Intentions in Interaction

Ben Rodenhäuser

University of Groningen, The Netherlands
b.rodenhauser@ai.rug.nl

Abstract. We observe that intentions play an important role for analyzing the dynamics of information and interaction in multi-agent settings in at least two sense: information about an agent's intentions allows an observer to draw useful conclusions from the former's actions; and intentions can be viewn as "programs", guiding the actions of the participants towards an outcome state. We present a Dynamic Epistemic Logic approach which formalizes the notions of "intending to take an action" and "taking an action based on a prior intention" to account for these phenomena. The proposal is relevant for understanding the reasoning processes of agents in strategic interaction.

Keywords: Logic of Intention, Logic for Multi-Agent Systems, Dynamic Epistemic Logic, Strategic Interaction, Theory of Action

1 Introduction

We introduce the modeling problems investigated in this paper by means of two simple scenarios. Subsequently, we discuss the issues raised by the examples, and the conceptual and formal approach taken to address the issues.

Scenario 1. *Jane sees Ian, the five-year old kid of her friend Karl, standing in the kitchen. There is a cupboard with two drawers. The top drawer contains several chocolate bars, the bottom drawer several copies of a small recipe book (incidentally, the book has the same dimensions as the chocolate bars). Jane doesn't know what is in what drawer, but she knows that Ian knows. She also knows that Ian intends to get himself some chocolate. Now Jane sees Ian opening the top drawer, retrieving an object (she does not see if it is a chocolate bar or the recipe book, because the two are easy to confuse), and leaving the kitchen. Jane enters the kitchen, takes, without hesitation, the recipe book from the bottom drawer, and starts to prepare dinner.*

The main intuition about this example we want to formalize is the idea that Jane's knowledge about Ian's intention is what "makes the difference" here, allowing her to infer from her observation where the chocolate bar and the recipe book are, respectively, to be found. That is, if Jane is correct in assuming that Ian is actually *acting on his intention* (an assumption we will make sure is justified in this paper), she is also correct in deducing what he is doing (retrieving a chocolate bar), and so she is correct in concluding that the recipe book is in the bottom drawer.

Scenario 2. *The referee explains the rules of a simple game to Ian and Jane: "I will assign consecutive numbers* n *and* $n+1$ *for certain* $n \in \mathbb{N}$ *to you. You will know your number, but not the other's. You then take turns in announcing whether you know the distribution of numbers or not. If at some point you have announced that you know the distribution, you should stop participating in the game." The referee then proceeds to distribute the numbers among Ian and Jane in closed envelopes.*

This game has an epistemic structure similar to that of the well-known Muddy Children puzzle, with announcements of ignorance gradually decreasing that very ignorance itself (van Ditmarsch et al. 2006). If Ian has 2, for instance, and announces he does not know the distribution, and if Jane does the same, then after these announcements Ian knows the distribution (since Jane cannot have 1, she must have 3, his reasoning goes). After he announces that, Jane knows as well (if Ian had 4, he could not already know, Jane reasons). For the purposes of the present paper, the interest of the scenario derives from analyzing the intentions of the both players. Consider the situation *after* the referee has explained the rules, but *before* Ian and Jane have been assigned their numbers. Suppose that Jane and Bob listen to the referee and decide both that they intend to play this game. It is not hard to come up with an informal version of the plan each of them should pursue: "Announce your ignorance as long as you do not know the distribution of the cards and as soon as you know the distribution, say that and drop out." The question now is how this intention should be formally represented. This includes encoding the program structure the informal description clearly has. As long as neither agent knows his or her own number, both have to be prepared to announce their ignorance an unbounded number of times, so we cannot simply represent their plans by a sequence of intended actions—we need to be able to encode iteration guarded by a kind of "knowledge test". We also have to ensure that agents terminate their intention at the appropriate moment. That is, after an agent has announced that he or she knows the distribution of the cards, he or she should *drop* the intention to make any further moves in the game.

Intentions in Multi-Agent Interaction. The two scenarios exhibit what seem to be significant and ubiquitous roles intentions play in multi-agent interaction. The *first scenario* illustrates that information about *another* agent's intention is a valuable resource an agent can draw on when observing the other's behavior, useful both for drawing conclusions from the observation, and for planning her own actions on that basis. The *second scenario* shows that intentions may be seen as instructions each agent has chosen to follow, that is, as a kind of "program", exhibiting a complex structure in the form of loops, conditional instructions, and conditions for termination (sticking to the metaphor, the actual *behavior* of interacting agents consists in generating an "execution trace" of these respective programs). It is these two characteristics that we want to get a hold on in this paper.

Intentions to Act. To capture them, we shall formalize a notion of "intention to act", that is, an *intention to act in a certain way.* Such an intention has as its content not a proposition (i.e. a set of states), but a dynamic object: we will represent intentions as sets of events consistent with an agent's motivation.[1] "Intentions to act" should be distinguished from "intentions to achieve" (sometimes called *outcome intentions*): The latter give an answer to the question "what state of affairs does an agent intend to arrive at?"; the former answer the question "what course of action does an agent intend to pursue?" To analyze our examples, intentions to act are what is needed.

Intentions as Commitments. Our modeling approach rests on a view of intentions as *commitments of an agent* (Harman 1986, Bratman 1987). Such commitments should not be understood as a moral obligation, nor, in general, as a *social* commitment towards other agents (in that respect, intentions qualitatively differ from promises). Rather, an agent is committed if she has *settled* on a particular course of action. This may be contrasted with the situation of an agent *having a desire*: desiring to have wine and ordering beer is consistent in a way in which intending to order wine and ordering beer is not. This is important from the point of view of our first scenario: Jane's conclusion that the chocolate is in the top drawer is misguided if she just knows that Ian *desires* chocolate—maybe Ian has decided to *resist* this very desire! Knowing his intention, however, Jane can be sure that Ian is actually *settled* on getting himself some chocolate. And that is what enables her to derive the information she needs once she observes Ian's action. As one philosopher puts it: "Intending to act is as close to acting as reasoning alone can get us" (Broome 2002); and this fact may be exploited by the observer of an action.

Actions Based on Intentions. As the preceding discussion makes clear, the second formal notion that is obviously needed here is that of an action based on a (prior) mental state. This idea is familiar from the philosophy of action: "According to a popular view, actions are, essentially, events with a suitable causal history, a causal history featuring pertinent mental events or states" (Mele 1997a). Philosophers that would deny a causal role of mental events in the genesis of actions usually agree that antecedent mental states are an important category for explaining the latter (Mele 1997b). There are many proposals for explicating just what mental states are the central ones here, and in precisely what way they contribute to explaining an action. In line with the focus given by our scenarios, we will focus on a notion of an *action based on a prior intention.*

Dynamic Epistemic Logic with Intentions. Our formal approach is situated in the tradition of Dynamic Epistemic Logic (Plaza 1989, Gerbrandy 1999, Baltag et al. 1999, Baltag and Moss 2004, van Benthem et al. 2006). Dynamic Epistemic Logic (DEL) is a modal logic framework concerned with modeling, understanding and explaining information flow in intelligent interaction (van Benthem 2009a,b). The "trademark" of DEL is its capability of encoding the dynamics of information flow in multi-agent settings in terms of model transfor-

[1] Our notion of an *event* is that used in the philosophy of action, where an event is understood as a "happening" (like "the flipping of the switch", or "the collapse of the building"). In game theory, events are sometimes defined to be sets of states. To these we refer, as is customary in modal logic, as *proposition.*

mations, or *updates*, making it an ideal candidate for a framework within which to tackle the modeling problems presented by our scenarios.

As it stands, however, the Baltag-Moss-Solecki (BMS) framework, the standard DEL setting, does not support a proper notion of agency. Indeed, the formal models DEL usually employs do not involve an agent taking action at all: no sense of agency is present in the formal DEL representations of social situations, actions are reduced to events that are "impersonal" (Baltag and Moss 2004), "detached from the agents assumed to perform them" (Kooi 2003). The challenge is thus to adapt DEL in such a way that it becomes possible to express the notions we are interested in. The following is a proposal how this might be achieved.

Intentions in Strategic Interaction. Both *intending to act* and *acting on an intention* are closely related to central questions a player in a game asks herself: "what will my opponent do?" and "with what intention is my opponent taking a particular move?" Our approach is thus linked to the study of strategic interaction, and ties in with recent research that seeks to elucidate the reasoning processes of players in games in a precise manner (Aumann and Dreze 2005, Brandenburger 2007, Ramanujam and Simon 2008).

Plan of the Paper. Sections 2 and 3 are introductory: we outline some formal background (section 2), and sketch our modeling approach (section 3). Sections 4 and 5 present the semantics and syntax of our logical framework, and section 6 returns to the scenarios introduced above. Section 7 derives reduction laws for our dynamic modalities. Section 8 states our conclusions.

2 Background

We briefly review the basics of epistemic logic and the BMS framework, the standard version of Dynamic Epistemic Logic, as a reference point for what follows. We are given a countable set Φ (the *atomic sentences*), and a finite set AG (the *agents*). All definitions to follow depend on these two sets.

Epistemic Models. An *epistemic model* is a structure $\mathfrak{s} = (S, \sim, V, s_0)$, where S is a non-empty set of objects (the *states*), $\sim = (\sim_i)_{i \in \text{AG}}$ is a collection of equivalence relations (one for each agent i) on S (the *accessibility relation* for agent i), $V : S \to \mathcal{P}(\Phi)$ is a function assigning a set of atomic sentences to each state (the *valuation*), and $s_0 \in S$ (the *designated state*).

For an epistemic model $\mathfrak{s}$, and a state $s \in S$, we introduce the notation $\mathfrak{s}^s := (S, \sim, V, s)$, i.e. $\mathfrak{s}^s$ is obtained from $\mathfrak{s}$ by replacing the designated state with s.

Working with equivalence relations is a standard move (van Ditmarsch et al. 2006). The notion of knowledge determined by this choice validates the familiar S5 axioms for veridicality, and positive and negative introspection. We adopt this notion here mainly because all the examples we consider can be dealt with within the scope of S5.

The Epistemic Language. Models are used to interpret the *epistemic language* $\mathcal{L}_\Phi$, where $\varphi \in \mathcal{L}_\Phi$ iff φ is built according to the following rule (in which p ranges over Φ, and i over $\textsc{ag}$):

$$\varphi := \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i \varphi$$

The truth definition for this language is the standard one, with the key clause

$$\mathfrak{s} \vDash \Box_i \varphi \text{ iff } \quad \forall s \in S: \text{If } s_0 \sim_i s, \text{ then } \mathfrak{s}^s \vDash \varphi$$

In addition, one may introduce common knowledge operators $\Box_\iota^*$ for any non-empty $\iota \subseteq \textsc{ag}$. We refer to the resulting language as $\mathcal{L}_\Phi^*$. In the following truth definition, $\sim_\iota^*$ is the reflexive transitive closure of $\bigcup_{i \in \iota} \sim_i$:

$$\mathfrak{s} \vDash \Box_\iota^* \varphi \text{ iff } \quad \forall s \in S: \text{If } s_0 \sim_\iota^* s, \text{ then } \mathfrak{s}^s \vDash \varphi$$

BMS Epistemic Events. The BMS framework extends epistemic logic with *BMS epistemic events* α that may affect the information states of the agents in various ways (epistemic events are sometimes also called action models, event models, epistemic actions, or update frames). In this paper, we also allow events to change the truth value of (finitely many) atomic sentences, using the approach of Baltag (2002). This means that both pure acts of communication and more "tangible" interventions into the state of the world may be represented as BMS epistemic events.

A *BMS epistemic event* (over $\mathcal{L}_\Phi$) is a structure $\alpha = (A, \sim, \textsc{con}, a_0)$, where A is a finite, non-empty set of objects (the *event tokens*), $\sim = (\sim_i)_{i \in \textsc{ag}}$ is a collection of binary equivalence relations (one for each agent i) on A (the *accessibility relation* for agent i), $\textsc{con}$ is a pair of total functions $\textsc{pre}: A \to \mathcal{L}_\Phi$ and $\textsc{flip}: A \to \mathcal{L}_\Phi$, and $a_0 \in A$ (the *designated event*). We require $\textsc{flip}_a$ to be a conjunction of $n \geq 0$ atomic sentences.

We write $p \in \textsc{flip}_a$ if p occurs as a conjunct in $\textsc{flip}_a$. We call $\textsc{pre}_a$ the *precondition* of a, and $\textsc{flip}_a$ the *fact change* of a. $\textsc{pre}$ and $\textsc{flip}$ induce a *content function* $\textsc{con}: A \to \mathcal{L}_\Phi \times \mathcal{L}_\Phi$ (note the reuse of notation), defined by $\textsc{con}_a := (\textsc{pre}_a, \textsc{flip}_a)$ for $a \in A$. We set $\textsc{con}_\alpha := \textsc{con}_{a_0}$, $\textsc{pre}_\alpha := \textsc{pre}_{a_0}$, and $\textsc{flip}_\alpha := \textsc{flip}_{a_0}$. As we have done for epistemic models, we introduce the notation $\alpha^a := (A, \sim, \textsc{con}, a)$ for any $a \in A$, i.e. α^a is obtained from α by replacing the designated state with a.

$\textsc{pre}_\alpha$ determines the "domain of applicability" of the epistemic event α, which is given by those epistemic models $\mathfrak{s}$ such that $\mathfrak{s} \vDash \textsc{pre}_\alpha$. Observe that preconditions are epistemic formulas, i.e. they are not allowed to contain occurrences of the event modalities introduced below. This is less general than the standard BMS approach; we stick to our restricted version here for reasons of simplicity. $\textsc{flip}_\alpha$ gives the atomic sentences whose truth value is *flipped* whenever the epistemic event α happens (contrast this with approaches where fact change is analyzed in terms of *substitutions*, as, e.g., in van Benthem et al. (2006)).

A special case of BMS epistemic events are *public BMS epistemic events*, defined as those BMS epistemic events α whose domain A is a singleton set. Public BMS epistemic events may be determined (up to renaming) by specifying their content. That is, they may be represented as pairs (φ, ψ), where $\varphi = \textsc{pre}_\alpha$ and

$\psi = \text{FLIP}_\alpha$.

Product Update of an Epistemic Model. To incorporate BMS epistemic events in epistemic logic, one observes that any BMS epistemic event induces a partial function on the proper class of all epistemic models, given by the operation of *product update*. As a preliminary notion, we write the *symmetric difference* of two sets of proposition letters P and Q as $P \Delta Q := (P \backslash Q) \cup (Q \backslash P)$. The symmetric difference of P and Q contains all and only the proposition letters that are contained in one of the sets, but not in the other. The notion is used to formalize the idea of an epistemic event flipping the truth values of certain atomic sentences.

Let $\mathfrak{s}$ be an epistemic model, and α a BMS epistemic event. The *product update of* $\mathfrak{s}$ *with* α (notation: $\mathfrak{s} \otimes \alpha$) is defined iff $\mathfrak{s} \vDash \text{PRE}_\alpha$. In that case, $\mathfrak{s} \otimes \alpha$ is the structure $(S^{\mathfrak{s}\otimes\alpha}, \sim^{\mathfrak{s}\otimes\alpha}, V^{\mathfrak{s}\otimes\alpha}, s_0^{\mathfrak{s}\otimes\alpha})$, where

$$\begin{aligned}
S^{\mathfrak{s}\otimes\alpha} &:= \{s.a \mid s \in S, a \in A, \mathfrak{s}^s \vDash \text{PRE}_{\alpha^a}\} \\
\sim_i^{\mathfrak{s}\otimes\alpha} &:= \{(s.a, t.b) \mid s \sim_i t, a \sim_i b\} \\
V^{\mathfrak{s}\otimes\alpha}(s.a) &:= V(s) \Delta \text{FLIP}_a \\
s_0^{\mathfrak{s}\otimes\alpha} &:= s_0.a_0
\end{aligned}$$

Truth Definition for Event Modalities. One now extends the language $\mathcal{L}_\Phi$ with formulas $[\alpha]\varphi$ containing event modalities $[\alpha]$ for any BMS epistemic event α, interpreted as follows:

$$\mathfrak{s} \vDash [\alpha]\varphi \ \text{ iff } \ \mathfrak{s} \otimes \alpha \vDash \varphi, \text{ if defined}$$

A main aspect of our work in the following will be to adapt the notion of an epistemic event and product update to accommodate intentions in our setting.

3 Modeling Approach

In this section, we outline our modeling approach more concretely, starting with some examples to sharpen intuitions.

1. Jane intends to go to the movies.
2. Jane intends to go to the movies or to the park.
3. Jane intends to go to the movies if it is raining, and to the park otherwise.
4. Karl intends to go shopping and work at the library afterwards.
5. Karl intends to tell Ian that he is getting married.
6. Jane intends to shout until the rescue team finds her.

A common feature of the examples is that the content of the intentions described can naturally be understood as a *set of (epistemic) events*. This is the basic idea underlying our formal model of intentions in this paper. Some such set may be a singleton, as in (1), meaning that the agent has settled on a unique action she intends to take; it may also contain more than one epistemic event, as in (2) and (3). In the latter case, the intention may still be deterministic, as in (3), but it may also present a state in which the agent has not settled on a unique course of actions, as in (2). In other words, (3) is *conditional* but *deterministic*, (1) is deterministic "by default" (because there is only one option still considered by the agent), while (2) is *non-deterministic*. Furthermore, intentions may be used to form *plans*, the simplest case being concatenation, or composition of two intentions to form a sequence, as in (4). Example (5) and (6) exhibit another important feature: an informal description of an intention often allows us to extract a criterion for determining under which conditions an action taken on the basis of the intention would count as "successful". In example (5), the criterion is Ian's knowing about the marriage; in example (6), it is "The rescue team notices Jane". Note that *several* actions may be required until the proposition is true—for instance, if the rescue team does not find Jane's position after one shout. We may thus treat the proposition formulating the "criterion of success" as a condition for terminating the intention: given the criterion is fulfilled, the intention may be safely dropped by the agent.

As a consequence of these considerations, we will model intentions as sets of epistemic events; plans will be taken to be *sequences* of intentions. For the sake of simplicity, we will not consider more complex, tree-like plans with cross-temporal dependencies. Event tokens within epistemic events will be equipped with a *termination condition* (in addition to the precondition and the fact change). The content function applied to an event token a then returns a triple $(\mathrm{PRE}_a, \mathrm{FLIP}_a, \mathrm{TRM}_a)$. Assuming that the intentions in our examples have *public* actions as their content, we can represent them as in the following list. $\cup$ stands for union, $\cdot$ for sequential composition, set brackets are omitted, and the expressions in typewriter font are proposition letters:

1. $(\top, \texttt{movies}, \top)$
2. $(\top, \texttt{movies}, \top) \cup (\top, \texttt{park}, \top)$
3. $(\texttt{rain}, \texttt{movies}, \top) \cup (\neg\texttt{rain}, \texttt{park}, \top)$
4. $(\top, \texttt{shopping}, \top) \cdot (\top, \texttt{library}, \top)$
5. $(\texttt{marriage}, \top, \Box_i\texttt{marriage})$
6. $(\top, \texttt{shout}, \texttt{find})$

Note that examples (1)-(4) are equipped with "trivial" termination conditions in the formal renderings.

Next, the question is how intentions should be integrated into an overall logical setting. Our proposal extends epistemic logic with intention modalities $\boxdot_i^n\vartheta$, which may be read as "agent i intends to take an action satisfying ϑ as the nth step in her plan". The intended semantics for this construct is that $\boxdot_i^n\vartheta$ is true at the nth stage of i's plan if all epistemic events constituting this step make ϑ true.

Furthermore, we will introduce action modalities $[\alpha]_i\varphi$, read as "if agent i takes action α, then afterwards φ". These new modalities *replace* the BMS event modalities of section 2 above. As explained in the introduction, "taking an action" will be analyzed in term of an event and an antecedent mental state, the overall idea being the following. In the BMS framework, the domain of applicability of an epistemic event is determined by its precondition. In contrast, the availability of an action for an agent in our setting is determined by the agent's mental state, which in turn, consists of her *information*, and her *intentions*. To be available for an agent, an action has to be *licensed* by both. An action α is licensed by an agent's information if the event's precondition is *known to be true* by the agent (given that we are working in S5, this *entails* that the precondition of α *is* actually true). And α is licensed by the agent's intention if α is bisimilar to an element of the set modeling the first intention in the agent's plan. To simplify matters, we will not consider the possibility of agents making mistakes, or otherwise deviating from their prior intentions in this paper (in fact, this was already implicitly presupposed in our informal discussion in the introduction). Bisimiliarity is a notion of *observational equivalence*—using it encodes the idea that agents have *full control* over the execution of their intentions. Put differently, any action an agent takes precisely *matches her intention* in taking it.

4 Structural Notions

We now present formal analogues of the notions discussed in the previous sections. This leads to the definition of the product update of an intention model with an action.

Epistemic Events. An *epistemic event* (over $\mathcal{L}_\Phi$) is a finite structure $\alpha = (A, \sim, \textsc{con}, a_0)$, where $\textsc{con}$ is a triplet of total functions $(\textsc{pre}, \textsc{flip}, \textsc{trm})$, $\textsc{trm} : A \to \mathcal{L}_\Phi$, and $(A, \sim, (\textsc{pre}, \textsc{flip}), a_0)$ is a BMS epistemic event (as defined in section 2).

We call $\textsc{trm}_a \in \mathcal{L}_\Phi$ the *termination condition* of a. As above, for an event token a, we put, reusing notation, $\textsc{con}_a := (\textsc{pre}_a, \textsc{flip}_a, \textsc{trm}_a)$. When discussing several epistemic events at the same time, we distinguish their components by using superscripts, writing A^α and A^β, $\to_i^\alpha$ and $\to_i^\beta$ etc.

Intentions and Plans. An *intention* is a set of epistemic events. A *plan* is a finite sequence (i.e. a string) of intentions.

An intention σ consisting of *public* epistemic events (defined as in section 2 above) may be written as $\textsc{con}_{\alpha_1} \cup \ldots \cup \textsc{con}_{\alpha_n}$, where $\sigma = \{\alpha_1, \ldots, \alpha_n\}$, and $\textsc{con}_{\alpha_k} = (\textsc{pre}_{\alpha_k}, \textsc{flip}_{\alpha_k}, \textsc{trm}_{\alpha_k})$ for $1 \le k \le n$.

Intention Models. An *intention model* is a structure $\mathfrak{m} = (S, \sim, V, \pi, s_0)$, where $(S, \sim, V, s_0)$ is an epistemic model, and $\pi = (\pi_i)_{i \in \textsc{ag}}$ is a collection of total functions assigning a plan to each state in S ($\pi_i(s)$ is called the *plan of i at s in* $\mathfrak{m}$).

We require that π *respects* $\sim$, that is, $s \sim_i t$ implies $\pi_i(s) = \pi_i(t)$ for all agents i. This means that *agents know their plans*.

Given an intention model $\mathfrak{m}$, and a state $s \in S$, we write $\pi_i(s)_n$ for the nth element of $\pi_i(s)$. Suppose $\pi_i(s) = \pi_i(s)_1 \cdot \ldots \cdot \pi_i(s)_n$ for some $n \in \mathbb{N}$. We call $\pi_i(s)_1$ the *proximal intention of agent i at s* in $\mathfrak{m}$ (notation: $\textsc{prox}_{\pi_i(s)}$), and $\pi_i(s)_2 \cdot \ldots \cdot \pi_i(s)_n$ the *distal intentions of agent i at s* in $\mathfrak{m}$ (notation: $\textsc{dist}_{\pi_i(s)}$).

Bisimilar Events. By way of preparation, for epistemic formulas φ_k and ψ_k, we may write $\bigwedge_{1\le k\le 3}(\varphi_k \leftrightarrow \psi_k)$ as $(\varphi_1,\varphi_2,\varphi_3) \leftrightarrow (\psi_1,\psi_2,\psi_3)$. Let α,β be epistemic events. A *bisimulation between α and β* is a binary relation $R \subseteq A^\alpha \times A^\beta$ such that $a_0^\alpha R a_0^\beta$ and

- If aRb, then $\vDash \textsc{con}_a^\alpha \leftrightarrow \textsc{con}_b^\beta$ (a and b *have equivalent atomic type*), and
- If aRb and $a \sim_i^\alpha a'$, then $\exists b' : b \sim_i^\beta b'$ and $a'Rb'$ and vice versa (a and b *satisfy back and forth*).

$\alpha \equiv \beta$ means that there is a bisimulation between α and β. In that case, we say that α and β are *bisimilar*.

Bisimilarity is a crucial ingredient in the following definition of licensing, which was informally discussed in section 2.

Licensing. Let $\mathfrak{m}$ be an intention model, and suppose $\textsc{prox}_{\pi_i(s_0)} = \sigma$. An epistemic event α is *licensed by agent i's intention in* $\mathfrak{m}$ if there exists an epistemic event $\beta \in \sigma$ such that $\alpha \equiv \beta$; α is *licensed by agent i's information in* $\mathfrak{m}$ if $\mathfrak{m} \vDash \Box_i \text{PRE}_\alpha$; α is *licensed for agent i in* $\mathfrak{m}$ (notation: $\mathfrak{m} \in \textsc{dom}_\alpha^i$) if α is licensed by agent i's intention in $\mathfrak{m}$ and α is licensed by agent i's information in $\mathfrak{m}$.[2]

Note that if α is licensed by agent i's information in $\mathfrak{m}$, then $\mathfrak{m} \vDash \textsc{pre}_\alpha$ by the truth definition of $\Box_i$, since $\sim_i$ is reflexive.

Product Update of an Intention Model. Let $\mathfrak{m}$ be an intention model, α an epistemic event and $i \in \textsc{ag}$. The *product update of* $\mathfrak{m}$ *with* α *for* i (notation: $\mathfrak{m} \otimes_i \alpha$) is defined iff α is licensed for agent i in $\mathfrak{m}$. In that case, $\mathfrak{m} \otimes_i \alpha := (S^{\mathfrak{m}\otimes_i\alpha}, \sim_i^{\mathfrak{m}\otimes_i\alpha}, V^{\mathfrak{m}\otimes_i\alpha}, \pi^{\mathfrak{m}\otimes_i\alpha}, s_0^{\mathfrak{m}\otimes_i\alpha})$, determined in two steps. First, we define:

$$\begin{aligned}
S^{\mathfrak{m}\otimes_i\alpha} &:= \{s.a \mid s \in S, a \in A, \mathfrak{m}^s \in \textsc{dom}^i_{\alpha^a}\}\\
\sim_i^{\mathfrak{m}\otimes_i\alpha} &:= \{(s.a, t.b) \mid s \sim_i t, a \sim_i b\}\\
V^{\mathfrak{m}\otimes_i\alpha}(s.a) &:= V(s) \Delta \textsc{flip}_a\\
s_0^{\mathfrak{m}\otimes_i\alpha} &:= s_0.a_0
\end{aligned}$$

This determines an epistemic model $\mathfrak{t}$. Note that this matches the original definition of product update given in section 2, modulo the use of $\textsc{dom}^i_{\alpha^a}$ instead of $\textsc{pre}_{\alpha^a}$. Second, it remains to update the plans of the agents. For agents j distinct from i we just copy their plans (observe that this means that we do not attempt to capture any systematic effects some agent's action might have on another agent's intentions). For agent i, we proceed conditional on the termination condition:

$$\begin{aligned}
\pi_j^{\mathfrak{m}\otimes_i\alpha}(s.a) &:= \pi_j(s) && \text{for all } j \neq i\\
\pi_i^{\mathfrak{m}\otimes_i\alpha}(s.a) &:= \text{DIST}_{\pi_i(s)} && \text{if } \mathfrak{t}^{s.a} \vDash \Box_i \text{TRM}_{\alpha^a}\\
\pi_i^{\mathfrak{m}\otimes_i\alpha}(s.a) &:= \pi_i(s) && \text{if } \mathfrak{t}^{s.a} \nvDash \Box_i \text{TRM}_{\alpha^a}
\end{aligned}$$

[2] As an anonymous referee pointed out, one might also define the notion of licensing by using van Eijck et al. (2008)'s notion of *action emulation* instead of bisimulation: this would be justified because, as shown by van Eijck et al., updating with two epistemic events that emulate each other yields bisimilar output models. We do not pursue this here because the notion of bisimulation is more standard and slightly simpler to formulate.

This means that agent i drops the proximal intention (as given by the old intention model) iff she knows that its termination condition applies (in the new intention model). Since termination conditions are epistemic formulas, this is well-defined.

5 Language and Semantics

We now formally introduce our language to talk about knowledge, intentions, and actions, and state its semantics.

The sublanguage $\mathcal{L}^*_\Theta$. Let $\Theta := \mathcal{L}_\Phi \times \mathcal{L}_\Phi \times \mathcal{L}_\Phi$ be the set of triples of epistemic formulas over Φ. As defined in section 2, this determines an epistemic language $\mathcal{L}^*_\Theta$, which has the elements $C \in \Theta$ as atomic formulas, and has common knowledge operators besides the operators for individual knowledge. Formulas $\vartheta \in \mathcal{L}^*_\Theta$ are interpreted at epistemic events (for brevity, we do not state the obvious Boolean clauses; we remind the reader that $(\varphi_1, \varphi_2, \varphi_3) \leftrightarrow (\psi_1, \psi_2, \psi_3)$ abbreviates $\bigwedge_{1 \le k \le 3} \varphi_k \leftrightarrow \psi_k$):

$$
\begin{aligned}
\alpha \vDash C \quad & \text{iff} \quad \vDash C \leftrightarrow \textsc{con}_\alpha \\
\alpha \vDash \Box_i \vartheta \quad & \text{iff} \quad \forall a \in A : \text{If } a_0 \sim_i a, \text{ then } \alpha^a \vDash \vartheta \\
\alpha \vDash \Box^*_\iota \vartheta \quad & \text{iff} \quad \forall a \in A : \text{If } a_0 \sim^*_\iota a, \text{ then } \alpha^a \vDash \vartheta
\end{aligned}
$$

The main language $\mathcal{L}_\Phi[\boxdot, \alpha]$. Let $i \in \textsc{ag}$, $n \in \mathbb{N}$, $p \in \Phi$, $\vartheta \in \mathcal{L}^*_\Theta$, and α an epistemic event. Then $\varphi \in \mathcal{L}_\Phi[\boxdot, \alpha]$ iff φ is obtained according to the following rule:

$$
\varphi := \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box_i \varphi \mid \boxdot^n_i \vartheta \mid [\alpha]_i \varphi
$$

We abbreviate $\boxdot^1_i$ as $\boxdot_i$. Formulas $\varphi \in \mathcal{L}_\Phi[\boxdot, \alpha]$ are interpreted in intention models, as follows:

$$
\begin{aligned}
\mathfrak{m} \vDash \Box_i \varphi \quad & \text{iff} \quad \forall s \in S : \text{If } s_0 \sim_i s, \text{ then } \mathfrak{m}^s \vDash \varphi \\
\mathfrak{m} \vDash \boxdot^n_i \vartheta \quad & \text{iff} \quad \forall \alpha \in \pi_i(s_0)_n : \alpha \vDash \vartheta \\
\mathfrak{m} \vDash [\alpha]_i \varphi \quad & \text{iff} \quad \mathfrak{m} \otimes_i \alpha \vDash \varphi, \text{ if defined}
\end{aligned}
$$

6 Examples

To illustrate how our setting works, we return to the two scenarios presented in the introduction. Our aim is to bring out the two roles of intentions in intelligent interaction we have stressed: information about another agent's intention allows to draw useful conclusions from observations; and intentions act as a certain kind of "program", generating longer-term patterns of interaction.

The Chocolate Bar. Recall that our main intuition about the chocolate bar example was that Jane's knowledge about Ian's prior intention allows her to conclude that she can find the recipe book in the bottom drawer. We present a formal version of the example in figure 1, using the following abbreviations:

- `choc-top` := "The chocolate is in the top drawer"
- `choc-bot` := "The chocolate is in the bottom drawer"
- `retrieve-top` := "Ian has retrieved an object from the top drawer"
- `retrieve-bot` := "Ian has retrieved an object from the bottom drawer"

We use double edges to indicate designated states and event tokens. Reflexive arrows are omitted (recall that we are working with equivalence relations in this paper).

The reasoning in the scenario can now be explained as follows: When Ian opens the top drawer, Jane cannot be sure which object he is retrieving from the way the action appears to her (the epistemic event thus consists of two event tokens). She does clearly see, however, that he is retrieving an object from the top drawer (both event tokens change the truth-value of `retrieve-top`). And her prior knowledge about Ian's intention tells her that, were the chocolate in the bottom drawer, Ian would have retrieved an object from that very bottom drawer (since the formula $\texttt{choc-bot} \to \boxdot_i(\texttt{choc-bot}, \texttt{retrieve-bot}, \top)$ is true in both states of the initial model). For this reason, Ian's action cannot be one of retrieving the recipe book from the top drawer—this uses the assumption that Ian's intention is a *commitment*; in the formal model, the assumption corresponds to the fact that neither of the two event tokens is compatible with the right state of the initial model. So the left state is the only state surviving the update, and in the output model, Jane knows where the chocolate is. From the description of the example it is clear that $\texttt{choc-top} \leftrightarrow \texttt{recipe-book-bot}$ is globally true in the initial model (and thus, in the output model as well), where `recipe-book-bot` stands for "The recipe book is in the bottom drawer". So Jane now knows where to find the recipe book, which explains why she can retrieve it without hesitation. And this is what the product update operation introduced in section 4 correctly predicts.

Consecutive Numbers. Figure 2 shows a model of the consecutive numbers scenario given in the introduction. In this model, the referee has already assigned the numbers to Ian and Jane, the actual distribution being 2 for Ian and 3 for Jane. We have proposition letters n_i for "Ian has number i", and h_j for "Jane's

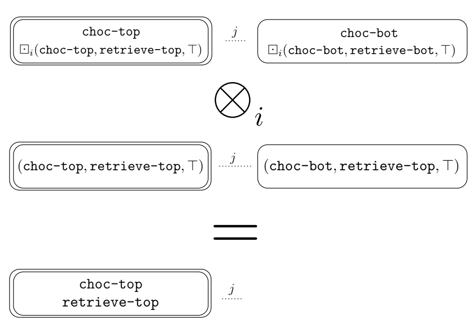

Fig. 1. *The chocolate bar.*

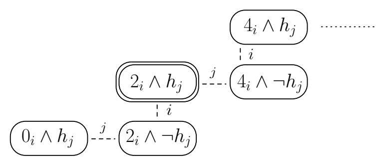

Fig. 2. *Consecutive numbers.*

number is Ian's number plus 1" ($\neg h_j$ thus represents "Jane's number is Ian's number minus 1"). As explained in the introduction, few moves suffice in this situation to arrive at knowledge for both agents, and end the game.

As pointed out, however, we are interested here in finding appropriate intentions that, understood as "programs" guiding the interaction, work for *all* possible distributions of numbers. There seem to be three major requirements on such a program: (1) it should express a conditional intention ("announce that you know *if you know*, and that you don't otherwise"); (2) it should allow for iterated announcements of ignorance ("announce that you do not know *as long as* you do not know"; and it should make sure termination at the right moment ("as soon as you have said that you know, drop out"). Using our framework, we can model Ian and Jane's respective intentions as follows:

- Jane's intention is $\alpha \cup \beta$, where $\alpha = (\Diamond_j h \wedge \Diamond_j \neg h, \top, \bot)$, and $\beta = (\Box_j h \vee \Box_j \neg h, \top, \top)$.
- Ian's intention is $\gamma \cup \delta$, where $\gamma = (\Diamond_i h \wedge \Diamond_i \neg h, \top, \bot)$ and $\delta = (\Box_i h \vee \Box_i \neg h, \top, \top)$.

The union corresponds to requirement (1), the termination condition $\bot$ gives rise to an unbounded number of iterated announcements of ignorance and corresponds to requirement (2), while the termination condition $\top$ ensures termination after one announcement of knowledge (corresponding to (3)).

Assuming that Ian and Jane have no further intentions (that is, their plan is just a sequence of length 1), in the model of figure 2, we have that

$$\langle \gamma \rangle_i \langle \alpha \rangle_j \langle \delta \rangle_i \langle \beta \rangle_j (\boxdot_i \bot \wedge \boxdot_j \bot)$$

is true. But more generally, this type of pattern "works" for any concrete distribution of numbers, in that repeatedly generating actions licensed for Jane and Ian will eventually lead them both to terminate their intentions: no further updates *of any kind* are licensed after one announcement of ignorance, since both agents have dropped all their intentions.

7 Reduction Laws

The goal of this section is to develop reduction laws to recursively characterize the action modalities. The main point is that licensing may be syntactically characterized in our setting, using well-known techniques from modal logic (Moss 2007, Balbiani and Herzig 2007, Baltag et al. 1999).

Characteristic formulas. For all epistemic events α, we inductively define formulas nf_α^n as follows:

$$\begin{aligned} \mathrm{nf}_\alpha^0 &:= (\mathrm{PRE}_\alpha, \mathrm{FLIP}_\alpha, \mathrm{TRM}_\alpha) \\ \mathrm{nf}_\alpha^{n+1} &:= \mathrm{nf}_\alpha^0 \wedge \bigwedge_{i \in \mathrm{AG}} (\Box_i \bigvee_{a_0 \sim_i a} \mathrm{nf}_{\alpha^a}^n \wedge \bigwedge_{a_0 \sim_i a} \Diamond_i \mathrm{nf}_{\alpha^a}^n) \end{aligned}$$

We set $\mathrm{nf}_\alpha := \mathrm{nf}_\alpha^{|A|}$, where $|A|$ is the size of A. One may now define formulas μ_α as follows:

$$\mu_\alpha := \bigwedge_{a \in A, i \in \mathrm{AG}} (\mathrm{nf}_{\alpha^a} \to \Box_i \bigvee_{a \sim_i b} \mathrm{nf}_{\alpha^b} \wedge \bigwedge_{a \sim_i b} \Diamond_i \mathrm{nf}_{\alpha^b})$$

It may be checked that this construction is sound, i.e. for all α and for all $a \in A$, we have $\alpha^a \vDash \mu_\alpha$ (Balbiani and Herzig 2007). Next, χ_α, *the characteristic formula of* α, is defined as:

$$\chi_\alpha := \mathrm{nf}_\alpha \wedge \Box_{\mathrm{AG}}^* \mu_\alpha$$

χ_α indeed *characterizes* α, in the following sense:

Lemma 1. *For all epistemic events α and β: $\beta \vDash \chi_\alpha$ iff $\alpha \equiv \beta$.*

Proof: *Assuming $\alpha \equiv \beta$ it follows that $\beta \vDash \chi_\alpha$, since bisimilarity implies modal equivalence. For the other direction, suppose $\beta \vDash \chi_\alpha$. Consider the relation $R \subseteq A^\alpha \times A^\beta$ given by aRb iff $\beta^b \vDash \mathrm{nf}_{\alpha^a} \wedge \Box_{\mathrm{AG}}^* \mu_\alpha$. We claim that R is a bisimulation. Since $a_0^\alpha R a_0^\beta$ by assumption, it then follows that $\alpha \equiv \beta$. To prove the claim, suppose aRb for some a and b. Since $\beta^b \vDash \mathrm{nf}_{\alpha^a}^0$, α^a and β^b have equivalent atomic type. It remains to show that α^a and β^b satisfy back and forth. Fix some agent i. Suppose there is an a' such that $a \sim_i^\alpha a'$. Since $\vDash \Box_{\mathrm{AG}}^* \mu_\alpha \to \mu_\alpha$, from the assumption $\beta^b \vDash \mu_\alpha$. Since $\beta^b \vDash \mathrm{nf}_{\alpha^a}$, using μ_α we see that $\beta^b \vDash \Diamond_i \mathrm{nf}_{\alpha^{a'}}$. Hence there exists b' such that $b \sim_i^\alpha b'$ and $\beta^{b'} \vDash \mathrm{nf}_{\alpha^{a'}}$. And since $\vDash \Box_{\mathrm{AG}}^* \mu_\alpha \to \Box_i \Box_{\mathrm{AG}}^* \mu_\alpha$, we have $\beta^{b'} \vDash \Box_{\mathrm{AG}}^* \mu_\alpha$. So $a'Rb'$. For the other half of the argument, suppose there exists b' such that $b \sim_i^\beta b'$. Since $\beta^b \vDash \Box_i \bigvee_{a \sim_i^\alpha a'} \mathrm{nf}_{\alpha^{a'}}$, we have $\beta^{b'} \vDash \mathrm{nf}_{\alpha^{a'}}$ for some a' such that $a \sim_i^\alpha a'$. And since $\beta^{b'} \vDash \mu_\alpha$, it follows that $a'Rb'$, hence R is a bisimulation, which completes the proof.* □

Domain Definitions. Consciously overloading notation, we define, for any agent i, a sentence DOM_α^i (the *domain definition of* α *for* i) as follows:

$$\mathrm{DOM}_\alpha^i := \Box_i \mathrm{PRE}_\alpha \wedge \hat{\Diamond}_i \chi_\alpha$$

Lemma 2. *Let $\mathfrak{m}$ be an intention model. Then $\mathfrak{m} \in \mathrm{DOM}_\alpha^i$ iff $\mathfrak{m} \vDash \mathrm{DOM}_\alpha^i$.*

Proof. *Using lemma 1.* □

Reduction Laws. Using domain definitions, we state the reduction laws in figure 3.

$$
\begin{aligned}
[\alpha]_i\, p &\leftrightarrow (\textsc{dom}^i_\alpha \rightarrow p) && (p \notin \textsc{flip}_\alpha)\\
[\alpha]_i\, p &\leftrightarrow (\textsc{dom}^i_\alpha \rightarrow \neg p) && (p \in \textsc{flip}_\alpha)\\
[\alpha]_i\, \Box_j \varphi &\leftrightarrow (\textsc{dom}^i_\alpha \rightarrow \bigwedge\{\Box_j [\alpha^a]_i \varphi \mid a_0 \sim_j a\})\\
[\alpha]_i\, \boxdot^n_i \vartheta &\leftrightarrow (\textsc{dom}^i_\alpha \rightarrow (([\alpha]_i\, \Box_i\, \textsc{trm}_\alpha \rightarrow \boxdot^{n+1}_i \vartheta)\\
&\qquad \wedge ([\alpha]_i \neg\, \Box_i\, \textsc{trm}_\alpha \rightarrow \boxdot^n_i \vartheta))\\
[\alpha]_i\, \boxdot^n_j \vartheta &\leftrightarrow (\textsc{dom}^i_\alpha \rightarrow \boxdot^n_j \vartheta) && (i \neq j)\\
[\alpha]_i\, \neg\varphi &\leftrightarrow (\textsc{dom}^i_\alpha \rightarrow \neg[\alpha_i]\varphi)\\
[\alpha]_i\, (\varphi \wedge \psi) &\leftrightarrow ([\alpha]_i \varphi \wedge [\alpha]_i \psi)
\end{aligned}
$$

Fig. 3. *Reduction Laws.*

Proposition 1. *The reduction laws are valid.*

Proof: *Let $\mathfrak{m}$ be an intention model. For all but the last reduction law, observe that $\mathfrak{m} \nvDash \textsc{dom}^i_\alpha$ implies that both sides of the respective law are true. For the left side, this follows from the definition of product update using lemma 2, for the right side it follows from the definition of material implication. In proving the respective equivalences, we may thus assume that $\mathfrak{m} \vDash \textsc{dom}^i_\alpha$. The remainder of the respective proofs are then analogous to standard arguments (van Ditmarsch et al. 2006). The validity of the last reduction law is a direct consequence of the semantics.* □

8 Conclusion

Summary. We have argued that intention is a useful notion in studying information flow and interaction in multi-agent settings, specifically focusing on how information about another agent's intentions drives reasoning about observations; and on the role of intentions as complex instructions agents adopt and act upon. We have modeled intentions and actions using one unifying formal concept, namely the notion of an epistemic event. This allowed us to account for the way actions, which we assumed to be perfectly controlled by the agents taking them, originate in mental states in terms of a formal notion of licensing. The dropping of intentions was encoded using special termination conditions our epistemic events come equipped with. The setting allowed us to formalize our initial examples. We have shown how the usual DEL technique of using reduction axioms transfers to our setting.

Related Research. Prominent lines of research in logics of intentions have been established in the 90s by Cohen and Levesque, Rao and Georgeff, and Meyer and collaborators. Rao and Georgeff use a temporal setting based on CTL, and study belief, desire and intention in a very expressive first-order modal language (Rao and Georgeff 1991). In Cohen and Levesque's work (which also uses first-order expressivity), time and PDL style events are primitive notions, as are beliefs and goals. Intention, on the other hand, is a derived concept (Cohen and Levesque 1990). The KARO framework of Meyer and collaborators is a modal action theory in the tradition of PDL. KARO has been extended to account for motivational attitudes, including goals and commitments (this work is contained in Meyer et al. (1999), which references the earlier work of the authors on the KARO formalism). Meyer and Veltman (2007) includes a survey of these three traditions, as does van der Hoek and Wooldridge (2003). Another important tradition in logics of action is constituted by the family of STIT logics (Belnap et al. 2001). A STIT analysis of the concept of an intentional action has been proposed in Broersen (2009). In the last several years, the question how agents revise their intentions has attracted growing interest (Shoham 2009, Icard et al. 2010, van der Hoek et al. 2007, Lorini and Herzig 2008). Roy (2008) is an extensive analysis of intentions originating in the same research tradition as the present paper, i.e. Dynamic Epistemic Logic. However, Roy approaches intention from an angle quite different from the present paper. On the formal level, the semantics of intentions (in so-called game models associated to strategic games) is given by neighborhood functions, i.e. on the level of states in the game model. BMS epistemic events represent not actions *in*, but announcements *about* a game. On the conceptual level, Roy's work is focused on studying rationality concepts in strategic games, not on studying the roles of intentions we have emphasized.

More generally, formalizing *intending to act* in terms of epistemic events is an original contribution of the present paper, as is the account of *acting on an intention* using an adapted product update. In terms of the scope of analysis, much previous research goes far beyond the work reported in this paper, as, e.g., questions of rationality and intention revision are formally addressed. On the other hand, the roles of intentions—as resources for information flow, and as "programs" structuring interaction—we have stressed have received less attention so far.

Directions for Future Work. The meta-logical properties of the system presented here obviously rank at the top of the list of topics for further research. Two desirable extensions of the present analysis have already been mentioned: The modeling of *more complex types of plans* that include cross-temporal dependencies; and the question how mistakes, or, more generally, *actions taken based on, but deviating from prior intentions* may be modeled. The above discussion of related work suggests two further topics: How can *intention revision* be modeled in the approach of this paper? And how to account for the various *rationality constraints on intentions* that have captured the attention of so many authors?

Beyond that, we mention three natural directions further investigations could take. *(1)* The understanding of how intentions structure interaction would be enhanced by modeling more precisely how actions are *scheduled* in advance. This would mean to consider indexical intentions ("i intends to take action α at time t", cf. Icard et al. (2010)). In the present setting, a natural way to do this would be by time-stamping event tokens (Baltag 2001). *(2)* A role of intentions we have not touched upon here is as an enabler for interpersonal coordination in cases of agents acting simultaneously in a coordinated fashion. Michael Bratman has argued that cases of *shared agency* (like Gilbert (1990)'s example of "walking together") rest on symmetrically "interlocking intentions" and common knowledge that this is the case (Bratman 2009). But simultaneous action may also be asymmetric, as in the case of an agent intervening in another's action (e.g. "breaking into a secret message-passing"). *(3)* We have focused on the rather idealized case of knowledge as formalized by the familiar S5 requirements of truthfulness and full introspection. It would be more realistic to work with a notion of (possibly false) belief. Using a belief-revision friendly of DEL then becomes a natural move (van Benthem 2007, Baltag and Smets 2008).

References

R.J. Aumann and J.H. Dreze. When all is said and done, how should you play and what should you expect? 2005. URL http://www.ma.huji.ac.il/raumann/pdf/dp387.pdf.

Philippe Balbiani and Andreas Herzig. Talkin'bout kripke models. In Jorgen Villadsen, Thomas Bolander, and Torben Braüner, editors, *Proceedings of the International Workshop on Hybrid Logic HyLo 2007*. 2007.

Alexandru Baltag. Logics for insecure communication. In Johan van Benthem, editor, *Proceedings of TARK'01 (Eight Conference on Rationality and Knowledge)*, pages 111–122, 2001.

Alexandru Baltag. A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54:1–46, 2002.

Alexandru Baltag and Lawrence Moss. Logics for epistemic programs. *Synthese*, 139(2):165–224, 2004.

Alexandru Baltag and Sonja Smets. A qualitative theory of dynamic interactive belief revision. In Giacomo Bonanno, Wiebe van der Hoek, and Michael Wooldridge, editors, *Texts in Logic and Games*, volume 3. Amsterdam University Press, 2008.

Alexandru Baltag, Lawrence Moss, and Slawomir Solecki. The logic of public announcements, common knowledge and private suspicions. Technical report, CWI, 1999.

Nuel Belnap, Michael Perloff, and Ming Xu. *Facing the Future. Agents and Choice in Our Indeterminist World*. Oxford University Press, 2001.

Adam Brandenburger. The power of paradox: some recent developments in interactive epistemology. *International Journal of Game Theory*, 35:465–492, 2007.

Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.

Michael E. Bratman. Modest sociality and the distinctiveness of intention. *Philosophical Studies*, 144:149–165, 2009.

Jan Broersen. First steps in the stit-logic analysis of intentional action. In Andreas Herzig and Emiliano Lorini, editors, *Proceedings ESSLLI 2009 workshop on Logical Methods for Social Concepts (LMSC'09)*. 2009.

John Broome. Practical reasoning. In J. Bermudez and A. Millar, editors, *Reasons and Nature: Essays in the Theory of Rationality*. Oxford University Press, 2002.

Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.

Jelle Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute for Logic, Language and Computation, 1999.

Margaret Gilbert. Walking together: A paradigmatic social phenomenon. *Midwest Studies*, 15:1–14, 1990.

Gilbert Harman. *Change in View*. MIT Press, 1986.

Thomas Icard, Eric Pacuit, and Yoav Shoham. Joint revision of belief and intention. Forthcoming in Proceedings of KR-2010, 2010.

Barteld Kooi. *Knowledge, Chance and Change*. PhD thesis, University of Groningen, 2003.

Emiliano Lorini and Andreas Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77, 2008.

Alfred R. Mele. Agency and mental action. *Noûs*, 31:231–249, 1997a.

Alfred R. Mele. Introduction. In Alfred R. Mele, editor, *The Philosophy of Action*. Oxford University Press, 1997b.

John-Jules Meyer and Frank Veltman. Intelligent agents and common sense reasoning. In Patrick Blackburn, Johan van Benthem, and Frank Wolter, editors, *Handbook of Modal Logic*. 2007.

John-Jules Meyer, Wiebe van der Hoek, and B. van Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113:1–40, 1999.

Lawrence Moss. Finite models constructed from canonical formulas. Unpublished, 2007.

Jan Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216. 1989.

R. Ramanujam and Sunil Simon. Dynamic logic on games with structured strategies. In *Proceedings of KR 2008*, pages 49–58. 2008.

Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-architecture. Technical Note 14, Australian Artificial Intelligence Institute, 1991.

Olivier Roy. *Thinking Before Acting. Intentions, Logic, Rational Choice*. PhD thesis, Institute for Logic, Language and Computation, 2008.

Yoav Shoham. Logical theories of intention and the database perspective. *Journal of Philosophical Logic*, 38:633–647, April 2009.

Johan van Benthem. Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 14:129–155, 2007.

Johan van Benthem. Logic and intelligent interaction. In Clark Glymour, Wei Wan, and Dag Westerstahl, editors, *Logic, Methodology and Philosophy of Science: Proceedings of the Thirteenth International Congress*. College Publications, 2009a.

Johan van Benthem. Logical dynamics of information and interaction. Book Manuscript, 2009b.

Johan van Benthem, Jan van Eijck, and Barteld Kooi. Logics of communication and change. *Information and Communication*, 204(11):1620–1662, 2006.

Wiebe van der Hoek and Michael Wooldridge. Towards a logic of rational agency. *Logic Journal of the IGPL*, 11(2):133157, 2003.

Wiebe van der Hoek, Wojciech Jamroga, and Michael Wooldridge. Towards a theory of intention revision. *Synthese*, 155(2):103–128, 2007.

Hans van Ditmarsch, Barteld Kooi, and Wiebe van der Hoek. *Dynamic Epistemic Logic*. Springer, 2006.

Jan van Eijck, Ji Ruan, and Tomasz Sadzik. Action emulation. Working Paper, 2008.