From Hypersets to Kripke Models in Logics of Announcements

Lawrence S. Moss

Abstract

This note discusses two semantics for a logic of group announcements and verifies that the two have the appropriate relation. The first semantics is the *hyperset semantics* of Gerbrandy [4, 5] and Gerbrandy and Groeneveld [6]. The second is the *Kripke model semantics* of Baltag, Moss and Solecki [2]. The relation between the two semantics was noted without proof in [2]. A proof does appear in Gerbrandy [5]. The presentation here is more algebraic in that it uses *coalgebras* and *final coalgebra maps* to give the semantics of [4, 5, 6], and the equivalence of the two semantics is shown without bisimulation.

Contents

1 Introduction

This note is concerned with two semantics for the logic of group-level updates. Such logics were first proposed by Plaza [8], and independently in Gerbrandy [4, 5] and Gerbrandy and Groeneveld [6]. The idea is to start with multi-agent modal logic and add propositional operators having to do with *conscious announcement* of propositions to groups of agents.

Syntax We fix a set AtProp of *atomic propositions* and a set $\mathcal{A}$ of *agents*. Then the usual multi-modal logic $\mathcal{L}$ over AtProp and $\mathcal{A}$ is generated in the following way:

$$\text{sentences } \varphi \qquad p \in \mathsf{AtProp} \quad | \quad \neg\varphi \quad | \quad \varphi \wedge \psi \quad | \quad \Box_A \varphi \ (A \in \mathcal{A})$$

We extend $\mathcal{L}$ to the language $\mathcal{L}(\Box^*)$ by adding the *common knowledge* operators $\Box^*_{\mathcal{B}}$, for all non-empty $\mathcal{B} \subseteq \mathcal{A}$; to $\mathcal{L}([\ \ ])$ by adding the binary *announcement operators* $[\varphi]_{\mathcal{B}}\psi$ for all non-empty $\mathcal{B} \subseteq \mathcal{A}$, and to $\mathcal{L}([\ \ ], \Box^*)$ by adding both constructs.

The key construct is the announcement operation, taking two propositions φ and ψ, and also a set $\mathcal{B}$ of agents, and returning the proposition $[\varphi]_{\mathcal{B}}\psi$. The intended meaning of this is if we take the group $\mathcal{B}$ and announce φ to this group in such a way that everyone in $\mathcal{B}$ is aware of the announcement and everyone outside of $\mathcal{B}$ is ignorant of it, then ψ holds. Of course, at this point we have not given a formal semantics. But this is the idea.

Our point This note does not discuss the uses of this logic and related variations. What we are interested in here are two semantics. The first semantics proposed uses hypersets, or non-wellfounded sets (see Aczel [1]). The second uses Kripke models. This second semantics is more general as we shall explain below. We will reverse the historical development and discuss the Kripke model semantics first because it requires less background to understand.

An example of an announcement To see what the subject is about, consider the case of a set $\mathcal{A}$ of three agents, say A, B, and C; two atomic sentences, p and q, and a Kripke model with four worlds w, x, y, and z depicted below. We have written out the accessibility relations in tabular form. In other notation, we have $w \to_A w$, $w \to_A x$, $x \to_C z$, $w \models p$, etc. We have some standard semantic facts, such as $w \models \neg\Box_B p$ (in w, B does not know p, since p is false in y), and $w \models \Box_A \neg\Box_B p$, etc.

Now suppose someone comes to each world $v \in W$ where p holds, takes A and B off to the side, and tells them (together) that indeed, p holds there. We want to *update* the worlds so that A and B's accessibility relations only include worlds where p is true. On the other hand, C was excluded from the announcement (and in fact at this point, we want to assume that C did not even know about it.) So C's accessibility relation should not change. We want to represent the updated version v' in a way that captures the epistemic alternatives available to each of the agents. Our proposal is that the worlds

w	
p	$A : w, x$
q	$B : w, y, z$
	$C : w, y$

x	
p	$A : w, x$
$\neg q$	$B : x$
	$C : x, z$

y	
$\neg p$	$A : y, z$
q	$B : w, y, z$
	$C : w, y$

z	
p	$A : y, z$
$\neg q$	$B : w, y, z$
	$C : x, z$

Figure 1: The Kripke model W

below represent the updated versions of the corresponding worlds above. Note first that in the updated worlds, we have kept C's accessibility relations the way they were, since C was not party to the communication.

w'	
p	$A : w', x'$
q	$B : w', z'$
	$C : w, y$

x'	
p	$A : w', x'$
$\neg q$	$B : x'$
	$C : x, z$

z'	
p	$A : z'$
$\neg q$	$B : w', z'$
	$C : x, z$

Figure 2: The updated model W' adds worlds w', x', and z' to W

Further, consider the update of w, and focus on the worlds accessible to B. Before the update, B used to think that w, y, and z were possible. It should be clear why there is no trace of y in the worlds accessible to B after the update of w. So we need to update w to *some* new world; this is why w' is needed. And w' should have the same propositional content, since announcements do not change facts. The main question might be: why in w' do A's accessibility relations point to w' and x' (and not w and x)? And why do B's point to w' and z' (and not w and z)? The reason is that announcing p to A and B should mean that not only do A and B think $\neg p$ is impossible, but also that $\neg p$ should be impossible *from all the worlds they think possible*, etc.

The main justification for our proposal on how to model actions comes from looking at semantic facts that hold in the updated worlds. For example, $\langle W', w' \rangle \models \Box^*_{\{A,B\}} p$. That is, in W' at the world in w', it is common knowledge among A and B that p holds. (We use the standard modeling of common knowledge via infinite iteration.) On the other hand, one can check that the updates do not change any knowledge facts for C: for all $v \in W$, $\langle W, v \rangle \models \Box_C \varphi$ iff $\langle W', v' \rangle \models \Box_C \varphi$. Both of these consequences seem right.

These considerations lead to what we call the *Kripke model* semantics for $\mathcal{L}([\ ])$. We interpret our languages on an arbitrary $\langle W, w \rangle$, where W is a multi-agent Kripke model with accessibilities $\rightarrow_A$ for each $A \in \mathcal{A}$, and w is a world of W. (We call such $\langle W, w \rangle$ *model-world pairs.*) However, to do this, we need to define $\langle W, w \rangle \models \varphi$ in terms of semantic facts about model-world pairs $\langle V, v \rangle$, where V differs from W. (This contrasts with the standard semantics of modal logic where W alone suffices.) Here are the main clauses in the definition:

$\langle W, w\rangle \models p$	if $w \models p$ in W in the usual way
$\langle W, w\rangle \models \Box_A\varphi$	if for all $v \in W$ such that $w \to_A v$, $\langle W, v\rangle \models \varphi$
$\langle W, w\rangle \models \Box^*_{\mathcal{B}}\varphi$	if $\langle W, w\rangle \models \Box_{A_1} \cdots \Box_{A_n}\varphi$ for all sequences $\langle A_1, \ldots, A_n\rangle$ from $\mathcal{B}^*$, including the empty sequence
$\langle W, w\rangle \models [\psi]_{\mathcal{B}}\varphi$	if $\langle W \oplus^{\psi,\mathcal{B}} W, \mathsf{new}(w)\rangle \models \varphi$, where $W \oplus^{\psi,\mathcal{B}} W$ and new are defined below

The worlds of $W \oplus^{\psi,\mathcal{B}} W$ are the elements of the disjoint sum $W + W$. (We take $W + W$ to be a set such as $(\{0\} \times W) \cup (\{1\} \times W)$.) We indicate the left injection of W into $W + W$ (the function taking w to $\langle 0, w\rangle$) by new, and the right injection by old. We make this into a Kripke model $W \oplus^{\psi,\mathcal{B}} W$ as follows, using the structure of W:

$\mathsf{new}(v) \to_A \mathsf{new}(u)$	if $v \to_A u$, $A \in \mathcal{B}$, and $u \models \psi$	$\mathsf{old}(v) \to_A \mathsf{old}(u)$	if $v \to_A u$
$\mathsf{new}(v) \to_A \mathsf{old}(u)$	if $v \to_A u$ and $A \notin \mathcal{B}$	$\mathsf{new}(v) \models p$	if $v \models p$
$\mathsf{old}(v) \to_A \mathsf{new}(u)$	never holds	$\mathsf{old}(v) \models p$	if $v \models p$

This completes the definition of $W \oplus^{\psi,\mathcal{B}} W$. So we have specified the semantics of our languages.

Once again, the idea is that $\langle W, w\rangle \models [\psi]_{\mathcal{B}}\varphi$ means that after all of the agents $A \in \mathcal{B}$ learn ψ via a public announcement in the world w, φ holds. The effect of this announcement in the semantics is that agents $A \in \mathcal{B}$ should discard all worlds where ψ fails; they are no longer "possible worlds." For $A \notin \mathcal{B}$, there is no effect of the announcement. Here is how this gets formalized in our two copies: The left copy is the updated copy and the right side is the "old world." It is clear that the right copy is isomorphic to the original W since no arrows depart from it. In the left copy, the agents not in $\mathcal{B}$ are not aware of the update, and so all of their possibilities are in the old world. But agents in $\mathcal{B}$ are restricted: if $A \in \mathcal{B}$, and $v \to_A u$, then $\mathsf{new}(u)$ is not a successor of $\mathsf{new}(v)$ in $W \oplus^{\psi,\mathcal{B}} W$ iff $u \not\models_A \psi$.

2 The hyperset update

The main point of this note is to compare the Kripke model update which we just studied with another one which we introduce now. This is the semantics via hypersets studied in Gerbrandy [4, 5] and in Gerbrandy and Groeneveld [6]. However, we are not going to work with the original formulations, since they involved some machinery from Barwise and Moss [3], especially the Corecursion Theorem. To understand the Corecursion Theorem from first principles would take some work, since it critically uses a set theory based on urelements. This would be new for practically everyone. It seems to me that it might be easier for some, and perhaps most, people who have an interest in these matters to see a different development.

So another goal of this paper is to give a new formulation of the hyperset semantics based on ideas coming from *coalgebra*. I believe that the importance of coalgebras for the study of transition systems is due to Aczel [1] (the same book which introduced non-wellfounded sets to a wide audience); of course there may be earlier references that I do not know of. There is some overhead to either view: one either has to learn some non-standard set theory (especially involving urelements), or some basic definitions from category theory. Part of the purpose of this note is to put down enough details so that someone familiar with the definitions of *category* and *functor* could understand our reformulation. A related purpose is to show that in order to work with the hyperset update, it is not really necessary to invoke even the concept of a bisimulation(!). One can get all of the needed results out of final coalgebra maps and their properties.

Here are a series of general definitions and remarks, in preparation for our formulation of the hyperset update:

Some general points of notation For any sets A and B, $A + B$ denotes the disjoint union of A and B. The exact definition of this set is irrelevant, but it brings along injections of A and B into it. We will write these as $\mathsf{new}_{A+B} : A \to A + B$ and $\mathsf{old}_{A+B} : B \to A + B$. We usually drop the subscripts. (It is more common to name these by something like inl and inr. We name them "old" and "new" to fit better with the metaphor of updating worlds.) If $f : A \to C$ and $g : B \to C$, then $\langle f, g\rangle : A + B \to C$ is the unique map such that $f = \langle f, g\rangle \circ \mathsf{new}$ and $g = \langle f, g\rangle \circ \mathsf{old}$. Also, if $f : A \to B$ and $g : C \to D$, then $f + g : A + C \to B + D$ is $\langle \mathsf{new}_{B+D} \circ f, \mathsf{old}_{B+D} \circ g\rangle$.

Re-packaging multi-modal Kripke structures as coalgebras A Kripke model W can be re-packaged as a coalgebra for the following functor F on sets or classes a and functions $f : a \to b$:

$$\begin{array}{rcl} F(a) & = & \mathcal{P}(\mathsf{AtProp}) \times \mathcal{A} \to \mathcal{P}(a) \\ F(f) & = & \langle S, \alpha\rangle \mapsto \langle S, f[\alpha(A)]\rangle \end{array}$$

An *F-coalgebra* is a pair $\langle A, e\rangle$, where A is a set or class and $e : A \to F(A)$. We drop the F since it is the only functor in the paper. To see that coalgebras correspond to Kripke structures, in one direction we take a model W to $\langle W, f\rangle$, where

$$f(w) \quad = \quad \langle \mathsf{AtProp}(w), A \mapsto \{v : w \to_A v\}\rangle$$

Here $\mathsf{AtProp}(w)$ is the set of atomic propositions true at w in W. In the other direction, a coalgebra $\langle E, e\rangle$ gives a Kripke model: the worlds are the elements $x \in E$; the atomic propositions true at x are the elements of $\pi_1(e(x))$, and $x \to_A y$ iff $y \in (\pi_2(e(x)))(A)$.

Continuing the correspondence, a *morphism of coalgebras $\langle A, e\rangle$ and $\langle B, f\rangle$* is a map $k : A \to B$ so that $f \circ k = F(k) \circ e$. Modulo re-packaging, this is a p-morphism of Kripke models.

A *final coalgebra* is a coalgebra $\langle C, c\rangle$ such that for all $\langle A, e\rangle$ there is a unique coalgebra morphism $k : C \to e$. The map k is called *the final coalgebra map for* $\langle A, e\rangle$. Concerning the functor F of this paper, there exists a final coalgebra $\langle C, c\rangle$. Assuming the Antifoundation Axiom, we can take C to be the greatest fixed point of F, considered as a monotone operator on sets. *AFA* also allows us to assume that c is the identity on C. To compare the two approaches, we therefore assume *AFA* and the consequences just noted.

A fact worth noting is that if we have a morphism $k : A \to B$ as above, and if the final coalgebra maps are $f : A \to C$ and $g : B \to C$, then $f = k \circ g$. To see this, one checks that $k \circ g$ is a morphism of coalgebras. So by the uniqueness part of finality, it must be identical to f.

Returning to hyperset update This is the map $g = g_{\psi,\mathcal{B}}$ defined on C in a few steps as follows: First, one considers $f : C \to C + C$ given by

$$\begin{array}{lll} f(c) & = & \langle \pi_1(c), \\ & & \qquad A \in \mathcal{B} \mapsto \{\mathsf{new}(b) : b \in \pi_2(c)(A) \ \& \ b \models_C \psi\} \\ & & \qquad A \notin \mathcal{B} \mapsto \{\mathsf{old}(b) : b \in \pi_2(c)(A)\} \\ & & \rangle \end{array}$$

Next, let $q : C \to F(C + C)$ be $F\mathsf{old} \circ c$. (Recall that $c : C \to F(C)$ is the identity.) Then $\langle f, q\rangle : C + C \to F(C + C)$ is a coalgebra. Notice that $F\mathsf{old} \circ c = q = \langle f, q\rangle \circ \mathsf{old}$, so $\mathsf{old} : C \to C + C$ is a coalgebra morphism from $\langle C, c\rangle$ to $\langle C + C, \langle f, q\rangle\rangle$. The final coalgebra map for c is of course id_C. Let $s : C + C \to C$ be the final coalgebra map. We define the *final coalgebra update* g to be $s \circ \mathsf{new}$. Note that the notation does not mention ψ and $\mathcal{B}$, but to remind ourselves of them we could write $g_{\psi,\mathcal{B}}$.

Remark In order to connect this more closely to the original formulations, it is probably useful to note the following result:

Proposition 2.1 *For any set or class A and any map $k : A \to F(A+C)$ there is a unique $h : A \to C$ such that $c \circ h = F(\langle h, \mathsf{id}_C\rangle)$.*

In fact, h is obtained by the same kind of construction which we used in defining g above: considering a bigger coalgebra and using its final coalgebra map. Proposition 2.1 is from [7]. That paper also contains other results along these lines and some applications.

Now, Proposition 2.1 is closely related to the ideas behind the original formulations from [4, 5, 6]. The idea is that we take $A = C$ and $k = f$. That is, we take an element c of C and return the pair consisting of the atomic propositions true at c, and also a function. That function takes an agent $A \notin B$ to set contains either elements of C accessible via A (the "old worlds" accessible from a by $\to_A$), and these worlds are especially tagged as such. The function also takes $A \in \mathcal{B}$ to the set of elements of C accessible via A which are *modified by this very operation*; again, these worlds are tagged as "new" so that they aren't confused with the old ones. The circularity in this formulation is a central ideological feature of the approach, and we won't discuss the reasons for it here. But we will mention that work on hypersets and coalgebras can clarify circular characterizations of this kind.

The model-world update Re-packaging $W \oplus_{\psi,\mathcal{B}} W$ from earlier gives a coalgebra $k : W + W \to F(W+W)$. In more detail,

$$\begin{array}{lcll} k(\mathsf{new}(w)) & = & \langle & \mathsf{AtProp}(w), \\ & & & \quad A \in \mathcal{B} \mapsto \{\mathsf{new}(v) : w \to_A v \ \& \ v \models_W \psi\} \\ & & & \quad A \notin \mathcal{B} \mapsto \{\mathsf{old}(v) : w \to_A v\} \\ & & \rangle & \\ k(\mathsf{old}(w)) & = & \langle & \mathsf{AtProp}(w), \quad A \mapsto \{\mathsf{old}(v) : w \to_A v\} \\ & & \rangle & \end{array}$$

We let $d_{W+W} : W + W \to C$ be the final coalgebra map.

Proposition 2.2 *Assume that ψ is preserved under final coalgebra maps. Then*

$$d_{W+W} \circ \mathsf{new} \quad = \quad g_{\psi,\mathcal{B}} \circ d_W.$$

This is the central result of this note. It is based on the following claim:

Claim $d_W + d_W$ is a coalgebra morphism from $\langle W + W, k\rangle$ to $\langle C + C, \langle f, q\rangle\rangle$.

This is the heart of the proof, we will go into the details. Consider some $\mathsf{old}(w) \in W + W$. Applying k gives $\langle \mathsf{AtProp}(w), A \mapsto \{\mathsf{old}(v) : w \to_A v\}\rangle$. Now applying $F(d_w + d_w)$ to this gives

$$\langle \mathsf{AtProp}(w), A \mapsto \{\mathsf{old}(d_w((v)) : w \to_A v\}\rangle.$$

Going the other way, $(d_W + d_W)(\mathsf{old}(w)) = \mathsf{old}_C(d_W(w))$. Then applying $\langle f, q\rangle$ to this gives $q(d_W(w)) = F\mathsf{old} \circ c(d_W(w))$. Now

$$c(d_W(w)) \quad = \quad d_W(w) \quad = \quad \langle \mathsf{AtProp}(d_W(w)), A \mapsto \{d_W(v) : w \to_A v\}\rangle,$$

so that

$$F\mathsf{old} \circ c(d_W(w)) \quad = \quad \langle \mathsf{AtProp}(d_W(w)), A \mapsto \{\mathsf{old}(d_W(v)) : w \to_A v\}\rangle.$$

Our proof in this case is completed by noticing that $\mathsf{AtProp}(w) = \mathsf{AtProp}(d_W(w))$. And this holds because coalgebra morphisms preserve the AtProp function, in view of the way such morphisms work.

The remaining case is some $\mathsf{new}(w) \in W+W$. Now $(F(d_W+d_W))(k(\mathsf{new}(w)))$ is

$$\begin{array}{ll} \langle & \mathsf{AtProp}(w), \\ & \quad A \in \mathcal{B} \mapsto \{\mathsf{new}(d_W(v)) : w \to_A v \ \& \ v \models_W \psi\} \\ & \quad A \notin \mathcal{B} \mapsto \{\mathsf{old}(d_W(v)) : w \to_A v\} \\ \rangle & \end{array}$$

And also $\langle f, q\rangle(d_W + d_W(\mathsf{new}(w)) = f(d_W(w))$. And

$$\begin{array}{lcl} f(d_W(w)) & = & \langle \pi_1(d_W(w)), \\ & & \quad A \in \mathcal{B} \mapsto \{\mathsf{new}(b) : b \in \pi_2(d_W(w))(A) \ \& \ b \models_C \psi\} \\ & & \quad A \notin \mathcal{B} \mapsto \{\mathsf{old}(b) : b \in \pi_2(d_W(w))(A)\} \\ & & \rangle \end{array}$$

These two expressions are equal: this involves tracing through our definitions. It is exactly here that we use the assumption that ψ is preserved under final coalgebra maps.

So now we verified the claim that $d_W + d_W$ is a coalgebra morphism from $\langle W + W, k\rangle$ to $\langle C + C, \langle f, q\rangle\rangle$. Recall that s is the final coalgebra map for $W + W$. It follows that $d_{W+W} = s \circ (d_W + d_W) = (g \circ d_W) + d_W$. Hence $d_{W+W} \circ \mathsf{new} = g \circ d_W$. This completes the proof of Proposition 2.2.

Proposition 2.3 *Every φ is preserved under final coalgebra maps.*

Proof By induction on φ. The only interesting step is when we assume that both ψ and φ are preserved under final coalgebra maps, and prove the same thing for $[\psi]\varphi$. Note that for all $\langle W, w\rangle$,

$$\begin{array}{llll} w \models_W [\psi]\varphi & \text{iff} & \mathsf{new}(w) \models_{W \oplus^{\psi,\mathcal{B}} W} \varphi & \\ & \text{iff} & ((d_W + d_W) \circ \mathsf{new})(w) \models_C \varphi & \text{ind. hyp. on } \varphi \\ & \text{iff} & (g_{\psi,\mathcal{B}} \circ d_W)(w) \models_C \varphi & \text{ind. hyp. on } \psi \text{ and Prop. 2.2} \\ & \text{iff} & d_W(w) \models_C [\psi]\varphi & \end{array}$$

$\dashv$

It follows that for all ψ and $\mathcal{B}$, $d_{W+W} \circ \mathsf{new} = g_{\psi,\mathcal{B}} \circ d_W$. This is the sense in which the two semantics for updates agree.

Conclusion The modest point of this note has been to give a coalgebraic formulation of the hyperset semantics, and to prove the equivalence of that semantics with the Kripke semantics. For more on both semantics, and on the area of announcement logics as a whole, one should see the papers [2, 4, 5, 6, 8].

References

[1] Peter Aczel, **Non-Well-Founded Sets**, CSLI Lecture Notes Number 14, CSLI Publications, 1988.

[2] Alexandru Baltag, Lawrence S. Moss, and Sławomir Solecki, Logics of Public Announcements, Common Knowledge, and Private Suspicions, to appear. Preliminary version in the *Proceedings of TARK-VII* , 1998.

[3] Jon Barwise and Lawrence Moss, *Vicious Circles.* CSLI Lecture Notes Number 60, CSLI Publications, Stanford, 1996.

[4] Jelle Gerbrandy, Dynamic Epistemic Logic, in **Logic, Language, and Information, vol. III**, CSLI, to appear 1999.

[5] Jelle Gerbrandy, **Bisimulations on Planet Kripke**, Ph.D. Dissertation, University of Amsterdam, 1999.

[6] Jelle Gerbrandy and Willem Groeneveld, Reasoning about information change, *Journal of Logic, Language, and Information* **6** (1997) (Special issue on Cognitive Actions in Focus (TARK 1996), edited by Johan van Benthem and Yoav Shoham), 147–169.

[7] Lawrence S. Moss, "Parametric Corecursion," to appear in *Theoretical Computer Science*.

[8] Jan Plaza, Logics of Public Communications, *Proceedings,* 4^{th} *International Symposium on Methodologies for Intelligent Systems*, 1989.