Multi-Agent Belief Revision
with Linked Preferences

Jan van Eijck[1] and Floor Sietsma[2]

[1] Centre for Mathematics and Computer Science (CWI)
Science Park 123, 1098 XG Amsterdam, The Netherlands

[2] ILLC, Plantage Muidergracht 24, 1018 TV Amsterdam, The Netherlands
`jve@cwi.nl`, `floorsietsma@gmail.com`

Abstract

In this paper we forge a connection between dynamic epistemic logics of belief revision on one hand and studies of collective judgement and multi-agent preference change on the other. Belief revision in the spirit of dynamic epistemic logic uses updating with relational substitutions to change the beliefs of individual agents. Collective judgement in social choice theory studies the collective outcomes of individual belief changes.

We start out from the logic of communication and change (LCC), which is basically epistemic propositional dynamic logic (PDL) extended with update action modalities. We show how epistemic PDL can be made to handle belief operators. Our new version of epistemic/doxastic PDL does not impose any constraints on the basic relations. Because of this it does not suffer from the drawback of LCC that constraints on epistemic relations, such as transitivity and reflexivity, may get lost under updates that are admitted by the system.

After introducing the base system without constraints, we study the effects of imposing a single constraint, namely the constraint that the agent's preference relations are linked. Linkedness is a natural extension of local connectedness to the multi-agent case. It assures that if Alice prefers y to x and Bob prefers z to x, then both Alice and Bob can make up their minds when given a choice between y and z. Since general belief changes may not preserve linkedness, we propose a recipe for belief change that does preserve it.

Finally, we show that the resulting framework can be used to model consensus seeking procedures. We focus on the case of plenary Dutch meetings. In Dutch meetings, a belief change (or rather: preference change) is performed for all agents in the meeting if a majority believes (or: is in favour of) the proposition that is under discussion. A special case of these meetings is judgement aggregation, and we apply our framework to the discursive dilemma in

this field. Our framework has obvious connections to coalition logic and social choice theory.

1   Introduction

In [17] the following problem in judgement aggregation is considered. There are three judges a, b, c with a, b agreeing that p, and b, c agreeing that q, so that both p and q command a majority, but $p \wedge q$ does not. The example shows that majority judgement is not closed under logical consequence. It seems that there is a collective belief in p and in q, but not in $p \wedge q$.

To see the relevance of the example for the practice of law, assume that p expresses that the defendant has done action X, and q expresses that the defendant is under a legal obligation not to do X. Then $p \wedge q$ expresses that the defendant has broken his contract not to do X. So two of the three judges agree that the defendant has done X, two of the three judges agree that the defendant is under a legal obligation not to do X, but there is no majority for the verdict of 'guilty'. This is a standard paradox in judgement aggregation called the discursive dilemma or doctrinal paradox.

The discursive dilemma is an example of a situation where multiple agents have different beliefs, and where a naive way of combining these beliefs — majority voting on each individual belief — leads to an implausible outcome. What seems to be needed is a more sophisticated way, or more sophisticated ways, to combine individual beliefs into collective beliefs.

We will present an epistemic/doxastic framework that can be used to model such situations, and present ways to update these frameworks with new beliefs. In the above example, this gives various protocols for judgement aggregation that are different from the above simple majority rule.

The framework of (our version of) doxastic/epistemic logic is quite abstract. It starts out from a basic preference relation for each agent, without any constraints on the preference relations. In particular, we do not insist that the preference relations are reflexive or transitive. The relations in our framework can be seen as plausibility relations, and used for modeling the beliefs of agents, as is the case in the above example. But we can also view them as modelling preference, and arrive at an interpretation that connects the framework to the area of social choice theory. Henceforth, we will stick to 'preference relation', but we do not intend to exclude other interpretations.

2   Preference Relations Without Constraints

A preference model $\mathbf{M}$ for set of agents Ag and set of basic propositions $Prop$ is a tuple (W, P, V) where W is a non-empty set of worlds, P is a function that maps each

agent a to a relation P_a (the preference relation for a), and V is a map from W to $\mathcal{P}(Prop)$ (a map that assigns to each world a *Prop*-valuation). The tuple (W, P) is called the frame of the model. There are no conditions at all on the P_a. Appropriate conditions will be imposed by constructing the operators for belief and knowledge by means of PDL operations.

We fix a PDL style language for talking about preference. Assume p ranges over a set of basic propositions *Prop* and a over a set of agents *Ag*. Then the language of epistemic/doxastic PDL is given by:

$$\phi \quad ::= \quad \top \mid p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid [\pi]\phi$$
$$\pi \quad ::= \quad a \mid a\breve{} \mid ?\phi \mid \pi_1; \pi_2 \mid \pi_1 \cup \pi_2 \mid \pi^*$$

We use *PROG* for the set of program expressions (expressions of the form π) of this language.

The intuitive meaning of $a\breve{}$ is 'converse of a', i.e., reversing the directions of the arrows. The construct $\pi_1; \pi_2$ expresses sequential composition of relations. In the standard interpretation of PDL, π^* expresses program iteration. In our epistemic/doxastic interpretation, it expresses transitive closure of a (possibly complex) preference relation.

The language is to be interpreted in the usual PDL manner, with $[\![\pi]\!]^{\mathbf{M}}$ giving the relation that interprets relational expression π in $\mathbf{M} = (W, V, P)$. $[\pi]\phi$ is true in world w of $\mathbf{M}$ if for all v with $(w, v) \in [\![\pi]\!]^{\mathbf{M}}$ it holds that ϕ is true in v. We adopt the usual abbreviations of $\bot$, $\phi_1 \vee \phi_2$, $\phi_1 \rightarrow \phi_2$, $\phi_1 \leftrightarrow \phi_2$ and $\langle\pi\rangle\phi$. The following additional abbreviations allow us to express knowledge, strong belief, conditional belief and plain belief:

knowledge $[\sim_a]\phi$ abbreviates $[(a \cup a\breve{})^*]\phi$.

> The formula $[\sim_a]\phi$ expresses that agent a knows that ϕ.

strong belief $[\geq_a]\phi$ abbreviates $[a^*]\phi$.

> The formula $[\geq_a]\phi$ expresses that agent a strongly believes that ϕ. The only difference between knowledge and strong belief is that knowledge has the property of negative introspection (if a does not know ϕ, then a knows that she does not know ϕ), while strong belief lacks this property.

conditional belief $[\rightarrow_a^\phi]\psi$ abbreviates $\langle\sim_a\rangle\phi \rightarrow \langle\sim_a\rangle(\phi \wedge [\geq_a](\phi \rightarrow \psi))$.

> The formula $[\rightarrow_a^\phi]\psi$ expresses that in all of the agent's preferred ϕ worlds, ψ holds.

plain belief $[\to_a]\psi$ abbreviates $[\to_a^\top]\psi$.

The formula $[\to_a]\psi$ expresses that in all of the agent's preferred worlds, ψ holds. Note that this can be expressed equivelently as $\langle\sim_a\rangle[\geq_a]\phi$.

The definition of $\to_a^\phi$ (conditional belief for a, with condition ϕ) is from Boutillier [10]. This definition, also used in [4], states that conditional to ϕ, a believes in ψ if either there are no accessible ϕ worlds, or there is an accessible ϕ world in which the belief in $\phi \to \psi$ is safe. The definition of $\to_a^\phi$ matches the well-known accessibility relations $\to_a^P$ for each subset P of the domain, given by:

$$\to_a^P := \{(x,y) \mid x\sim_a y \wedge y \in \mathrm{MIN}_{\leq_a} P\},$$

where $\mathrm{MIN}_{\leq_a} P$, the set of minimal elements of P under $\leq_a$, is defined as

$$\{s \in P : \forall s' \in P(s' \leq_a s \Rightarrow s \leq_a s')\}.$$

This logic is completely axiomatized by the standard PDL rules and axioms ([21; 16]) plus the following axioms that describe the relation between the basic programs a and $a\breve{}$:

$$\vdash \phi \to [a]\langle a\breve{}\rangle\phi \qquad \vdash \phi \to [a\breve{}]\langle a\rangle\phi$$

Any preference relation P_a can be turned into a pre-order by taking its reflexive transitive closure $P_a{}^*$. The abbreviation for strong belief introduces the $\geq_a$ as names for these pre-orders. The knowledge abbreviation introduces the $\sim_a$ as names for the equivalences given by $(P_a \cup P_a\breve{})^*$.

If the P_a are well-founded, $\mathrm{MIN}_{\leq_a} P$ will be non-empty for non-empty P. The canonical model construction for PDL yields finite models. Therefore, the P_a relations in the canonical model will be well-founded, for each relation on a finite model is well-founded. Thus, there is no need to impose well-foundedness as a relational condition. We get it for free by the fact that PDL has the finite model property.

This yields a very expressive complete and decidable PDL logic for preference, to which we can add mechanisms for belief update and for belief revision.

Our logic is able to express knowledge and strong belief as single modalities (as $(a\cup a\breve{})^*$ and a^*, respectively). We were unable to find single modalities for conditional belief and plain belief, however. Note that expressing conditional belief *in terms of single unary modalities* is different from expressing conditional belief *as a single unary modality*. So we have the following conjecture, which we were unable to prove:

Conjecture It is impossible to describe conditional belief and plain belief as single unary modalities in converse PDL.

Intermezzo: Conditional Belief, Complement and Intersection

Instead of using PDL with converse as a basis for constructing logics of belief revision, one might wish to use PDL with converse and complement, i.e., the result of adding operations $a^{\smile}$ and $\overline{\pi}$ to the language, with $\overline{\pi}$ to be interpreted as relational complementation. Unfortunately, the resulting logic is undecidable [15]. Still, the advantage of having complement available is that it allows us to express all our knowledge and belief constructs as *single* unary modalities, as follows:

unary modality for knowledge $(a \cup a^{\smile})^*$.

unary modality for strong belief a^*.

unary modality for conditional belief $(a \cup a^{\smile})^*; ?(\phi \wedge [a^* - a^{*\smile}]\neg\phi)$.

unary modality for plain belief $(a \cup a^{\smile})^*; ?[a^* - a^{*\smile}]\bot$.

This uses $a^* - a^{*\smile}$, which can be expressed with complement as $\overline{\overline{a^*} \cup a^{*\smile}}$. Note that $a^* - a^{*\smile}$ represents the moves to a strictly a-better world: w is related to w' by $a^* - a^{*\smile}$ iff there is an a-path from w to w' but not vice versa. It is easy to see that $?(\phi \wedge [a^* - a^{*\smile}]\neg\phi)$ is true in a world w iff w is an a-best ϕ world.

But perhaps this shift to an undecidable logic is unnecessary. Maybe adding intersection and converse to PDL is enough. Intersection PDL with converse is like PDL, but with a construct for atomic converse and an operation $\pi_1 \cap \pi_2$ for relation intersection. Like PDL with converse, intersection PDL with converse is decidable [18]. Here is an open question:

Question Is it possible to describe conditional belief and plain belief as single modalities in intersection PDL with converse?

3 Belief Revision with Linked Preference Relations

The preference relations that serve as the basis for construction of a preference preorder in Section 2 leave something to be desired. Compare an optometrist who collects answers for a number of lenses she tries out on you: "Better or worse?", (change of lens), "Better or worse?" (change of lens), "Better or worse?".... If you reply "worse" after a change of x to y, and "worse" after a change from y to z, she will most probably not bother to collect your reaction to a change from x to z. But what if you answer "better" after the second swap? Then, if she is reasonable, she will try to find out how x compares to z. It does make sense to impose this as a requirement on preference relations.

There are several ways to do this. Recall that we did not impose a requirement of transitivity on the basic preference relations. Here is a definition that does not imply transitivity, but yields that the transitive closures of the basic preferences are well-behaved. See Figure 1. The intended meaning is that if the black arrows are present, either the grey arrow on the left or the grey arrow on the right has to be present as well.
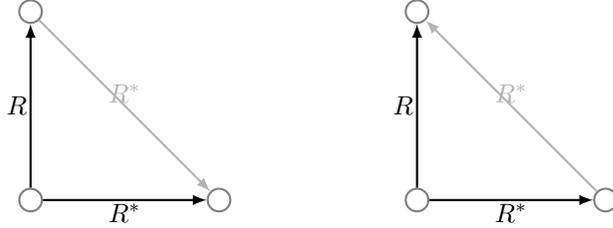

Figure 1. Forward linkedness of a binary relation R

Definition 3.1. A binary relation R is **forward linked** if the following holds:

$$\forall x, y, z((xRy \wedge xR^*z) \rightarrow (yR^*z \vee zR^*y)).$$

R is **linked** if both R and $R^{\smile}$ are forward linked.

Note that this is different from the notion of **weak connectedness** (terminology of [13]): a relation R is weakly connected if $\forall x, y, z((xRy \wedge xRz) \rightarrow (yRz \vee y = z \vee zRy))$.

Theorem 3.2. R is forward linked iff R^* is weakly connected.

Proof. The right to left direction is immediate. For the left to right direction, assume R is forward linked. Let wR^*w_1 and wR^*w_2. Then there is an $n \in \mathbb{N}$ with wR^nw_1. We prove the claim by induction on n. If $n = 0$ then $w = w_1$ and $w_1R^*w_2$, and we are done. Otherwise, assume the claim holds for n. We have to show it holds for $n + 1$. Suppose $wR^{n+1}w_1$. Then for some w', $wRw'R^nw_1$. By forward linking of R, either $w'R^*w_2$ or w_2R^*w'. In the first case, use the induction hypothesis to get $w_1R^*w_2$ or $w_2R^*w_1$. In the second case, it follows from w_2R^*w' and $w'R^nw_1$ that $w_2R^*w_1$. Q.E.D.

Starting from relations that are linked, one can use the method of constructing knowledge and belief relations by means of the regular operations that was presented in the previous section to construct 'preference models' in the style of Grove

[14], Board [9], and Baltag and Smets [3; 4] (who call them 'multi-agent plausibility frames').

It is well-known that the following principle characterizes weak connectedness of P_a (cf. [13]):

$$[a]((\phi \wedge [a]\phi) \to \psi) \vee [a]((\psi \wedge [a]\psi) \to \phi).$$

The notion of forward linking is characterized by:

$$[a]((\phi \wedge [a^*]\phi) \to \psi) \vee [a^*]((\psi \wedge [a^*]\psi) \to \phi). \tag{*}$$

Theorem 3.3. Principle (*) holds in a belief revision frame iff P_a is forward linked.

Proof. Let (W, P) be a frame where P_a is forward linked, and let $\mathbf{M} = (W, P, V)$ be some model based on the frame. We show that (*) holds. Let w be a world in $\mathbf{M}$. Assume $\mathbf{M} \not\models_w [a]((\phi \wedge [a^*]\phi) \to \psi)$. We have to show: $\mathbf{M} \models_w [a^*]((\psi \wedge [a^*]\psi) \to \phi)$. From the fact that $\mathbf{M} \not\models_w [a]((\phi \wedge [a^*]\phi) \to \psi)$, we get that there is a world w_1 with $wP_a w_1$ and $\mathbf{M} \models_{w_1} \phi \wedge [a^*]\phi \wedge \neg\psi$.

Let w_2 be an arbitrary world with $wP_a^* w_2$. Then by forward linking of P_a, either $w_1 P_a^* w_2$ or $w_2 P_a^* w_1$. In the first case, it follows from $\mathbf{M} \models_{w_1} [a^*]\phi$ that $\mathbf{M} \models_{w_2} \phi$, and therefore $\mathbf{M} \models_{w_2} (\psi \wedge [a^*]\psi) \to \phi$. In the second case, it follows from $\mathbf{M} \models_{w_1} \neg\psi$ that $\mathbf{M} \models_{w_2} \neg[a^*]\psi$, and therefore $\mathbf{M} \models_{w_2} (\psi \wedge [a^*]\psi) \to \phi$. So in both cases, $\mathbf{M} \models_{w_2} (\psi \wedge [a^*]\psi) \to \phi$, and since w_2 was an arbitrary world with $wP_a^* w_2$, it follows that $\mathbf{M} \models_w [a^*]((\psi \wedge [a^*]\psi) \to \phi)$.

Next, assume a frame (W, P) where P_a is not forward linked. We will construct a model $\mathbf{M} = (W, P, V)$ and an instance of (*) that does not hold. If P_a is not forward linked, there are w, w_1, w_2 with $wP_a w_1$, $wP_a^* w_2$, and neither $w_1 P_a^* w_2$ nor $w_2 P_a^* w_1$. We proceed to make an instance of (*) fail in w. Construct the valuation of $\mathbf{M}$ by setting p true in w_1 and in all worlds w' with $w_1 P_a^* w'$, and false everywhere else, and setting q true in w_2 and in all worlds w'' with $w_2 P_a^* w''$, and false everywhere else. Note that since not $w_1 P_a^* w_2$, p will be false in w_2, and that since not $w_2 P_a^* w_1$, q will be false in w_1. So we get $\mathbf{M} \models_{w_1} p \wedge [a^*]p \wedge \neg q$ and $\mathbf{M} \models_{w_2} q \wedge [a^*]q \wedge \neg p$. It follows that

$$\mathbf{M} \models_w \langle a \rangle(p \wedge [a^*]p \wedge \neg q) \wedge \langle a^* \rangle(q \wedge [a^*]q \wedge \neg p),$$

i.e.,

$$\mathbf{M} \not\models_w [a]((p \wedge [a^*]p) \to q) \vee [a^*]((q \wedge [a^*]q) \to p),$$

showing that this instance of (*) does not hold in $\mathbf{M}$.   Q.E.D.

Remark. As one of the referees pointed out, (*) is a so-called Sahlqvist formula. Such formulas are canonical for the properties they define. Canonicity of (*) for the

property of forward linkedness means that the canonical frame for a normal modal logic containing (*) is forward linked, and that (*) is valid on any class of forward linked frames. See [8] for further details.

In the multi-agent case there is a further natural constraint. Consider a situation where Alice and Bob have to decide on the chairperson of a program committee. Carol is mediator. Alice says she prefers y to x. Bob counters by saying that he prefers z to x. What should Carol do? Clearly, she should urge *both* of them to compare y and z.

Translating this example to our logic of belief, we want to require that if $x \geq_a y$ and $x \geq_b z$, then either $y \geq_a z$ or $z \geq_a y$ and either $y \geq_b z$ or $z \geq_b y$.

Definition 3.4 (Connectedness). A set of binary relations $\mathbf{R}$ on a domain W is **forward connected** if for all R, S in $\mathbf{R}$,

$$\text{if } xRy \text{ and } xSz, \text{ then either } ySz \text{ or } zSy.$$

$\mathbf{R}$ is **backward connected** if the set $\{R^\smile \mid R \in \mathbf{R}\}$ is forward connected. $\mathbf{R}$ is **connected** if $\mathbf{R}$ is both forward and backward connected.

Clearly, our motivating requirement that if $x \geq_a y$ and $x \geq_b z$ then either $y \geq_a z$ or $z \geq_a y$ expresses that $\leq_a$ and $\leq_b$ are forward connected.

Another natural requirement is that if Alice prefers x to y and Bob prefers x to z, then Alice and Bob should both compare y and z. If $x \leq_a y$ and $x \leq_b z$ then either $y \leq_a z$ or $z \leq_a y$ (and similarly for $\leq_b$). This expresses that $\leq_a$ and $\leq_b$ are backward connected.

The following definition extends the notion of linkedness to the multi-agent case (see Figure 2).
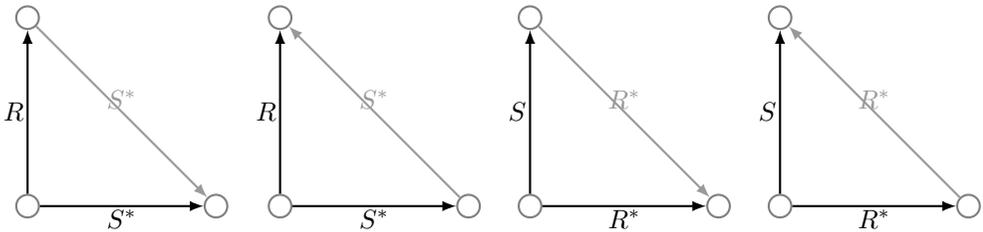

Figure 2. Forward linkedness of a pair of relations $\{R, S\}$.

Definition 3.5 (Linkedness). A set of binary relations $\mathbf{R}$ on a domain W is **forward linked** if for all R, S in $\mathbf{R}$, if xRy and xS^*z, then either yS^*z or zS^*y. $\mathbf{R}$ is **backward linked** if the set $\{R^{\smile} \mid R \in \mathbf{R}\}$ is forward linked. $\mathbf{R}$ is **linked** if $\mathbf{R}$ is both forward and backward linked.

It follows from Definition 3.5 that the set $\{R\}$ is forward linked iff R is forward linked according to Definition 3.1. So Definition 3.5 gives a natural extension of linking (and of local connectedness) to the multi-agent case.

The following theorem shows that Definition 3.5 can indeed replace Definition 3.4.

Theorem 3.6. R and S are linked iff R^* and S^* are connected.

Proof. The right to left direction is immediate.

For the left to right direction, assume R and S are linked. Suppose xR^*y and xS^*z. We will prove that for any w on the R path from x to y, either wS^*z or zS^*w. This clearly holds for $w = x$. Suppose w is the successor of w' on the path, and the result holds for w'. Suppose $w'S^*z$. Since $w'Rw$ the result holds by forward linking of R and S. Suppose yS^*w'. From $w'Rw$ and $w'S^*w'$ we get that either $w'S^*w$ or wS^*w'. In the first case trivially yS^*w. Assume the second case. We will show that for any v on the S path from w to w', either vS^*y or yS^*v. This clearly holds for $v = w'$. Suppose v is the predecessor of v' on the path, and the result holds for v'. Suppose $v'S^*y$. Then since vSv', trivially vS^*y. Suppose yS^*v'. Then the result holds by backward linking of S.

This shows that R^* and S^* are forward connected. The proof that R^* and S^* are backward connected is similar. $\qquad$ Q.E.D.

If R and S are preference relations, $(R \cup R^{\smile})^*$ and $(S \cup S^{\smile})^*$ are knowledge relations. The common knowledge relation (for the agents corresponding to R and S) is the reflexive transitive closure of the two knowledge relations, so common knowledge is expressed by:

$$((R \cup R^{\smile})^* \cup (S \cup S^{\smile})^*)^*$$

This is easily seen to be equivalent to:

$$(R \cup R^{\smile} \cup S \cup S^{\smile})^*$$

As we will now show, if R and S are *linked* relations, then this can be simplified even further. If R and S are linked then common knowledge equals the union of strong common belief and strong reverse common belief:

Theorem 3.7. If R and S are linked, then

$$(R \cup R^{\smile} \cup S \cup S^{\smile})^* = (R \cup S)^* \cup (R^{\smile} \cup S^{\smile})^*.$$

Proof. The inclusion from right to left is obvious. For the inclusion from left to right, assume $x(R \cup R^\smile \cup S \cup S^\smile)^* y$. Letting X and Y range over R and S, observe that each $X \circ Y^{\smile *}$ link can be replaced by either a Y^* or a $Y^{\smile *}$ link, and similarly for $X^\smile \circ Y^*$ links, by linking of R and S. Continuing this process until all one-step links are of the form $R \cup S$ or of the form $R^\smile \cup S^\smile$, this yields $x(R \cup S)^* y$ or $x(R^\smile \cup S^\smile)^* y$. Q.E.D.

The theorem shows that linking of relations simplifies the notion of common knowledge.

The modal characterization of relation linking is given by:

$$[a]((\phi \wedge [b^*]\phi) \rightarrow \psi) \vee [b^*]((\psi \wedge [b^*]\psi) \rightarrow \phi) \qquad \text{(LINK)}$$

Theorem 3.8. The set of LINK principles (with a, b ranging over the set of all agents) holds in a preference model iff the basic preference relations in the model are forward linked.

Proof. Analogous to the proof of Theorem 3.3. Q.E.D.

4 Belief Update and Belief Change

For the definition of action models $\mathbf{A}$ and of the update product operation $\otimes$ that combines an epistemic model and an action model, to produce an updated epistemic model, we refer to [2]. In [7] it is shown how extending the PDL language with an extra modality $[\mathbf{A}, e]\phi$, where $\mathbf{A}$ is an action model and e is a designated action of $\mathbf{A}$, does not change its expressive power. Interpretation of the new modality: $[\mathbf{A}, e]\phi$ is true in w in $\mathbf{M}$ if success of the update of $\mathbf{M}$ with action model $\mathbf{A}$ to $\mathbf{M} \otimes \mathbf{A}$ implies that ϕ is true in (w, e) in $\mathbf{M} \otimes \mathbf{A}$. The logic of PDL with action modalities added was called LCC or logic of communication and change.

Figure 3. Belief about h, belief update, and its result.

LCC, as proposed in [7], has a design flaw: it starts out from individual agent accessibity relations with the appropriate properties for knowledge or for belief. But updates with non S5 action models may destroy these properties: the update product of an S5 model with a non-S5 action model is in general not an S5 model. This

means that the resulting relations cannot be interpreted as epistemic accessibilities any longer. Similarly, if one interprets the agent accessibilities as belief. A KD45 model can easily change into a non KD45 model by an appropriate action update. This issue is remedied in [12], where it was first proposed to construct the relational properties for knowledge from more basic relations, by means of the PDL operations. The present paper proposes the same for (conditional) belief. Action update by means of the product construction can now be viewed as belief update. In Figure 3 belief update about a coin toss is modelled. Agent a initially believes $\neg h$, while h (heads) is actually the case,. for h is true in the actual world, shown in grey. Next, h is announced. The result of the belief update is that a now believes h.

Belief change is something different from belief update. Update can never reverse the directions of relational arrows, or introduce new arrows, for the arrows in the update result are the arrows that are present in the situation before the update and in the action model. Belief change is something more radical than this: replacing existing preference relations by new ones.

Here, we will focus on belief change rather than belief update. Belief change can be compared to factual change. A coin shows head, you toss it again, and now it shows tails. This can be modeled as a propositional substitution $\{h \mapsto \neg h\}$. Similarly for belief change. Agent a prefers x to y, she changes her preference, and now she prefers y to x. Or suppose she reverses *all* her preferences. This can also be handled as a substitution, $\{a \mapsto a^{\smile}\}$.

In [7], it was proposed to handle factual change by propositional substitution. Relational substitutions were proposed for belief change in [5], and it was shown in [11] that adding relational substitutions for preference change to epistemic PDL makes no difference for expressive power: the resulting system still reduces to PDL.

A preference substitution is a map from agents to $PROG$ expressions that can be represented by a finite set of bindings

$$\{a_1 \mapsto \pi_1, \ldots, a_n \mapsto \pi_n\}$$

where the a_j are agents, all different, and where the π_i are $PROG$ expressions. It is assumed that each a that does not occur in the left hand side of a binding is mapped to a. Call the set $\{a \in Ag \mid \rho(a) \neq a\}$ the *domain* of ρ. If $\mathbf{M} = (W, P, V)$ is a preference model and ρ is a preference substitution, then $\mathbf{M}^{\rho}$ is the result of changing the preference map P of $\mathbf{M}$ to P^{ρ} given by:

$$P^{\rho}(a) \quad := \quad \begin{cases} P_a \text{ for } a \text{ not in the domain of } \rho, \\ [\![\rho(a)]\!]^{\mathbf{M}} \text{ for } a \text{ in the domain of } \rho. \end{cases}$$

Now extend our PDL language with a modality $[\rho]\phi$ for preference change, with the

following interpretation:

$$\mathbf{M}, w \models [\rho]\phi \quad :\Longleftrightarrow \quad \mathbf{M}^\rho, w \models \phi.$$

An important thing to note is that since there are constraints on the preferences P_a (they are linked), we need to ensure that the belief changing substitutions satisfy these constraints. We leave the general problem of finding a precise characterization of linking-preserving substitutions for future research. Here we focus on a particular kind of belief change. Consider the suggestive upgrade $\sharp_a\phi$ discussed in Van Benthem and Liu [6]:

$$\sharp_a\phi =_{\text{def}} ?\phi; a; ?\phi \ \cup ?\neg\phi; a; ?\neg\phi \ \cup ?\neg\phi; a; ?\phi.$$

This is a variation on what is called lexicographic upgrade in the belief revision community (see e.g., [19]). The suggestive upgrade removes all relations from ϕ-worlds to $\neg\phi$-worlds. Clearly, belief revisions with suggestive upgrade do not preserve linking of relations. For consider a case where wP_aw_1 and $wP_a^*w_2$ and $w_1P_a^*w_2$, with ϕ true in w_1 but not in w and w_2. Then suggestive upgrade will remove the a-path from w_1 to w_2 (see Figure 4).
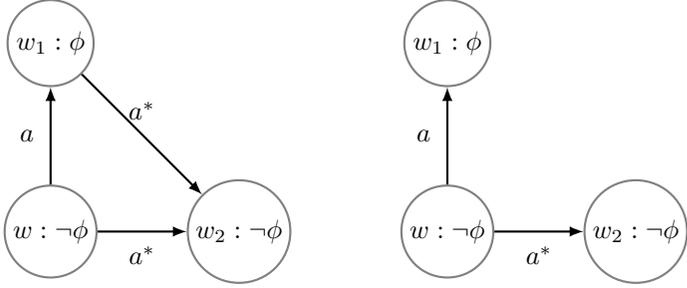

Figure 4. Suggestive upgrade does not preserve linking.

However, if we revise the upgrade procedure so that it adds extra links instead of removing them, as follows, we get a variation that preserves linking:

$$\natural_a\phi =_{\text{def}} ?\phi; a^*; ?\phi \ \cup ?\neg\phi; a^*; ?\neg\phi \ \cup ?\neg\phi; (a^* \cup a^{\smile *}); ?\phi.$$

Thus, links from ϕ worlds x to $\neg\phi$ worlds y get reversed, and extra links to x get added to 'support' the new link from y to x. Moreover, ϕ to ϕ links and $\neg\phi$ to $\neg\phi$ links are strengthened by means of transitive closure, in order to deal with the problem of detours through worlds that assign a different truth value to ϕ.

We can prove the following.

Theorem 4.1. If $\mathbf{M} = (W, P, V)$ is a preference model where P_a and P_b are linked, and ϕ is a PDL formula, then $[\![\natural_a\phi]\!]^{\mathbf{M}}$ and P_b are also linked.

Proof. Write a for P_a, b for P_b, and $\natural_a\phi$ for $[\![\natural_a\phi]\!]^{\mathbf{M}}$. First note that for any worlds x and y, if xa^*y then either $x(\natural_a\phi)y$ or $y(\natural_a\phi)x$.

Suppose xby and $x(\natural_a\phi)^*z$. We will show that either wa^*y or ya^*w for all w on the path from x to z. Firstly let $w = x$. Since xby and xa^*x, either xa^*y or ya^*x by linking of a and b. Now let w' be the predecessor of w on the path, so $x(\natural_a\phi)^*w'$ and $w'(\natural_a\phi)w$. Suppose either ya^*w' or $w'a^*y$. Since $w'(\natural_a\phi)w$, either $w'a^*w$ or wa^*w'. If ya^*w' and $w'a^*w$ or wa^*w' and $w'a^*y$, then trivially ya^*w or wa^*y. Suppose $w'a^*y$ and $w'a^*w$. By forward linking of a and Theorem 3.2, wa^*y or ya^*w. Suppose ya^*w' and wa^*w'. By backward linking of a and Theorem 3.2, ya^*w or wa^*y. So then for any w on the path wa^*y or ya^*w, so za^*y or ya^*z, so $z(\natural_a\phi)y$ or $y(\natural_a\phi)z$.

Suppose $x(\natural_a\phi)y$ and xb^*z. Then either xa^*y or ya^*x. In the first case the result follows by Theorem 3.6. Suppose ya^*x. We will show that for any w on the path from x to z, yb^*w or wb^*y. Firstly let $w = x$. ya^*x and yb^*y so by Theorem 3.6 the result holds. Suppose w' is the predecessor of w on the path and the result holds for w'. Suppose yb^*w'. Then since $w'bw$, trivially yb^*w. Suppose $w'b^*y$. Then the result holds by linkedness of b.   Q.E.D.

Now call a substitution where all bindings are of the form $a \mapsto \natural_a\phi$ a linked substitution. Then we get a complete logic for belief change with linked substitutions, by means of reduction axioms that 'compile out' the belief changes (see [11]):

Theorem 4.2. The logic of epistemic preference PDL with belief change modalities for linked substitutions is complete.

Proof. The preference change effects of $[\rho]$ can be captured by a set of reduction axioms for $[\rho]$ that make $[\rho]$ act as identity on basic propositions,

$$\begin{aligned}
[\rho]\top &\leftrightarrow \top \\
[\rho]p &\leftrightarrow p,
\end{aligned}$$

make $[\rho]$ commute with the boolean operators,

$$\begin{aligned}
[\rho]\neg\phi &\leftrightarrow \neg[\rho]\phi \\
[\rho](\phi_1 \wedge \phi_2) &\leftrightarrow [\rho]\phi_1 \wedge [\rho]\phi_2
\end{aligned}$$

and that handle formulas of the form $[\rho][\pi]\phi$ by means of reduction axioms of the form

$$[\rho][\pi]\phi \quad \leftrightarrow \quad [F_\rho(\pi)][\rho]\phi,$$

with F_ρ given by:

$$
\begin{array}{lcl}
F_\rho(a) & := & \rho(a) \\
F_\rho(?\phi) & := & ?[\rho]\phi, \\
F_\rho(\pi_1; \pi_2) & := & F_\rho(\pi_1); F_\rho(\pi_2), \\
F_\rho(\pi_1 \cup \pi_2) & := & F_\rho(\pi_1) \cup F_\rho(\pi_2), \\
F_\rho(\pi^*) & := & (F_\rho(\pi))^*.
\end{array}
$$

It is easily checked that these reduction axioms are sound, and that for each formula of the extended language the axioms yield an equivalent formula in which $[\rho]$ occurs with lower complexity, which means that the reduction axioms can be used to translate formulas of the extended language to PDL formulas. Completeness then follows from the completeness of PDL. Q.E.D.

5 Analyzing Plenary Dutch Meetings

A plenary Dutch meeting (Dutch: 'Vergadering') is a simultaneous preference/belief change event where the following happens. Assume an epistemic situation $\mathbf{M}$ with actual world w, and assume proposition ϕ is on the agenda.

- If a majority prefers ϕ to $\neg\phi$, i.e., if

$$
|\{i \in Ag \mid \mathbf{M} \models_w [\rightarrow_i]\phi\}| \;>\; |\{i \in Ag \mid \mathbf{M} \models_w [\rightarrow_i]\neg\phi\}|
$$

  then simultaneous belief change $\{i \mapsto \natural_i\phi \mid i \in Ag\}$ takes place.

- If a majority prefers $\neg\phi$ to ϕ, i.e., if

$$
|\{i \in Ag \mid \mathbf{M} \models_w [\rightarrow_i]\phi\}| \;<\; |\{i \in Ag \mid \mathbf{M} \models_w [\rightarrow_i]\neg\phi\}|
$$

  then simultaneous belief change $\{i \mapsto \natural_i\neg\phi \mid i \in Ag\}$ takes place.

- If there is no majority either way, nothing happens.

In fact, Dutch meetings are procedures for judgement aggregation [17]. Let us return to our example of three judges a, b, c with a, b agreeing that p, and b, c agreeing that q, so that both p and q command a majority, but $p \wedge q$ does not. Using our logic, we can picture the situation as a preference model. We assume that every agent has greater belief in worlds that match her beliefs in more propositions. Then we get the model of Figure 5.

So a has the greatest belief in the world where p and not q hold, but after that she has more belief in a world where p and q both hold than in the world where q and not p hold, because in the first world at least her belief in p is right. Similarly
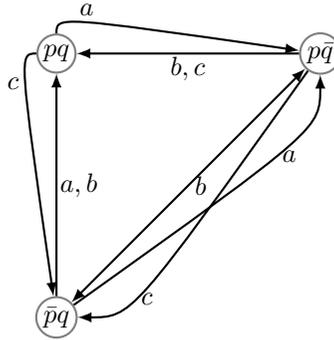
Figure 5. Preference model for the case of the three judges.

for c. For b, she believes in the world where p and q hold, and values the other worlds equally plausible.

In this model the following formulas hold:

$$[\to_a]p, [\to_b]p, [\to_b]q, [\to_c]q, [\to_a]\neg(p \wedge q), [\to_c]\neg(p \wedge q).$$

This shows that there are majority beliefs in p and in q, but there is also a majority belief in $\neg(p \wedge q)$. If the judges decide to have a Dutch meeting about p, the result will be unanimous belief in p (see Figure 6).
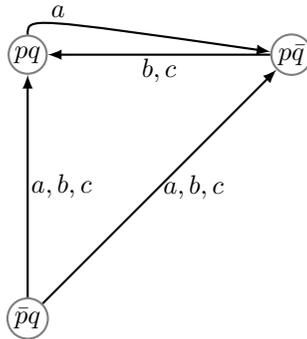
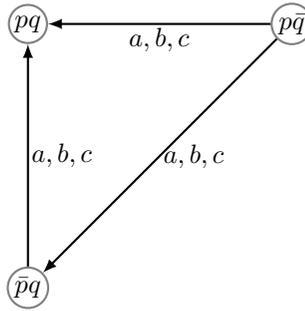Figure 6. Result of Dutch meeting about p.

Figure 7. Result of Dutch meeting about q after Dutch meeting about p.

Now if the judges hold a subsequent Dutch meeting about q, the result will be unanimous belief in q (see Figure 7). Now the judges unanimously believe in $p \wedge q$, so the defendant will be judged guilty.
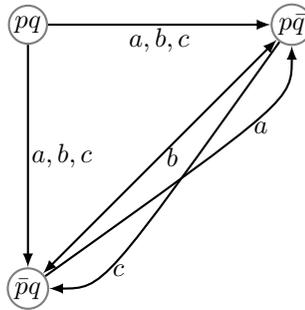

Figure 8. Result of Dutch meeting about $p \wedge q$.

However, if a Dutch meeting about $p \wedge q$ was held in the first place, the result would be belief in $\neg(p \wedge q)$ (see Figure 8). Clearly, in this case the defendant would be acquitted.

Experienced judges are of course familiar with this phenomenon. Procedural discussions about how to decompose a problem, and in which order to discuss the component problems may seem beside the point of a legal issue, but they turn out to be highly relevant for the outcome of the legal deliberations. Our logic can be used

to model the various procedural approaches for arriving at a unanimous conclusion. See [20] for an alternative analysis of procedural aspects in processes of judgement aggregation, in terms of coalition logic.

6 Conclusion

We have shown how propositional dynamic logic with converse can be used as a basis for developing a very expressive system of multi-agent belief revision and belief change, and we have studied linking of beliefs as a natural requirement on multi-agent belief change.

In future work, we would like to answer the open technical questions that were mentioned at the end of Section 2. Also on our research agenda is the problem of finding natural classes of substitutions that preserve relation linking.

Next, since our logic provides a general mechanism for simultaneous belief change, it can be used to describe and analyse topics in judgement aggregation, the effects of agenda setting, the effects of subgroup meetings to create general belief, and many further issues of collective rationality.

Finally, we would like to compare our logic to other proposals to give a modal analysis of judgement aggregation, such as [1], and we would like to find out if our logic can be put to work in practical social choice analysis.

Acknowledgments.

References

[1] Thomas Agotnes, Wiebe van der Hoek, and Michael Wooldridge. Reasoning about judgement and preference aggregation. In *The Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 07)*, pages 567–575. IFAAMAS, May 2007.

[2] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge, and private suspicions. In I. Bilboa, editor, *Proceedings of TARK'98*, pages 43–56, 1998.

[3] A. Baltag and S. Smets. Conditional doxastic models: A qualitative approach to dynamic belief revision. *Electronic Notes in Theoretical Computer Science (ENTCS)*, 165:5–21, 2006.

[4] A. Baltag and S. Smets. A qualitative theory of dynamic interactive belief revision. In G. Bonanno, W. van der Hoek, and M. Wooldridge, editors, *Texts in Logic and Games*. Amsterdam University Press, 2008. To appear.

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

[6] J. van Benthem and F. Liu. Dynamic logic of preference upgrade. *Journal of Applied Non-Classical Logics*, 14(2):157–182, 2007.

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

[8] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.

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

[10] C. Boutilier. Toward a logic of qualitative decision theory. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Proceedings of the 4th International Conference on Principle of Knowledge Representation and Reasoning (KR-94)*, pages 75–86. Morgan Kaufmann, 1994.

[11] Jan van Eijck. Yet more modal logics of preference change and belief revision. In K.R. Apt and R. van Rooij, editors, *New Perspectives on Games and Interaction*, volume 4 of *Texts in Logic and Games*, pages 81–104. Amsterdam University Press, 2008.

[12] Jan van Eijck and Yanjing Wang. Propositional Dynamic Logic as a logic of belief revision. In Wilfrid Hodges and Ruy de Queiros, editors, *Proceedings of Wollic'08*, number 5110 in Lecture Notes in Artificial Intelligence, pages 136–148. Springer, 2008. `http://dx.doi.org/10.1007/978-3-540-69937-8_13`.

[13] R. Goldblatt. *Logics of Time and Computation, Second Edition, Revised and Expanded*, volume 7 of *CSLI Lecture Notes*. CSLI, Stanford, 1992 (first edition 1987). Distributed by University of Chicago Press.

[14] A. Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17:157–170, 1988.

[15] D. Harel. Dynamic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, pages 497–604. Reidel, Dordrecht, 1984. Volume II.

[16] D. Kozen and R. Parikh. An elementary proof of the completeness of PDL. *Theoretical Computer Science*, 14:113–118, 1981.

[17] C. List and P. Pettit. On the many as one. *Philosophy and Public Affairs*, 33(4):377–390, 2005.

[18] Carsten Lutz. PDL with intersection and converse is decidable. In C.-H. L. Ong, editor, *Proceedings of the 19th International Workshop on Computer Science Logic (CSL 2005)*, number 3634 in Lecture Notes in Computer Science, pages 413–427. Springer, 2005.

[19] A. C. Nayak. Iterated belief change based on epistemic entrenchment. *Erkenntnis*, 41:353–390, 1994.

[20] Marc Pauly. *Logic for Social Software*. PhD thesis, ILLC, Amsterdam, 2001.

[21] K. Segerberg. A completeness theorem in the modal logic of programs. In T. Traczyck, editor, *Universal Algebra and Applications*, pages 36–46. Polish Science Publications, 1982.