Dynamics of Public Commitments in Dialogue

Antoine Venant
Université Toulouse 3, IRIT
antoine.venant@irit.fr

Nicholas Asher
CNRS, IRIT
asher@irit.fr

Friday 20th March, 2015

Abstract

In this paper, we present a dynamic semantics for dialogue in terms of commitments. We use this to provide a model theoretic treatment of ambiguity and its effects on the evolutions of commitments as a dialogue proceeds. Our first semantics ensures common commitments and has a simple logic for which we provide a complete axiomatization. On the other hand, our semantics poses difficulties for the analysis of particular dialogue moves, in particular acknowledgments, and of disputes. We provide a second semantics that addresses these difficulties.

1 Introduction

Ambiguity arises in dialogue content at various levels of granularity—lexical, syntactic, semantic levels and at the level of discourse structure. In context, these ambiguities trigger pragmatic inferences. These different mechanisms interact in an especially complex way in computing a semantics in terms of commitments, which is for many reasons an attractive idea (Hamblin, 1987; Traum and Allen, 1994; Traum, 1994). To see why, assume as most do that conversation is a rational activity designed to achieve certain goals that the dialogue's participants aim to accomplish by talking with their interlocutors. Pragmatic inferences are drawn by rational conversationalists by reasoning on the basis of these conversational objectives and the dialogue context. Assume further that the *coherence* of a dialogue agent i's contribution is tied to the possibility of inferring coherence relations between i's utterances which often constrain in return the possible disambiguisation of those utterances, and force or cancel scalar implicatures (Asher, 2013). More particularly, a particular discourse move m typically presupposes a particular commitment on the part of m's agent concerning the commitments that *other* agents have made on a move n. This commitment may not be what the the author of n intended for innocuous or strategic reasons. Here is an example (from Venant et al., 2014).

(1) a. C: N. isn't coming to the meeting. It's been cancelled.
b. A: That's not why N. isn't coming. He's sick.
c. C: I didn't say that N. wasn't coming because the meeting was cancelled. The meeting is cancelled because N. isn't coming.

C's initial contribution contains a discourse ambiguity. A has taken C to be committed one of its possible disambiguations when C turns out to have committed to the other. But A is not wrong to take C to be committed to what he takes him to commit to, we think; and so there is a question of how to represent the meaning of this exchange (Venant et al., 2014).

We address these considerations by providing a *dynamic* notion of public commitments. By performing a dialogue act X, an agent A *commits* to some content, potentially ambiguous. By responding with another dialogue act Y, a second agent B might commit to having interpreted X in a particular way. (or else to be incoherent). What is central in putting this notion of meaning at work, and the object of this paper is the dynamics of such commitments. We provide an axiomatization for a simple, first kind

Proceedings of the 11th International Conference on Computational Semantics, pages 272–282,
London, UK, April 15-17 2015.

of dynamics, look at some problems this simple dynamics has with the semantics of dialogue acts like acknoweldgments, and produce a second dynamics that resolves the problems of the first.

2 Public Commitments, Ambiguities and Strategic Context

A conversation is (ideally at least) a sequential exchange of messages. As stated in the introduction, it is also a rational activity. Messages are exchanged for some purpose; conversationalists expect something out of the conversation. In a fully cooperative settings, they typically seek an exchange of information and update their beliefs accordingly with the information they receive. In such case the conversational process closely follows the process of successive belief-state updates. It can nonetheless not be *equated* with such cognitive updates even in the cooperative case. Agents can pretend to believe in the responses of others for various purposes, even when they know their contributions are incorrect. In (partially) non-cooperative settings, nothing guarantees that a message will be believed. One can add uncertainty to the picture and, for instance, model the effect of a message as modifying a probability distribution over possible states of the world. However in a fully non-cooperative setting, *e.g.* where one agent i does not believe anything that agent j says, and this is known by everyone else, the reception of any message from j should leave the i's uncertainty exactly as it was. Nevertheless, some conversations actually take place in such settings (political debates or discussions between adversaries with opposing views). Thus, even if conversationalists' interests are opposed in these cases, there must be additional constraints that i) make it rational for them to have the conversation and ii) provide some effect to the sending of a message. Following (Venant et al., 2014), we formulate a general theory of dialogue content even for such cases using public commitments: even when agents do not communicate beliefs to their interlocutor, they communicate commitments. Performing an utterance publicly commits its author to the content of that utterance (Hamblin, 1987). Therefore, the conversational objectives of an agent are not solely expressed in terms of the sole informational content of the messages, but in terms of the public commitments of every participants as well. Typically, i may ask j "Did you eat all of my cookies?", knowing perfectly well that j did and has no incentive to tell the truth anyway, but with a conversational objective of just having i commit to an answer (either have him admit the fact, or gather material that will allow to confront him later.

Sending or not sending a message may thus have strategic consequences while leaving the agents' belief states unchanged. While much of game theory applied to language assigns utilities and thus preferences to belief states, we can also think of preferences over commitments. Player i's goal may be to extract a certain commitment from j; that is, i will be happy with her conversational performance if j commits to some proposition φ—for instance, in a philosophical debate, i might hope to show that j commits to a contradiction or some absurd proposition. Conversely, it may then be part of j's *winning condition* to avoid a commitment to φ. This makes ambiguity an essential strategic tool: by uttering an ambiguous message j may on one reading not commit to φ, while on another reading she does. Our example (1) from the introduction already reveals how ambiguities lead to different commitments. (1-b) reveals that A takes C in (1-a) to have committed to a particular rhetorical connection between N's not coming to the meeting and the meetings cancellation—namely, one of explanation: that N isn't coming *because* the meeting's been cancelled. We know this because A's contribution in (1-b) has the form of a Correction (Asher and Lascarides, 2003). However, (1-a) is genuinely ambiguous: it also has the reading on which N isn't coming and so *as a result* the meeting's been cancelled. And (1-c) reveals that C commits to having committed to the result reading with (1-a). Now suppose A's goal was to get C to commit to an attackable, thereby perhaps impugning his credibility. C's message looks like it satisfied A's goal, but because it was ambiguous, C can avoid the attack that A might have planned.

In light of this discussion, our analysis of commitments must involve commitments to ambiguous propositions. On the other hand, our informal analysis of (1) show that a proper analysis of the dynamic of commitments must also involve *nested commitments*. For instance, (1-b) implies that A commits that C commits to the explanation reading of (1-a). In fact Venant et al. (2014) show that such nested commitments are a consequence of the semantics of rhetorical relations. In what follows we develop

a dynamic account of nested commitments with ambiguous signals. Our approach is compatible with but does not assume any compact representation of the ambiguous signal as in, e.g., Reyle (1993), as we represent all disambiguations model-theoretically. We hope in future work to investigate compact representations of our models.

3 A Language for the Dynamics of public Commitment with Ambiguities

To model the dynamic of public commitment with ambiguous signals, we assume here an abstract, simplified view of conservations as sequences of $\langle (linguistic)action, speaker\rangle$ pairs. We will build ambiguity into the linguistic actions recursively: in the base case, an action is an unambiguous utterance, whose content we simplify to be a propositional formula. Ambiguous actions are recursively constructed from a set of (lower-level) actions (representing its possible disambiguations). In order to explain this in more formal terms, we introduce some preliminary definitions: Let PROP denote a set of propositional variables (at most countably infinite) and I a set of agents. We define simultaneously the set of actions $\mathcal{A}$ and formulas $\mathcal{L}_0$:

Definition 1 (Actions and formulas)**.** $\mathcal{A}$ and $\mathcal{L}_0$ are the smallest sets such that:

$$\begin{array}{ll} \forall p \in \text{PROP}\ p \in \mathcal{L}_0 & \forall \varphi \in \mathcal{L}_0\ \varphi! \in \mathcal{A} \\ \forall \varphi, \psi \in \mathcal{L}_0 \forall i \in I\ \neg\varphi, C_i\varphi, \varphi \wedge \psi \in \mathcal{L}_0 & \text{for any finite collection of actions } (\alpha_s)_{s=1\ldots n} \text{ in } \mathcal{A} \\ \forall \varphi \in \mathcal{L}_0 \forall \alpha \in \mathcal{A} \forall i \in I\ [\alpha^i]\varphi \in \mathcal{L}_0 & (\sim \alpha_s)_{s=1\ldots n} \in \mathcal{A} \end{array}$$

Additional logical constants and connectors are defined as usual: $\varphi \vee \psi \equiv \neg(\neg\varphi \wedge \neg\psi)$, $\varphi \rightarrow \psi \equiv \neg\varphi \vee \psi$, $\bot \equiv p \wedge \neg p$, $\bot \equiv p \vee \neg p$.

The semantics of our language is based on that for Public Announcements logic (PAL) with private suspicions introduced in Baltag et al. (1998). More specifically, we translate each of our actions in (Baltag et al., 1998)'s *action structures* and then rely on their semantics.

Recalling some basic definitions, a *frame* is a tuple $\langle W, (R_i)_{i\in X}\rangle$ with W a set of worlds and for each $i \in I$, R_i is a binary relation over W, and a *model* $\mathcal{M}$ is a pair $\langle \mathcal{F}, \nu\rangle$ with $\mathcal{F}$ a Kripke frame and $\nu : W \mapsto \wp(\text{PROP})$ an assignment at each world w of propositional variables true at w. We will sometimes use models as superscripts for set of worlds $W^{\mathcal{M}}$, or accessibility relations $R_i^{\mathcal{M}}$ to refer to the set of worlds or the relation of that particular model or frame. We will also abuse notation and write $w \in \mathcal{M}$ as a shortcut for $w \in W^{\mathcal{M}}$. A *pointed model* is a pair $\langle \mathcal{M}, w\rangle$ with $w \in W^{\mathcal{M}}$.

The semantics of action-free formulas is as usual with respect to a pointed model:

Definition 2 (Semantics of static formulas)**.**

$$\begin{array}{l} \langle \mathcal{M}, w\rangle \models p \text{ iff } p \in \nu^{\mathcal{M}}(w) \\ \langle \mathcal{M}, w\rangle \models \neg\varphi \text{ iff } \langle \mathcal{M}, w\rangle \not\models \varphi \\ \langle \mathcal{M}, w\rangle \models \varphi \wedge \psi \text{ iff } \langle \mathcal{M}, w\rangle \models \varphi \text{ and } \langle \mathcal{M}, w\rangle \models \psi \\ \langle \mathcal{M}, w\rangle \models C_i\varphi \text{ iff } \forall w', R_i^{\mathcal{M}}(w, w') \rightarrow \langle \mathcal{M}, w'\rangle \models \varphi \end{array}$$

In order to provide a semantics for terms with actions, we need (Baltag et al., 1998)'s definition of an *action structure*:

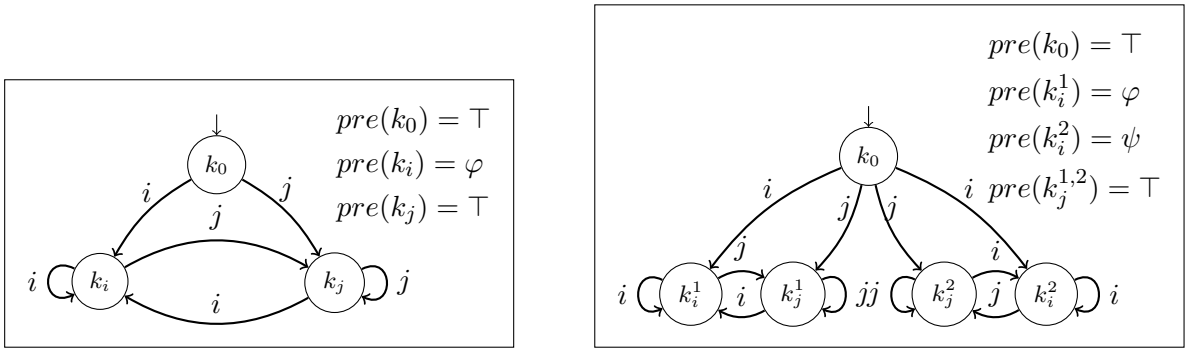
Definition 3 (Action Structures)**.** An action structure is a pair $\langle \mathcal{F}, pre\rangle$ where $pre : W^{\mathcal{F}} \mapsto \mathcal{L}_0$ associates to each world in $\mathcal{F}$ a formula, called the *precondition* of this world.

Interpreting formulas with actions require us to first update the model with the action, then to evaluate the formulas with respect to the updated model. As mentioned earlier, we proceed in two steps; we first associate a pointed action-structure with each action in $\mathcal{A} \times I$, and then classically update the model with this action. We first recall (Baltag et al., 1998)'s informal definition of the update operation. The

update of a model $\mathcal{M}$ through an action a is obtained by taking, for each world in a's structure a different copy of $\mathcal{M}$'s world that satisfy the precondition, then allowing a transition for agent i from a world to another iff i)the two worlds were initialy i-related and ii)the two copies they belong to are i-related in a's structure. More precisely:

Definition 4 (Action updates)**.** Let $S = \langle \mathcal{F}, pre \rangle$ be an action structure. Let $k \in S$ and let $\langle \mathcal{M}, w_0 \rangle$ be a pointed model. Let $|\varphi|^{\mathcal{M}} = \{w \in \mathcal{M} \mid \mathcal{M}, w \models \varphi\}$. If $w_0 \notin pre(k)$, the update $\langle \mathcal{M}, w_0 \rangle \star \langle S, k \rangle$ *fails*. Otherwise, it is defined as $\langle \mathcal{M}^S, (w_0, k) \rangle$ the model with $W^S = \bigcup_{l \in S} |pre(l)|^{\mathcal{M}} \times l$ as set of worlds, accessibility relations defined as $R_i^{\mathcal{M}^S}((w,l),(w',l'))$ iff i)$R^{\mathcal{M}}(w,w')$ and ii) $R_i^S(l,l')$, and valuations left unchanged *i.e.* $\nu((w,l)) = \nu(w)$.

We now provide the translation of conversational moves of our language (*i.e.* elements of $\mathcal{A} \times I$) into pointed action-structures:



(a) Unambiguous move $\varphi!^i$

(b) Ambiguous move $(\varphi! \sim \psi!)^i$

Figure 1: Some action structures

A simple unambiguous discourse move by i will generate a *common commitment* to $C_i\varphi$. An ambiguous move, on the other hand, will not but will involve a disjunction of common commitments. A common commitment for a group G towards a proposition φ, $C^*_G\varphi$, has the effect that $C_G\varphi \wedge C_G C_G\varphi \wedge \ldots C_G(C_G)^n\varphi \wedge \ldots$ (analogously to common knowledge). Semantically, we define common commitments for a group G, $C^*_G\varphi$, as

$$\langle \mathcal{M}, w \rangle \models C^*_G\varphi \text{ iff } \forall w'(\bigcup_{x \in G} R_x)^+(w,w') \rightarrow \langle \mathcal{M}, w \rangle \models \varphi,$$

where the union $\cup$ of two relations is defined as $(R \cup R)(w,w')$ iff $R(w,w')$ or $R'(w,w')$, and R^+ denotes the transitive closure of the binary relation R ($R^+(w,w')$ iff $\exists n > 0\, \exists w_1, \ldots w_n\ w_n = w' \wedge R(w,w_1) \wedge \ldots R(w_{k-1},w_k) \ldots R(w_{n-1},w_n)$).

Definition 5 (Interpretation of conversational moves)**.** The interpretation function $[\![\cdot]\!]$ interprets conversational moves of $\mathcal{A} \times I$) as pointed action-structures. Let $m = \alpha!^i \in \mathcal{A} \times I$. $[\![m]\!]$ is defined inductively over α:

- If $\alpha = \varphi!$ then $[\![\alpha]\!] = \langle K, pre, k_0 \rangle$ with $K = \{k_0, k_i, k_j\}$, accessibility relation is defined as $R_i^K(k_{\{0,i,j\}}, k_i)$, $R_j^K(k_{\{0,i,j\}}, k_j)$ and no other transitions; preconditions are defined as $pre(k_0) = pre(k_j) = \top$ and $pre(k_i) = \varphi$. The pointed world is k_0 and the action-structure is depicted in figure 1a.
- If $\alpha = (\sim \alpha_s)_{s=1\ldots n}$, let $\langle K^s, pre^s, k_0^s \rangle = [\![\alpha_s!^i]\!]$ be the action structure recursively computed for $\alpha_s!^i$. Assuming the K^s- and $K^{s'}$-worlds are disjoint for $s \neq s'$ (otherwise, first take disjoint copies of the K^ss), define $[\![\alpha_m]\!] = \langle K, pre, k_0 \rangle$ with $K = \bigcup_s K^s \setminus \{k_0^s\}$, accessibility relations defined as i)$\forall k \in K^s\, x \in \{i,j\}\quad R_x^K(k_0,k)$ iff $R_x^{K^s}(k_0^s,k)$, ii) $\forall k,k' \in K_s \setminus \{k_0^s\}\ R_x^K(k,k')$ iff

$R_x^{K^s}(k, k')$ and iii) there are no other transitions than the one previously listed. pre is defined as $pre(k_0) = \top$ and for $pre(k) = pre^s(k)$ for $k \in K^s$. The pointed world is k_0. Figure 1b shows the action-structure $[\![(\varphi! \sim \psi!)^i]\!]$ for a move by i which is ambiguous between a commitment to φ and one to ψ.

Note that given this definition, $\sim$ is "associative" in the sense that

$$[\![((\alpha_1 \sim \alpha_2) \sim \alpha_3)^i]\!] = (\alpha_1 \sim (\alpha_2 \sim \alpha_3))^i]\!] = [\![(\alpha_1 \sim \alpha_2 \sim \alpha_3)^i]\!] \text{ (up to renaming of the worlds).}$$

Armed with these definitions, we can now complete the semantics of $\mathcal{L}_0$ providing the semantics for action terms:

Definition 6. Semantics of dynamic formulas:

$$\langle \mathcal{M}, w \rangle \models [\alpha^i]\varphi \text{ iff } \langle \mathcal{M}, w \rangle \star [\![\alpha^i]\!] \models \varphi$$

Note that due to the fact $[\![\alpha^i]\!]$'s pointed world always has $\top$ as precondition, the update $\langle \mathcal{M}, w \rangle \star [\![\alpha^i]\!]$ cannot fail and the definition is correct.

Worked out example We illustrate our dynamics by providing an abstract but principled view of the evolving commitments in (2):

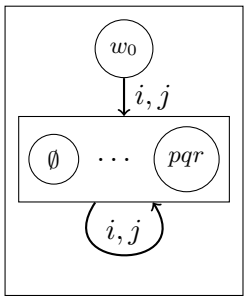
(2)
a. i : I have my piano lesson in ten minutes. When I get back the shop will be closed.
b. i : And there is no more beer.
c. j : I am not going to get you beer. Go get it yourself.
d. i : I did not say that. I am not asking you to get it.
e. j : Oh yes you did.

What is central to the picture here? i commits to some proposition (we abbreviate it as p), and then to something else, that in its context of utterance might be interpreted as a commitment on a request for j to get beer. i makes an utterance that entails that he takes j to be committed to the request. j then this dispute this commitment of his. i refuses the correction. Assume that we can refer to an external semantic/pragmatic theory that licenses or rejects possible interpretations of a sentence in context, and that such a linguistic theory tells us that (2-b) as (at least) an assertion that there is no more beer ($\neg b$) licenses a pragmatic inference to a request for i to get some beer ($\neg b \wedge r$). We can then correctly describe (2) as involving these action sequences:

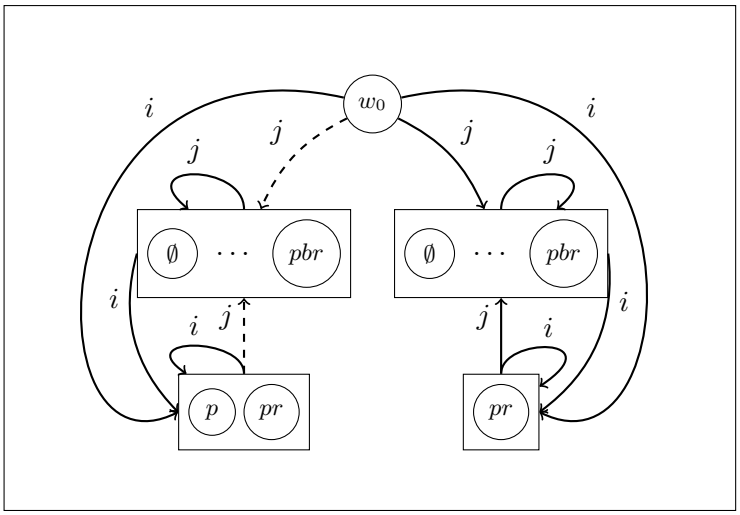
(3)
a. $i : p!^i$
b. $i : (\neg b! \sim (\neg b \wedge r)!)^i$
c. $j : (C_i r)!^j$
d. $i : (C_j C_i r \wedge \neg C_i r \wedge \neg r)!^i$
e. $j : (C_i(C_j C_i r \wedge \neg C_i r) \wedge C_i r)!^i$

Figures 2, show how the dynamics transform the initial model. In order to keep the figures readable, we graphically group nodes into clusters, edges going in and out of these clusters are to be understood as distributing over each inner node. We also omit some isolated worlds that therefore have no impact (*i.e.* they are present in the definition of action update, but not reachable from any other world). Nodes are labelled by their valuations, except for the actual world labelled as w_0. The initial model of the conversation is depicted in figure 2a.

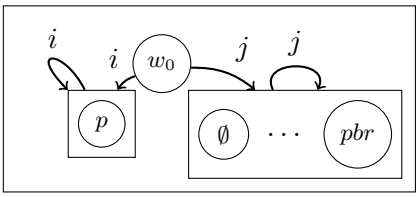
The initial model in figure 2a shows that neither speaker commits to anything. Figure 2 shows how i's assertions in (3-b)–(3-d) have transformed the commitment space for j and i: after updating with (3-b), i's public commitments and j's commitments concerning i's commitments are ambiguous as to whether the implicature to go get beer holds; but after the update with (3-c), only i's commitments remain ambiguous. j's commitments concerning i''s commitments are no longer ambiguous; he commits



(a) Initial Model for example (3)



(b) Models after (3-b) and (3-c), respectively with and without the dashed edges.



(c) Model after (3-d)

Figure 2: Models at different stages of example (3). Arrows should be understood as distributing over all inner nodes.

to i's having committed to the implicature that he should go get the beer. After (3-d), j's commitments concerning i have become inconsistent. This is a consequence of our strong modeling assumptions about a perfect communication channel leading to common commitments. We will see in section 5 yet another reason to weaken our proposal.

4 Complete Deduction System for $\mathcal{L}_0$

One of the interests in keeping the base language of our analysis simple is to be able to investigate the logical properties of the dynamics of commitments. Accordingly, in this section, we present a complete deduction system for $\mathcal{L}_0$. The system and completness proof follow from the general picture drawn in (Baltag et al., 1998), where the authors provide a complete deduction system for the language allowing any kind of action-structure. It turns out however, that the restricted action-structures that are the interpretations of our conversational moves $\mathcal{A}$(see definition 5) allow nice simplifications, most notably the elimination of any reference to action-structures in the syntactic rules. This allows us to have deduction system for L_0 which does not require embedding of L_0 into an larger language with additional syntactic constructions. The deduction system is presented on figure 3.

In order to proof completeness of the above system, we adapt step by step (Baltag et al., 1998)'s proof to the simplified system. The proof function by reduction of the logic to the static logic K. The idea behind the proof is, once soundness is established, to see our system's axioms as rewrite rules (rewriting the left-hand sides of the equivalences into the right-hand sides), and show that the system is able to proof the equivalence of any given formula to an action-free formula. From there it is quite straightforward to reduce provability of a formula to provability of an action free formula, which is granted as $K-$axioms are part of our system.

Lemma 1. *The deduction rules are sound.*

All propositional validities

from $\vdash \varphi \to \psi$ and $\vdash \varphi$ to infer $\vdash \psi$ (MP)

$\vdash [\alpha^i](\varphi \to \psi) \to ([\alpha^i](\varphi) \to [\alpha^i](\psi))$ ($[\alpha^i]$-normality)

$\vdash C_i(\varphi \to \psi) \to (C_i(\varphi) \to C_i(\psi))$ (C-normality)

from $\vdash \varphi$ to infer $\vdash C_i\varphi$ (C-necessitation)

from $\vdash \varphi$ to infer $\vdash [\alpha^i]\varphi$ ($[\alpha^i]$-necessitation)

$\vdash [\alpha^i]p \leftrightarrow p$ (rw1)

$\vdash [\alpha^i]\neg\psi \leftrightarrow \neg[\alpha^i]\psi$ (rw2)

$\vdash [\alpha^i](\psi_1 \wedge \psi_2) \leftrightarrow [\alpha^i]\psi_1 \wedge [\alpha^i]\varphi_2$ (rw3)

$\vdash [\varphi!^i]C_j\psi \leftrightarrow C_j[\varphi!^i]\psi \qquad (\text{for } j \neq i)$ (rw4)

$\vdash [\varphi!^i]C_i\psi \leftrightarrow C_i(\varphi \to [\varphi!^i]\psi)$ (rw5)

$\vdash [\sim(\alpha_s)^i_{s\in S}]C_x\varphi \leftrightarrow \bigwedge_S [\alpha^i_s]C_x\varphi$ (rw6)

Figure 3: Deduction system for $\mathcal{L}_0$
$i, j, x \in I$, $p \in$ PROP

Proof. The proof is trivial for Modus Ponens, necessitation and normality rules. (rw1)'s, (rw2)'s and (rw3) soundness follows directly from definitions (and the fact that our actions never fail). (rw4),(rw5) and (rw6) requires a little more work:

– Let for a world $w \in \mathcal{M}$ $\langle \mathcal{M}^\alpha, (w_0, k_0)\rangle$ denote $\langle \mathcal{M}, w\rangle \star [\![\varphi!^i]\!]$, the update of $\mathcal{M}$ by action φ^i at w (recall that the set of worlds and relations of $\mathcal{M}^\alpha$ does not depend on w). Notice first that the following is true for any formula γ, $k \in \{k_0, k_i, k_j\}$ and w such that $(w, k) \in \mathcal{M}^\alpha$:

$$\langle \mathcal{M}^\alpha, (w, k_0)\rangle \text{ and } \langle \mathcal{M}^\alpha, (w, k)\rangle \text{ are bissimilar.}$$

by definition the valuations of (w, k_0) and (w, k) are the same. Since the worlds accessible from k_0 in $[\![\varphi]\!]$ are exactly those accessible from k, it follows from the definition of $\mathcal{M}^\alpha$ that the worlds accessible from (w, k_0) are exactly those accessible from (w, k) which is sufficient to establish the bissimulation.

Let us now prove the soundness of (rw5). The proof for (rw4) is similar. Let $\langle \mathcal{M}, w_0\rangle$ be a pointed model. $\langle \mathcal{M}, w\rangle \models [\varphi!^i]C_i\psi$ iff $\langle \mathcal{M}^\alpha, (w_0, k_0)\rangle \models C_i\psi$ iff $\forall (w, k) \in \mathcal{M}^\alpha\ R_i((w_0, k_0), (w, k)) \to \langle \mathcal{M}^\alpha, (w, k)\rangle \models \psi$. Since we have shown that $\langle \mathcal{M}^\alpha, (w, k)\rangle$ is bissimilar to (w, k_0) and using the definition of $\mathcal{M}^\alpha$, we find the above to be further equivalent to $\langle \mathcal{M}, w\rangle \models \varphi$ and $R_i^{\mathcal{M}}(w_0, w) \to \langle \mathcal{M}^\alpha, (w, k_0)\rangle \models \psi$. But since by definition $\langle \mathcal{M}, w\rangle \models [\varphi!^i]\psi$ iff $\langle \mathcal{M}^\alpha, (w, k_0)\rangle \models \psi$, satisfaction of the initial formula is finally equivalent to $\langle \mathcal{M}, w_0\rangle \models C_i(\varphi \to [\varphi!^i]\psi)$.

– Let $\alpha = ((\sim\alpha_s)_{s\in S})$ and $x \in I$. by construction, for any world k x-accessible from k_0 in $[\![\alpha]\!]$ there is world k_s in $[\![\alpha_s]\!]$ x-accessible from k_0 and such that $\langle K^\alpha, k\rangle$ and $\langle K^{\alpha_s}, k_s\rangle$ are bissimilar. This implies that for any world in (w, k) x-accessible from (w_0, k_0) in $\mathcal{M}^\alpha$ there is a $s \in S$ and a world (w, k_s) in $\mathcal{M}^{\alpha_s}$ such that $\langle \mathcal{M}^\alpha, (w, k)\rangle$ and $\langle \mathcal{M}^{\alpha_s}, (w, k_s)\rangle$ are bissimilar. Conversely, for any world $(w, k_s) \in \mathcal{M}^{\alpha_s}$ x-accessible from (w_0, k_0) there is a bissimilar world $(w, k) \in \mathcal{M}^\alpha$ x-accessible from (w_0, k_0).

Assume that for each α_s, $\langle \mathcal{M}^{\alpha_s}, (w_0, k_0)\rangle \models C_x\varphi$. Let (w, k) be x-accessible from $(w_0, k_0) \in \mathcal{M}^\alpha$. We me must have $\langle \mathcal{M}^{\alpha_s}, (w, k_s)\rangle \models \varphi$ and by bissimilarity $\langle \mathcal{M}^\alpha, (w, k)\rangle \models \varphi$, hence $\langle \mathcal{M}^\alpha, (w_0, k_0)\rangle \models C_x\varphi$.

Conversely, assume that $\langle \mathcal{M}^{\alpha}, (w_0, k_0)\rangle \models C_x\varphi$. Let $(w, k_s) \in \mathcal{M}^{\alpha_s}$ be a world x-accessible from (w_0, k_0), there is a $(w, k) \in \mathcal{M}^{\alpha}$ bissimilar to (w, k_s) and therefore $\langle \mathcal{M}^{\alpha_s}, (w, k_s)\rangle \models \varphi$ and $\langle \mathcal{M}^{\alpha_s}, (w_0, k_0)\rangle \models C_x\varphi$.

All together we can conclude to $\langle \mathcal{M}, w_0\rangle \models [(\sim \alpha_s)_{s\in S}]C_x\varphi$ iff $\langle \mathcal{M}, w_0\rangle \models \bigwedge_S [\alpha_s]C_x\varphi$, *i.e.* (rw6) is sound.

□

Lemma 2. *Rules* (rw1)–(rw6) *seen as rewrite rules rewriting the left-hand sides of the equivalences into the right-hand sides form a terminating rewriting system.*

This is classicaly obtained from (for instance) the technique of lexicographic path ordering. We do not detail the proof here for sake of space.

Since a rewrite-rule can always be applied to a formula starting with an action, a direct corollary of lemma 2 is that any formula can be rewritten into an action-free formula by the by the rewrite system obtained from the deduction rules.

Lemma 3. *If* $\vdash \varphi \leftrightarrow \psi$ *then for all well formed formula* γ *of* $\mathcal{L}_0$, $\vdash \gamma[\varphi/p] \leftrightarrow \gamma[\psi/p]$.

This can be achieved by induction over the length of γ.

Proposition 1. *The deduction system is strongly complete.*

Together with lemma 2, lemma 3 yields through a quick induction over the rewrite steps, that for any formula $\varphi \in \mathcal{L}_0$, there is an action-free formula φ_0 (one of φ normal forms w.r.t the rewrite system) such that $\vdash \varphi \leftrightarrow \varphi_0$, from there the strong completeness is reduced to the one of modal logic K.

5 Acknowledgments and corrections

Next we look at two particular dialogue moves that affect commitments in complex ways: acknowledgments and corrections. For many researchers Clark (1996); Ginzburg (2012); Traum and Allen (1994), *inter alia*, an acknowledgment as in (4)c by 0 of a discourse move m by 1 can signal that 0 has understood what 1 has said, or that 0 has committed that 1 has committed to a content p with m, and serve to "ground" or to establish a mutual belief that 1 has committed to p. Corrections, and self-corrections, as in (4)d, on the other hand, serve to remove commitments.

(4)
a. 0: Did you have a bank account in this bank?
b. 1: No sir.
c. 0: OK. So you're saying that you did not have a bank account at Credit Suisse?
d. 1: No. sorry, in fact, I had an account there.
e. 0: OK thank you.

We believe that acknowledgments perform an important grounding function in a commitment based semantics for dialogue: they serve to produce *common commitments*, the commitment analogue to mutual beliefs. There is, however, a problem with our semantics when it comes to treating acknowledgments: grounding acknowledgments are semantically superfluous; if m entails p, then i's making m entails $C^*_G C_i p$. Rational speakers should never acknowledge in a grounding sense; i's acknowledgment of j can only mean that i agrees with the content of j's move, which manifestly it does not, as in (4)c (such acknowledgments are often present in legal questioning).

We have other indications that our dialogue semantics so far is not quite right. For instance, saying "φ" is not the same as saying "I commit to φ", and simply i's saying "φ" should not induce via the logic alone a common commitment that $C_i\varphi$. Of course if i says "φ" and then "I did not say φ" he his ultimately saying something *false*. But this is not the same as him committing to an *absurdity*, *i.e.* an inconsistency not just with the actual state of the world, but in its own right. As already illustrated,

the dynamics of sections 4 and 5 validates $\langle \mathcal{M}, w\rangle \models [\varphi^i]C^*_{i,j}C_i\psi$ which indeed makes it impossible for one to consistently perform such a sequence of utterances. This hypothesis can be seen either as a consequence of perfect linguistic knowledge and a communication channel, and mutual commitment of the agents thereto.

To treat acknowledgments, we first enrich our language into a language $\mathcal{L}_{ack}$ with actions for acknowledgments. We do that by adding the recursive construction $Ack(\alpha^x)$ to the set of linguistic action $\mathcal{A}$, for any $\alpha \in \mathcal{A}$ and $x \in I$. Defining the semantics of $\mathcal{L}_{ack}$ just requires us to define the interpretation of acknowledgment-actions into action-structures. Let $\alpha \in \mathcal{A}$ be a linguistic action. Let $\langle K^\alpha, k^\alpha_0, pre^\alpha\rangle = [\![\alpha^x]\!]$. Let k_0 and k_j be "fresh" symbols not appearing in K^α.

$$[\![Ack(\alpha^x)^i]\!] = \langle \{k_0, k_j\} \cup K_\alpha, k_0, pre\rangle$$

Accessibility relations are defined as $R_i(k_0, k^\alpha_0)$, $R_j(k_0, k_j)$, $R_{i,j}(k_j, k_j)$, $\forall k, k' \in K^\alpha, \forall x \in \{i, j\}\ R_x(k, k')$ iff $R^{K^\alpha}_x(k, k')$ and no other transitions. $pre(k_0) = pre(k_j) = \top$ and pre coincide with pre^α on K^α.

It is easy to check that effects of action $Ack(\alpha^x)^i$ commit i to the effects of α^x and that, given the dynamics of sections 4 and 5, acknowledgments of previous actions have no effect in the sense that $\langle \mathcal{M}, w\rangle \models [\alpha^x][Ack(\alpha^x)^i]\varphi$ iff $\langle \mathcal{M}, w\rangle \models [\alpha^x]\varphi$. This formalizes the problem. To address the problem, we provide an alternative semantics for $\mathcal{L}_{ack}$, in which we redefine the interpretation of linguistic actions as action structures. Only unambiguous utterance-actions need a new definition, as the recursive computation mechanism of action-structures for ambiguous utterances- and acknoledgments-actions stays the same.

Definition 7 (Weak action interpretation)**.** Define $[\![\cdot]\!]^1$ by

$$[\![\varphi!^i]\!]^1 = \langle k_0, k_i, k_1, k_0, pre\rangle$$

with $R_i(k_0, k_i), R_j(k_0, k_1), R_{i,j}(\{k_i, k_1\}, k_1)$ and no other transitions. $pre(k_0) = pre(k_1) = \top$ and $pre(k_i) = \varphi$

$$[\![\sim(\alpha_s)^i_{s\in S}]\!]^1 \text{ and } [\![Ack(\alpha^x)]\!]^1 \text{ are computed as before}$$

Define finally $\models^1$ as the new truth-maker operator defined as $\models$ was, but this time based on the interpretation $[\![\cdot]\!]^1$ of linguistic actions.

Under $\models^1$ action $[\varphi!^i]$ has i commits to φ, but changes neither i's second order commitments (in general $\langle \mathcal{M}, w\rangle \not\models^1 C_iC_i\varphi$) nor anyone else's commitments. This now fixes our problem of the liar who denies commitments he has previously made; someone can now commit to φ but then later say *I never said* φ and remain consistent.

This weaker semantics, however, makes grounding impossible in finite conversations. The situation is analogous in other models where a discourse move m by i entails only (a)C_ip and (b)that all the conversational participants believe C_ip, see for instance (Traum, 1994; Ginzburg, 2012). Then j's acknowledgment of m would entail $C_jC_ip \wedge Bel_GC_jC_ip$. We can show using a game theoretic framework, that common commitments are achievable only after an infinite sequence of acknowledgment moves between i and j.

Can we do without common commitments in conversation? We think not; common commitments are essential (see also Clark (1996)) for strategic reasons and can be present even when mutual beliefs about a shared task are not. Suppose that i's goal is that $C_j\varphi$ and that j cannot consistently deny the commitment. If i only extracts from j a move m that $C_j\varphi$, j has a winning strategy for denying i victory. She simply denies committing to φ (*I never said that*), since $C_j\neg C_j\varphi$ is consistent with $C_j\varphi$, even if $Bel_jC_j\varphi$. Player j lies, but she is consistent. If i manages to achieve $C_jC_j\varphi$, j can still similarly counter i maintaining consistency. *Only if* i achieves the *common commitment* $C^*_GC_j\varphi$, with G the group of conversational participants) does j not have a way of denying her commitment without becoming inconsistent, as $C^*C_j\varphi \rightarrow (C_jC_j\varphi \wedge C_jC_jC_j\varphi \wedge \cdots)$.

Our proposal is that a particular sort of acknowledgment and confirming question licenses the move to common commitment. It is the one in (4)c, where 0 asks a confirming question after an acknowledgment of a move m. If 1's answer to the confirming question is consonant with m, then $C^*_{\{0,1\}}C_1\varphi$, and 0

has achieved her goal. We can explain this using our notion of ambiguous commitments. An acknowledgment is in fact ambiguous. One reading comes from our simple semantics where an acknowledgment adds one layer of commitment—i.e. if j acknowledges i's commitment to φ with a simple *OK*, we have $C_jC_i\varphi$. The other reading is that it indeed implies a common commitment of the form $C^*_{i,j}C_j\varphi$, following our second semantics for assertions. The clarification question, when answered in the affirmative, selects the common commitment formulation. (Clark and Brennan, 1991) acknowledges that grounding may seem to require conversationalist to give infinitely many positive bits of evidence—(*Requiring positive evidence of understanding seems to lead to an infinite regress*), and claims that some form of evidence such as *continued attention* solves the situation as it can occur continuously and does not require a separate presentation. Our proposal is compatible but distinct from Clark's (ours is also formally worked out), and interestingly survives in non-cooperative settings.

We quickly now turn to corrections. Speakers can not only deny prior commitments but also "undo" or "erase" them with *self-corrections*. For instance, if in (4)b 1 commits to not having a bank account; in (4)d 1 no longer has this commitment (See Ginzburg (2012) for a detailed account of repair). Conversational goals of the form $C^*_GC_ip$ are unstable if i may correct herself; they may be satisfied on one finite sequence but not by all its continuations. j's being able to correct a previous turn's commitments increases the complexity of i's goalsSerre (2004), which affects the existence of a winning strategy for i; an unbounded number of correction moves will make any stable $C^*_GC_ip$ goal unattainable, if p is not a tautology. We observe, however, a sequence of self-corrections is only a good strategy for achieving j's conversational goals if she is prepared to provide an explanation for her shift in commitments (and such explanations must come to an end). As (Venant et al., 2014) argues, conversationalists are constrained to be credible in a certain sense if they are to achieve their conversational goals. Constantly shifting one's commitments with self-corrections leads to non-credibility, thus avoiding the problem of unbounded erasures.

To provide a semantics for corrections, we begin from Lascarides and Asher (2009), who provide a *syntactic* notion of revision over the logical form of the discourse structure. Using the correction of m as an action update on the commitment slate prior to m yields a semantics for corrections. Our formal semantics captures the dynamic effects of announcements, corrections and acknowledgments; common commitments are important conversational goals and that particular conditions must obtain if they are to be achieved.

6 Conclusions

We have presented two semantics for dialogue in terms of commitments that is general enough to handle non-cooperative and cooperative dialogues. The first one is conceptually simple and has a straightforward axiomatization but fails to give a sensible semantics for acknowledgments and is also too restrictive concerning denials of commitments, which our semantics makes inconsistent instead of simply a lie. Finally, we discussed corrections as another problem for the semantics of dialogue and offered a solution.

References

Asher, N. (2013). Implicatures and discourse structure. *Lingua 132*(0), 13 – 28. SI: Implicature and Discourse Structure.

Asher, N. and A. Lascarides (2003). *Logics of Conversation*. Cambridge University Press.

Baltag, A., L. S. Moss, and S. Solecki (1998). The logic of public announcements, common knowledge, and private suspicions. In *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '98, San Francisco, CA, USA, pp. 43–56. Morgan Kaufmann Publishers Inc.

Clark, H. (1996). *Using Language*. Cambridge, England: Cambridge University Press.

Clark, H. H. and S. E. Brennan (1991). Grounding in communication. In L. Resnick, J. Levine, and S. Teasley (Eds.), *Perspectives on Socially Shared Cognition*, pp. 127–149. American Psychological Association.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Hamblin, C. (1987). *Imperatives*. Blackwells.

Lascarides, A. and N. Asher (2009). Agreement, disputes and commitment in dialogue. *Journal of Semantics 26*(2), 109–158.

Reyle, U. (1993). Dealing with ambiguities by underspecification: Construction, interpretation and deduction. *Journal of Semantics 10*, 123–179.

Serre, O. (2004). Games with winning conditions of high borel complexity. In *ICALP*, pp. 1150–1162.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Ph. D. thesis, Computer Science Department, University of Rochester.

Traum, D. and J. Allen (1994). Discourse obligations in dialogue processing. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL94)*, Las Cruces, New Mexico, pp. 1–8.

Venant, A., N. Asher, and C. Degremont (2014). Credibility and its attacks. In V. Rieser and P. Muller (Eds.), *The 18th Workshop on the Semantics and Pragmatics of Dialogue*, pp. 154–162.