On the Dynamics of Institutional Agreements

Andreas Herzig[1], Tiago de Lima[2], and Emiliano Lorini[1]

[1] IRIT, Toulouse, France
Andreas.Herzig@irit.fr
Emiliano.Lorini@irit.fr
[2] Eindhoven University of Technology, The Netherlands
T.d.Lima@tue.nl

Abstract. In this work we continue the work initiated in [1], in which a logic of individual and collective acceptance was introduced. Our aim in this paper is to investigate the extension of the logic of acceptance by *public announcements* of formulas. The function of public announcements is to diminish the space of possible worlds accepted by agents and sets of agents while functioning as members of a given group, team, organization, institution, etc., x. If a set of agents C ends up with an empty set of worlds that they accept while functioning as members of x, then the agents in C do not identify themselves any longer with x. In such a situation the agents in C should have the possibility to join x again. To that aim we discuss at the end of the paper an operation which consists of an agent (or set of agents) joining a given group, team, organization, institution, etc.

1 Introduction

The concept of *collective acceptance* has been studied in social philosophy in opposition to group attitudes such as *common belief* and *common knowledge* that are popular in artificial intelligence and theoretical computer science [2, 3]. As suggested in [4], the main difference between collective acceptance and common belief (or common knowledge) is that a collective acceptance by a set of agents C is based on the fact that the agents in C identify and recognize themselves as members of the same *social context*, such as a group, team, organization, institution, etc. Common belief (and common knowledge) does not necessarily entail this aspect of mutual recognition and identification with respect to a social context. In this sense, according to [4, 5], collective acceptance rather than common belief is the more appropriate concept to characterize a proper notion of *group belief*. For example, in the context of the organization Greenpeace the agents in a set C (collectively) accept that their mission is to protect the Earth *qua* members of Greenpeace. The state of acceptance *qua* members of Greenpeace is the kind of acceptance the agents in C are committed to when they are functioning together as members of Greenpeace.

It has been emphasized that a similar distinction between acceptance and belief exists at the *individual* level. While an agent's belief that p is an attitude of the agent constitutively aimed at the truth of p, an agent's acceptance is not necessarily connected to the actual truth of the proposition. In order to better distinguish these two notions, it has been suggested in [6] that while an agent's beliefs are not subject to the agent's will, its acceptances are voluntary; while its beliefs aim at truth, its acceptances are sensitive to pragmatic considerations; while its beliefs are shaped by evidence, its acceptances need not be; finally, while its beliefs are context-independent, its acceptances might depend on context. Often the acceptances of an agent depend on social contexts, that is, while identifying itself as a member of a group (or team, organization, institution, etc.) an agent reasons and accepts things *qua* member of this group. In these situations it may happen that the agent's acceptances are in conflict with its beliefs. For instance, a lawyer who is trying to defend a client in a murder case accepts *qua* lawyer that the client is innocent, even if he believes the contrary.

The aim of this paper is to continue the work initiated in [1, 7]. There, a logic of individual and collective acceptance was introduced.[3] One of the notable features of that logic is that the accessibility relation associated to the acceptance operator is not necessarily serial: an empty set of possible worlds associated to a group C in a context x just means that C does not identify itself with x.

Our aim here is to investigate the extension of the logic of acceptance by *public announcements* of formulas, noted $x!\psi$. Modal operators of type $[x!\varphi]$ are intended to express that the members of a certain group, team, organization, institution, etc., x learn that φ is true in that institution in such a way that their acceptances, *qua* members of x, are updated. The function of public announcements is to diminish the space of possible worlds accepted by agents and groups of agents. It might also happen that a given set of agents C ends up with an empty set of possible worlds that they accept while functioning as members of a certain social context x. As we have said, this means that C quits x: the agents in C do not identify themselves any longer with x. In such a situation C should have the possibility to join x again. To that aim we discuss at the end of the paper an operation which consists of an agent (or set of agents) joining a given social context x.

The main contribution of this paper is to extend the logic presented in [1] to public announcements and show that, differently from common belief and common knowledge, reduction axioms can be given. As usual, the addition of these axioms to the Hilbert axiomatics of the logic of acceptance provides a complete axiomatization of the logic of acceptance and announcements. As far as we know, this constitutes the first attempt to build up a logic of acceptance and public announcements.

The paper is organized as follows. In Section 2 we present the syntax and semantics of acceptance logic together with its axiomatization. In Section 3 we extend it with announcements, and show that our extension also allows for reduction axioms and thereby a complete axiomatization. In Section 4 we formalize an example which illustrates the dynamics of acceptance based on announcements.

---

[3] This logic has some similarities with the logic of *group belief* that we have developed in [8, 9].

In Section 5 we briefly discuss the operation which consists of an agent (or set of agents) joining a social context. This section is not intended to provide a solution to the logical characterization of this social phenomenon though. In Section 6 we draw conclusions.

2 The Logic of Acceptance $\mathcal{AL}$

We now present a variant of the *Acceptance Logic* ($\mathcal{AL}$) that was introduced in [1]. $\mathcal{AL}$ enables expressing that certain agents identify themselves as members of a social context x and, reasoning about what agents and groups of agents accept while functioning together as members of a certain social context. The axioms of $\mathcal{AL}$ clarify the relationships between individual acceptance (acceptances of individual agents) and collective acceptance (acceptances of groups of agents).

2.1 Syntax

The syntactic primitives of $\mathcal{AL}$ are the following: a finite non-empty set of agents AGT; a countable set of atomic formulas ATM; and a finite set of labels $CTXT$ denoting social contexts such as groups, teams, organizations, institutions, etc. The language $\mathcal{L}_{\mathcal{AL}}$ of the logic $\mathcal{AL}$ is given by the following BNF:

$$\varphi \ ::= \ p \mid \neg\varphi \mid \varphi \vee \varphi \mid \mathcal{A}_{C:x}\varphi$$

where p ranges over ATM, C ranges over 2^{AGT}, and x ranges over $CTXT$. The formula $\mathcal{A}_{C:x}\varphi$ reads "the agents in C accept that φ while functioning together as members of x". We write $i{:}x$ instead of $\{i\}:x$.

For example, $\mathcal{A}_{C:Greenpeace}$*protectEarth* expresses that the agents in C accept that the mission of Greenpeace is to protect the Earth while functioning as activists in the context of Greenpeace; and $\mathcal{A}_{i:Catholic}$*PopeInfallibility* expresses that agent i accepts that the Pope is infallible while functioning as a Catholic in the context of the Catholic Church.

The intuition is that in two different contexts the same agent may accept contradictory propositions. For example, while functioning as a Catholic, agent i accepts that killing is forbidden, and while functioning as a soldier i accepts that killing is allowed. The CEO of Airbus accepts that Airbus is in good health while functioning as a member of Airbus Industries, and privately accepts the contrary.

The classical boolean connectives $\wedge$, $\rightarrow$, $\leftrightarrow$, $\top$ (tautology) and $\bot$ (contradiction) are defined from $\vee$ and $\neg$ in the usual manner.

The formula $\mathcal{A}_{C:x}\bot$ has to be read "agents in C are not functioning together as members of x", because we assume that functioning as a group member is, at least in this minimal sense, a rational activity. Conversely, $\neg\mathcal{A}_{C:x}\bot$ has to be read "agents in C are functioning together as members of x". Thus, $\neg\mathcal{A}_{C:x}\bot \wedge \mathcal{A}_{C:x}\varphi$ stands for "agents in C are functioning together as members of x and they accept that φ while functioning together as members of x" or simply "agents in C accept that φ *qua* members of x". This is a case of *group acceptance*. For the individual case, formula $\neg\mathcal{A}_{i:x}\bot \wedge \mathcal{A}_{i:x}\varphi$ has to be read "agent i accepts that φ *qua* member of x". This is a case of *individual acceptance*.

2.2 Semantics and Axiomatization

We use a standard possible worlds semantics. Let the set of all couples of non-empty subsets of agents and social contexts be

$$\Delta = \{C{:}x \;:\; C \in 2^{AGT} \text{ and } x \in CTXT\}.$$

An *acceptance model* is a triple $\mathcal{M} = \langle W, \mathscr{A}, \mathscr{V}\rangle$ where:

- W is a non-empty set of possible worlds;
- $\mathscr{A} : \Delta \to W \times W$ maps every $C{:}x \in \Delta$ to a relation $\mathscr{A}_{C:x}$ between possible worlds in W; and
- $\mathscr{V} : ATM \to 2^W$ is valuation function associating a set of possible worlds $\mathscr{V}(p) \subseteq W$ to each atomic formula p of ATM.

We write $\mathscr{A}_{C:x}(w)$ for the set $\{w' : \langle w, w'\rangle \in \mathscr{A}_{C:x}\}$. $\mathscr{A}_{C:x}(w)$ is the set of worlds that is accepted by the agents in C while functioning together as members of x.

Given $\mathcal{M} = \langle W, \mathscr{A}, \mathscr{V}\rangle$ and $w \in W$, the couple $\langle \mathcal{M}, w\rangle$ is a *pointed acceptance model*. The satisfaction relation $\models$ between formulas of $\mathcal{L}_{\mathcal{AL}}$ and pointed acceptance models $\langle \mathcal{M}, w\rangle$ is defined as usual for atomic propositions, negation and disjunction. The satisfaction relation for acceptance operators is the following:

$$\mathcal{M}, w \models \mathcal{A}_{C:x}\varphi \quad \text{iff} \quad \mathcal{M}, w' \models \varphi \text{ for every } w' \in \mathscr{A}_{C:x}(w)$$

Validity of a formula φ (noted $\models_{\mathcal{AL}} \varphi$) is defined as usual.

The axiomatization of $\mathcal{AL}$ is given in Fig. 1. It is not meant to reflect the semantics the way it has been presented up to this point. Instead, these axioms are meant as postulates, and we are going to present the corresponding semantic constraints later.

(K)	All K-principles for the operators $\mathcal{A}_{C:x}$
(4*)	$\mathcal{A}_{C:x}\varphi \to \mathcal{A}_{B:y}\mathcal{A}_{C:x}\varphi$, if $B \subseteq C$
(5*)	$\neg\mathcal{A}_{C:x}\varphi \to \mathcal{A}_{B:y}\neg\mathcal{A}_{C:x}\varphi$, if $B \subseteq C$
(Inc)	$(\neg\mathcal{A}_{C:x}\bot \wedge \mathcal{A}_{C:x}\varphi) \to \mathcal{A}_{B:x}\varphi$, if $B \subseteq C$
(Una)	$\mathcal{A}_{C:x}(\bigwedge_{i\in C}\mathcal{A}_{i:x}\varphi \to \varphi)$

Fig. 1. Axiomatization of $\mathcal{AL}$.

As usual, the K-principles are the axioms and inference rules of the basic modal logic K. Axioms **4*** and **5*** are introspection axioms: when the agents in a set C function together as members of x, then, for all $y \in CTXT$ and B such that $B \subseteq C$, the agents in B have access to all the facts that are accepted (or that are not accepted) by the agents in C. In particular, if the agents in C

(do not) accept that φ while functioning together as members of x then, while functioning together as members of y, the agents of every subset B of C accept that agents in C (do not) accept that φ.

Example 1. Suppose that three agents i, j, k, while functioning together as members of the UK trade union, accept that their mission is to try to increase teachers' wages, but they do not accept *qua* members of the trade union that their mission is to try to increase railway workers' wages: $\mathcal{A}_{\{i,j,k\}:Union} IncrTeacherWage$ and $\neg\mathcal{A}_{\{i,j,k\}:Union} IncrRailwayWage$. By axiom **4*** we infer that, while functioning as a UK citizen, i accepts that i, j, k accept that their mission is to try to increase teachers' wages, while functioning together as members of the trade union: $\mathcal{A}_{i:UK}\mathcal{A}_{\{i,j,k\}:Union} IncrTeacherWage$. By Axiom **5*** we infer that, while functioning as a UK citizen, i accepts that i, j, k do not accept, *qua* members of the trade union, that their mission is to try to increase railway workers' wages: $\mathcal{A}_{i:UK}\neg\mathcal{A}_{\{i,j,k\}:Union} IncrRailwayWage$.

Axiom **Inc** says that, if the agents in C accept that φ *qua* members of x then every subset B of C accepts φ while functioning together as members of x. This means that things accepted by the agents in C *qua* members of a certain social context x are necessarily accepted by agents in all of C's subsets with respect to the same context x. Axiom **Inc** describes the *top down* process leading from C's collective acceptance to the individual acceptances of C's members.[4]

Example 2. Imagine three agents i, j, k that, *qua* players of the game Clue, accept that someone called Mrs. Red, has been killed:

$$\neg\mathcal{A}_{\{i,j,k\}:Clue}\bot \wedge \mathcal{A}_{\{i,j,k\}:Clue} killedMrsRed.$$

By axiom **Inc** we infer that also the two agents i, j, while functioning as Clue players, accept that someone called Mrs. Red has been killed:

$$\mathcal{A}_{\{i,j\}:Clue} killedMrsRed.$$

Axiom **Una** expresses a unanimity principle according to which the agents in C, while functioning together as members of x, accept that if each of them individually accepts that φ while functioning as member of x, then φ is the case. This axiom describes the *bottom up* process leading from individual acceptances of the members of C to the collective acceptance of the group C.

In order to make our axioms valid we impose the following constraints on acceptance models, for any world $w \in W$, context $x, y \in CTXT$, and coalitions $C, B \in 2^{AGT}$ such that $B \subseteq C$:

(S.1) if $w' \in \mathscr{A}_{B:y}(w)$ then $\mathscr{A}_{C:x}(w') = \mathscr{A}_{C:x}(w)$;

---

[4] Note that the more general

$$(\neg\mathcal{A}_{C:x}\bot \wedge \mathcal{A}_{C:x}\varphi) \rightarrow \mathcal{A}_{B:y}\varphi, \text{ if } B \subseteq C$$

would lead to unwanted consequences: the group of Catholics' acceptance *qua* members of the Catholic church that the Pope is infallible does not entail that Catholics privately accept that the Pope is infallible.

Also note that for $B \subseteq C$, neither $\mathcal{A}_{C:x}\bot \rightarrow \mathcal{A}_{B:x}\bot$ nor $\mathcal{A}_{B:x}\bot \rightarrow \mathcal{A}_{C:x}\bot$ should hold.

(S.2) if $\mathscr{A}_{C:x}(w) \neq \emptyset$ then $\mathscr{A}_{B:x}(w) \subseteq \mathscr{A}_{C:x}(w)$;
(S.3) if $w' \in \mathscr{A}_{C:x}(w)$ then $w' \in \bigcup_{i \in C} \mathscr{A}_{i:x}(w')$.

Axioms **4*** and **5*** together correspond to the constraint **S.1**; axiom **Inc** corresponds to **S.2**, and axiom **Una** to **S.3** (in the sense of correspondence theory). As all our axioms are in the Sahlqvist class we obtain straightforwardly:

Theorem 1. *The axiomatization of $\mathcal{AL}$ of Fig. 1 is sound and complete w.r.t. the class of $\mathcal{AL}$ models satisfying constraints **S.1**, **S.2**, and **S.3**.*

Proof. It is a routine task to check that all the axioms of the logic $\mathcal{AL}$ correspond to their semantic counterparts. It is routine, too, to check that all $\mathcal{AL}$ axioms are in the Sahlqvist class, for which a general completeness result exists [10]. □

Example 3. It follows from axioms **4***, **5*** and **Inc** that if $B \subseteq C$ then $\models_{\mathcal{AL}} \mathcal{A}_{B:y}\mathcal{A}_{C:x}\varphi \leftrightarrow \mathcal{A}_{B:y}\bot \vee \mathcal{A}_{C:x}\varphi$ and $\models_{\mathcal{AL}} \mathcal{A}_{B:y}\neg\mathcal{A}_{C:x}\varphi \leftrightarrow \mathcal{A}_{B:y}\bot \vee \neg\mathcal{A}_{C:x}\varphi$. We also have $\models_{\mathcal{AL}} \mathcal{A}_{C:x}(\mathcal{A}_{C:x}\varphi \rightarrow \varphi)$.

3 The Logic of Acceptance and Public Announcements $\mathcal{ALA}$

In its nature, acceptance comes by communication: if a group accepts that one of its members i accepts that φ then this is often the result of a speech act performed by i. Acceptance is therefore closely related to the notion of commitment that has been studied in agent communication languages [11–13].

In this paper we study the combination of acceptance logic $\mathcal{AL}$ with a rather simple communicative act, viz. public announcements as defined in public announcement logic ($\mathcal{PAL}$) [14]. Basically, when formula ψ is publicly announced, all agents learn that ψ is true. Our truth condition is that of Kooi [15], that is slightly different from the standard one in public announcement logic: it does not require announcements to be truthful.

3.1 Language and Models

The language $\mathcal{L}_{\mathcal{ALA}}$ of acceptance logic with announcements ($\mathcal{ALA}$) extends $\mathcal{L}_{\mathcal{AL}}$ by modal formulas of the form $[x!\psi]\varphi$, where $\varphi, \psi \in \mathcal{L}_{\mathcal{AL}}$. Such formulas are read "φ holds after the public announcement of ψ in context x". Modal operators of type $[x!\psi]$ are intended to express that the agents learn that ψ is true in the social context x, in such a way that their acceptances, *qua* members of x, are updated.

The announcement $x!\mathcal{A}_{i:x}\psi$ is an *event*. It approximates i's *action* of announcing that ψ in context x. (This is an assertion in speech act theory and in Walton and Krabbe's dialogue games [16].)

It is worth noting that when x denotes an institution, events of type $x!\psi$ can be used to describe the event of issuing or promulgating a certain norm ψ (e.g. obligation, permission) within the context of the institution x.[5]

[5] For a logical characterization of the act of *proclaiming* or *promulgating* a norm, see also [17].

Formulas of $\mathcal{L}_{\mathcal{ALA}}$ are interpreted in pointed acceptance models. The satisfaction relation $\models$ of Section 2 is extended by the following clause:

$$\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\varphi \quad \text{iff} \quad \langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \varphi$$

with

- $\mathscr{A}^{x!\psi}_{C:y}(w) = \mathscr{A}_{C:y}(w)$, for all $C{:}y \in \Delta$, $w \in W$ and $y \neq x$;
- $\mathscr{A}^{x!\psi}_{C:y}(w) = \mathscr{A}_{C:y}(w) \cap ||\psi||_{\mathcal{M}}$, for all $C{:}y \in \Delta$, $w \in W$ and $y = x$,

where as usual $||\psi||_{\mathcal{M}} = \{w : \mathcal{M}, w \models \psi\}$ is the extension of ψ in $\mathcal{M}$, i.e., the set of worlds where ψ is true. Thus, in a way similar to [15], the agents take into account the announcement of ψ in the social context x and modify their acceptances *qua* members of x by eliminating all arrows leading to $\neg\psi$ worlds (instead of eliminating the worlds themselves, as in $\mathcal{PAL}$). On the contrary, when x and y are different, the accessibility relations associated to the acceptances *qua* members of y are not modified, after the announcement of ψ in the social context x.

Validity of a formula φ (noted $\models_{\mathcal{ALA}} \varphi$) is defined as before. For example, $\models_{\mathcal{ALA}} [x!p]\mathcal{A}_{C:x}p$, and $\models_{\mathcal{ALA}} \mathcal{A}_{C:x}\neg p \rightarrow [x!p]\mathcal{A}_{C:x}\bot$. The latter means that coalition C quits all social contexts within which C's acceptances are inconsistent with what is announced.

Note that contrarily to standard common knowledge and belief, the modified accessibility relations for acceptances are not computed from the modified accessibility relations for individuals, but are first-class citizens here: they are changed 'on their own'.

Proposition 1. *If $\mathcal{M}$ is an acceptance model then $\mathcal{M}^{x!\psi}$ is an acceptance model.*

Proof. We show that $\mathcal{M}^{x!\psi} = \langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle$ satisfies **S.1**, **S.2** and **S.3**. In what follows let $B \subseteq C$.

(S.1): Let $w_2 \in \mathscr{A}^{x!\psi}_{B:y}(w_1)$. If the latter is true then $w_2 \in \mathscr{A}_{B:y}(w_1)$, which implies $\mathscr{A}_{C:x}(w_2) = \mathscr{A}_{C:x}(w_1)$, because $\mathcal{M}$ respects **S.1**.
Now, we show that $\mathscr{A}^{x!\psi}_{C:x}(w_2) \subseteq \mathscr{A}^{x!\psi}_{C:x}(w_1)$. Consider a possible world $w_3 \in \mathscr{A}^{x!\psi}_{C:x}(w_2)$. This means that $w_3 \in \mathscr{A}_{C:x}(w_2) \cap ||\psi||_{\mathcal{M}}$. Then, in particular, $w_3 \in \mathscr{A}_{C:x}(w_2)$, which implies $w_3 \in \mathscr{A}^{x!\psi}_{C:x}(w_1)$, because $\mathscr{A}_{C:x}(w_2) = \mathscr{A}_{C:x}(w_1)$.
By using an analogous argument, we show that $\mathscr{A}^{x!\psi}_{C:x}(w_1) \subseteq \mathscr{A}^{x!\psi}_{C:x}(w_2)$.

(S.2): Let $\mathscr{A}^{x!\psi}_{C:x}(w_1) \neq \emptyset$ and $w_2 \in \mathscr{A}^{x!\psi}_{B:x}(w_1)$. We show that $w_2 \in \mathscr{A}^{x!\psi}_{C:x}(w_1)$. The hypothesis implies $w_2 \in \mathscr{A}_{B:x}(w_1) \cap ||\psi||_{\mathcal{M}}$. Then, in particular, $w_2 \in \mathscr{A}_{B:x}(w_1)$. Also note that the hypothesis implies $\mathscr{A}_{C:x}(w_1) \neq \emptyset$. Then, $w_2 \in \mathscr{A}_{C:x}(w_1)$, because $\mathcal{M}$ respects **S.2**. We conclude that $w_2 \in \mathscr{A}_{C:x}(w_1) \cap ||\psi||_{\mathcal{M}}$. The latter is true if and only if $w_2 \in \mathscr{A}^{x!\psi}_{C:x}(w_1)$.

(S.3): Let $w_2 \in \mathscr{A}^{x!\psi}_{C:x}(w_1)$. We show that $w_2 \in \mathscr{A}^{x!\psi}_{i:x}(w_2)$ for some $i \in C$. The hypothesis is equivalent to $w_2 \in \mathscr{A}_{C:x}(w_1) \cap ||\psi||_{\mathcal{M}}$. Then, in particular, $w_2 \in \mathscr{A}_{C:x}(w_1)$, which implies $w_2 \in \mathscr{A}_{i:x}(w_2)$ for some $i \in C$, because $\mathcal{M}$ respects **S.3**. Then, $w_2 \in \mathscr{A}_{i:x}(w_2) \cap ||\psi||_{\mathcal{M}}$ for some $i \in C$. The latter is true if and only if $w_2 \in \mathscr{A}^{x!\psi}_{i:x}(w_2)$ for some $i \in C$.

$\square$

3.2 Reduction Axioms and Completeness

Just as in dynamic epistemic logics without common belief, $\mathcal{ALA}$ has reduction axioms for all cases (individual and collective acceptance). This contrasts with logics having the common belief operator, for which such axioms do not exist [18].

Proposition 2. *The following equivalences are $\mathcal{ALA}$ valid.*

(R.1) $[x!\psi]p \leftrightarrow p$
(R.2) $[x!\psi]\neg\varphi \leftrightarrow \neg[x!\psi]\varphi$
(R.3) $[x!\psi](\varphi_1 \wedge \varphi_2) \leftrightarrow [x!\psi]\varphi_1 \wedge [\psi!\varphi]_2$
(R.4) $[x!\psi]\mathcal{A}_{C:y}\varphi \leftrightarrow \mathcal{A}_{C:y}[x!\psi]\varphi$ *(if $y \neq x$)*
(R.5) $[x!\psi]\mathcal{A}_{C:y}\varphi \leftrightarrow \mathcal{A}_{C:y}(\psi \to [x!\psi]\varphi)$ *(if $y = x$)*

Proof. **(R.1):**
$\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]p$
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle \models p$
iff $w \in \mathscr{V}(p)$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models p$.
(R.2):
$\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\neg\varphi$
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \neg\varphi$
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \not\models \varphi$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \not\models [x!\psi]\varphi$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models \neg[x!\psi]\varphi$.
(R.3):
$\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi](\varphi_1 \wedge \varphi_2)$
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \varphi_1 \wedge \varphi_2$
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \varphi_1$ and $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \varphi_2$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\varphi_1$ and $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\varphi_2$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\varphi_1 \wedge [x!\psi]\varphi_2$.
(R.4): We show that the equivalent $\neg[x!\psi]\mathcal{A}_{C:y}\varphi \leftrightarrow \neg\mathcal{A}_{C:y}[x!\psi]\varphi$ is valid.
$\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models \neg[x!\psi]\mathcal{A}_{C:y}\varphi$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\neg\mathcal{A}_{C:y}\varphi$, by **(R.3)**,
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \neg\mathcal{A}_{C:y}\varphi$
iff there is $w' \in \mathscr{A}^{x!\psi}_{C:y}(w)$ such that $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w' \models \neg\varphi$
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle \models \neg\varphi$, because $y \neq x$,
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}, \mathscr{V}\rangle \models [x!\psi]\neg\varphi$
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}, \mathscr{V}\rangle \models \neg[x!\psi]\varphi$, by **(R.3)**,
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle \models \neg\mathcal{A}_{C:y}[x!\psi]\varphi$.
(R.5): We show that $\neg[x!\psi]\mathcal{A}_{C:y}\varphi \leftrightarrow \neg\mathcal{A}_{C:y}(\psi \to [x!\psi]\varphi)$ is valid.
$\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models \neg[x!\psi]\mathcal{A}_{C:y}\varphi$
iff $\langle W, \mathscr{A}, \mathscr{V}\rangle, w \models [x!\psi]\neg\mathcal{A}_{C:y}\varphi$, by **(R.3)**,
iff $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w \models \neg\mathcal{A}_{C:y}\varphi$
iff there is $w' \in \mathscr{A}^{x!\psi}_{C:y}(w)$ such that $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle, w' \models \neg\varphi$
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}, \mathscr{V}\rangle \models \psi$ and $\langle W, \mathscr{A}^{x!\psi}, \mathscr{V}\rangle \models \neg\varphi$,

because $y = x$,
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}, \mathscr{V} \rangle \models \psi$ and $\langle W, \mathscr{A}, \mathscr{V} \rangle \models [x!\psi]\neg\varphi$
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}, \mathscr{V} \rangle \models \psi \wedge [x!\psi]\neg\varphi$
iff there is $w' \in \mathscr{A}_{C:y}(w)$ such that $\langle W, \mathscr{A}, \mathscr{V} \rangle \models \psi \wedge \neg[x!\psi]\varphi$, by **(R.3)**,
iff $\langle W, \mathscr{A}, \mathscr{V} \rangle \models \neg\mathcal{A}_{C:y}(\psi \rightarrow [x!\psi]\varphi)$.

□

These equivalences are called reduction axioms because they allow to rewrite every formula by successively eliminating the announcement operators, ending up with a formula that contains none.

Theorem 2. *For every $\mathcal{ALA}$ formula there is an equivalent $\mathcal{AL}$ formula.*

Proof. The proof goes just as for public announcement logic (found in [15]): each of the above $\mathcal{ALA}$ valid equivalences **R.2**–**R.5**, when applied from the left to the right, yields a simpler formula, where 'simpler' roughly speaking means that the announcement operator is pushed inwards. Once the announcement operator attains an atom it is eliminated by the first equivalence **R.1**. □

Theorem 3. *The formulas that are valid in $\mathcal{ALA}$ models are completely axiomatized by the axioms and inference rules of $\mathcal{AL}$ together with the reduction axioms of Proposition 2.*

Proof. This a straightforward consequence of Theorem 1 and Theorem 2. □

Here are some examples of reductions.

Example 4. The formula $[x!p]\mathcal{A}_{C:x}p$ is successively rewritten as follows:
$\mathcal{A}_{C:x}(p \rightarrow [x!p])$ by **R.5**
$\mathcal{A}_{C:x}(p \rightarrow p)$ by **R.1**
The latter is a theorem of every normal modal logic (and therefore also of acceptance logic $\mathcal{AL}$). It follows that the initial formula is valid, too.

Example 5. The formula $\mathcal{A}_{i:x}\neg p \rightarrow [x!p]\mathcal{A}_{i:x}\bot$ is rewritten as follows:
$\mathcal{A}_{i:x}\neg p \rightarrow \mathcal{A}_{i:x}(p \rightarrow [x!p]\bot)$ by **R.5**
$\mathcal{A}_{i:x}\neg p \rightarrow \mathcal{A}_{i:x}(p \rightarrow \bot)$ by **R.1**
The latter is a theorem of every normal modal logic (and therefore also of acceptance logic $\mathcal{AL}$).

Example 6. The formula $[x!(\mathcal{A}_{i:x}p)]\mathcal{A}_{C:x}\mathcal{A}_{i:x}p$ is rewritten as follows:
$\mathcal{A}_{C:x}(\mathcal{A}_{i:x}p \rightarrow [x!(\mathcal{A}_{i:x}p)]\mathcal{A}_{i:x}p)$ by **R.5**
$\mathcal{A}_{C:x}(\mathcal{A}_{i:x}p \rightarrow \mathcal{A}_{i:x}(\mathcal{A}_{i:x}p \rightarrow [x!(\mathcal{A}_{i:x}p)]p))$ by **R.5**
$\mathcal{A}_{C:x}(\mathcal{A}_{i:x}p \rightarrow \mathcal{A}_{i:x}(\mathcal{A}_{i:x}p \rightarrow p))$ by **R.1**
The latter is an $\mathcal{AL}$ theorem (because $\mathcal{A}_{i:x}(\mathcal{A}_{i:x}p \rightarrow p)$ is an $\mathcal{AL}$ theorem, cf. Example 3 of Section 2). It follows that the initial formula is valid, too.

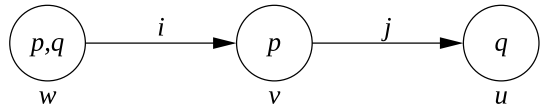

Fig. 2. An epistemic model.

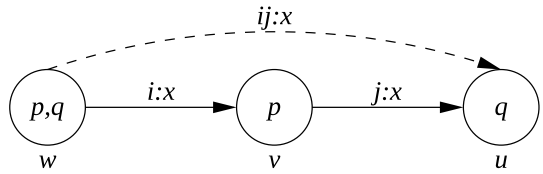

Fig. 3. The acceptance model corresponding to the epistemic model in Fig. 2.

3.3 Discussion

As said before, $\mathcal{ALA}$ has reduction axioms for all cases (individual and collective acceptance), while it has been shown in [18] that logics of common belief (and common knowledge) do not. Technically, it happens because acceptance models have more arrows. Let $\mathcal{M}$ be the epistemic model in Fig. 2. Note that even though $\mathcal{M}, w \not\models \mathbf{C}_{ij}(q \rightarrow [!q]p)$, we still have $\mathcal{M}, w \models [!q]\mathbf{C}_{ij}p$. In words, it is not common belief among i and j that q implies that after the public announcement of q, we have that p, but after the public announcement of q it is common belief among i and j that p. That is, common belief may appear 'out of the blue', i.e., it was not foreseeable by the agents, it just 'pops up'.

However, when we build the acceptance model that corresponds to the model $\mathcal{M}$, it looks like the structure in Fig. 3. Semantic constraint **S.1** obliges the corresponding acceptance model to have more arrows, in particular, it must have the dashed arrow from w to u labelled by $ij{:}x$. Let $\mathcal{M}'$ be such acceptance model, note that $M', w \not\models \mathbf{A}_{ij:x}(q \rightarrow [!q]p)$, and also $\mathcal{M}', w \not\models [x!q]\mathbf{A}_{ij:x}p$. That is, contrary to common belief, common acceptances cannot just 'pop up' without be foreseeable by the agents.

None the less, we argue that the reduction axiom **R.5** is an intuitive property of collective acceptance. This is due to the fact that, differently from the standard notions of common belief and common knowledge, collective acceptance entails an aspect of mutual identification and recognition with respect to a group.

Consider the left to right direction of the reduction axiom **R.5**. When the agents in a set C identify themselves with a group x and recognize each other as members of this group, they accept certain rules and principles to stand for the the rules and principles of the group. That is, the agents in C *share a common body* of rules and principles. Among these shared rules and principles, there are the rules and principles which describe how the world should evolve when an announcement occurs. They govern how the acceptance of the agents in the group will be changed after an announcement. Suppose that a certain fact ψ is publicly announced. After this announcement, the agents in C accept

φ, while identifying themselves with a group x and recognizing each other as members of this group: $[x!\psi]\mathcal{A}_{C:x}\varphi$. This collective acceptance of the agents in C is not created from scratch after the announcement of ψ. On the contrary, the creation of this acceptance depends on what the agents in C accepted (before the announcement) as a principle of group x. In particular, the creation of C's acceptance that φ rests on the fact that, before ψ is announced, the agents in C, while identifying themselves and recognizing each other as members of x, accept a principle saying that "if ψ is true then, after ψ is announced in x, φ will be true": $\mathcal{A}_{C:x}(\psi \rightarrow [x!\psi]\varphi)$.

For example, imagine that the agents in a set C identify themselves and recognize each other as members of the Lilliputian pacifist movement. Let ψ denote the proposition "the government of Lilliput has decided to attack the neighboring nation of Blefuscu".[6] After ψ is publicly announced the agents in C accept that φ = "they should start to protest against the Lilliput government", while functioning as members of the Lilliputian pacifist movement: $[LilliputPacifist!\psi]\mathcal{A}_{C:LilliputPacifist}\varphi$. This implies that (before the announcement) the agents in C, while identifying themselves and recognizing each other as members of the Lilliputian pacifist movement, accept a principle saying that "if ψ is true then, after ψ is announced, φ will be true": $\mathcal{A}_{C:LilliputPacifist}(\psi \rightarrow [LilliputPacifist!\psi]\varphi)$. That is, the creation of C's acceptance to protest against the Lilliput government depends on the fact that, before the announcement, the agents in C accept to protest against the Lilliput government in case it will announce its decision to attack the neighboring nation of Blefuscu. This means that C's acceptance to protest depends on the fact that, before the announcement, the agents in C accept a principle which specifies what to do in case the Lilliput government will manifest its intention to attack Blefuscu.

4 An Example

Until now we only considered that group acceptances emerge from consensus, by admitting axiom **Una**. One can go further and also consider other kinds of group acceptances, as shown in the next example. The example is inspired by Pettit [19].

Example 7. Imagine a three-member court which has to make a judgment on whether a defendant is liable (noted l) for a breach of contract. The three judges i, j and k accept a majority rule to decide on the issue. That is, i, j and k, while functioning as members of the court, accept that if the majority of them accepts that the defendant is liable (resp. not liable), then the defendant is liable (resp. not liable). Formally, for any B such that $B \subseteq \{i, j, k\}$ and $|B| = 2$ we have:

(Maj) $$\mathcal{A}_{\{i,j,k\}:court}(\bigwedge_{i \in B} \mathcal{A}_{i:court} l \rightarrow l) \quad \wedge \quad \mathcal{A}_{\{i,j,k\}:court}(\bigwedge_{i \in B} \mathcal{A}_{i:court} \neg l \rightarrow \neg l)$$

[6] Lilliput and Blefuscu are the two fictional nations, permanently at war, that appear in the novel "Gulliver's Travels" by Jonathan Swift.

Given the previous majority rule, we can prove that: after the announcement that both i and j accept l (the defendant is liable) while functioning as members of the court, the agents in $\{i, j, k\}$ accept l while functioning together as members of the court. Indeed, from the previous majority rule we can derive the formula $[court!\mathcal{A}_{i:court}l \wedge \mathcal{A}_{j:court}l]\mathcal{A}_{\{i,j,k\}:court}l$. To prove this, it is sufficient to note that, by means of the reduction axioms, the formula $[court!\mathcal{A}_{i:court}l \wedge \mathcal{A}_{j:court}l]\mathcal{A}_{\{i,j,k\}:court}l$ is successively rewritten as follows:
$\mathcal{A}_{\{i,j,k\}:court}((\mathcal{A}_{i:court}l \wedge \mathcal{A}_{j:court}l) \rightarrow [court!\mathcal{A}_{i:court}l \wedge \mathcal{A}_{j:court}l]l)$ by **R.5**
$\mathcal{A}_{\{i,j,k\}:court}((\mathcal{A}_{i:court}l \wedge \mathcal{A}_{j:court}l) \rightarrow l)$ by **R.1**
The latter is entailed by the majority rule **Maj**.

In the previous example, we have considered a majority rule as a principle which is responsible for the creation of collective acceptances from individual acceptances. This is stronger than the basic axiom of unanimity (**Una**) of $\mathcal{AL}$. One can imagine other kinds of rules. For instance, one can consider social contexts with leaders (see also [7]). In such contexts, one can formalize the rule according to which everything that the leaders accept is universally accepted in the social context. Let the set of leaders of x be $L_x \in 2^{AGT}$. Then one can formalize that everything that the leaders accept is universally accepted in the social context by:

$$\textbf{(Leader)} \quad \mathcal{A}_{C:x}(\mathcal{A}_{L_x:x}\varphi \rightarrow \varphi)$$

5 Adding Retractions to $\mathcal{ALA}$: Some General Insights

According to our semantics, $\mathcal{A}_{i:x}\neg p \rightarrow [x!p]\mathcal{A}_{i:x}\bot$ is an $\mathcal{ALA}$ theorem (cf. Example 5). In words, when p is publicly announced then i quits all contexts x where he accepted p: agent i is no longer part of the institution, is kicked out of the group, etc. In $\mathcal{ALA}$ there is no means for i to get out of that situation and re-integrate context x. At the present stage, our logic of acceptance does not include an operation which consists of an agent (or set of agents) joining a certain social context.

Semantically, what we need is the opposite of the previous model restrictions: an operation of adding arrows labelled by $i{:}x$ to the model. Syntactically, what we need is a new form of announcements $i{\leftarrow}C{:}x$ and corresponding modal operators of type $[i{\leftarrow}C{:}x]$, meaning that agent i adopts C's acceptances in context x. In terms of Kripke models, the accessibility relation $\mathscr{A}_{i:x}$ is identified with $\mathscr{A}_{C:x}$. This kind operation of adding arrows is reminiscent of the logic of preference upgrade of van Benthem and Liu [20], and the logic of granting and revoking permissions of Pucella and Weissman [21].[7] More intuitively, $i{\leftarrow}C{:}x$ represents the operation of agent i's joining the social context x by adopting the acceptances of group C of members of x. After this operation, agent i should start to function again as members of x.

Other kinds of retraction operations can be devised. For example, one might want to consider the operation of creating a supergroup D of a given group C,

[7] See [22] for a systemic study of these operators.

where D takes over all of C's acceptances. The logical form of such an operation might be expressed by the operator $[D{:=}C:x]$. This operation should allow in particular to express that the agents in D start to function as members of x (i.e., to move from $\mathcal{A}_{D:x}\bot$ to $\neg\mathcal{A}_{D:x}\bot$), by taking over all acceptances of the agents in the subgroup C.

We are currently working on the technical issue of providing a semantic characterization and axiomatics of the previous operations $i{\leftarrow}C{:}x$ and $D{:=}C:x$ and corresponding modal operators $[i{\leftarrow}C{:}x]$ and $[D{:=}C:x]$.

6 Conclusion

In this paper we continued the studies initiated in [1], where the logic $\mathcal{AL}$, intended to formalize group (and individual) acceptances, was proposed. Here we extend $\mathcal{AL}$ by public announcements. As far as we know, our approach es novel and there is no other attempt to build up a logic of acceptance and public announcements.

The public announcement of ψ is an event that results in all agents learning that ψ is true. The public announcement of $\mathcal{A}_{C:x}\psi$ can be understood as a speech act. It simulates the announcement made by the group C itself, that they accept ψ while functioning as members of x. Therefore, as seen in Example 7, public announcements can be used to reason about the acceptances of agents when they express their own acceptances to each other. For instance, in that particular example we saw that a public announcement makes one of the agents quit the group, since he learns that the acceptances of the other agents are contrary to his own acceptances in the same context. As noted in Section 3.1, when the social context x denotes an institution, announcements of the form $x!\psi$ can be used to describe the event of issuing or promulgating a certain norm ψ (e.g. obligation, permission) within the context of the institution x.

We also provide a complete axiomatization for the logic of acceptances and announcements $\mathcal{ALA}$. As well as for epistemic logic with public announcements, the axiomatization given for $\mathcal{ALA}$ uses reduction axioms. In $\mathcal{ALA}$, group acceptances are related to individual acceptances, but they are not computed from them. It contrasts with epistemic logics where the concept of common knowledge (or common belief) is completely defined in terms of individual knowledge (or belief). Due to this difference, it is possible to have reduction axioms for group acceptances, while it is known to be impossible for common knowledge. Still, in Section 3.3 we argue that this is an intuitive feature of group acceptances.

7 Acknowledgements

We would like to thank the three anonymous reviewers for their very helpful comments.

The contribution by Andreas Herzig and Emiliano Lorini was supported by the French ANR project 'ForTrust: Social Trust Analysis and Formalization'.

The contribution by Tiago de Lima is part of the research program Moral Responsibility in R&D Networks, supported by the Netherlands Organisation for Scientific Research (NWO), under grant number 360-20-160.

References

1. Gaudou, B., Longin, D., Lorini, E., Tummolini, L.: Anchoring institutions in agents' attitudes: Towards a logical framework for autonomous multi-agent systems. In Padgham, L., Parkes, D.C., eds.: Proceedings of AAMAS 2008, IFAAMAS (2008) 728–735
2. Fagin, R., Halpern, J., Moses, Y., Vardi, M.: Reasoning about Knowledge. The MIT Press, Cambridge (1995)
3. Lewis, D.K.: Convention: a philosophical study. Harvard University Press, Cambridge (1969)
4. Gilbert, M.: On Social Facts. Routledge, London and New York (1989)
5. Tuomela, R.: The Philosophy of Sociality. Oxford University Press, Oxford (2007)
6. Hakli, P.: Group beliefs and the distinction between belief and acceptance. Cognitive Systems Research (7) (2006) 286–297
7. Lorini, E., Longin, D.: A logical approach to institutional dynamics: from acceptances to norms via legislators. In Brewka, G., Lang, J., eds.: Proceedings of KR 2008, AAAI Press (2008)
8. Gaudou, B., Herzig, A., Longin, D.: Grounding and the expression of belief . In: Proceedings of KR 2006, AAAI Press (2006) 211–229
9. Gaudou, B., Herzig, A., Longin, D., Nickles, M.: A new semantics for the FIPA agent communication language based on social attitudes. In Brewka, G., Coradeschi, S., Perini, A., Traverso, P., eds.: Proceedings of ECAI 2006, IOS Press (2006) 245–249
10. Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, Cambridge (2001)
11. Fornara, N., Colombetti, M.: Operational specification of a commitment-based agent communication language. In Castelfranchi, C., Johnson, W.L., eds.: Proceedings of AAMAS 2002, Bologna, ACM Press (2002) 535–542
12. Verdicchio, M., Colombetti, M.: A Logical Model of Social Commitment for Agent Communication. In: Proceedings of AAMAS 2003, ACM (2003) 528–535
13. Singh, M.P.: Agent communication languages: Rethinking the principles. IEEE Computer **31**(12) (December 1998) 40–47
14. Plaza, J.: Logics of public communications. In Emrich, M.L., Hadzikadic, M., Pfeifer, M.S., Ras, Z.W., eds.: Proceedings of ISMIS 1989. (1989) 201–216
15. Kooi, B.: Expressivity and completeness for public update logic via reduction axioms. Journal of Applied Non-Classical Logics **17**(2) (2007) 231–253
16. Walton, D.N., Krabbe, E.C.: Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning. State University of New-York Press (1995)
17. Gelati, J., Rotolo, A., Sartor, G., Governatori, G.: Normative autonomy and normative co-ordination: Declarative power, representation, and mandate. Artificial Intelligence and Law **12(1-2)** (2004) 53–81
18. Kooi, B., Van Benthem, J.: Reduction axioms for epistemic actions. In Schmidt, R., Pratt-Hartmann, I., Reynolds, M., Wansing, H., eds.: Proceedings AiML 2004, King's College Publications (2004) 197–211

19. Pettit, P.: Deliberative democracy and the discursive dilemma. Philosophical Issues **11** (2001) 268–99
20. van Benthem, J., Liu, F.: Dynamic logic of preference upgrade. Journal of Applied Non-Classical Logics **17**(2) (2007) 157–182
21. Pucella, R., Weissman, V.: Reasoning about dynamic policies. In: Proceedings of FOSSACS 2004. Number 2987 in LNCS, Springer Verlag (2004)
22. Aucher, G., Balbiani, P., Farias Del Cerro, L., Herzig, A.: Global and local graph modifiers. In: Methods for Modalities 5 (M4M-5). ENTCS, Elsevier (2007)