}$ are given by

$$\varphi, \psi ::= p \mid \mathbf{a} \mid \mathbf{i} \mid \neg\varphi \mid \varphi \vee \psi \mid @_{\mathbf{i}}\varphi \mid @_{\mathbf{a}}\varphi \mid \langle\pi\rangle\varphi,$$
$$\pi, \rho ::= 1_W \mid \leq \mid \geq \mid 1_A \mid \sqsubseteq_{\mathbf{k}} \mid \sqsupseteq_{\mathbf{k}} \mid -\alpha \mid \pi \cup \rho \mid \pi \cap \rho \mid (\pi, \mathbf{j}) \sqcap_{\mathbf{i}} (\rho, \mathbf{k}) \mid ?(\varphi, \psi),$$

where $p \in \mathsf{P}$, $\mathbf{a} \in \mathsf{N}_1$, $\mathbf{i}, \mathbf{j}, \mathbf{k} \in \mathsf{N}_2$, $\alpha \in \{1_W, \leq, \geq\} \cup \{1_A, \sqsubseteq_{\mathbf{k}}, \sqsupseteq_{\mathbf{k}} \mid \mathbf{k} \in \mathsf{N}_2\}$. Propositional constants $(\top, \bot)$, other Boolean connectives $(\wedge, \rightarrow, \leftrightarrow)$ and the dual modal universal operators $[\pi]$ are defined as usual, e.g. $[\pi]\varphi := \neg\langle\pi\rangle\neg\varphi$. Moreover, we define $\langle< \rangle\varphi$ as $\langle\leq \cap -\geq\rangle\varphi$ and $\langle\sqsubset_{\mathbf{k}}\rangle\varphi$ as $\langle\sqsubseteq_{\mathbf{k}} \cap -\sqsupseteq_{\mathbf{k}}\rangle\varphi$, respectively. Finally, $?\varphi$ is defined as the program term $?(\varphi, \varphi)$.

We note that $@_{\mathbf{a}}\varphi$ is read as "the current agent satisfies φ in the world named by $\mathbf{a}$"and $@_{\mathbf{i}}\varphi$ as "agent $\mathbf{i}$ satisfies φ in the current world." The set of relational expressions contains the constants $1_W, 1_A$ (the global relations, whose corresponding operators mean "for all states" and "for all agents", respectively), the preference and reliability relations $(\leq, \sqsubseteq_{\mathbf{k}})$, their respective converse relations $(\geq, \sqsupseteq_{\mathbf{k}}$; cf. (Burgess, 1984; Goldblatt, 1992)), all the complements of the atomic relations, and an additional construct of the forms $(\pi, \mathbf{j}) \sqcap_{\mathbf{i}} (\pi', \mathbf{k})$ (needed for defining distributed preference later in Section 3, explained below) and $?(\varphi, \psi)$ (a generalization of the test operator in (Harel et al., 2000), also explained below), and it is closed under union and intersection operations over relations.

The formulas are interpreted in terms of *world-agent* pairs below, and we may read $[\leq]\varphi$ as "in all states which the current agent considers as least as good as the current state, the current agent satisfies φ". Moreover, we may read $\langle\sqsubseteq_{\mathbf{k}}\rangle\varphi$ as "there is a more or equally reliable agent j than the current agent such that j satisfies φ, from agent $\mathbf{k}$'s perspective." For example, $@_{\mathbf{i}}\langle\sqsubseteq_{\mathbf{k}}\rangle\mathbf{j}$ can be read as "agent $\mathbf{j}$ is more or equally reliable than agent $\mathbf{i}$ from agent $\mathbf{k}$'s perspective." $\langle\sqsupseteq_{\mathbf{k}}\rangle\varphi$ is read as "there is a less or equally reliable agent j than the current agent such that j satisfies φ, from agent $\mathbf{k}$'s perspective."

We note that the program construction $?(\varphi, \psi)$ (check if the first element of a given pair of states satisfies φ and if the second does ψ) is a generalization of the test operator in the standard (regular) propositional dynamic logic (Harel et al., 2000). So $?\varphi := ?(\varphi, \varphi)$ enables us to check if both elements of a given pair satisfies φ. Moreover, the program construction $(\pi, \mathbf{j}) \sqcap_{\mathbf{i}} (\pi', \mathbf{k})$ enables us to define, as agent $\mathbf{i}$'s relation between states, the distributed preference between agents $\mathbf{j}$ and $\mathbf{k}$, i.e., the intersection of $\mathbf{j}$'s preference and $\mathbf{k}$'s preference. Together with the other program constructions, it is useful for providing the axiom system for the preference and reliability changing operations to be introduced in Section 3. The following two definitions establish what a model is and how formulas of $\mathcal{HL}$ are interpreted over them.

Definition 3 (PR model)**.** A *PR* model is a tuple $M = (F, V)$ where $F = (W, \{\leq_i, \preccurlyeq_i\}_{i \in A})$ is a *PR*-frame and V is a valuation function from $\mathsf{P} \cup \mathsf{N}_1 \cup \mathsf{N}_2$ to $\mathcal{P}(W \times A)$ assigning a subset of the form $\{w\} \times A$ to a world-nominal $\mathbf{a} \in \mathsf{N}_1$ and a subset of the form $W \times \{i\}$ to an agent-nominal $\mathbf{i} \in \mathsf{N}_2$. Throughout the paper, we denote the unique element in the first coordinate of $V(\mathbf{a}) = \{w\} \times A$ and the second coordinate of $V(\mathbf{i}) = W \times \{a\}$ by $\underline{\mathbf{a}}$ and $\underline{\mathbf{i}}$, respectively.

Definition 4 (Truth definition)**.** Given a *PR*-model M, a satisfaction relation $M, (w, i) \Vdash \varphi$, and relations $R_\pi \subseteq (W \times A)^2$ are defined by simultaneous induction by:

$$\begin{aligned}
M, (w,i) \Vdash p \quad &\text{iff} \quad (w,i) \in V(p),\\
M, (w,i) \Vdash \mathbf{a} \quad &\text{iff} \quad w = \underline{\mathbf{a}},\\
M, (w,i) \Vdash \mathbf{i} \quad &\text{iff} \quad i = \underline{\mathbf{i}},\\
M, (w,i) \Vdash \neg\varphi \quad &\text{iff} \quad M, (w,i) \nVdash \varphi,\\
M, (w,i) \Vdash \varphi \vee \psi \quad &\text{iff} \quad M, (w,i) \Vdash \varphi \text{ or } M, (w,i) \Vdash \psi,\\
M, (w,i) \Vdash @_{\mathbf{a}}\varphi \quad &\text{iff} \quad M, (\underline{\mathbf{a}}, i) \Vdash \varphi,\\
M, (w,i) \Vdash @_{\mathbf{i}}\varphi \quad &\text{iff} \quad M, (w, \underline{\mathbf{i}}) \Vdash \varphi,\\
M, (w,i) \Vdash \langle\pi\rangle\psi \quad &\text{iff} \quad \text{For some } (v,j) \in W \times A,\\
&\qquad (w,i) R_\pi (v,j) \text{ and } M, (v,j) \Vdash \psi,\\
(w,i) R_{1_W} (v,j) \quad &\text{iff} \quad w, v \in W \text{ and } i = j,\\
(w,i) R_{\leq} (v,j) \quad &\text{iff} \quad w \leq_i v \text{ and } i = j,\\
(w,i) R_{\geq} (v,j) \quad &\text{iff} \quad v \leq_i w \text{ and } i = j,
\end{aligned}$$

$$
\begin{array}{rll}
(w,i)R_{-\alpha}(v,j) & \text{iff} & ((w,i),(v,i)) \notin R_\alpha \text{ and } i = j \\
 & & (\alpha \in \{1_W, \leq, \geq\}), \\
(w,i)R_{1_A}(v,j) & \text{iff} & w = v \text{ and } i,j \in A, \\
(w,i)R_{\sqsubseteq_{\mathbf{k}}}(v,j) & \text{iff} & w = v \text{ and } i \preccurlyeq_{\underline{\mathbf{k}}} j, \\
(w,i)R_{\sqsupseteq_{\mathbf{k}}}(v,j) & \text{iff} & w = v \text{ and } j \preccurlyeq_{\underline{\mathbf{k}}} i, \\
(w,i)R_{-\beta}(v,j) & \text{iff} & w = v \text{ and } ((w,i),(w,j)) \notin R_\beta \\
 & & (\beta \in \{1_A, \sqsubseteq_{\mathbf{k}}, \sqsupseteq_{\mathbf{k}} \mid \mathbf{k} \in \mathsf{N}_2\}), \\
(w,i)R_{\pi\cup\rho}(v,j) & \text{iff} & (w,i)R_\pi(v,j) \text{ or } (w,i)R_\rho(v,j), \\
(w,i)R_{\pi\cap\rho}(v,j) & \text{iff} & (w,i)R_\pi(v,j) \text{ and } (w,i)R_\rho(v,j), \\
(w,i)R_{(\pi,\mathbf{j})\sqcap_{\mathbf{i}}(\rho,\mathbf{k})}(v,j) & \text{iff} & i = j = \underline{\mathbf{i}} \text{ and } (w,\underline{\mathbf{j}})R_\pi(v,\underline{\mathbf{j}}) \\
 & & \text{and } (w,\underline{\mathbf{k}})R_\rho(v,\underline{\mathbf{k}}) \\
(w,i)R_{?(\varphi,\psi)}(v,j) & \text{iff} & M,(w,i) \Vdash \varphi \text{ and } M,(v,j) \Vdash \psi.
\end{array}
$$

We say that φ is *valid* in a *PR*-model M (written: $M \Vdash \varphi$) if $M,(w,i) \Vdash \varphi$ for all pairs (w,i) in M.

The logic $\mathcal{HL}$ is so expressive that we can formalize the notion of belief as well as our preference and reliability dynamics introduced in the later sections. For example, following the idea found in (Boutilier, 1994), we can define the conditional belief operator $B(\psi,\varphi)$ (read "under the condition that the current agent satisfies ψ, the current agent believes that she satisfies φ" or "the current agent desires (or has a goal) that she satisfies φ under the condition that she satisfies ψ") by

$$B^{\psi}\varphi := [1_W]((\psi \wedge \varphi) \rightarrow \langle\leq\rangle(\psi \wedge \varphi \wedge [\leq](\psi \rightarrow \varphi))).$$

Then the unconditional belief operator $B\varphi$ is defined as $B(\top,\varphi)$, which read as "the current agent believes that she satisfies φ" or "in the most preferred states for the current agent, she satisfies φ." We can also define the conditional reliability operator $R_{\mathbf{k}}(\psi,\varphi)$ (read "the most reliable ψ-agents for agent $\mathbf{k}$ satisfy φ.") by

$$R_{\mathbf{k}}(\psi,\varphi) := [1_A](\psi \rightarrow \langle\sqsubseteq_{\mathbf{k}}\rangle(\psi \wedge [\sqsubseteq_{\mathbf{k}}](\psi \rightarrow \varphi))),$$

where we can simplify the clause because of connectedness as noted in (Boutilier, 1994). The unconditional version $R_{\mathbf{k}}\varphi$ of $R_{\mathbf{k}}(\psi,\varphi)$ is defined as $R_{\mathbf{k}}(\top,\varphi)$ which read as "the most reliable agents for agent $\mathbf{k}$ satisfy φ." We may also define the "diamond"-version of $R_{\mathbf{k}}\varphi$ as $\neg R_{\mathbf{k}}\neg\varphi$ to denote $\langle R_{\mathbf{k}}\rangle\varphi$. Then $\langle R_{\mathbf{k}}\rangle\mathbf{j}$ means that agent $\mathbf{j}$ is one of the most reliable agents for $\mathbf{k}$.

Example 2. Let us represent "the current agent likes to put Monet's picture on wall x" by a state-nominal $\mathbf{a}_x$ in the setting of Example 1. On the *PR*-frame F_{exp} of Example 1, we define $V(\mathbf{a}_x) = \{(w_x,i),(w_x,j),(w_x,k)\}$ where $x = l$ or r. We use $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ as syntactic names (i.e., agent nominals) for i, j and k, where we interpret, e.g., $\underline{\mathbf{i}} = i$ in terms of our valuation function V. Define $M_{exp} := (F_{exp}, V)$. For example, the preference $w_l <_i w_r$ can be formalized as a formula $@_{\mathbf{i}}@_{\mathbf{a}_l}\langle<\rangle\mathbf{a}_r$, which is valid on M_{exp}. We can formalize Isabella's reliability of $i \approx_i k <_i j$ as $@_{\mathbf{i}}\langle\sqsubseteq_{\mathbf{i}}\rangle\mathbf{k} \wedge @_{\mathbf{k}}\langle\sqsubseteq_{\mathbf{i}}\rangle\mathbf{i} \wedge @_{\mathbf{k}}\langle\sqsubset_{\mathbf{i}}\rangle\mathbf{j}$, which is valid on M_{exp}. Moreover, $@_{\mathbf{i}}B\mathbf{a}_x$ formalizes "Isabella believes that she likes to put Monet's picture on wall x" and, when $x = r$, $@_{\mathbf{i}}B\mathbf{a}_r$ is valid on M_{exp}. Similarly, $@_{\mathbf{j}}B\mathbf{a}_l$ and $@_{\mathbf{k}}B\mathbf{a}_r$ are also valid in M_{exp}. We can see that, from Isabella's perspective, Ken is one of the most reliable agents who believes that $\mathbf{a}_r$. This can be formalized as $R_{\mathbf{i}}(B\mathbf{a}_r,\mathbf{k})$.

The static axiom systems $\mathsf{H}\mathbf{PR}$ and $\mathsf{H}\mathbf{PR}_{(m,n)}$ are given as in Table 1, where *uniform substitution* means a substitution that uniformly replaces propositional variables by formulas and nominals from N_i by nominals from N_i ($i = 1$ or 2).

Theorem 1 (Soundness and completeness)**.** φ is valid in all (possibly infinite) *PR*-models iff φ is derivable in $\mathsf{H}\mathbf{PR}$. Moreover, φ is valid in all *PR*-models with fixed m worlds and fixed n agents iff φ is derivable in $\mathsf{H}\mathbf{PR}_{(m,n)}$. Therefore, $\mathsf{H}\mathbf{PR}_{(m,n)}$ is decidable.

We note that the, as far as the authors know, decidability is still unknown for $\mathsf{H}\mathbf{PR}$, even the fragment of $\mathsf{H}\mathbf{PR}$ without program constructions (cf. (Sano, 2010)). So related computational properties of such fragment has not been yet well-studied (for purely bimodal logic fragment with a slightly different semantics, the reader is referred to (Marx and Mikulás, 2001)).

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 ordering she assigns to the set of agents.[2] For example, an agent might adopt the preferences of the set of agents on whom she relies the most, or might use the strict preferences of her most reliable agents for 'breaking ties' among her equally-preferred zones. In (Ghosh and Velázquez-Quesada, 2015a) the authors introduced the *general lexicographic upgrade* operation, which creates a preference ordering following a priority list of orderings. We generalize those operations in the following, where we consider the reliability orderings to be *pre-orders*, rather than being total orders (that is, also anti-symmetric and connected) as they are in the

[2]Note that this work, in line with its predecessor, (Ghosh and Velázquez-Quesada, 2015a), also does not focus on the formal representation of such announcement, but rather on the formal representation of its effects.

Table 1: Axiomatizations H**PR** and H**PR**$_{(m,n)}$

Bi-Hybrid Logical Axioms of H**PR**
All classical tautologies $(\mathtt{Dual}_\pi)$ $\langle\pi\rangle p \leftrightarrow \neg[\pi]\neg p$
$(\mathtt{K}_\pi)$ $[\pi](p \to q) \to ([\pi]p \to [\pi]q)$
Let $\mathbf{n} \in \mathsf{N}_i$ and $(\mathbf{n}, \mathbf{m}) \in \mathsf{N}_i^2$ $(i = 1, 2)$ below in this group
$(\mathtt{K}_@)$ $@_\mathbf{n}(p \to q) \to (@_\mathbf{n} p \to @_\mathbf{n} q)$
$(\mathtt{SelfDual}_@)$ $\neg@_\mathbf{n} p \leftrightarrow @_\mathbf{n}\neg p$ $\quad(\mathtt{Ref})$ $@_\mathbf{n}\mathbf{n}$
$(\mathtt{Intro})$ $\mathbf{n} \wedge p \to @_\mathbf{n} p$ $\quad(\mathtt{Agree})$ $@_\mathbf{n}@_\mathbf{m} p \to @_\mathbf{m} p$
$(\mathtt{Back})$ $\langle\pi\rangle@_\mathbf{a}@_\mathbf{i} p \to @_\mathbf{a}@_\mathbf{i} p$
Inference Rules of H**PR**
Modus Ponens, Uniform Substitutions,
Necessitation Rules for $[\pi]$, $@_\mathbf{i}$ and $@_\mathbf{a}$
$(\mathtt{Name})$ From $\mathbf{n} \to \varphi$ infer φ,
where $\mathbf{n} \in \mathsf{N}_1 \cup \mathsf{N}_2$ is fresh in φ
$(\mathtt{BG}_\pi)$ From $@_\mathbf{a}@_\mathbf{i}\langle\pi\rangle(\mathbf{b} \wedge \mathbf{j}) \to @_\mathbf{b}@_\mathbf{j}\varphi$ infer $@_\mathbf{a}@_\mathbf{i}[\pi]\varphi$,
where $\mathbf{b}$ and $\mathbf{j}$ are fresh in $@_\mathbf{a}@_\mathbf{j}[\pi]\varphi$
Interaction Axioms of H**PR**
$(\mathtt{Com}@)$ $@_\mathbf{i}@_\mathbf{a} p \leftrightarrow @_\mathbf{a}@_\mathbf{i} p$
$(\mathtt{Red}@_1)$ $\mathbf{a} \leftrightarrow @_\mathbf{i}\mathbf{a}$ $\quad(\mathtt{Red}@_2)$ $\mathbf{i} \leftrightarrow @_\mathbf{a}\mathbf{i}$
$(\mathtt{Dcom}\langle W\rangle@_2)$ $@_\mathbf{i}\langle\alpha\rangle p \leftrightarrow @_\mathbf{i}\langle\alpha\rangle@_\mathbf{i} p$ $\quad(\alpha \in \{1_W, \leq, \geq\})$
$(\mathtt{Com}\langle A\rangle@_1)$ $@_\mathbf{a}\langle\beta\rangle p \leftrightarrow \langle\beta\rangle@_\mathbf{a} p$ $\quad(\beta \in \{1_A, \sqsubseteq_\mathbf{k}, \sqsupseteq_\mathbf{k}\})$
Axioms for Atomic Programs of H**PR**
$(\mathtt{U}_W)$ $@_\mathbf{a}\langle 1_W\rangle\mathbf{b}$ $\quad(\mathtt{Cnv}_\leq)$ $@_\mathbf{a}\langle\leq\rangle\mathbf{b} \leftrightarrow @_\mathbf{b}\langle\geq\rangle\mathbf{a}$
$(\mathtt{U}_A)$ $@_\mathbf{i}\langle 1_A\rangle\mathbf{j}$ $\quad(\mathtt{Cnv}_\sqsubseteq)$ $@_\mathbf{i}\langle\sqsubseteq_\mathbf{k}\rangle\mathbf{j} \leftrightarrow @_\mathbf{j}\langle\sqsupseteq_\mathbf{k}\rangle\mathbf{i}$
$(\mathtt{Eq}_\sqsubseteq)$ $@_\mathbf{i}\mathbf{j} \to ([\sqsubseteq_\mathbf{i}]p \leftrightarrow [\sqsubseteq_\mathbf{j}]p)$
Axioms for Compounded Programs of H**PR**
$(\cup)$ $\langle\pi \cup \rho\rangle p \leftrightarrow \langle\pi\rangle p \vee \langle\rho\rangle p$
$(?)$ $\langle?(\varphi,\psi)\rangle p \leftrightarrow \varphi \wedge \langle 1_A\rangle\langle 1_W\rangle(\psi \wedge p)$
$(\cap)$ $@_\mathbf{a}@_\mathbf{i}\langle\pi \cap \rho\rangle(\mathbf{b} \wedge \mathbf{j}) \leftrightarrow @_\mathbf{a}@_\mathbf{i}(\langle\pi\rangle(\mathbf{b} \wedge \mathbf{j}) \wedge \langle\rho\rangle(\mathbf{b} \wedge \mathbf{j}))$
$(-_W)$ $@_\mathbf{a}\langle -\alpha\rangle\mathbf{b} \leftrightarrow @_\mathbf{a}\neg\langle\alpha\rangle\mathbf{b}$ $\quad(\alpha \in \{1_W, \leq, \geq\})$
$(-_A)$ $@_\mathbf{i}\langle -\beta\rangle\mathbf{j} \leftrightarrow @_\mathbf{i}\neg\langle\beta\rangle\mathbf{j}$ $\quad(\beta \in \{1_A, \sqsubseteq_\mathbf{k}, \sqsupseteq_\mathbf{k}\})$
$(\sqcap_\mathbf{i})$ $@_\mathbf{a}@_\mathbf{k}\langle(\pi,\mathbf{j}) \sqcap_\mathbf{i} (\pi',\mathbf{j}')\rangle(\mathbf{b} \wedge \mathbf{k}')$
$\leftrightarrow @_\mathbf{i}(\mathbf{k} \wedge \mathbf{k}') \wedge @_\mathbf{a}@_\mathbf{j}\langle\pi\rangle(\mathbf{b} \wedge \mathbf{j}) \wedge @_\mathbf{a}@_{\mathbf{j}'}\langle\pi'\rangle(\mathbf{b} \wedge \mathbf{j}'))$
Axioms for *PR*-frames of H**PR**
$(4_\leq)$ $@_\mathbf{a}\langle\leq\rangle\mathbf{b} \wedge @_\mathbf{b}\langle\leq\rangle\mathbf{c} \to @_\mathbf{a}\langle\leq\rangle\mathbf{c}$
$(4_\sqsubseteq)$ $@_\mathbf{j}\langle\sqsubseteq_\mathbf{i}\rangle\mathbf{k} \wedge @_\mathbf{k}\langle\sqsubseteq_\mathbf{i}\rangle\mathbf{l} \to @_\mathbf{j}\langle\sqsubseteq_\mathbf{i}\rangle\mathbf{l}$
$(\mathtt{Ref}_\leq)$ $@_\mathbf{a}\langle\leq\rangle\mathbf{a}$ $\quad(\mathtt{Cmp}_\sqsubseteq)$ $@_\mathbf{j}\langle\sqsubseteq_\mathbf{i}\rangle\mathbf{k} \vee @_\mathbf{k}\langle\sqsubseteq_\mathbf{i}\rangle\mathbf{j}$
Additional Axioms and Rules for H**PR**$_{(m,n)}$
$(\lvert W\rvert \leq m)$ $\bigvee_{0\leq k\neq l\leq m} @_{\mathbf{a}_k}\mathbf{a}_l$ $\quad(\lvert A\rvert \leq n)$ $\bigvee_{0\leq k\neq l\leq n} @_{\mathbf{i}_k}\mathbf{i}_l$
$(\lvert W\rvert \geq m)$ From $(\bigwedge_{1\leq k\neq l\leq m}\neg@_{\mathbf{a}_k}\mathbf{a}_l) \to \psi$ infer ψ,
where $\mathbf{a}_k$s are fresh in ψ
$(\lvert A\rvert \geq n)$ From $\left(\bigwedge_{1\leq k\neq l\leq n}\neg@_{\mathbf{i}_k}\mathbf{i}_l\right) \to \psi$ infer ψ,
where $\mathbf{i}_k$s are fresh in ψ.

earlier work, which was quite an artificial assumption on agents' reliabilities. Agent i's preference ordering *after* an 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 inspired by (van Benthem, 2007; Ghosh and Velázquez-Quesada, 2015a).

Definition 5. Given a set $X \subseteq A$ of agents, $u <_X v$ if $u <_k v$ holds for all agents $k \in X$. Moreover, $u \asymp_X v$ is used to mean $u <_X v$ or $v <_X u$ and $\mathrm{dom}(\asymp_X) := \{u \in A | u \asymp_X v \text{ for some } v \in A\}$.

Note that $\mathrm{dom}(\asymp_X)$ allows us to specify the connected components by the relation $\asymp_X$. Recall that $mr(i)$ denotes the set of all maximally reliable agents for i.

Definition 6 (Conservative Upgrade)**.** Agent i takes the strict preference ordering of her most reliable agents, and leaves the rest undecided (equipreferable). More precisely, the upgraded ordering $\leq_i'$ is defined by: $u \leq_i' v$ iff ($u <_{mr(i)} v$ or $u = v$) or ($u, v \notin \mathrm{dom}(\asymp_X)$).

Definition 7 (Radical Upgrade)**.** Agent i takes the strict preference ordering of her most reliable agents, and in the remaining disjoint zones she uses her old ordering. More precisely, the upgraded ordering $\leq_i'$ is defined by: $u \leq_i' v$ iff ($u <_{mr(i)} v$ or $u = v$) or ($u, v \notin \mathrm{dom}(\asymp_X)$ and $u \leq_i v$).

Note that both the conservative and radical upgrades preserve preorders (and thus upgraded models belong to our class of semantic models).

3.1 Expressing the preference dynamics

To formalize preference dynamics from the previous section, we add the following dynamic operators to the static syntax $\mathcal{HL}$. First of all, we regard all the agents involved in our two preference upgrade above as agent nominals (syntactic names of agents) and so let us denote agent i's syntactic name as $\mathbf{i}$ of boldface and the set of all syntactic names in $mr(i)$ as $\mathbf{mr(i)}$. $\mathcal{HL}_{\{\mathrm{pu}\}}$ is defined to be an expansion of $\mathcal{HL}$ with all operators $\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle$, where $\mathbf{i}$ be an agent-nominal and $\mathcal{R}$ is a list of sets of agent-nominals defined as $\mathcal{R} = \mathbf{mr(i)}$ (conservative upgrade) or $\mathcal{R} = (\mathbf{mr(i)}; \{\mathbf{i}\})$ (radical upgrade).

Definition 8 (Operators)**.** A formula $\mathrm{Req}(\mathcal{R})$, representing requirements for the list $\mathcal{R}$ is defined as the conjunction $\bigwedge_{\mathbf{j}\neq\mathbf{k}\in\mathbf{mr(i)}}\neg@_\mathbf{j}\mathbf{k}$ (mutual disjointness of agents involved in $mr(i)$) and $\bigwedge_{\mathbf{j}\in\mathbf{mr(i)}}\langle R_\mathbf{i}\rangle\mathbf{j}$ ($mr(i)$ is the set of maximally reliable agents for $\mathbf{i}$). Given a

PR-model $M = (W, \{\leq_i, \preccurlyeq_i\}_{i\in A}, V)$, define:

$$M,(w,j) \Vdash \langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi \quad \text{iff} \quad M,(w,j) \Vdash \mathrm{Req}(\mathcal{R}) \text{ and } \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}(M),(w,j) \Vdash \varphi,$$

where $\mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}(M)$ is the same model as M except $\leq_{\underline{\mathbf{i}}}$ is replaced by $\leq_{\mathcal{R}}$ where $\mathcal{R} = \mathbf{mr(i)}$ or $\mathcal{R} = (\mathbf{mr(i)};\{\mathbf{i}\})$ and corresponding $\leq_{\mathcal{R}}$'s are given by Definitions 6 and 7, respectively.

For an axiom system for the modality $\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle$, 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. In this case, the modalities can take the form of any relational expression. So we provide a 'matching' relational expression in the original model M by defining relational transformers similar to those in (Ghosh and Velázquez-Quesada, 2015a; Ghosh and Velázquez-Quesada, 2015b), in spirit of the program transformers of (van Benthem et al., 2006).

Before going into the notion of relational transformer, we have two observations. Firstly, when $\pi := ?\mathbf{j}\cap \leq$, we note that $(w,i)R_{\pi}(v,k)$ is equivalent to the conjunction of $i = k = \underline{\mathbf{j}}$ and $w \leq_{\underline{\mathbf{j}}} v$. Similarly, when we put $\pi' := ?\neg\mathbf{j}\cap\ \leq$, we remark that $(w,i)R_{\pi'}(v,k)$ is equivalent to the conjunction of $w \leq_i v$ and $i = k$ and $i \neq \underline{\mathbf{j}}$. Secondly, to reflect the relation $<_X$ in Definition 5, we need our program construction $(\pi,\mathbf{j})\sqcap_{\mathbf{i}}(\rho,\mathbf{k})$ to taking the intersection of (strict) preference relations of the possibly different agents than **i**. These observations allow us to capture the idea behind Definitions 6 and 7 syntactically in the following definition.

Definition 9 (Relational transformer)**.** Let us introduce the following abbreviations for relational expressions: We define $<_{\mathbf{mr(i)}} := \bigsqcap_{\mathbf{i}}\{(\leq\cap -\geq,\mathbf{j}) \mid \mathbf{j}\in\mathbf{mr(i)}\}$. Then $>_{\mathbf{mr(i)}}$ is similarly defined and $\succcurlyeq_{\mathbf{mr(i)}}$ is defined to be $<_{\mathbf{mr(i)}} \cup >_{\mathbf{mr(i)}}$. Moreover, a formula $\mathbf{d}(\succcurlyeq_{\mathbf{mr(i)}})$ is defined as $\langle\succcurlyeq_{\mathbf{mr(i)}}\rangle\top$.

A relational transformer $Tu^{\mathbf{i}}_{\mathcal{R}}$ is a function from relational expressions to relational expressions defined as follows. When $\mathcal{R} = \mathbf{mr(i)}$ (conservative upgrade),

$$Tu^{\mathbf{i}}_{\mathcal{R}}(\alpha) := \alpha \quad (\alpha \in \{1_A, 1_W, \sqsubseteq_{\mathbf{k}}, \sqsupseteq_{\mathbf{k}} \mid \mathbf{k}\in \mathsf{N}_2\}),$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}(\leq) := \Big(?\mathbf{i}\cap(<_{\mathbf{mr(i)}}\cup 1_A \cup ?\neg\mathbf{d}(\succcurlyeq_{\mathbf{mr(i)}}))\Big)\cup(?\neg\mathbf{i}\cap\leq)$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}(\leq) := \Big(?\mathbf{i}\cap(>_{\mathbf{mr(i)}}\cup 1_A \cup ?\neg\mathbf{d}(\succcurlyeq_{\mathbf{mr(i)}}))\Big)\cup(?\neg\mathbf{i}\cap\geq)$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}(\pi\cup\rho) := Tu^{\mathbf{i}}_{\mathcal{R}}(\pi)\cup Tu^{\mathbf{i}}_{\mathcal{R}}(\rho),$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}(\pi\cap\rho) := Tu^{\mathbf{i}}_{\mathcal{R}}(\pi)\cap Tu^{\mathbf{i}}_{\mathcal{R}}(\rho),$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}(?(\varphi,\psi)) := ?(\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi, \langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\psi).$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}((\pi,\mathbf{k})\sqcap_{\mathbf{j}}(\rho,\mathbf{l})) := ((Tu^{\mathbf{i}}_{\mathcal{R}}(\pi),\mathbf{k})\sqcap_{\mathbf{j}}(Tu^{\mathbf{i}}_{\mathcal{R}}(\rho),\mathbf{l}))$$
$$Tu^{\mathbf{i}}_{\mathcal{R}}(-\beta) := -Tu^{\mathbf{i}}_{\mathcal{R}}(\beta),$$

where $\beta \in \{1_W, \leq, \geq\}\cup\{1_A, \sqsubseteq_{\mathbf{k}}, \sqsupseteq_{\mathbf{k}} \mid \mathbf{k}\in\mathsf{N}_2\}$. When $\mathcal{R} = (\mathbf{mr(i)};\{\mathbf{i}\})$, we replace the occurrence of "$?\neg\mathbf{d}(\succcurlyeq_{\mathbf{mr(i)}})$" in $Tu^{\mathbf{i}}_{\mathcal{R}}(\leq)$ or $Tu^{\mathbf{i}}_{\mathcal{R}}(\geq)$ with

$$"?\neg\mathbf{d}(\succcurlyeq_{\mathbf{mr(i)}}))\langle\top\cap\leq'' \text{ or } "?\neg\mathbf{d}(\succcurlyeq_{\mathbf{mr(i)}})\cap\leq,''$$

respectively.

Theorem 2. The axioms and rules below together with those of H**PR** (or, those of H$\mathbf{PR}_{(m,n)}$) provide sound and complete axiom systems for $\mathcal{HL}_{\{\mathrm{pu}\}}$ with respect to possibly infinite *PR* models (or, *PR* models with m worlds and n agents, respectively).

$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle p \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge p,$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle(\varphi\vee\psi) \leftrightarrow \langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi\vee\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\psi,$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\neg\varphi \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge\neg\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\mathbf{j} \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge\mathbf{j}, \quad \langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\mathbf{a} \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge\mathbf{a},$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle@_{\mathbf{j}}\varphi \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge@_{\mathbf{j}}\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi,$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle@_{\mathbf{a}}\varphi \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge@_{\mathbf{a}}\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\langle\pi\rangle\varphi \leftrightarrow \mathrm{Req}(\mathcal{R})\wedge\langle Tu^{\mathbf{i}}_{\mathcal{R}}(\pi)\rangle\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi,$$
From $\varphi\to\psi$, we may infer $\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\varphi \to \langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle\psi$.

Proof. Soundness of the new axioms are straightforward. Completeness follows from the completeness of the static system H**PR** (cf. Chapter 7 of (van Ditmarsch et al., 2008), for an extensive explanation of this technique). □

Example 3. In our running example of Section 1, each agent is regarded to employ conservative upgrades to change his or her preference. Let us write the corresponding upgrade operators of $\{i,j,k\}$ by $\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}_i}\rangle$, $\langle \mathrm{pu}^{\mathbf{j}}_{\mathcal{R}_j}\rangle$, $\langle \mathrm{pu}^{\mathbf{k}}_{\mathcal{R}_k}\rangle$, respectively. Then, three flatmates did not reach an agreement after conservative upgrades of all agents, i.e.,

$$@_{\mathbf{i}}B\mathbf{a}_r \wedge @_{\mathbf{j}}B\mathbf{a}_l \wedge @_{\mathbf{k}}B\mathbf{a}_r \wedge$$
$$\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}_i}\rangle\langle \mathrm{pu}^{\mathbf{j}}_{\mathcal{R}_j}\rangle\langle \mathrm{pu}^{\mathbf{k}}_{\mathcal{R}_k}\rangle(@_{\mathbf{i}}B\mathbf{a}_l \wedge @_{\mathbf{j}}B\mathbf{a}_r \wedge @_{\mathbf{k}}B\mathbf{a}_r).$$

is valid in M_{exp}, because upgraded preferences are given by $w_r <'_i w_l$, $w_l <'_j w_r$ and $w_l <'_k w_r$.

4 RELIABILITY DYNAMICS

A public announcement of the agents' individual preferences may change the agents' reliability attributions as well: for example, an agent may consider more reliable those agents whose preferences coincide (or, for some reason, differ) from her own. In such cases, agent i's new reliability ordering, $\preccurlyeq'_i$,

can be given in terms of the agents' current preferences, $\leq_1, \ldots, \leq_n$, and i's current reliability ordering, $\preccurlyeq_i$. Thus, $\leq_i' := g(\leq_1, \ldots, \leq_n, \preccurlyeq_i)$ for some function g. We now provide formal definitions of some such possibilities.

4.1 Reliability change operations

The notion of "matching preference orders" will form the basis for the reliability dynamics. The idea is that two preference orderings match each other to a certain extent if the orderings are identical on some subset of the state space. A full match indicates that the orderings coincide on the whole domain; a partial match indicates that they coincide up to some proper subset of the domain.

Definition 10 (Matching preferences)**.** Let F be a *PR* frame given by $F = (W, \{\leq_i, \preccurlyeq_i\}_{i \in A})$ and let $i \in A$ be an agent. If $\leq_i$ is identical with $\leq_j$ on $W' \subseteq W$, then W' is said to be a set of match for i and j (notation: $\leq_i \sim_{W'} \leq_j$).

- Preference orders $\leq_i$ and $\leq_j$ are said to *fully match* each other iff $\leq_i \sim_W \leq_j$.[3] $FullMat(i)$ denotes the set of agents in $A \setminus \{i\}$ having full match with i.
- Preference orders $\leq_i$ and $\leq_j$ have *zero match* with each other iff there is no $W' \subseteq W$ with $|W'| \geq 2$ such that $\leq_i \sim_{W'} \leq_j$.[4] $ZeroMat(i)$ denotes the set of agents in $A \setminus \{i\}$ having zero match with i.

With these definitions, we can define some operations for reliability change.

Definition 11 (Full, Zero matching upgrade)**.** Agent i puts those agents that have full/zero match with her own preference ordering above those that do not, keeping her old reliability ordering within each of the two zones. More precisely, if $\preccurlyeq_i$ is agent i's current reliability ordering, then her new reliability ordering $\preccurlyeq_i'$ is defined by:

$$j \preccurlyeq_i' k \text{ iff } (j,k \in V \text{ and } j \preccurlyeq_i k) \text{ or } (k \in V \text{ and } j \notin V) \text{ or } (j,k \notin V \text{ and } j \preccurlyeq_i k).$$

Here $V = FullMat(i) \cup \{i\}, ZeroMat(i)$, respectively.

Once again, we can consider more generalized definitions for upgrade policies as well, but we just stick to simple definitions to give the main idea. Note that both the full matching and zero matching upgrades preserve total preorders (and thus upgraded models belong to our class of semantic models).

[3] Note how, by the finiteness of W (the reflexivity of the preference relations), there is always a maximal $X \subseteq W$ such that $\leq_i \sim_X \leq_j$ for every agent i, j.

[4] For the same reason, there is always a minimal $X \subseteq W$ such that $\leq_i \sim_X \leq_j$ for every agent i, j.

4.2 Expressing the reliablity dynamics

To describe reliability dynamics from the previous section, the following dynamic operators are added to the static syntax of $\mathcal{HL}$. $\mathcal{HL}_{\{\mathrm{rc}\}}$ is defined to be an expansion of $\mathcal{HL}$ with all operators $\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}} \rangle$, where $\mathbf{i}$ is an agent-nominal and $\mathcal{E}$ a pair of $\mathcal{HL}$-formulas of the form $@_{\mathbf{a}}\chi$ (recall: $\mathbf{a}$ is a world-nominal). An underlying semantic intuition for $\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}} \rangle$ is: Given a *PR*-model M, the pair $\mathcal{E} = (@_{\mathbf{a}_1}\chi_1, @_{\mathbf{a}_2}\chi_2)$ can be regarded as a partition (i.e., an equivalence relation on agents) in the sense that $(\{ i \in A \mid M, (i, \mathbf{a}_k) \Vdash \chi_k \})_{1 \leq k \leq 2}$ forms a partition of A, and the reliability ordering $\preccurlyeq_{\mathbf{i}}$ of the original *PR* model M is rewritten into the updated reliability ordering $\preccurlyeq_{\mathbf{i}}'$ as in Definition 11 of the Section 4.1.

Definition 12 (Operators)**.** Given any pair $\mathcal{E} = (\varphi_1, \varphi_2)$ of formulas of the form $@_{\mathbf{a}}\chi$, a formula $\mathrm{Eq}(\mathcal{E})$ is defined as the conjunction of $[1_A](\varphi_1 \vee \varphi_2)$ (exhaustiveness for agents) and $\neg\langle 1_A \rangle(\varphi_1 \wedge \varphi_2)$ (pairwise disjointness for agents). Given a pair $\mathcal{E} = (@_{\mathbf{a}_1}\chi_1, @_{\mathbf{a}_2}\chi_2)$ and a *PR*-model $M = (W, \{\leq_i, \preccurlyeq_i\}_{i \in A}, V)$, define:

$$M, (w, j) \Vdash \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}} \rangle \varphi \quad \text{iff} \quad M, (w, j) \Vdash \mathrm{Eq}(\mathcal{E}) \text{ and } \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}(M), (w, j) \Vdash \varphi,$$

where $\mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}(M)$ is the same model as M except $\preccurlyeq_{\mathbf{i}}$ is replaced by $\preccurlyeq_{\mathbf{i}}'$ of Definition 11.

Definition 13 (Relational transformer)**.** Let $\mathcal{E} = (\varphi_1, \varphi_2)$ be a pair. A relational transformer $Tr^{\mathbf{i}}_{\mathcal{E}}$ is a function from relational expressions to relational expressions defined as follows.

$$\begin{aligned}
&Tr^{\mathbf{i}}_{\mathcal{E}}(\alpha) := \alpha \quad (\alpha \in \{1_A, 1_W, \leq, \geq\}),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsubseteq_{\mathbf{i}}) := (\sqsubseteq_{\mathbf{i}} \cap (?\varphi_1 \cup ?\varphi_2)) \cup (1_A \cap ?(\varphi_1, \varphi_2)),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsupseteq_{\mathbf{i}}) := (\sqsupseteq_{\mathbf{i}} \cap (?\varphi_1 \cup ?\varphi_2) \cup (1_A \cap ?(\varphi_1, \varphi_2)),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsubseteq_{\mathbf{k}}) := (?@_{\mathbf{i}}\mathbf{k} \cap Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsubseteq_{\mathbf{i}})) \cup (?\neg@_{\mathbf{i}}\mathbf{k} \cap \sqsubseteq_{\mathbf{k}}) \ (\mathbf{k} \neq \mathbf{i}),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsupseteq_{\mathbf{k}}) := (?@_{\mathbf{i}}\mathbf{k} \cap Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsupseteq_{\mathbf{i}})) \cup (?\neg@_{\mathbf{i}}\mathbf{k} \cap \sqsupseteq_{\mathbf{k}}) \ (\mathbf{k} \neq \mathbf{i}),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(\pi \cup \rho) := Tr^{\mathbf{i}}_{\mathcal{E}}(\pi) \cup Tr^{\mathbf{i}}_{\mathcal{E}}(\rho),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(\pi \cap \rho) := Tr^{\mathbf{i}}_{\mathcal{E}}(\pi) \cap Tr^{\mathbf{i}}_{\mathcal{E}}(\rho),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(?(\varphi, \psi)) := ?(\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}} \rangle \varphi, \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}} \rangle \psi).\\
&Tr^{\mathbf{i}}_{\mathcal{E}}((\pi, \mathbf{k}) \sqcap_{\mathbf{j}} (\rho, \mathbf{k})) := (Tr^{\mathbf{i}}_{\mathcal{E}}(\pi), \mathbf{k}) \sqcap_{\mathbf{j}} (Tr^{\mathbf{i}}_{\mathcal{E}}(\rho), \mathbf{k}),\\
&Tr^{\mathbf{i}}_{\mathcal{E}}(-\alpha) := -Tr^{\mathbf{i}}_{\mathcal{E}}(\alpha),
\end{aligned}$$

where $\alpha \in \{1_W, \leq, \geq\} \cup \{1_A, \sqsubseteq_{\mathbf{k}}, \sqsupseteq_{\mathbf{k}} \mid \mathbf{k} \in \mathsf{N}_2\})$.

When $\mathbf{k} \neq \mathbf{i}$, i.e., $\mathbf{k}$ and $\mathbf{i}$ are syntactically distinct agent nominals, the reader may wonder why we should have generalized test operators "$?@_{\mathbf{i}}\mathbf{k}$" and "$?\neg@_{\mathbf{i}}\mathbf{k}$" in the definitions $Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsubseteq_{\mathbf{k}})$ and $Tr^{\mathbf{i}}_{\mathcal{E}}(\sqsubseteq_{\mathbf{k}})$. This is because the

same agent might have two distinct (syntactic) names. Based on a similar strategy for Theorem 2, we can now prove the following theorem.

Theorem 3. The axioms and rules below together with those of H**PR** (or, those of H**PR**$_{(m,n)}$) provide sound and complete axiom systems for $\mathcal{HL}_{\{\mathrm{rc}\}}$ with respect to possibly infinite *PR* models (or, *PR* models with m worlds and n agents, respectively).

$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle p \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge p,\ \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle(\varphi \vee \psi) \leftrightarrow \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\varphi \vee \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\psi,$$
$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\neg\varphi \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge \neg\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\varphi$$
$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\mathbf{j} \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge \mathbf{j},\quad \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\mathbf{a} \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge \mathbf{a},$$
$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle@_{\mathbf{j}}\varphi \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge @_{\mathbf{j}}\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\varphi,$$
$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle@_{\mathbf{a}}\varphi \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge @_{\mathbf{a}}\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\varphi$$
$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\langle\pi\rangle\varphi \leftrightarrow \mathrm{Eq}(\mathcal{E}) \wedge \langle Tr^{\mathbf{i}}_{\mathcal{E}}(\pi)\rangle\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\varphi,$$
From $\varphi \to \psi$, we may infer $\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\varphi \to \langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle\psi$.

Example 4. After Isabella and John know others' preferences, we regard, in our running example, that Isabella uses full-match reliability change $\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}_i}\rangle$ and John employs zero-match reliability change $\langle \mathrm{rc}^{\mathbf{j}}_{\mathcal{E}_j}\rangle$. Unlike Example 3, let us first consider reliability changes of Isabella and John and then take the conservative upgrades of all agents. This process and the resulting agreements among agents are formalized as

$$\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}_i}\rangle\langle \mathrm{rc}^{\mathbf{j}}_{\mathcal{E}_j}\rangle\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}_i}\rangle\langle \mathrm{pu}^{\mathbf{j}}_{\mathcal{R}_j}\rangle\langle \mathrm{pu}^{\mathbf{k}}_{\mathcal{R}_k}\rangle$$
$$(@_{\mathbf{i}}B\mathbf{a}_r \wedge @_{\mathbf{j}}B\mathbf{a}_r \wedge @_{\mathbf{k}}B\mathbf{a}_r),$$

which is valid in M_{exp}, because Isabella's reliability is changed into $j \prec'_i i \approx'_i k$ and John's reliability does not change.

We note here that while the main focus of the work is to model joint deliberation in form of simultaneous preference and reliability upgrades, the model operations and modalities of Sections 3.1 and 4.2 deal with single agent upgrades. This presentation style has been chosen in order to simplify notation and readability, but the provided definitions can be easily extended in order to match our goals. In particular, the model operations of Definitions 8 and 12 can be extended to simultaneous upgrades by asking for a list $\mathcal{R}$ of lexicographic lists (with $\mathcal{R}_i$ the list for agent i), and asking for a list $\mathcal{E}$ of partition lists (with $\mathcal{E}_i$ the list for agent i), respectively. Then the corresponding modalities, $\langle \mathrm{pu}^{\mathbf{i}}_{\mathcal{R}}\rangle$ and $\langle \mathrm{rc}^{\mathbf{i}}_{\mathcal{E}}\rangle$ can still be axiomatised by the presented system with some simple modifications.

5 CONCLUSION

This work continues the line of study in (Ghosh and Velázquez-Quesada, 2015a; Ghosh and Velázquez-Quesada, 2015b) and provides a further interplay between the preferences that the agents have about the world around and the reliability attributions they have with respect to one another. We deal with both preference change based on reliability, and reliability change based on preferences, and propose two-dimensional dynamic hybrid logics to express such changes. The main technical results that we have are sound and complete axiomatizations which lead to decidability (provided the numbers of agents and of states are *fixed* finite numbers) as well. In process, we also discuss about agent beliefs in such situations, e.g. relating reliability attributions with the notions of belief (cf. the running example in the text). The novel contribution of the work is the study of change in reliability attribution of agents based on their preferences.

To conclude, let us provide some pointers towards future work: (1) What other reasonable preference and reliability upgrade policies can there be and how to model them? (2) How to investigate the role of knowledge in such changes, especially if manipulation comes into play? (3) What would be the characterizing conditions for reaching consensus in such deliberative processes? We endeavor to provide answers to such questions in future. [5]

REFERENCES

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

Boutilier, C. (1994). Conditional logics of normality: A modal approach. *Artificial Intelligence*, 68(1):87–154.

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

Demolombe, R. (2001). To trust information sources: A proposal for a modal logic framework. In Castelfranchi, C. and Tan, Y.-H., editors, *Trust*

[5]The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. The work of the second author was partially supported by JSPS KAKENHI Grant-in-Aid for Young Scientists (B) Grant Number 15K21025 and JSPS Core-to-Core Program (A. Advanced Research Networks).

and Deception in Virtual Societies. Kluwer Academic, Dordrecht.

Demolombe, R. (2004). Reasoning about trust: A formal logical framework. In Jensen, C. D., Poslad, S., and Dimitrakos, T., editors, *iTrust*, volume 2995 of *Lecture Notes in Computer Science*, pages 291–303. Springer.

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

Falcone, R., Barber, K. S., Sabater-Mir, J., and Singh, M. P., editors (2008). *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.

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

Gargov, G., Passy, S., and Tinchev, T. (1987). Modal environment for boolean speculations, preliminary report. In Skordev, D., editor, *Mathematical Logic and Its Applications*, pages 253–263. Plenum Press.

Ghosh, S. and Velázquez-Quesada, F. R. (2015a). Agreeing to agree: Reaching unanimity via preference dynamics based on reliable agents. In Bordini, R., Elkind, E., Weiss, G., and Yolum, P., editors, *AAMAS 2015*, pages 1491–1499.

Ghosh, S. and Velázquez-Quesada, F. R. (2015b). A note on reliability-based preference dynamics. In van der Hoek, W., Holliday, W. H., and fan Wang, W., editors, *LORI 2015*, pages 129 –142.

Goldblatt, R. (1992). *Logics of Time and Computation*. Number 7 in CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, CA, 2nd edition.

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

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

Harel, D., Kozen, D., and Tiuryn, J. (2000). *Dynamic Logic*. MIT Press, Cambridge, MA.

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

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

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

Lorini, E., Jiang, G., and Perrussel, L. (2014). Trust-based belief change. In Schaub, T., editor, *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.

Marx, M. and Mikulás, S. (2001). Products, or how to create modal logics of high complexity. *Logic Journal of IGPL*, 9(1):71–82.

Rodenhäuser, B. (2014). *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. ILLC Dissertation Series DS-2014-04.

Sano, K. (2010). Axiomatizing hybrid products: How can we reason many-dimensionally in hybrid logic? *Journal of Applied Logic*, 8(4):459–474.

Seligman, J., Liu, F., and Girard, P. (2013). Knowledge, friendship and social announcement. In van Benthem, J. and Liu, F., editors, *Logic Across the University: Foundations and Applications*, volume 47 of *Studies in Logic*, pages 445–469. College Publications.

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

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

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