C&L Intention Revisited*

Andreas Herzig and **Dominique Longin**
IRIT - CNRS
118 route de Narbonne
F-31062 Toulouse cedex 04 (France)
{herzig,longin}@irit.fr

Abstract

The 1990 papers of Cohen and Levesque (C&L) on rational interaction have been most influential. Their approach is based on a logical framework integrating the concepts of belief, action, time, and choice. On top of these they define notions of achievement goal, persistent goal, and intention.

We here revisit their approach in a simplified, propositional logic, for which we give complete axiomatization.

Within that logic we study the definition of achievement goals, refining C&L's analysis. Our analysis allows us to identify the conditions under which achievement goals persist. We then discuss the C&L definition of intention as well as a variant that has been proposed by Sadek and Bretier. We argue that both are too strong and propose a weakened version.

Introduction

The fundamental role of intention in communication and more generally in interaction has been stressed by Bratman (1987; 1990). Bratman's analysis has inspired most of the authors in the literature, starting with Cohen & Levesque (1990a; 1990b) (C&L henceforth). Their approach has been taken up by Perrault (1990), Rao and Georgeff (1991; 1992), Sadek (1992), Konolige and Pollack (1993), and is the standard reference on BDI logics (Wooldridge 2000).

C&L and Sadek reduce intention to primitive concepts of *belief*, *choice*, *action*, and *time*. In contrast, intention is primitive in the other approaches. This is probably due to C&L's rather complex framework, which requires a modal predicate logic with equality and quantification over sequences of events, and includes a temporal logic with a binary 'before' operator. Moreover there is only part of the semantics: syntactical assumptions are postulated that have no semantical counterpart. Finally, the frame problem remains unsolved, and attempts to fill that gap (Perrault 1990) (Appelt & Konolige 1989) have turned out to be unsatisfactory (Herzig & Longin 2000).

In this paper we simplify and perfect C&L's approach. We first define and study a minimal propositional logic of action, time, belief, and choice (that we call ABC logic) able to support C&L's approach. We here take advantage of recent progress in reasoning about actions and beliefs and in product logics, and give a complete axiomatization. We then study the definition of achievement goals, refining the C&L analysis. Our analysis allows us to identify the conditions under which achievement goals persist. We then discuss the C&L definition of intention as well as a variant that has been proposed by Sadek. We argue that both are too strong and propose a weakened version.

The components of ABC logic are introduced in the next three sections. We then give a complete axiomatization. Within ABC logic we define achievement goals, and show under which conditions their persistence can be deduced. Finally we discuss how intentions can be defined from achievement goals.

Action and time

We here introduce a simple logic of action and time. Generally speaking, events and actions can be interpreted as transition relations on states, be it states of the world, mental states, dialogue states, or a blend of them. This is the kind of model that Dynamic Logic offers. We add to this logic a unary modal operator "henceforth".

Semantics of events and actions

We suppose there is a set of *events* $EVT = \{\alpha, \beta, \ldots\}$ and a set of *agents* $AGT = \{i, j, \ldots\}$. Actions are events that are brought about by agents. We sometimes write $i{:}\alpha$ to identify the agent of α. EVT contains *purely epistemic events* which do not change the physical world, but only the agents' mental states. Epistemic events include observations and communication actions.

The formula $[\alpha]\phi$ expresses that if α happens then ϕ holds after α. The dual $\langle\alpha\rangle\phi = \neg[\alpha]\neg\phi$ expresses that α happens and ϕ is true afterwards. Hence $[\alpha]\bot$ expresses that α does not happen, and $\langle\alpha\rangle\top$ expresses that α happens.

*Our work has benefitted from numerous discussions with colleagues, in particular with Robert Demolombe, Jérôme Lang, Philippe Balbiani, Jacques Virbel, Olivier Gasquet, Yves Lespérance, Daniel Vanderveken, Mehdi Dastani, Jan Broersen, Leon van der Torre, Joris Hulstijn. Thanks are due to Maarten Marx and Tinko Tinchev for information on the complexity of product logics, and to Hector Levesque for clarifications on the C&L approach. Part of the material in this paper has been presented at the Seventh Workshop on the Semantics and Pragmatics of Dialogue (DiaBruck 2003).

Semantics of time

To speak about sequences of more than one event we use a temporal operator $\Box$. $\Box\phi$ expresses that henceforth ϕ holds. A dual operator $\Diamond$ is defined by $\Diamond\phi = \neg\Box\neg\phi$ ('eventually ϕ').

Models have a set of possible worlds W, and a mapping

$$V : W \to (ATM \to \{0,1\})$$

associating a valuation V_w to every $w \in W$. There are mappings

$$\mathcal{R}_\Box : W \to 2^W$$

and

$$\mathcal{R} : EVT \to (W \to 2^W)$$

associating sets of possible worlds $\mathcal{R}_\Box(w)$ and $\mathcal{R}_\alpha(w)$ every possible world w. We indentify such mappings with accessibility relations: $w\mathcal{R}_\Box w'$ iff $w' \in \mathcal{R}_\Box(w)$, etc. As usual,

$$w \models [\alpha]\phi \text{ if } w' \models \phi \text{ for every } w' \in \mathcal{R}_\alpha(w)$$

and

$$w \models \Box\phi \text{ if } w' \models \phi \text{ for every } w' \in \mathcal{R}_\Box(w)$$

With C&L we suppose:

- if $w\mathcal{R}_\alpha w'$ and $w\mathcal{R}_\beta w''$ then $w' = w''$;
- $\mathcal{R}_\Box$ is reflexive[1], transitive[2], and confluent[3];
- if $w\mathcal{R}_\alpha w'$ then $w\mathcal{R}_\Box w'$;
- if $w\mathcal{R}_\alpha w'$, $w\mathcal{R}_\Box w''$ and $w \neq w''$ then $w'\mathcal{R}_\Box w''$.

It follows from the last two conditions that events are organized in histories: if $w\mathcal{R}_\alpha w'$ and $w\mathcal{R}_\beta w''$ then $w' = w''$. From that it follows that events are deterministic. (To see this put $\beta = \alpha$.)

Our semantics is slightly weaker than C&L's. First, $\mathcal{R}_\Box$ is not necessarily linear. Second, w might be possible in the future without there being a particular sequence of actions leading to w: ϕ will be eventually true without necessarily having a sequence of actions which will achieve ϕ. This will be relevant when it comes to intentions, because an agent might believe w can be achieved without having a plan to reach w.

Mental attitudes

We now add the basic mental attitudes of belief and choice to the picture.

Semantics of belief

Under the *doxastic logics* denomination, modal logics of belief are popular in philosophy and AI, and the system KD45 is widely accepted.[4] In the models, for each agent i and possible world w there is an associated set of possible worlds $\mathcal{B}_i(w) \subseteq W$: the worlds that are compatible with i's beliefs.

[1] For every $w \in W$, $w\mathcal{R}_\Box w$.

[2] If $w_1\mathcal{R}_\Box w_2\mathcal{R}_\Box w_3$ then $w_1\mathcal{R}_\Box w_3$.

[3] If $w\mathcal{R}_\Box w_1$ and $w\mathcal{R}_\Box w_2$ then there is a w_3 such that $w_1\mathcal{R}_\Box w_3$ and $w_2\mathcal{R}_\Box w_3$.

[4] The most important criticism that has been made to KD45 is that it accepts omniscience, i.e. an agent's beliefs are closed under tautologies, conjunction, and logical consequences. In particular the latter point, viz. that an agent believes all the consequences of his beliefs, has been considered to be unrealistic. We here accept omniscience to simplify the framework.

Hence every $\mathcal{B}_i$ is a mapping

$$\mathcal{B}_i : W \to 2^W$$

For every $i \in AGT$ there is a modal operator Bel_i, and $Bel_i\phi$ expresses that agent i believes that ϕ. The truth condition for the modal operator Bel_i stipulates that $w \models Bel_i\phi$ if ϕ holds in all worlds that are compatible with i's beliefs, i.e.

$$w \models Bel_i\phi \text{ if } v \models Bel_i\phi \text{ for every } v \in \mathcal{B}_i(w)$$

$\mathcal{B}_i$ can be seen as an accessibility relation, and it is standard to suppose that

- every relation $\mathcal{B}_i$ is serial[5], transitive, and euclidian[6].

$BelIf_i\phi$ abbreviates $Bel_i\phi \vee Bel_i\neg\phi$.

Semantics of choice

Among all the worlds in $\mathcal{B}_i(w)$ that are possible for agent i, there are some that i prefers. C&L say that i *chooses* some subset of $\mathcal{B}_i(w)$. Semantically, these worlds are identified by yet another accessibility relation

$$\mathcal{C}_i : W \to 2^W$$

$Choice_i\phi$ expresses that agent i chooses that ϕ. We sometimes also say that i *prefers* that ϕ.[7] Without surprises, $w \models Choice_i\phi$ if ϕ holds in all preferred worlds, i.e.

$$w \models Choice_i\phi \text{ if } w' \models \phi \text{ for every } w' \in \mathcal{C}_i(w)$$

We suppose that

- $\mathcal{C}_i$ is serial, transitive, and euclidian.

This differs from C&L, who only have supposed seriality, and follows Sadek's approach. The latter has argued that choice is a mental attitude which obeys to principles of introspection that correspond with transitivity and euclideanity.

Choice and belief

What is the relation between choice and belief? As said above, an agent only chooses worlds he considers possible:

- $\mathcal{C}_i(w) \subseteq \mathcal{B}_i(w)$.

Hence belief implies choice, and choice is a mental attitude that is weaker than belief. This corresponds to validity of the (Inc_{Choice_i}) principle $Bel_i\phi \to Choice_i\phi$. We moreover require that worlds chosen by i are also chosen from i's possible worlds, and vice versa:

- if $w\mathcal{B}_i w'$ then $\mathcal{C}_i(w) = \mathcal{C}_i(w')$.

(See Figure 1.)

Such a semantics validates the equivalences

$$Choice_i\phi \leftrightarrow Bel_i Choice_i\phi \quad (1)$$

$$\neg Choice_i\phi \leftrightarrow Bel_i\neg Choice_i\phi \quad (2)$$

$$Choice_i\phi \leftrightarrow Choice_i Choice_i\phi \quad (3)$$

$$\neg Choice_i\phi \leftrightarrow Choice_i\neg Choice_i\phi \quad (4)$$

The implication $Choice_i Bel_i\phi \to Choice_i\phi$ is also valid, but not the converse.

[5] For every $w \in W$, $\mathcal{B}_i \neq \emptyset$

[6] for all $w \in W$, if $v, v' \in \mathcal{B}_i(w)$ then $v' \in \mathcal{B}_i(v)$ and $v \in \mathcal{B}_i(v')$.

[7] While C&L use a modal operator 'goal' (probably in order to have a uniform denomination w.r.t. the different versions of goals they study), it seems more appropriate to us to use the term 'choice'.

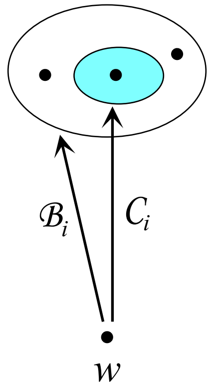

Figure 1: Belief and choice

The kinematics of mental attitudes

Several proposals were made in the beginning of the 90s concerning the relation between action and belief. They built on what was state of the art in the reasoning-about-actions field in the 80s, and used complex default or autoepistemic logics (Perrault 1990; Appelt & Konolige 1989). In the beginning of the 90s, Scherl and Levesque (1993) have proposed simple principles that can be integrated easily into the original C&L framework, which is what we undertake here.

We first make some hypotheses on the perception of events. Then we state general principles governing relationships between belief, choice, action and time.

Hypotheses on perception

We suppose that an event occurs iff every agent i perceives it. More precisely, we suppose that i's perception is correct (in the sense that if i believes that α has occurred then α indeed occurred) and complete (in the sense that if α occurs then α is perceived by i). Hence event occurrences are public.

HYPOTHESIS. *All event occurrences are perceived correctly and completely by every agent.*

We note that this hypothesis just aims at simplifying our exposition, and that misperception can be integrated following ideas of Bacchus et al. (1995; 1999) and Baltag et col. (1998; 2000).

While an agent perceives the occurrence of an event, or more precisely of an event token, we suppose that he does not learn anything beyond that about the event's particular effects. We therefore define *uninformative events* as event tokens whose outcome is not perceived by the agents. When an agent learns that such an event has occurred, he is nevertheless able to predict its results according to the action laws he believes to hold. Consider e.g. the action of tossing a coin. Suppose the agent learns that toss has occurred. As he cannot observe the effects, he predicts them in an *a priori* way, according to his mental state and the action laws. The agent might thus be said to 'mentally execute' toss. After toss he believes that $Heads \vee Tails$ holds, but neither believes $Heads$ nor $Tails$. It is only the observation that the coin fell heads which may make the agent start to believe that $Heads$.

We suppose the observation of ϕ never occurs when ϕ is false. To learn that the observation of ϕ has occurred means to learn that ϕ (supposing observations are reliable). Thus, observation actions are uninformative: all the relevant information is encoded in the notification of the event occurrence. Then to take into account the observation of ϕ amounts to incorporate ϕ into $\mathcal{B}_i(w)$.

In the same way, we can suppose that i's action of informing that ϕ is uninformative (both for the speaker i and the hearer). There are perception actions which do not satisfy our hypothesis, such as *testing-if-*ϕ. Such tests can nevertheless be reduced to uninformative actions: *testing-if-*ϕ is the nondeterministic composition of *observing-that-*ϕ and *observing-that-*$\neg\phi$.

HYPOTHESIS. *All events are uninformative.*

Our second hypothesis is deeper than the first: without presenting a formal proof here, we suppose that every event can be constructed from uninformative events by means of dynamic logic nondeterministic composition "$\cup$" and sequencing ";". For example the everyday action of tossing corresponds to the complex toss; $(\text{observe}Heads \cup \text{observe}Tails)$. In fact such a hypothesis is often made in reasoning about actions, e.g. in (Scherl & Levesque 1993) or (Shapiro *et al.* 2000, footnote 10).

Mental attitudes and action

Suppose the actual world is w, and some event α occurs leading to a new actual world w'. Which worlds are possible for agent i at w'? According to Moore (1985) and Scherl and Levesque (1993; 2003), i makes 'mentally happen' α in all his worlds $v \in \mathcal{B}_i(w)$, and then collects the resulting worlds $\mathcal{R}_\alpha(v)$ to form the new belief state. We thus have $\mathcal{B}_i(w') = (\mathcal{R}_\alpha \circ \mathcal{B}_i)(w) = \bigcup_{v \in \mathcal{B}_i(w)} \mathcal{R}_\alpha(v)$. This identity must be restricted in order to keep i's beliefs consistent, i.e. to avoid $\mathcal{B}_i(w') = \emptyset$. We thus obtain:

- If $w\mathcal{R}_\alpha w'$ and $(\mathcal{R}_\alpha \circ \mathcal{B}_i)(w) \neq \emptyset$ then $\mathcal{B}_i(w') = (\mathcal{R}_\alpha \circ \mathcal{B}_i)(w)$.

This relies on our hypothesis that events are uninformative: apart from the mere occurrence of α agent i should learn nothing about α's particular effects that obtain in w', and $\mathcal{B}_i(w')$ only depends on $\mathcal{B}_i(w)$ and α.

Note that such an explanation is in accordance with our hypotheses. Syntactically, this makes the principle of no forgetting (NF_{Bel_i}) $Bel_i[\alpha]\phi \wedge \neg Bel_i[\alpha]\bot \rightarrow [\alpha]Bel_i\phi$ valid, as well as the dual principle of no learning (NL_{Bel_i}) $[\alpha]Bel_i\phi \wedge \neg[\alpha]\bot \rightarrow Bel_i[\alpha]\phi$.

How do an agent's choices evolve? We recall that for each possible world there is an associated temporal structure (its history). Therefore agent i's choices concern not only possible states of the world, but also possible histories. We therefore suppose that i's preferences after α are just the images

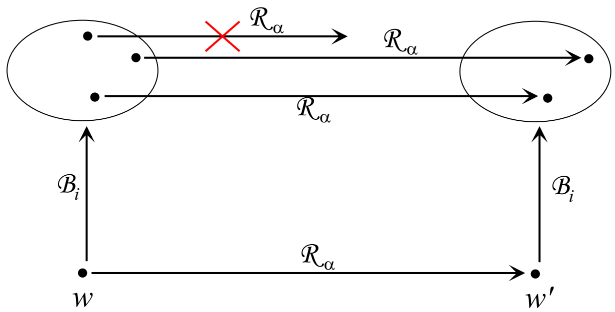

Figure 2: Action and belief

by α of its preferred worlds before α. Just as for belief, this identity must be restricted in order to keep i's choices consistent. We thus obtain the constraint:

- If $w\mathcal{R}_\alpha w'$ and $(\mathcal{R}_\alpha \circ \mathcal{C}_i)(w) \neq \emptyset$ then $\mathcal{C}_i(w') = (\mathcal{R}_\alpha \circ \mathcal{C}_i)(w)$.

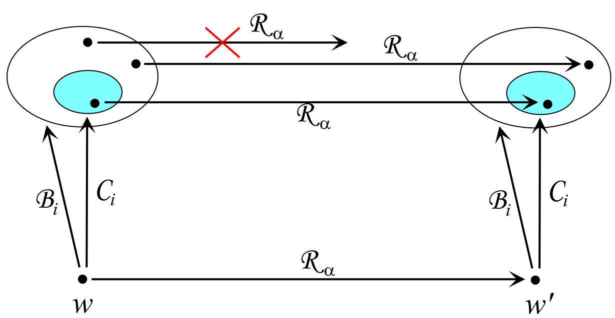

Figure 3: Action, belief, and choice

Again, note that such an explanation is in accordance with our hypotheses. Syntactically, this makes valid the principle (NF$_{Choice_i}$) $Choice_i[\alpha]\phi \wedge \neg Choice_i[\alpha]\bot \rightarrow [\alpha]Choice_i\phi$, and (NL$_{Choice_i}$) $[\alpha]Choice_i\phi \wedge \neg[\alpha]\bot \rightarrow Choice_i[\alpha]\phi$.

Mental attitudes and time

Which constraints can be formulated on Bel_i and $\Box$?

First, note that from (NF$_{Bel_i}$) it follows that $Bel_i\Box\phi \wedge \neg Bel_i[\alpha]\bot \rightarrow [\alpha]Bel_i\Box\phi$, i.e. beliefs about invariants persist as long as there are no surprises.

What about a 'no forgetting' principle for the temporal operator $Bel_i\Box\phi \rightarrow \Box Bel_i\phi$? In fact this would be too strong: suppose that for some reason, i wrongly believes that some object is broken and cannot be repaired. We thus have $Bel_i\Box\neg Broken$, which together with such a principle would imply $\Box Bel_i\neg Broken$. Which is absurd: imagine e.g. i learns that the object is in fact not broken. Then such a no forgetting principle would forbid any belief revision.

Only weaker identities can be motivated here: for each of i's possible worlds v, if u' is possible for i in some world u in the future of v then there is a world v' possible for i such that u' is in its future. And vice versa:

- if $w\mathcal{B}_i v$ then $(\mathcal{R}_\Box \circ \mathcal{B}_i)(v) = (\mathcal{B}_i \circ \mathcal{R}_\Box)(v)$

This constraint can also be interpreted as a form of introspection through time. Indeed, the introspection principles for belief correspond to $\mathcal{B}_i \circ \mathcal{B}_i = \mathcal{B}_i$, and it can be shown that due to transitivity and euclideanity of $\mathcal{B}_i$ our condition is equivalent to $\mathcal{B}_i \circ \mathcal{R}_\Box \circ \mathcal{B}_i = \mathcal{B}_i \circ \mathcal{R}_\Box$. Note that corresponding principles of negative introspection cannot be motivated.

Similar to belief we impose for choice:

- if $w\mathcal{C}_i v$ then $(\mathcal{R}_\Box \circ \mathcal{C}_i)(v) = (\mathcal{C}_i \circ \mathcal{R}_\Box)(v)$

This makes the principle (Inv$_{Choice_i}$) $Choice_i(\Box Choice_i\phi \leftrightarrow Choice_i\Box\phi)$ valid. It follows that $Choice_i\Box Choice_i\phi \leftrightarrow Choice_i\Box\phi$, which says that if an agent prefers ϕ to be invariant then he chooses that he will always prefer ϕ, and vice versa.

Comments: revision of beliefs and choices

Our conditions say nothing about i's beliefs after a surprising action occurrence, i.e. when $(\mathcal{R}_\alpha \circ \mathcal{B}_i)(w) = \emptyset$. In this case i must revise his beliefs. Integrations of belief revision into a logic of action and belief have been proposed in (Shapiro *et al.* 2000). In (Herzig & Longin 2002) we have proposed an alternative based on updating by the preconditions of α. It amounts to suppose that our language contains not only modal action operators $[\alpha]$, but also *update operators* $[upd(\phi)]$, for every formula ϕ. In the original paper such operations were seen as particular actions. Here we have to separate them because our semantics is in terms of histories, and at most one action happens at a given w, while we would like to allow several updates leaving w.

Our conditions do not constrain either i's choices when $(\mathcal{R}_\alpha \circ \mathcal{C}_i)(w) = \emptyset$, i.e. after an unwanted action occurrence. Then i has to revise his choices.

There are two cases. First, if $Choice_i[\alpha]\bot$ and $Bel_i[\alpha]\bot$ then a surprising event has occurred, and the agent has to revise both his beliefs and his choices. We think that in this case our account of belief revision in (Herzig & Longin 2002) can be extended to choice revision. In the second case we have $Choice_i[\alpha]\bot$ and $\neg Bel_i[\alpha]\bot$. Then i did not believe the event was impossible, but preferred so. Devices such as a preference relation have to be integrated here, and we leave a more detailed investigation to future work.

Completeness theorem

We have defined the semantics of a basic logic of action, belief, and choice. To sum it up, our models have the form $\langle W, \mathcal{B}, \mathcal{C}, \mathcal{R}, \mathcal{R}_\Box, V\rangle$, where W is a set of possible worlds, $\mathcal{B}$ and $\mathcal{C}$ associate accessibility relations to every agent, $\mathcal{R}$ associates an accessibility relation to every action, $\mathcal{R}_\Box$ is the accessibility relation for $\Box$, and V associates a valuation to every possible world. We call ABC *models* the set of models satisfying all the constraints imposed in the three preceding sections, and write $\models_{ABC} \phi$ if ϕ is valid in ABC models. We write $\mathcal{S} \models_{ABC} \phi$ if ϕ is a logical consequence of the set of formulas $\mathcal{S}$ in ABC models.

We give now an axiomatization of ABC. We suppose the axioms and inference rules of the basic normal modal logic

K for every modal operator,[8] plus the following:

$$\neg(Bel_i\phi \wedge Bel_i\neg\phi) \quad (\mathsf{D}_{Bel_i})$$
$$Bel_i\phi \rightarrow Bel_i Bel_i\phi \quad (\mathsf{4}_{Bel_i})$$
$$\neg Bel_i\phi \rightarrow Bel_i\neg Bel_i\phi \quad (\mathsf{5}_{Bel_i})$$
$$\neg(Choice_i\phi \wedge Choice_i\neg\phi) \quad (\mathsf{D}_{Choice_i})$$
$$Choice_i\phi \rightarrow Bel_i Choice_i\phi \quad (\mathsf{PI}_{Choice_i})$$
$$\neg Choice_i\phi \rightarrow Bel_i\neg Choice_i\phi \quad (\mathsf{NI}_{Choice_i})$$
$$Bel_i\phi \rightarrow Choice_i\phi \quad (\mathsf{Inc}_{Choice_i})$$
$$\Box\phi \rightarrow \phi \quad (\mathsf{T}_\Box)$$
$$\Box\phi \rightarrow \Box\Box\phi \quad (\mathsf{4}_\Box)$$
$$\Diamond\Box\phi \rightarrow \Box\Diamond\phi \quad (\mathsf{Confl}_\Box)$$
$$\Box\phi \rightarrow [\alpha]\phi \quad (\mathsf{Inc}_{[\alpha]})$$
$$\langle\alpha\rangle\phi \rightarrow [\beta]\phi \quad (\mathsf{Hist}_1)$$
$$\Diamond\phi \rightarrow (\phi \vee [\alpha]\Diamond\phi) \quad (\mathsf{Hist}_2)$$
$$Bel_i[\alpha]\phi \wedge \neg Bel_i[\alpha]\bot \rightarrow [\alpha]Bel_i\phi \quad (\mathsf{NF}_{Bel_i})$$
$$[\alpha]Bel_i\phi \wedge \neg[\alpha]\bot \rightarrow Bel_i[\alpha]\phi \quad (\mathsf{NL}_{Bel_i})$$
$$Choice_i[\alpha]\phi \wedge \neg Choice_i[\alpha]\bot \rightarrow [\alpha]Choice_i\phi \quad (\mathsf{NF}_{Choice_i})$$
$$[\alpha]Choice_i\phi \wedge \neg[\alpha]\bot \rightarrow Choice_i[\alpha]\phi \quad (\mathsf{NL}_{Choice_i})$$
$$Bel_i(\Box Bel_i\phi \leftrightarrow Bel_i\Box\phi) \quad (\mathsf{Inv}_{Bel_i})$$
$$Choice_i(\Box Choice_i\phi \leftrightarrow Choice_i\Box\phi) \quad (\mathsf{Inv}_{Choice_i})$$

Some comments are in order.

(PI_{Choice_i}) is an axiom of positive introspection for choice similar to ($\mathsf{4}_{Bel_i}$) and (NI_{Choice_i}) is the negative version.

Axiom (Hist_1) implies determinism of every α: $\langle\alpha\rangle\phi \rightarrow [\alpha]\phi$. ($\mathsf{Hist}_2$) is similar to the first of the Segerberg axioms (Harel 1984).

Axioms (NF_{Bel_i}) and (NL_{Bel_i}) can be put together into the single $(\neg[\alpha]\bot \wedge \neg Bel_i[\alpha]\bot) \rightarrow ([\alpha]Bel_i\phi \leftrightarrow Bel_i[\alpha]\phi)$. Equivalences of this kind have been called successor state axioms for belief in (Scherl & Levesque 1993).

(NF_{Choice_i}) and (NL_{Choice_i}) are their analogues for choice. Such axioms for choice have not been studied before.

(Inv_{Bel_i}) is a subjective version of a successor state axiom for belief and time. (Inv_{Choice_i}) is a similar axiom for choice and time. As far as we know they have not been studied before either.

From (NF_{Bel_i}) it follows that

$$Bel_i\Box\phi \wedge \neg Bel_i[\alpha]\bot \rightarrow [\alpha]Bel_i\Box\phi,$$

i.e. beliefs about invariants persist as long as there are no surprises.

From (Inv_{Bel_i}) it can be deduced in KD45 that

$$Bel_i\Box\phi \leftrightarrow Bel_i\Box Bel_i\phi$$

i.e. if i believes ϕ to be an invariant then he believes that he will always be aware of ϕ.

Moreover,

$$Bel_i\Box(Bel_i\phi \rightarrow \phi)$$
$$Bel_i\Diamond Bel_i\phi \rightarrow Bel_i\Diamond\phi$$
$$Choice_i\Diamond Bel_i\phi \rightarrow Choice_i\Diamond\phi$$

are valid.

The other way round, $Bel_i\Diamond\phi \rightarrow Bel_i\Diamond Bel_i\phi$ and $Choice_i\Diamond\phi \rightarrow Choice_i\Diamond Bel_i\phi$ should not hold. Here is an example illustrating that, inspired by Heisenberg's uncertainty principle. Let p mean that some electron is in a particular place. Suppose you believe that it will eventually be in that place: $Bel_i\Diamond p$. According to Heisenberg it is impossible to know that at the same point in time: $\Box\neg Bel_i p$. Now if we suppose that i is aware of that principle, we obtain $Bel_i\neg\Diamond Bel_i p$.

A similar argument can be made against $Choice_i\Diamond\phi \rightarrow Choice_i\Diamond Bel_i\phi$. This is opposed to Sadek and colleagues' approach (Sadek 1992; Bretier 1995; Louis 2003), where the principle $Choice_i\Diamond\phi \rightarrow Choice_i\Diamond Bel_i\phi$ is accepted.

We call ABC *logic* the logic thus axiomatized, and write $\vdash_{ABC} \phi$ if ϕ is a theorem of ABC.

THEOREM. $\models_{ABC} \phi$ iff $\vdash_{ABC} \phi$.

It is a routine task to check that all the axioms correspond to their semantic counterparts. It is routine, too, to check that all of our axioms are in the Sahlqvist class, for which a general completeness result exists (Sahlqvist 1975; Blackburn, de Rijke, & Venema 2001).

We conjecture that Marx's proof (1999) of decidability and EXPSPACE complexity of the problem of satisfiability in the product logic S5×K extends straightforwardly to ABC logic in the case of a single agent.[9]

In the rest of the paper, we apply ABC logic to investigate the notions of achievement goal, persistent goal, and intention.

Achievement goals

C&L view goals and intentions as particular future-oriented choices which take the form $Choice_i\Diamond\phi$.

If ϕ is already believed to be true then there is no point in maintaining the goal or the intention that ϕ. C&L therefore concentrate on goals which require some change in order to make them true. Basically such goals are of the form $Choice_i\Diamond\phi\wedge\neg\psi$, where ψ is a condition triggering the abandonment of the goal.

Which forms do ϕ and ψ take? First of all ϕ and ψ should be equivalent: when ϕ obtains then the goal can be abandoned, and whenever the goal is abandoned then ϕ holds.

[8] for example for $[\alpha]$:

$$\text{from } \phi \leftrightarrow \psi \text{ infer } [\alpha]\phi \leftrightarrow [\alpha]\psi \quad (\mathsf{RE}_{[\alpha]})$$
$$[\alpha](\phi \wedge \psi) \rightarrow [\alpha]\phi \wedge [\alpha]\psi \quad (\mathsf{M}_{[\alpha]})$$
$$[\alpha]\phi \wedge [\alpha]\psi \rightarrow [\alpha](\phi \wedge \psi) \quad (\mathsf{C}_{[\alpha]})$$
$$[\alpha]\top \quad (\mathsf{N}_{[\alpha]})$$

[9] We are indebted to Maarten Marx for pointing this out.

(This is at least expected by i.) Second, ψ should not be factual, but rather about i's mental state: else the agent has no means to decide when to abandon his goal. Hence achievement goals take the following form.

DEFINITION. Agent i has the achievement goal that ϕ if (1) in his preferred worlds ϕ is believed later and (2) i does not believe ϕ:

$$AGoal_i\phi \stackrel{\text{def}}{=} Choice_i\Diamond Bel_i\phi \wedge \neg Bel_i\phi \quad (\mathsf{Def}_{AGoal_i})$$

The only basic modal principle our definition of achievement goals validates is

$$\frac{\phi \leftrightarrow \psi}{AGoal_i\phi \leftrightarrow AGoal_i\psi}.$$

For the rest, just as in the C&L account none of the standard principles is valid.

The so-called *side effect problem* is to avoid to systematically adopt the consequences of our goals. Formally $AGoal_i\phi \wedge Bel_i(\phi \rightarrow \psi) \rightarrow AGoal_i\psi$ should not be valid. Just as for C&L, this formula is not valid in ABC logic. Even if we strengthen the condition $Bel_i(\phi \rightarrow \psi)$ in various ways, $AGoal_i\phi$ does not imply $AGoal_i\psi$. The reason is that the side effect might be believed, which makes that ψ cannot be an achievement goal. And just as C&L, if we add the condition $\neg Bel_i\psi$ then we validate

$$AGoal_i\phi \wedge Bel_i\Box(\phi \rightarrow \psi) \wedge \neg Bel_i\psi \rightarrow AGoal_i\psi.$$

(The proof makes use of the Axiom (Inv_{Bel_i}).) We also validate and the inference rule

$$\frac{\phi \rightarrow \psi}{AGoal_i\phi \wedge Bel_i \wedge \neg Bel_i\psi \rightarrow AGoal_i\psi}.$$

Finally, the valid equivalences

$$AGoal_i\phi \leftrightarrow Bel_i AGoal_i\phi$$

and

$$\neg AGoal_i\phi \leftrightarrow Bel_i\neg AGoal_i\phi$$

express that an agent is aware of his achievement goals. The equivalence

$$AGoal_i\phi \leftrightarrow AGoal_i Bel_i\phi$$

is valid as well (while only the left-to-right direction is valid for C&L).

Comparison with C&L

C&L's original definition of achievement goals is

$$AGoal_i^{CL}\phi \stackrel{\text{def}}{=} Choice_i\Diamond\phi \wedge Bel_i\neg\phi.$$

THEOREM. $AGoal_i\phi \leftrightarrow AGoal_i^{CL} Bel_i\phi$.

This can be proved using introspection properties of belief.

C&L satisfy Axiom D: $\neg(AGoal_i\phi \wedge AGoal_i\neg\phi)$, while we do not.[10] Thus, while an agent's choices are consistent, his achievement goals are not necessarily so. This can be justified by the same temporal considerations that lead to rejection of axiom C: i might want ϕ to be true at some point in the future, and ϕ to be false at some other point in the future. But note that $AGoal_i\Box\phi \wedge AGoal_i\Box\neg\phi$ is unsatisfiable due to the confluence of time.

In their definition, C&L stipulate that i should believe ϕ is false. We have preferred the weaker $\neg Bel_i\phi$ because it is more natural: in general goals are abandoned only when they are believed to be true, and therefore absence of belief is sufficient to maintain the goal (but see our Byzantine example below for a counterexample).

C&L only require $Choice_i\Diamond\phi$. We have seen in the previous section that $Choice_i\Diamond Bel_i\phi \rightarrow Choice_i\Diamond\phi$ is a theorem. We have also said there that the other sense of the implication should not hold. So let us consider a situation where $Choice_i\Diamond\phi \wedge \neg Choice_i\Diamond Bel_i\phi$ holds. The following example seems to motivate the need for achievement goals in C&L's sense.

Let r mean that a message of i has been received by j, and let i believe initially that j has not received the message yet. Suppose we are in a Byzantine-generals-style scenario where i is not guaranteed that his message will eventually be received by j, and where i believes that in any case he will never know whether j received the message or not. (In the original scenario it is just possible for i that he will never know.) Hence we have $Bel_i\neg r \wedge Choice_i\Diamond r \wedge Bel_i\Box\neg BelIf_i r$. From the latter it follows that $\neg Choice_i\Diamond Bel_i r$. In summary, we have $Bel_i\neg r \wedge AGoal_i^{CL} r \wedge \neg AGoal_i r$.

Now in such a context it seems reasonable that i acts by nevertheless posting the message. C&L can account for this case by stating $AGoal_i^{CL} r$. What would be i's achievement goal in our account? We argue that in the example i has the achievement goal that $\neg Bel_i\neg r$: such an achievement goal can first motivate i to post the message, and then trigger abandonment (say after the time period i esteems necessary for the message travelling under favorable conditions). Note that $AGoal_i\neg Bel_i\neg r$ is consistent with the scenario description.

Consider another example where there is only one action of toggling a switch, and suppose that in the initial world $w_0 \models \neg Bel_i Light \wedge \neg Bel_i\neg Light$, i.e. i ignores whether the light is on or off: for i there is at least one possible world where $Light$ holds, and there is at least one possible world where $\neg Light$ holds. As toggling is the only available action we have $w_0 \models Bel_i\Box(\neg Bel_i Light \wedge \neg Bel_i\neg Light)$, i.e. i believes he will always ignore whether the light is on or off. According to C&L agent i can nevertheless have the

[10] As C&L's admit, this is 'for the wrong reasons': their stronger definition of achievement goals is responsible for $AGoal_i\phi \rightarrow Bel_i\neg\phi$, which warrants axiom D for $AGoal_i$. Note that they do not validate the stronger but equally intuitive principle $\frac{\neg(\phi\wedge\psi)}{\neg(AGoal_i\phi \wedge AGoal_i\psi)}$. Apparently this has not been noted in the literature.

achievement goal $AGoal_i^{CL} Light$ in w_0, while he cannot have such a goal with our definition. Thus i is aware that he will never be able to abandon his goal that $Light$ in the expected way, viz. by coming to believe that $Light$.

Persistent goals

C&L have defined persistent goals to be achievement goals that are kept until they are achieved, or are abandoned for some other reasons. We can show that persistence can be *deduced* from our no forgetting principle for choice as long as the event is not unwanted:

THEOREM. $\models_{ABC} (AGoal_i\phi \wedge \neg Choice_i[\alpha]\bot) \rightarrow [\alpha](AGoal_i\phi \vee Bel_i\phi)$

PROOF. We prove $\neg Bel_i\phi \wedge Choice_i\Diamond Bel_i\phi \rightarrow Choice_i[\alpha]\bot \vee [\alpha]Choice_i\Diamond Bel_i\phi$. This can be deduced from (NL_{Choice_i}), (Hist_2), (Inc_{Choice_i}) as follows.

First, axiom (Hist_2) tells us that

$$\Diamond Bel_i\phi \rightarrow (Bel_i\phi \vee [\alpha]\Diamond Bel_i\phi)$$

for any action α. Therefore

$$Choice_i\Diamond Bel_i\phi \rightarrow Choice_i(Bel_i\phi \vee [\alpha]\Diamond Bel_i\phi).$$

As by (5_{Bel_i}) and (Inc_{Choice_i}) we have

$$\neg Bel_i\phi \rightarrow Choice_i\neg Bel_i\phi,$$

the left hand side implies

$$Choice_i[\alpha]\Diamond Bel_i\phi.$$

From that we get with (NL_{Choice_i}) that

$$Choice_i[\alpha]\bot \vee [\alpha]Choice_i\Diamond Bel_i\phi.$$

■

We inherit the properties of achievement goals concerning logical principles, the side effect problem, and persistence.

Comparison with C&L

C&L's original definition is that a persistent goal that ϕ is an achievement goal that ϕ that can only be abandoned if

1. ϕ is achieved, or
2. the agent learns that ϕ can never be achieved, or
3. for some other reason.

This leads to their principle

$$PGoal_i\phi \rightarrow [\alpha](PGoal_i\phi \vee Bel_i\phi \vee Bel_i\Box\neg\phi \vee \psi),$$

where ψ is an unspecified condition accounting for case (3). Our theorem makes (3) more precise by identifying it with the occurrence of an unwanted event, which is the only case when achievement goals have to be revised.[11] Indeed, the theorem tells us that C&L's case (2) is excluded when $\neg Choice_i[\alpha]\bot$ holds: in this case we are guaranteed that i will not learn through α that ϕ will be false henceforth. Given our hypothesis that events are uninformative, this is as it should be.

[11] In the case where i is the agent of α (noted $i{:}\alpha$) one might reasonably suppose that $Choice_i[i{:}\alpha]\bot \rightarrow [i{:}\alpha]\bot$, i.e. there are no such unwanted action occurrences. We then get unconditioned persistence of achievement goals: $AGoal_i\phi \rightarrow [i{:}\alpha](AGoal_i\phi \vee Bel_i\phi)$. This is related to intentional actions as discussed in C&L's (1990a, section 4.2.1), where moreover $Bel_i[i{:}\alpha]\bot \vee Bel_i\neg[i{:}\alpha]\bot$ is assumed. We just note that such principles are of the Sahlqvist type, and can be added to ABC logic without harm.

Intentions

C&L have distinguished intentions-to-do and intentions-to-be. We here only consider the latter, which, following Bratman, C&L have defined as particular persistent goals: the agent must be committed to achieve the goal, in the sense that he must believe that he will perform an action which will lead to the goal.

DEFINITION. Agent i has the intention that ϕ if (1) i has the achievement goal that ϕ, and (2) i does not believe $Bel_i\phi$ will obtain anyway:

$$Int_i\phi \stackrel{\text{def}}{=} AGoal_i\phi \wedge \neg Bel_i\Diamond Bel_i\phi \qquad (\mathsf{Def}_{Int_i})$$

Hence intentions are achievement goals which do not automatically obtain in the future. As $\neg Bel_i\Diamond Bel_i\phi$ implies $\neg Bel_i\phi$, it follows that $Int_i\phi \leftrightarrow Choice_i\Diamond Bel_i\phi \wedge \neg Bel_i\Diamond Bel_i\phi$. If not explicitly, this implicitly links i's intending that ϕ to i's choosing actions that get him closer to ϕ: $Int_i\phi$ triggers i's planning for ϕ. Therefore it seems justified to say that our definition captures the spirit of Bratman's intentions.

What is the status of achievement goals when $Bel_i\Diamond Bel_i\phi$ holds? In this case, $AGoal_i\phi \wedge Bel_i\Diamond Bel_i\phi$ is equivalent to $Bel_i\Diamond Bel_i\phi \wedge \neg Bel_i\phi$: i believes ϕ will be achieved in the future, no matter what continuation of his possible histories occurs. Then according to our definition i has to abandon $Int_i\phi$ at w_1. This is reminiscent of McDermott's Little Nell example: suppose that i intends that ϕ at w_0, and that i successfully plans and acts in a way such that later on at w_1 he is sure ϕ will be achieved in the future, i.e. $Bel_i\Diamond Bel_i\phi$ holds at w_1. According to McDermott i then abandons his intention that ϕ too early, and will never achieve ϕ. We believe the problem can be solved by separating planning-oriented (future-oriented) intention from intention-in-action: at w_1 agent i switches from the planning-oriented intention $Int_i\phi$ to the intention-in-action to execute the plan (alias complex action) which he believes ensures that ϕ will obtain. i will stick to this plan from w_1 on and as long as no unforeseen events occur.[12]

Again, we inherit the properties of achievement goals concerning logical principles, the side effect problem, and in particular persistence:

THEOREM. $\models_{ABC} (Int_i\phi \wedge \neg Choice_i[\alpha]\bot) \rightarrow [\alpha](Int_i\phi \vee Bel_i\Diamond Bel_i\phi)$

PROOF. The theorem of the previous section establishing that achievement goals are also persistence goals, a look at the proof tells us that

$$(AGoal_i\phi \wedge \neg Choice_i[\alpha]\bot) \rightarrow [\alpha]Choice_i\Diamond Bel_i\phi$$

Therefore by classical principles

$$(AGoal_i\phi \wedge \neg Choice_i[\alpha]\bot) \rightarrow [\alpha]((Choice_i\Diamond Bel_i\phi \wedge \neg Bel_i\Diamond Bel_i\phi) \vee Bel_i\Diamond Bel_i\phi)$$

[12] We could pursue this and define future-directed intention-to-do α as $Choice_i\Diamond\langle i{:}\alpha\rangle\top$.

from which the present theorem follows by the definition of intention. ■

Hence intentions persist as long as there are no unwanted action occurrences.

Comparison with C&L

Our definition of $Int_i\phi$ differs from C&L's in a fundamental way because it does not mention actions: C&L basically stipulate that in every preferred history there must be some action α whose author is i and which brings about ϕ.

Using quantification over actions this could be approximated by:

$$Int_i^{CL}\phi \stackrel{\text{def}}{=} \neg Bel_i\phi \wedge Choice_i\Diamond\exists i{:}\alpha\langle i{:}\alpha\rangle Bel_i\phi.$$

But as pointed out by Sadek (2000) and Bretier (1995), such a definition is too strong in particular in cooperative contexts, where it often suffices for i to *trigger* actions of some other agent j which will achieve the goal. They have advocated a correction, which we roughly approximate here by:

$$Int_i^{S}\phi \stackrel{\text{def}}{=} \neg Bel_i\phi \wedge Choice_i\Diamond Bel_i\phi \wedge$$
$$Choice_i\forall i{:}\alpha(Bel_i\langle i{:}\alpha\rangle\Diamond Bel_i\phi \rightarrow Choice_i\Diamond\langle i{:}\alpha\rangle\top).$$

Again, this is too strong: my intention to go to Vancouver in june here would force me to choose the action of hiring an aircraft. In another sense, both C&L's and Sadek's definitions are too weak because they lack a causal connection between the action and the goal: basically they entitle me to entertain the intention that it be sunny in Vancouver in june if each of my preferred histories has some action of mine leading to a state where this holds.

As our definition of intention does not mention events at all, this example also illustrates that our definition is also too weak in this respect.

Conclusion

We have integrated action, time, belief, and choice in a simple propositional modal logic that is sound, complete and decidable, and which we think provides the basic framework for the logical analysis of interaction. We have shown how different notions of goal and intention can be expressed in it, and have identified the conditions under which such motivational attitudes persist.

Although Cohen and Levesque's papers are standard references, to the best of our knowledge such a simplification has never been undertaken. Our completeness, decidability and complexity results pave the way for methods of mechanical deduction.

In ABC logic we have also in part solved the frame problem for belief and intention. While the frame problem for belief has been investigated extensively in the literature, there is not too much work in the literature on the frame problem for intentions, and the only references we are aware of are (Shapiro & Lespérance 2000; Shapiro, Lespérance, & Levesque 1997; 1998). These accounts are preliminary, in particular they lead to fanatic agents.

What is lacking for a comprehensive solution to the frame problem for intention is the integration of belief and choice revision (sometimes called intention reconsideration in agent theories (Thomason 2000; Schut & Wooldridge 2001)). We leave this important issue to future work.

What remains also to be addressed is the question of how intentions lead to actions. This is is the topic of plan generation, which still has to be integrated in our logic.

References

Appelt, D., and Konolige, K. 1989. A nonmonotonic logic for reasoning about speech acts and belief revision. In Reinfrank, M.; de Kleer, J.; Ginsberg, M.; and Sandewall, E., eds., *Proc. 2nd Int. Workshop on Non-monotonic Reasoning*, number 346 in LNAI, 164–175. Springer Verlag.

Bacchus, F.; Halpern, J.; and Levesque, H. 1995. Reasoning about noisy sensors in the situation calculus. In *Proc. 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, 1933–1940.

Bacchus, F.; Halpern, J.; and Levesque, H. 1999. Reasoning about noisy sensors in the situation calculus. *Artificial Intelligence* 111:131–169.

Baltag, A.; Moss, L. S.; and Solecki, S. 1998. The logic of public announcements, common knowledge, and private suspicions. In *Proc. TARK'98*, 43–56. Morgan Kaufmann.

Baltag, A. 2000. A logic of epistemic actions. Technical report, CWI. `http://www.cwi.nl/~abaltag/papers.html`.

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

Bratman, M. E. 1987. *Intentions, plans, and practical reason*. Harvard University Press, MA.

Bratman, M. E. 1990. What is intention? In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. Cambridge, MA: MIT Press. chapter 2, 15–31.

Bretier, P. 1995. *La communication orale coopérative : contribution à la modélisation logique et à la mise en œuvre d'un agent rationnel dialoguant*. Ph.D. Dissertation, Université Paris Nord, Paris, France.

Cohen, P. R., and Levesque, H. J. 1990a. Intention is choice with commitment. *Artificial Intelligence J.* 42(2–3):213–261.

Cohen, P. R., and Levesque, H. J. 1990b. Persistence, intentions, and commitment. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. Cambridge, MA: MIT Press. chapter 3, 33–69.

Harel, D. 1984. Dynamic logic. In Gabbay, D. M., and Günthner, F., eds., *Handbook of Philosophical Logic*, volume II. D. Reidel, Dordrecht. 497–604.

Herzig, A., and Longin, D. 2000. Belief dynamics in cooperative dialogues. *J. of Semantics* 17(2). vol. published in 2001.

Herzig, A., and Longin, D. 2002. Sensing and revision in a modal logic of belief and action. In van Harmelen, F., ed., *Proc. ECAI2002*, 307–311. IOS Press.

Konolige, K., and Pollack, M. E. 1993. A representationalist theory of intention. In *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*, 390–395. Chambery, France: Morgan Kaufmann.

Louis, V. 2003. *Conception et mise en œuvre de modles formels du calcul de plans d'action complexes par un agent rationnel dialoguant*. Ph.D. Dissertation, Université de Caen, France.

Marx, M. 1999. Complexity of products of modal logics. *J. of Logic and Computation* 9(2):221–238.

Moore, R. C. 1985. A formal theory of knowledge and action. In Hobbs, J., and Moore, R., eds., *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex. 319–358.

Perrault, C. R. 1990. An application of default logic to speech act theory. In Cohen, P. R.; Morgan, J.; and Pollack, M. E., eds., *Intentions in Communication*. Cambridge, MA: MIT Press. chapter 9, 161–185.

Rao, A. S., and Georgeff, M. P. 1991. Modeling rational agents within a BDI-architecture. In Allen, J. A.; Fikes, R.; and Sandewall, E., eds., *Proc. 3rd Int. Conf. on Knowledge Representation and Reasoning (KR'91)*, 473–484. San Mateo, CA: Morgan Kaufmann.

Rao, A. S., and Georgeff, M. P. 1992. An abstract architecture for rational agents. In Nebel, B.; Rich, C.; and Swartout, W., eds., *Proc. 4th Int. Conf. on Knowledge Representation and Reasoning (KR'92)*, 439–449. Cambridge, Massachusetts: Morgan Kaufmann.

Sadek, M. D. 1992. A study in the logic of intention. In Nebel, B.; Rich, C.; and Swartout, W., eds., *Proc. 4th Int. Conf. on Knowledge Representation and Reasoning (KR'92)*, 462–473. Cambridge, Massachusetts: Morgan Kaufmann.

Sadek, M. D. 2000. Dialogue acts are rational plans. In Taylor, M.; Nel, F.; and Bouwhuis, D., eds., *The structure of mutimodal dialogue*, 167–188. Philadelphia/Amsterdam: John Benjamins publishing company. From ESCA/ETRW, Workshop on The Structure of Multimodal Dialogue (Venaco II), 1991.

Sahlqvist, H. 1975. Completeness and correspondence in the first and second order semantics for modal logics. In Kanger, S., ed., *Proc. 3rd Scandinavian Logic Symposium 1973*, number 82 in Studies in Logic. North Holland.

Scherl, R., and Levesque, H. J. 1993. The frame problem and knowledge producing actions. In *Proc. Nat. Conf. on AI (AAAI'93)*, 689–695. AAAI Press.

Scherl, R., and Levesque, H. J. 2003. The frame problem and knowledge producing actions. *Artificial Intelligence* 144(1-2).

Schut, M., and Wooldridge, M. 2001. Principles of intention reconsideration. In *Proc. AGENTS'01*. ACM Press.

Shapiro, S., and Lespérance, Y. 2000. Modeling multiagent systems with the cognitive agents specification language - a feature interaction resolution application. In Castelfranchi, C., and Lespérance, Y., eds., *Intelligent Agents Vol. VII - Proc. 2000 Workshop on Agent Theories, Architectures, and Languages (ATAL-2000)*. Springer-Verlag.

Shapiro, S.; Pagnucco, M.; Lespérance, Y.; and Levesque, H. J. 2000. Iterated belief change in the situation calculus. In *Proc. KR2000*, 527–538.

Shapiro, S.; Lespérance, Y.; and Levesque, H. J. 1997. Specifying communicative multi-agent systems with ConGolog. In *Working notes of the AAAI fall symposium on Communicative Action in Humans and Machines*, 75–82. AAAI Press.

Shapiro, S.; Lespérance, Y.; and Levesque, H. J. 1998. Specifying communicative multi-agent systems. In Wobcke, W.; Pagnucco, M.; and Zhang, C., eds., *Agents and Multi-Agent Systems - Formalisms, Methodologies, and Applications*, 1–14. Springer-Verlag, LNAI 1441.

Thomason, R. H. 2000. Desires and defaults: A framework for planning with inferred goals. In *Proc. of the Seventh International Conference Knowledge Representation and Reasoning (KR'2000)*, 702–713. Morgan Kaufmann Publishers.

Wooldridge, M. 2000. *Reasonning about Rational Agent*. Cambridge, Massachusetts: MIT Press.