Synthese (2009) 171:291–320
Knowledge, Rationality & Action 893–922
DOI 10.1007/s11229-009-9644-3

A dynamic-epistemic hybrid logic for intentions and information changes in strategic games

O. Roy

Received: 9 February 2009 / Accepted: 15 July 2009 / Published online: 1 August 2009

Abstract In this paper I present a dynamic-epistemic hybrid logic for reasoning about information and intention changes in situations of strategic interaction. I provide a complete axiomatization for this logic, and then use it to study intentions-based transformations of decision problems.

Keywords Intentions · Rationality · Interaction · Transformation of decision problems · Fixed points

The capacity of human agents to deliberate in advance and to form intentions about future actions is a central aspect of rational agency, and at least since (Harman 1976) and (Bratman 1987) this aspect has been studied extensively in philosophy of action. Among the distinguishing features of intentions identified in this literature is the fact that these states generate specific *expectations*, both about oneself and in situations of interaction. The expectations are, in turn, an important anchor for inter-temporal and inter-personal coordination, and they are the building blocks of many theories of shared agency (Velleman 1997; Bratman 1999). Philosophers of action have also repeatedly pointed out that intentions bring in a specific *dynamic* in deliberation. They induce *transformations* of decision problems by filtering the set of options, and by triggering deliberation on means (Bratman 1987; Bratman et al. 1988; Roy 2009).

These philosophical insights on the role of intentions in rational agency find a natural environment to be developed further in contemporary epistemic game theory and dynamic epistemic logic. Epistemic game theory offers an extensive toolbox with which to study the relation between mutual expectations and interactive rationality

O. Roy (✉)
Faculty of Philosophy, University of Groningen, Oude Boteringestraat 52,
9712 GL Groningen, The Netherlands
e-mail: o.roy@rug.nl

(Aumann 1999; Brandenburger 2007). Dynamic epistemic logic, on the other hand, provides a unifying framework for the analysis of dynamic phenomena in interaction (van Benthem to appear), in particular those pertaining to transformations of models of mutual expectations (van Benthem 2007a,b; Baltag et al. 2009) and preferences (Liu 2008; Girard 2008).

Despite this obvious congeniality of interest between the philosophical theory of intentions, on the one hand, and epistemic game theory and dynamic epistemic logic on the other, there has so far been very little contact between the two. Intentions are absent from contemporary epistemic game theory, and the existing literature on intention in decision theory has focused on the rationality of making and keeping commitments; see for instance (McClennen 1990) and (Gul and Pesendorfer 2001). Logical theories of intentions, on the other hand, have mostly considered the relations between intentions and other propositional attitudes, such as desires, goals and beliefs, in the single-agent case. See for instance (Cohen and Levesque 1990; Georgeff et al. 1999; Wooldridge 2000). These theories have not touched on the role of mutual expectations generated by intentions, nor on the transformations of decision problem that these attitudes induce in situations of interaction.

In this paper I draw from these three traditions in the foundations of rational agency to develop a dynamic epistemic (hybrid) logic with which to reason about information and intention changes in strategic interaction. I provide a complete axiomatization for this logical system and use it to further the investigation set out in (van Hees and Roy 2008) and (Roy 2009) on the role of intentions in strategic interaction. I answer four questions that have been left open in these earlier two papers. I first provide an epistemic characterization of the operation there called "cleaning" of strategic games, showing that it corresponds to a public announcement of "intention-rationality"—the fact that the agents do not choose strategies that would exclude the achievement of their intentions—under the condition that these intentions are common knowledge. Second, I show how this operation behaves in situations where intentions are not common knowledge, i.e. when there is uncertainty about intentions. Third, I define alternative transformations of decision problems, and study their relation with van Hees and Roy's original cleaning operation. Finally, again using the resources of the dynamic formalism, I study the iterated behaviour of these two operations. I show that they "stabilize" in the long run, and that they lead to related but nevertheless different decision situations. Along the way I briefly touch the question of intention revision. All in all this shows that the logical framework proposed here brings in new insights, both to the philosophical theory of intentions and the epistemic theory of games.

The paper is organized as follows. In Sect. 1, I develop the dynamic-epistemic hybrid logical system. In Sect. 1.1, I introduce the structures I study, namely games in strategic forms and epistemic frames with intentions. In Sect. 1.2, I introduce the logical language to talk about the static features of these structure—strategies, intentions and information—and provide a complete axiomatization for it. In Sect. 1.3, I add public announcement operators to this language, and provide a complete axiomatization for these as well. In Sect. 2, I move to applications of this logical system. I show in Sect. 2.1 that the notion of cleaning of strategy sets proposed by van Hees and Roy (2008) corresponds, in the present logical framework, to a public announcement of "intention-rationality" under the condition that intentions are common knowledge.

En route to relaxing this assumption, in Sect. 2.2, I study an alternative announcement or way to clean a strategy set; one that takes the information of the agents into account. In Sect. 2.3, I use the logical system to compare the two forms of cleaning, in both their one-shot and iterated versions. Section 3 concludes: the lengthier proofs are presented in the Appendix.

1 Logic for intentions and information dynamics in strategic games

1.1 Games in strategic form and epistemic frames with intentions

In the background of the present analysis are *agents*, *actions*, *outcomes* and *preferences*, which are the basic ingredients of *games in strategic form* (Osborne and Rubinstein 1994, p. 11). Such games are situations in which a number of agents must choose simultaneously among a set of actions or *strategies*. The combination of all these choices determines the *outcome* of the game, and agents have their own *preferences* over all possible outcomes. Here, I take combinations of strategies, one for each agent, as outcomes of the game. It is common to assume that the preferences of the agents induce a reflexive, transitive and total ordering of the outcome.

Definition 1 (*Games in strategic form*) A game in strategic form is a tuple $\mathbb{G} = \langle \mathcal{A}, \{S_i, \leq_i\}_{i\in\mathcal{A}}\rangle$ such that:

- $\mathcal{A}$ is a finite set of agents.
- S_i is a non-empty and finite set of strategies for i. A tuple $\sigma \in \Pi_{i\in\mathcal{A}} S_i$ is called a strategy profile. I write $\sigma(i)$ for i's component in σ.
- $\leq_i$ is a binary preference relation on $\Pi_{i\in\mathcal{A}} S_i$.

An example of a game in strategic form, here a coordination game, is displayed in Table 1. In this simple scenario there are two players, Ann and Bob, the former choosing a row and the latter a column. Each has two strategies available: either go to Restaurant A or Restaurant B. Their preferences are represented by the numerical utility values in each cell of the table, i.e. at each strategy profile of the game, with $\sigma \leq_i \sigma'$ if and only is the value for i of σ is greater or equal than the value of σ' for i. In this example Ann's utility values are represented on the left side, and Bob's on the right side of each cell.

Preferences stay in the background throughout this paper. I do agree that investigating the relation between intentions and preferences is of great importance for any theory of interactive decision making with intentions, and that game-theoretical analysis of interaction can hardly start without preferences. In this paper, however, I focus on another important aspect of intention-based reasoning in interaction, namely the

Table 1 A coordination game

Ann\Bob	Restaurant A	Restaurant B
Restaurant A	1, 1	0, 0
Restaurant B	0, 0	1, 1

fact that intentions induce "transformations" of decision problems. As we shall see, this aspect raises interesting issues for the theory of intentions in interaction, and so even without taking preferences explicitly into account. What is more, the present logical framework allows for a straightforward extension to preferences, along the line of van Benthem et al. (2008) and van Hees and Roy (2008).

The real loci of the present analysis are not games in strategic forms but rather what I call *epistemic frames with intentions*, which supplement the basic ingredients of the former with a representation of the *information* available to the agents and of the *intentions* they might have formed before entering the game. The epistemic frames I use stem from two sources: on the one hand qualitative models of the agents' *epistemic states*, most developed in the epistemic logic and epistemic game theory literature (Fagin et al. 1995; Aumann 1999; van Benthem to appear), and on the other, the representation of the agent's intentions proposed by van Hees and Roy (2008).

Definition 2 (*Epistemic frames with intentions*) An *epistemic frame with intentions* for the strategic game $\mathbb{G}$ is a tuple $\mathbb{F} = \langle W, f, \{\iota_i, \sim_i\}_{\in \mathcal{A}}\rangle$ such that:

- W is a non-empty and finite set of states.
- $f : W \to \Pi_{i \in \mathcal{A}} S_i$ is a strategy function that assigns to each state $w \in W$ a strategy profile σ. For convenience I write $w(i)$ for $f(w)(i)$.
- $\sim_i \subseteq W^2$ is an epistemic equivalence relation on W such that, for all $w \in W$ and all agents $i \in \mathcal{A}$:
  - if $w \sim_i w'$ then $w(i) = w'(i)$.

  I use $[w]_i$ to denote $\{w' : w \sim_i w'\}$.
- $\iota_i \subseteq W^2$ is an intention compatibility relation on W such that, for all $w \in W$ and all agents $i \in \mathcal{A}$:
  - $\iota_i(w) \neq \emptyset$ and,
  - if $w \sim_i w'$ then $\iota_i(w) = \iota_i(w')$, with $\iota_i(w) = \{w' : w \; \iota_i \; w'\}$.

A state of an epistemic frame with intentions represents the *ex interim stage* (Aumann and Dreze 2008) in the play of a strategic game $\mathbb{G}$ where, to use a well-known metaphor from (von Neumann and Morgenstern 1944), the players have passed their envelopes to the umpire, i.e. they have made their decision, but they are still uninformed about the decisions and intentions of the others. The decision made by each agent at a state, which I also call the choice of an agent at a state, is given by the function f, which assigns to all elements of W a strategy profile in $\mathbb{G}$. Agent i's information at a state w is represented by the relations $\sim_i$, which gives all states w' which i considers possible at w. This describes not only the first-order information of the agents, which choices and intentions of the others they consider possible, but also their higher-order information, i.e. what they consider possible that the others are considering possible. I shall return briefly to higher-order information in Sect. 2.1. The relation ι_i gives for each state w the set of states which are compatible with i's intentions at that state.

At this stage I only impose minimal constraints on the information and intentions of the agents, and these are intended to reflect similar constraints frequently used in the game-theoretical and philosophical literature. I first require each set $\iota_i(w)$ to be non-empty, so that intentions stay "internally consistent" (Bratman 2009). I also assume that an agent "knows" what he chooses and intends, if $w \sim_i w'$ then

Fig. 1 An epistemic frame (without intentions) for the coordination game of Table 1. The solid and dashed arrows represent Ann and Bob's relation $\sim_i$, respectively. An assignment of intentions and strategy choice at each state is described in Table 2

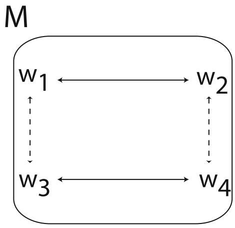


Table 2 An assignment of intentions and strategy choice at each state for Ann and Bob in the epistemic frame of Fig. 1, with $A - A$ meaning that Ann chooses Restaurant A, and Bob as well

State	$f(w)$	$\iota_{Ann}(w)$	$\iota_{Bob}(w)$
w_1	$A - A$	$\{w_1\}$	$\{w_1\}$
w_2	$A - B$	$\{w_1\}$	$\{w_1, w_2, w_3, w_4\}$
w_3	$B - A$	$\{w_1, w_3\}$	$\{w_1\}$
w_4	$B - B$	$\{w_1, w_3\}$	$\{w_1, w_2, w_3, w_4\}$

$w(i) = w'(i)$ and $\iota_i(w) = \iota_i(w')$, but at a given state he might consider that many different choices and intentions are possible for the others (Aumann 1999; van Benthem 2007a,b; Brandenburger 2007). In cases where, for instance, there are states $w', w'' \in [w]_i$ such that $w'(j) \neq w''(j)$, with $i \neq j$, I say that agent i is *uncertain* about the choices of agent j at w, and similarly about j's intentions if there are states $w', w'' \in [w]_i$ such that $\iota_j(w') \neq \iota_j(w')$, with $i \neq j$. Observe that I do not, at least for now, impose a constraint relating what an agent intends to his information at a given state. Such a constraint is discussed in Sect. 2.2.

An example of an epistemic frame with intentions for the coordination game of Table 1 is shown in Fig. 1 and Table 2. At state w_1, for instance, Ann chooses to go to Restaurant A, $w_1(Ann) = A$, and meeting Bob there is the only state compatible with her intentions: $\iota_{Ann}(w_1) = \{w_1\}$. At that state she is uncertain about Bob's choice and intentions. She considers it possible both that Bob is choosing Restaurant A, which he does at w_1, and that he is choosing Restaurant B, which he does at w_2. In the first case she thinks that meeting her at this restaurant is also the only state compatible with his intention, $\iota_{Bob} = \{w_1\}$, but in the second case she thinks that all states are compatible with his intentions, so he does not intend anything in particular.

In this example the set of strategy profiles is in 1-to-1 correspondence with the set of states, but in general this need not to be so. A given strategy profile might be assigned to more than one state in a given model, and some profiles might not be assigned to any state at all. In the first case, where a given profile is assigned to more than one state, one can represent situations where some agents think that some strategy choices of others might be supported by different intentions or information. When a given profile σ is not assigned to any state in the model, we are in a situation where there is absolutely no uncertainty, either first- or higher-order, regarding the fact that σ is not played in this particular situation.

This generality allows one not only to model cases where the players have intentions to reach certain outcomes in the game, i.e. to achieve certain strategy profiles, but also to model cases where the agents' intentions bear on the intentions and information of others as well. In other words, these models are well-suited to study cases of higher-order intentions, just as they are well-suited to study higher-order information. At state w_3, one could for instance see Ann's intentions as bearing on Bob's, in the sense that the states compatible with her intentions at w_3 are exactly those where he intends to meet her at Restaurant A. As the syntax of the logical language which will be used to talk about epistemic frames with intentions makes clear, one can naturally read such case as stating that, at w_3, Ann "intends" that Bob intend to meet her at Restaurant A. Observe that Ann's intention at that state is different from Bob's, as in w_3 they do not meet at Restaurant A. Of course is it open to debate whether agents can have such genuine higher-order intentions, and the present models do not take a stance on this issue. Rather, they provide a general framework for studying the consequences of allowing, or not, such intentions, which I view as one of their assets.

1.2 Modal logic for intentions and information in strategic games

Each state of an epistemic frame with intentions provides us with a description of the agents' information, intentions and strategy choices, and in this section I introduce the language which describes these *static* features. I show how to interpret this language in epistemic frames with intentions and provide a complete axiomatization for the class of such frames.

1.2.1 Language

I use a normal propositional modal language (Blackburn et al. 2006) to talk about intentions and information in epistemic game frames. I also need to talk about the strategy choices of agents at certain states. One way to approach this would be to introduce specific proposition letters to denote strategies, or more precisely that an agent chooses a certain strategy a state. This would provide us with the required additional expressive power to talk about strategy choices, but would remain an *ad hoc* solution. Instead, I opt for hybrid logic (Areces and ten Cate 2007). Not only does this extended modal logic provide us with the adequate level of expressive power to study the relation between strategy choices, information and intentions, but it comes with a well-behaved, and well-understood model-theory (ten Cate 2005). Furthermore, such an extended modal language turns out to be sufficiently expressive to capture Nash equilibrium in games in strategic forms (van Benthem et al. 2006; van Benthem et al. 2008), something which is of importance in connecting the present work with existing game-theoretic analysis.

Definition 3 (*Language for epistemic frames with intentions*) Given a set of atomic propositions PROP and a set of nominals S, let $\mathcal{L}_{SG}$ be the language recursively defined as:

$$\phi ::= p \mid \sigma \mid \phi \wedge \phi \mid \neg\phi \mid K_i\phi \mid I_i\phi \mid E\phi$$

The propositional fragment of this $\mathcal{L}_{\mathcal{SG}}$ is standard. I write $\top$ and $\bot$ for propositional tautologies and contradictions, respectively. Formulas of the form $K_i\phi$ should be read "i knows that ϕ" and those of the form $I_i\phi$ as "i intends that ϕ." These connectives have duals: for $\neg K_i\neg\phi$ I use $\Diamond_i\phi$, which means "i considers ϕ possible", and for $\neg I_i\neg\phi$ I use $\mathsf{i}_i\phi$, meaning "ϕ is compatible i's intentions." Nominals σ are intended as names for states in epistemic game frames. I often write $\sigma(i)$ to refer to the strategy that agent i plays at σ, i.e. as a shorthand for $f(V(\sigma))(i)$. The global modality $E\phi$, which should be read as "there is a state where ϕ holds", is part of the standard hybrid toolkit. Its dual, which I write $A\phi$, is intended to mean "ϕ holds in all states".

Mutual knowledge, "everybody knows that ϕ," is definable in this language as a conjunction of $K_i\phi$ for all agents in $\mathcal{A}$. *Common knowledge* of ϕ is usually defined as the (infinite) conjunction of "everybody knows that everybody knows that .... ϕ" for all finite iterations of "everybody knows that". The notion of common knowledge will be useful in Sect. 2.1, to understand the epistemic assumption behind an existing formalization of the "cleaning" of decision problems, but besides that it plays no role in this paper. On the other hand, introducing the common knowledge operator brings in technical complications in the axiomatization of the static and the dynamic languages I use. For these two reasons I thus choose not to introduce common knowledge in the logical syntax. I do, however, give a model-theoretic definition of this notion in Appendix 4.1.

1.2.2 Epistemic models with intentions

I interpret $\mathcal{L}_{\mathcal{SG}}$ over epistemic *models* with intentions, which are epistemic frames with intentions together with a valuation for the propositions and nominals. I often write "models" instead of "epistemic models with intentions".

Definition 4 (*Models*) A model $\mathbb{M}$ is an epistemic game frame $\mathbb{F}$ together with a valuation function $V : (\textsc{prop} \cup S) \to \mathcal{P}(W)$ that assigns to each propositional atom and nominal the set of states where it is true, with the condition that for all $\sigma \in S$, $V(\sigma)$ is a singleton. A pointed model is a pair $\mathbb{M}, w$.

Definition 5 (*Truth in $\mathcal{L}_{\mathcal{SG}}$*) $\mathbb{M}, w \models x$ iff $w \in V(x)$ for $x \in \textsc{prop} \cup S$.
$\mathbb{M}, w \models \phi \wedge \psi$ iff $\mathbb{M}, w \models \phi$ and $\mathbb{M}, w \models \psi$
$\mathbb{M}, w \models \neg\phi$ iff $\mathbb{M}, w \not\models \phi$
$\mathbb{M}, w \models E\phi$ iff there is a w' such that $\mathbb{M}, w' \models \phi$
$\mathbb{M}, w \models K_i\phi$ iff for all w' such that $w \sim_i w'$, $\mathbb{M}, w' \models \phi$.
$\mathbb{M}, w \models I_i\phi$ iff for all $w' \in \iota_i(w)$, $\mathbb{M}, w' \models \phi$.
I write $||\phi||_{\mathbb{M}}$ for $\{w : \mathbb{M}, w \models \phi\}$, and omit the $\mathbb{M}$ altogether when no confusion can arise.

Definition 6 (*Validity in $\mathcal{L}_{\mathcal{SG}}$*) A formula ϕ is valid in a model $\mathbb{M}$, denoted $\mathbb{M} \models \phi$, whenever $\mathbb{M}, w \models \phi$ for all $w \in W$. A formula is valid in a frame $\mathbb{F}$ whenever it is valid in all models $\mathbb{M} = \langle \mathbb{F}, V \rangle$. Finally, a formula is valid in a class of models M whenever it is valid in all models $\mathbb{M} \in \mathsf{M}$. Validity with respect to classes of frames is defined in the same way.

The nominals and propositions are interpreted in the same way. It is the special clause on the valuation function V that turns the former into real "names" for states. The epistemic operator K_i is interpreted just as in standard epistemic logic, $K_i\phi$ being true at a state w whenever it is true in all states that i considers possible at w. The intention operator I_i is interpreted similarly: it is true at a state w whenever ϕ is true in all states that are compatible with i's intentions at w.

$\mathcal{L}_{\mathcal{SG}}$ is quite an expressive language, but a standard argument shows that it is invariant under hybrid bisimulation (Areces and ten Cate 2007, p. 839). I make extensive use of this expressiveness in Sect. 2, where I generalize van Hees and Roy's (2008) analysis of intention-based transformations of strategic games. For the present it is sufficient to point out that $\mathcal{L}_{\mathcal{SG}}$ can express strategy choices and knowledge thereof at a given state, which is important when axiomatizing the class of epistemic models with intentions.

$$\bigvee_{\sigma(i)=s_i} \sigma \quad i \text{ plays strategy } s_i$$

$$s_i \rightarrow K_i s_i \quad \text{Knowledge of strategy choice}$$

The reader can check that $\bigvee_{\sigma(i)=s_i} \sigma$ is true in a model exactly at those states where i plays strategy s_i, and I abbreviate this formula by s_i. In turn, the formula $s_i \rightarrow K_i s_i$ corresponds, in the technical sense (Blackburn et al. 2001), to the fact that for any two states, if $w \sim_i w'$ then $w(i) = w'(i)$.

1.2.3 Axiomatization

The set of valid formulas of $\mathcal{L}_{\mathcal{SG}}$ over the class of epistemic models with intentions is completely axiomatized by logic $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$, generated by the axioms and inference rules presented in Table 3. That the epistemic fragment is axiomatized by the $S5$ axioms does not come as a surprise, given that $\sim_i$ is an equivalence relation. The intention fragment is in turn axiomatized by the basic modal axiom K and (Ser), which ensures that the relation ι_i is "serial", in the sense that it is never the case that $\iota_i(w)$ is the empty set. A brief syntactic derivation shows that (Ser) does indeed ensure "internal consistency" of intentions, as it implies $\neg I_i \bot$. Finally, the interactions axioms enforce the required interplay between the various components of epistemic frames with intentions. (Inc$_E$) makes E a global modality, while (K-I) and (K-strat) respectively ensure knowledge of one's own intentions and own strategy choice.

I only sketch the proof of Theorem 1 below, the details are given in Appendix 4.2.

Theorem 1 *The logic $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$ is complete with respect to the class of epistemic models with intentions.*

Proof The proof is a collage of known techniques for the various fragments of $\mathcal{L}_{\mathcal{SG}}$. The first part amounts to ensuring that we can build a *named* and *pasted* model for any consistent set of formulas in $\mathcal{L}_{\mathcal{SG}}$. In such models all properties definable by a *pure formula*, i.e. a formula with only nominals as atoms, are cannonical (ten Cate 2005, p. 69). During the construction of the named model we also make sure that it

Table 3 The axiom system for $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$. Here $<>_i$ is $\Diamond_i$ or $¡_i$

- All propositional tautologies.
- Intention fragment:
  - (K) $I_i(\phi \to \psi) \to (I_i\phi \to I_i\psi)$
  - (Ser) $I_i\phi \to ¡_i\phi$
- Epistemic fragment: $S5$ axioms.
  - (K) $K_i(\phi \to \psi) \to (K_i\phi \to K_i\psi)$
  - (4) $K_i\phi \to K_iK_i\phi$
  - (5) $\neg K_i\phi \to K_i\neg K_i\phi$
  - (T) $K_i\phi \to \phi$
- Hybrid fragment: $S5$ axioms for E and:
  - (Inc$_{E-\sigma}$) $E(\sigma \wedge \phi) \to A(\sigma \to \phi)$
  - (Exists$_\sigma$) $E(\sigma)$
- Interaction axioms.
  - (Inc$_E$) $<> \phi \to E\phi$
  - (K-I) $I_i\phi \to K_iI_i\phi$
  - (K-strat) $s_i \to K_is_i$
- (Nec) for all modal connective, and the following additional inference rules. In both cases $\sigma \neq \sigma'$ and the former does not occur in ϕ.

• (Name) From $\sigma \to \phi$ infer ϕ.

• (Paste) From $(E(\sigma' \wedge <> \sigma) \wedge E(\sigma \wedge \phi)) \to \psi$ infer $E(\sigma' \wedge <> \phi) \to \psi$

contains enough states to prove an existence lemma for E. All this is routine for hybrid logic completeness. Most definitions and lemmas come from (Blackburn et al. 2001, pp. 434–445) and (Gargov and Goranko 1993). For the other fragments of $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$, standard arguments show the required existence and truth lemmas for K_i and I_i. The proof is completed by ensuring that the model can be seen as an epistemic model with intentions. This is a more or less direct consequence of the aforementioned canonicity of pure formulas and the various interaction axioms.

1.3 Dynamic-epistemic logic for information and intentions changes in strategic games

In this section, I expand $\mathcal{L}_{\mathcal{SG}}$ with the operator for "public announcements" (van Ditmarsch et al. 2007) in order to study the dynamic interplay of intentions and information in epistemic models with intentions. Just as in the preceding section, I first define the syntax and the semantics of this dynamic extension, and then provide a complete axiomatization for the intended class of frames and its dynamics.

In the process I enter the realm of intention *revision*, as the intentions of an agents might have to be changed drastically in light of new information. Intention revision is a complex issue, which I cannot attempt to cover here in any depth. Rather, I here adopt a simple intention revision policy in order to illustrate how the dynamic language for epistemic frames provides a unifying umbrella for research on intentions in strategic interaction.

1.3.1 Language

The dynamic extension of the language $D\mathcal{L}_{\mathcal{GF}}$ is defined in a way that is now standard in the dynamic-epistemic logic literature, i.e., by simply adding "public announcement" operators to the existing syntax. The resulting logical system comes close to the one proposed by ten Cate (2002). More sophisticated dynamic-epistemic logics have been developed, see for instance van Ditmarsch et al. (2007) and van Benthem (2007a), but as the reader will see, the current framework already provides sufficient tools to capture important aspects of information and intentions in strategic interaction.

Definition 7 ($D\mathcal{L}_{\mathcal{GF}}$) $D\mathcal{L}_{\mathcal{GF}}$, the dynamic extension of $\mathcal{L}_{\mathcal{SG}}$, is recursively defined as follows:

$$\phi ::= p \mid \sigma \mid \phi \wedge \phi \mid \neg\phi \mid K_i\phi \mid I_i\phi \mid \ E\phi \mid [\phi!]\phi$$

The only new formulas in this language are of the form $[\phi!]\psi$, and should be read as "after truthfully announcing that ϕ, it is the case that ψ".

1.3.2 Contracted models and simple intention revisions

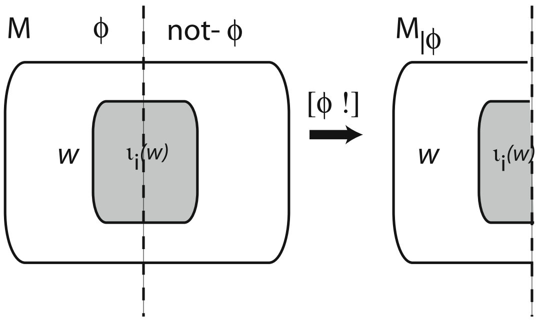
Public announcements are "epistemic actions" which bring new *hard information* (van Benthem 2007a) in a given epistemic model. Information is said to be hard when it is true, and unquestionably so. It overrides any prior information that an agent might have. In contrast, *soft* information (*idem*), is information that the agents only accept with caution, if at all. Here I focus on public announcements of hard information, leaving aside softer forms of epistemic action.

The effect of such incoming hard information on an agent's epistemic states is relatively uncontroversial and quite well understood. In an epistemic model the information of an agent is always truthful, in the sense that if $K_i\phi$ holds at a state w then ϕ also holds at w. Such information can thus never be contradicted by a public announcement of hard information, whatever that is, because such announcements are by definition truthful. In view of this the natural way to update an agent's epistemic state after an announcement of ϕ is simply to discard all the states which were considered possible before this announcement in which ϕ did not hold (Plaza 1989; Gerbrandy 1999).

The case of intention *revision* in the light of a public announcement is more intricate because such an announcement, despite its truthfulness, *can* contradict an agents intentions. The situation here is similar to the one studied in belief revision theory (Alchourron et al. 1985; Rott 2001). Just as a belief that ϕ can be mistaken, and thus contradicted by a truthful announcement of $\neg\phi$, an agent can have the intention that ϕ at state w, in the technical sense of $\mathbb{M}, w \models I_i\phi$, even when $\mathbb{M}, w \models \neg\phi$. In such a case simply discarding from i's intentions all the states where ϕ does not hold would leave him with inconsistent intentions. To avoid this one must look for more subtle intention revision policies.

In what follows I use the following simple policy, which is designed to preserve intention consistency while avoiding the introduction of more technical machinery.

Fig. 2 The intention restriction when $\iota_i(w) \cap ||\phi|| \neq \emptyset$



This policy divides into two cases: the consistent case, when what is announced is consistent with one's intentions; and the revision case, when the announcement contradicts one's intentions. In the consistent case I shall assume that the agent retains in his intentions only those states that satisfy the hard information he just received. In the revision case I shall assume that the agent throws away the old, unachievable intentions but refrains from committing to anything other than what he already knows to be the case.

Definition 8 (*Contracted models*) Given a model $\mathbb{M}$ and a formula $\phi \in D\mathcal{L}_{\mathcal{GF}}$, the contracted model $\mathbb{M}_{|\phi}$ is defined as follows.

1. $W_{|\phi} = ||\phi||_{\mathbb{M}}$.
2. The relation $\sim_{i\ |\phi}$ is the restriction of $\sim_i$ to $W_{|\phi}$.
3. For all $w \in W_{|\phi}$. The set $\iota_{i|\phi}(w)$ is defined as $||\phi|| \cap \iota_i(w)$ if $||\phi|| \cap \iota(w) \neq \emptyset$, and as $W_{|\phi}$ otherwise.
4. $V_{|\phi}$ is the restriction of V to $W_{|\phi}$.

The domain $W_{|\phi}$ of a model restricted to ϕ is just what one would expect: the set of states where ϕ was the case before the announcement, with the epistemic relations modified accordingly.

The restriction of the intention relation ι_i corresponds to the revision policy just described. On the one hand, if what is announced was compatible with the agent's intention, that is if $||\phi|| \cap \iota_i(w) \neq \emptyset$, then the agent adapts his intention to the announcement. Formally, the new intention set is built by taking the restriction of the agent's intention to the states compatible with the formula announced: $\iota_{i|\phi}(w) = ||\phi|| \cap \iota_i(w)$. This case is illustrated in Fig. 2. In the revision case, where the announcement is *not* compatible with what the agent intends, that is when $\iota_i(w) \cap ||\phi|| = \emptyset$, the agent's intention revision boils down to his not forming any new specific intentions, which formally gives $\iota_{i|\phi}(w) = W_{|\phi}$. This is illustrated in Fig. 3.

This simple revision policy is a starting point which has the advantage of using only existing resources of epistemic models with intentions. More sophisticated revisions would draw from the extensive formal literature on belief revision, see e.g. the two publications cited above, as well as from the discussions in philosophy of action, e.g. in Bratman (1987). See van der Hoek et al. (2007) for a recent attempt in this direction.

It is important to observe that $V_{|\phi}$ might be undefined for certain nominals, as the state they denote might have been removed by the announcement of ϕ. The model $\mathbb{M}_{|\phi}$,

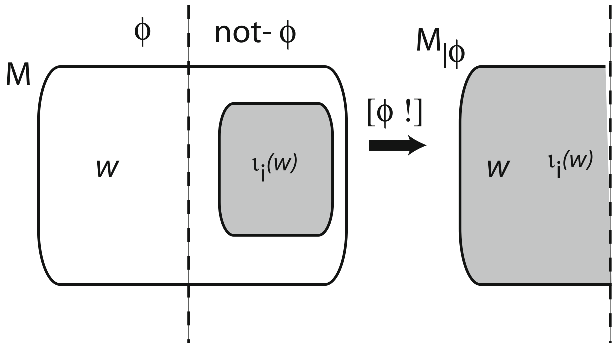


Fig. 3 The intention restriction when $\iota_i(w) \cap ||\phi|| = \emptyset$

in other words, might not fall within the class of models delimited by Definition 1.2.2. This complication is dealt with in the next section by a slight modification of the truth conditions for nominals after public announcements.

1.3.3 Truth and validity in $D\mathcal{L}_{\mathcal{GF}}$

Formulas of $D\mathcal{L}_{\mathcal{GF}}$ are evaluated on epistemic models with intentions precisely like those of $\mathcal{L}_{\mathcal{SG}}$, except for the public announcement formulas, which are evaluated in contracted models.

Definition 9 (*Truth for public announcement formulas*)

$$\mathbb{M}, w \models [\phi!]\psi \quad \textit{iff} \quad \textit{If } \mathbb{M}, w \models \phi \text{ then } \mathbb{M}_{|\phi}, w \models \psi.$$

The condition "If $\mathbb{M}, w \models \phi$ then..." ensures that we are dealing with *truthful* announcements.

As noted above, the truth conditions for nominals need to be revisited in the context of public announcements, as the valuation function V might become partial in $\mathbb{M}_{|\phi}$.

Definition 10 (*Truth for nominals in contracted models*) Given an epistemic model with intentions $\mathbb{M}$, a formula ϕ of $D\mathcal{L}_{\mathcal{GF}}$ and the contracted model $\mathbb{M}_{|\phi}$:

$$\mathbb{M}_{|\phi}, w \models \sigma \quad \textit{iff} \quad V_{|\phi}(\sigma) \text{ is well-defined and } w \in V_{|\phi}(\sigma).$$

Validity with respect to (classes of) frames and models is defined precisely as for $\mathcal{L}_{\mathcal{SG}}$.

1.3.4 Axiomatization

The logic $\Lambda_{D\mathcal{L}_{\mathcal{GF}}}$ is generated by all axioms and inference rules of $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$, together with the "reduction" formulas (van Ditmarsch et al. 2007, p. 88) of Table 4.

These formulas show how to decompose post-announcement conditions in terms of pre-announcement ones. Formulas 1–4 are known from the dynamic-epistemic logic literature.

Table 4 The reduction axioms for $\Lambda_{D\mathcal{L}_{\mathcal{GF}}}$. Here $[\cdot]$ is either A or K_i

1. $[\phi!]x \leftrightarrow \phi \rightarrow x$ for $x \in \text{PROP} \cup S$.
2. $[\phi!]\neg\psi \leftrightarrow \phi \rightarrow \neg[\phi!]\psi$.
3. $[\phi!]\psi \wedge \xi \leftrightarrow \phi \rightarrow ([\phi!]\psi \wedge [\phi!]\xi)$.
4. $[\phi!][\cdot]\psi \leftrightarrow \phi \rightarrow [\cdot](\phi \rightarrow [\phi!]\psi)$
5. $[\phi!]I_i\psi \leftrightarrow \phi \rightarrow (\mathsf{i}_i\phi \wedge I_i(\phi \rightarrow [\phi!]\psi) \vee (\neg\mathsf{i}_i\phi \wedge [\phi!]A\psi))$

Clause 1 is of special importance for the completeness result below. It shows that any formula whose evaluation potentially involves moving outside the class of epistemic model with intentions, that is to move to models where the valuation function is undefined for some nominals, is equivalent to a formula of $\mathcal{L}_{\mathcal{SG}}$, whose evaluation is done completely within the intended class of models. In other words, by the completeness result of Sect. 1.2.3, any formula satisfiable in a contracted model is equivalent to a formula of $\mathcal{L}_{\mathcal{SG}}$, satisfiable in an epistemic model with intentions. Note that this is crucially due to the fact that I only allow the valuation function to become undefined for some nominals *after* a public announcement. A fully general treatment, involving undefined or "partially denoting" nominals even before model contraction, requires some modification of the axiomatic system. See Hansen (2009) for details.

Formula 4 covers the epistemic and the global modalities, and formula 5 encodes the intention revision policy just discussed. Not surprisingly, the latter matches the two cases of the update rule for ι_i. If the intentions of i were already compatible with the announcement of ϕ at state w in a model $\mathbb{M}$, that is if $\mathbb{M}, w \models \mathsf{i}_i\phi$, then all ϕ-states compatible with i's intentions should satisfy ψ after the announcement, which is essentially what $I_i(\phi \rightarrow [\phi!]\psi)$ states. On the other hand, if the announcement of ϕ was not compatible with i's intentions, i.e. if $\mathbb{M}, w \models \neg\mathsf{i}_i\phi$, then i intends that ψ after the announcement if and only if ψ is true everywhere in the restricted model, i.e. $[\phi!]A\psi$, which is exactly what the intention revision rule for ι_i prescribes.

The argument in Appendix 4.3 shows that the formulas of Table 4 are valid with respect to the class of epistemic models with intentions and their transformation via public announcements, which is enough to show the following:

Theorem 2 *The logic $\Lambda_{D\mathcal{L}_{\mathcal{GF}}}$ is complete with respect to the class of epistemic game models with intentions.*

2 Applications

In this section I apply $D\mathcal{L}_{\mathcal{GF}}$ and its associated logic $\Lambda_{D\mathcal{L}_{\mathcal{GF}}}$ to generalize the investigation of van Hees and Roy (2008) and Roy (2009) concerning the role of intentions in the transformation of decision problems. More precisely, I answer the four open questions mentioned in the introduction: provide an epistemic characterization of the operation they call "cleaning"; analyze the behaviour of this operation in situations where there is uncertainty about intentions; relate this to other intention-based transformations of interactive situations; and analyze their iterated behaviour.

2.1 Cleaning as public announcement of "Intentions-Rationality"

Van Hees and Roy (2008; 2009) define the cleaning operation as a tool for the formal study of the way intentions impose a "filter of admissibility" on the options in a given decision problem, a phenomenon which has attracted much attention in the philosophical literature since the work of Bratman (1987). The idea behind cleaning is that an agent should discard from a given deliberation all the options which would rule out the achievement of his intentions.

In their framework there is no model of the agents' information and thus no uncertainty about the agents' intentions. Taking the present notation, they define cleaning for a strategic game $\mathbb{G}$ with respect to *one* intention profile ι, where each ι_i is a set of strategy profiles in $\mathbb{G}$. The operation boils down to discarding from each strategy set S_i the elements which could never lead to a profile in the intention set ι_i.

In the models of interactive situations I use here, which also includes the agents' information, assuming that there is no uncertainty about intentions translates into assuming that these intentions are the same throughout a given epistemic frame with intentions, or that they are *uniform*.

Definition 11 (*Uniform intentions*) I say that intentions are uniform in an epistemic frame with intentions $\mathbb{F}$ whenever, for all $w, w' \in W$ and $i \in \mathcal{A}$, $\iota_i(w) = \iota_i(w')$.

Uniform intentions are always common knowledge. When intentions are uniform then at each state all agents not only know the intentions of others, but also know that everybody knows the intentions of others, and so on for any number of iteration of "everybody knows". With this to hand I can readily transpose van Hees and Roy's notion of cleaning into epistemic models with intentions:

Definition 12 (*Cleaning*) For a game in strategic form $\mathbb{G}$ and an epistemic model with intentions $\mathbb{M}$ where intentions are uniform, the cleaned strategy set $cl(S_i)$ from the perspective of ι_i is defined as

$$cl(S_i) = \{s_i \mid \text{there is a } w \in \ \iota_i \text{ such that } w(i) = s_i\}$$

The *cleaned version* of $\mathbb{F}$, denoted $cl_\iota(\mathbb{F})$, is the tuple $\langle cl(W), \{\sim_i^{cl}, \iota_i^{cl}\}_{i \in N}\rangle$ where:

- $cl(W) = \{w \mid \exists i \text{ such that } w(i) \in cl(S_i)\}$.
- $\sim_i^{cl}$ and V^{cl} are restrictions of $\sim_i$ and V to $cl(W)$.
- For all i, $\iota_i^{cl} = cl(W) \cap \iota_i$.

Cleaning is intended to capture the idea that agents remove from their strategy sets those strategies that would never lead to an intended state. Consider again the game in Table 1, with the model as in Fig. 1 and the intentions as follows: $\iota_{Ann}(w) = \{w_1\}$ and $\iota_{Bob}(w) = \{w_1, w_4\}$ for all $w \in W$. A first round of cleaning for Ann removes $Restaurant\ B$ from her strategy set, and by the same token w_3 and w_4. Intentions are uniform here, and so Bob knows that Ann cleans her strategy set this way, and he adapts his own intentions accordingly: $\iota_{Bob}^{cl} = \{w_1\}$. In this reduced model $Restaurant\ B$ becomes incompatible with Bob's intentions, and so one further round of cleaning leaves only w_1, the only state compatible with Ann and Bob's intentions.

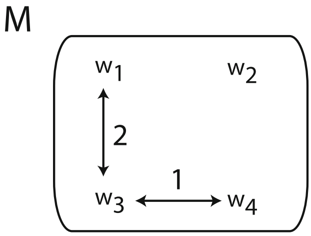


Fig. 4 The model for the proof of Fact 4, based on the game in Table 1. Only the epistemic relations are represented

Table 5 The intentions and strategy choice assignment for Fig. 4

State	$f(w)$	$\iota_1(w)$	$\iota_2(w)$
w_1	$A - A$	w_2, w_4	w_3, w_4
w_2	$A - B$	w_3	w_1
w_3	$B - A$	w_2	w_4
w_4	$B - B$	w_4	w_4

The idea that a strategy might lead to an intended state is also expressible directly in $\mathcal{L}_{\mathcal{SG}}$, and I call it *intention-rationality*:

$$\bigvee_{\sigma(i)=w(i)} \mathsf{i}_i \sigma \quad i\text{'s intention-rationality } (\mathrm{IR}_i) \text{ at state } w$$

Intention-rationality boils down to saying that an agent does not choose a strategy that excludes the achievement of his intentions: the formula above is true at a state w whenever there is a state compatible with i's intentions in which he plays the same strategy as in w. Notice that nominals are crucial in this definition, as they allow one to refer directly to the states and, in turn, to the strategy that the agents choose at these states.

Intention-rationality relates the agent's choice to what he intends, but does not take into account his information at a particular state. In Fig. 4 and Table 5, for instance, Ann is intention-rational at w_2 even though she knows that her intention is not realizable. This is surely a motivation for considering stronger form of intentions-rationality, which I do in the next section.

In general, that w' is compatible with i's intention at w does not mean that i is intention-rational at w', even if in both states this agent plays the same strategy, because his intentions at w' might be different from those at w. In models with uniform intentions, however, such a situation cannot arise: agents are intention-rational at least in the states they intend.

Lemma 1 *The following holds for all agents i in any model $\mathbb{M}$ where intentions are uniform*:

$$\iota_i \subseteq \left\| IR_i \right\|$$

Proof Take any $i \in \mathcal{A}$. We know that $\iota_i \neq \emptyset$. So take any $w \in \iota_i$. We have that $\mathbb{M}, w \models \mathsf{i}_i\sigma$ for $V(\sigma) = w$. But by uniformity of intentions we know that $\iota_i(w) = \iota_i$, and so that $\mathbb{M}, w \models \bigvee_{\sigma(i) = w(i)} \mathsf{i}_i\sigma$.

Corollary 1 *The following holds for all agents i in any model $\mathbb{M}$ where intentions are uniform*:

$$\iota_i \cap \left\| \bigvee_{i \in \mathcal{A}} IR_i \right\| \neq \emptyset$$

Lemma 1 shows that, without uncertainty about intentions, intention-rationality is always ensured somewhere in the model. Corollary 1, in turn, permits a straightforward rendering of cleaning in terms of public announcements of intention-rationality.

Fact 1 For any model $\mathbb{M}$ where intentions are uniform:

$$cl(\mathbb{M}) = \mathbb{M}_{|\bigvee_{i \in \mathcal{A}} IR_i}.$$

Proof We first show that $cl(W) = W_{|\bigvee_{i \in \mathcal{A}} IR_i}$. We know that $w' \in cl(W)$ iff there is an i such that $w'(i) \in cl(S_i)$. This, in turn, happens iff there is an i and a $w'' \in \iota_i$ such that $w''(i) = w'(i)$, which is also the same as saying that there is an i such that $\mathbb{M}, w' \models \bigvee_{\sigma(i)=w'(i)} \mathsf{i}_i\sigma$. This is equivalent to $\mathbb{M}, w' \models \bigvee_{i \in \mathcal{A}} IR_i$, which finally boils down to $w' \in W_{|\bigvee_{i \in \mathcal{A}} IR_i}$. It should then be clear that the restricted cleaned relations and valuation correspond to those obtained from the announcement of $\bigvee_{i \in \mathcal{A}} IR_i$, and vice versa. It remains to consider the update of the intention sets. Observe that by Corollary 1 we know that the second clause of the update rule of ι_i is never used when intentions are uniform, from which we get directly that $\iota_i^{cl} = \iota_{i|\bigvee_{i \in \mathcal{A}} IR_i}$.

The cleaned version of an epistemic model with intentions is thus the same as the contracted model which results from a public announcement of intention-rationality, under the condition that intentions are uniform, and thus common knowledge. Cleaning in the sense of van Hees and Roy corresponds to a public announcement that the players do not choose strategies that rule out the achievement of their intentions.

This answers a first question left open by van Hees and Roy, namely, understanding the epistemic assumptions underlying their cleaning operation: it can be seen as being performed under common knowledge of intention. This is indeed a very strong condition on what the agents know about each others' intentions. It presupposes that all agents are absolutely certain not only about their own intentions and the intentions of others, but also about what the others know about their intentions. Such hard information about intentions and information will, however, rarely obtain in interactive situations, and in the next sections I relax this assumption, which will at the same time provide an answer to two other questions left open by van Hees and Roy, namely how cleaning behaves when there is uncertainty about intentions, and how it is related to other intentions-based transformations of option sets.

2.2 Intention-rationality and "knowledge-consistency"

The first step towards generalizing van Hees and Roy's approach to situations where intentions are not common knowledge is to observe that the notion of intention-rationality defined above takes no account of what the agent considers possible at a given state. As observed earlier, one can easily design a state in a model for a given game in strategic form where one agent is intention-rational even though he *knows* that he cannot realize what he intends.

The following notion, which I call *epistemic* intention-rationality, parallels the one defined in the previous section but takes special care of the agent's information at a state.

$$\bigvee_{\sigma(i) = w(i)} (\mathsf{i}_i\sigma \wedge \Diamond_i\sigma) \quad i\text{'s epistemic intention-rationality } (\mathrm{IR}_i^*) \text{ at state } w$$

An agent is epistemically intention-rational at a state w whenever there is a state compatible with his intentions which he considers possible and where he plays the same strategy as in w. That this formula connects the agents' intentions and information is even more explicit once one considers the following formula, which states that an agent should not have intentions which he considers are impossible to realize.

$$I_i\phi \rightarrow \Diamond_i\phi \quad i\text{'s knowledge-consistency of intentions } (\mathrm{IK}_i)$$

This notion recalls the "strong belief consistency" constraint discussed in the philosophical literature Bratman (1987, p. 31). In contrapositive, it says that an agent's intentions should always be at least compatible with his hard information at a state, i.e. with what he knows to be the case. Knowledge-consistency and epistemic intention-rationality turn out to be semantically equivalent.

Fact 2 At any pointed model $\mathbb{M}, w$, $\mathbb{M}, w \models IK_i$ iff $\mathbb{M}, w \models IR_i^*$.

Proof It should be clear that $IK_i \rightarrow IR_i$, which means that all that remains to be shown is the right-to-left direction. We show it in contrapositive. Assume that $\mathbb{M}, w \models I_i\phi \wedge K_i\neg\phi$ for some ϕ. This means that $\iota_i(w) \subseteq ||\phi||$ and that $||\phi|| \cap [w]_i = \emptyset$. We thus have $\iota_i(w) \cap [w]_i = \emptyset$. But this means that for all σ' and $w'(i) = w(i)$ such that $V(\sigma') = w'$ and $w' \in [w]_i$, $\mathbb{M}, w \models \neg\mathsf{i}_i\sigma'$. In other words, $\mathbb{M}, w \models \bigwedge_{\sigma'(i)=s_i} \Diamond_i\sigma' \rightarrow \neg\mathsf{i}_i\sigma'$.

Knowledge-consistency clearly implies intention-rationality, but not the other way around. Look for example at the state w_1 in Fig. 4. Ann is intention-rational but her intentions are knowledge-inconsistent at that state.

Intention-rationality and its epistemic variant are both fully introspective in epistemic game models with intentions, essentially because agents know their own intentions and strategy choices. This, again, reveals how intentions and information interact in such models, which becomes crucial in the analysis of the iterated announcements of these two notions.

Fact 3 (*Positive and negative introspection of IR_i and IR_i^**) For all pointed game models, $\mathbb{M}, w \models IR_i$ implies $\mathbb{M}, w \models K_i IR_i$ and $\mathbb{M}, w \models \neg IR_i$ implies $\mathbb{M}, w \models K_i \neg IR_i$. The same hold for IR_i^*.

Proof We only show the positive introspection for IR_i; the arguments for the other claims are similar. Assume that $\mathbb{M}, w \models IR_i$. This happens if and only if there is a $w' \in \iota_i(w)$ such that $w'(i) = w(i)$. Take any $w'' \in [w]_i$. We know that $\iota_i(w'') = \iota_i(w)$, which means that w' is also in $\iota_i(w'')$, and $w''(i) = w(i)$, which means that $w''(i) = w'(i)$, and so that $\mathbb{M}, w'' \models IR_i$.

2.3 Iterated announcements of intention-rationality and knowledge-consistency

By showing that cleaning corresponds to a specific public announcement in $DL_{\mathcal{GF}}$, I have in fact opened the door to using the expressive power of this language to analyze a whole range of alternative intentions-based transformations of interactive situations, i.e. alternative public announcements related to what agents intend and know in a specific interactive situation. In this section I focus on an obvious candidate: announcement of knowledge consistency. I compare it to cleaning, in both the one-shot and iterated cases, and both with and without uncertainty about intentions, thus answering the three remaining open questions mentioned at the beginning of this section.

A first thing to observe about the announcements of intention-rationality, i.e. cleaning, and of knowledge consistency, which I call *epistemic* cleaning, is that they are both *self-fulfilling* when intentions are common knowledge. Under this assumption, intention-rationality and knowledge consistency hold everywhere in the model that results from a single public announcement of one or the other. In the general case, however, this is only true of knowledge-consistency.

Definition 13 (*Self-fulfilling announcements*) An announcement that ϕ is said to be *self-fulfilling* at a pointed model $\mathbb{M}, w$ if $\mathbb{M}, w \models [\phi!]A\phi$.

Fact 4 The announcement of $\bigvee_{i \in \mathcal{A}} IR_i$ is self-fulfilling for any pointed model $\mathbb{M}, w$ when intentions are common knowledge, but this not the case in general.

Proof For the first part, we have to show that for any pointed model with common knowledge of intentions, if $\mathbb{M}, w \models \bigvee_{i \in I} IR_i$ then $\mathbb{M}_{|\bigvee_{i \in \mathcal{A}} IR_i}, w \models A \bigvee_{i \in \mathcal{A}} IR_i$. We show something stronger, namely that for all $w \in ||IR_i||$, $\mathbb{M}_{|IR_i}, w \models A\ IR_i$. Take any $w' \in W_{|IR_i}$. We have to show that there is a w'' in $\iota_{i|IR_i}(w')$ such that $w''(i) = w'(i)$. Because $w' \in W_{|IR_i}$ we know that w' was in $||IR_i||$ before the announcement. But this means that there was a $w'' \in \iota_i(w')$ such that $w''(i) = w'(i)$. But since $\iota_i(w)$ is the same for all $w \in W$, by assumption, we know that w'' was also in $||IR_i||$. This means, in turn, that $\iota_{i|IR_i}(w') = \iota_{i|IR_i}(w') \cap ||IR_i||$, and so that $w'' \in \iota_{i|IR_i}(w')$, as required.

For the second part, take the set of states in Fig. 4 with the intentions and strategy choice as in Table 5. The announcement of $\bigvee_{i \in \mathcal{A}} IR_i$ removes w_2 and w_3, making both agents intention-irrational at w_1.

Fact 5 The announcement of $\bigvee_{i\in\mathcal{A}} IR_i^*$ is self-fulfilling for any pointed model $\mathbb{M}, w$.

Proof The proof follows the same line as in the previous fact. Namely, we show that if $\mathbb{M}, w \models IR_i^*$ then $\mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR_i^*}, w \models A\ IR_i^*$. The reasoning is entirely similar. Take any $w' \in W_{|IR^*}$. We know that $\iota_i(w') \cap [w']_i \neq \emptyset$. Now take any w'' in this intersection. Because $w'' \sim_i w'$, we know that $\iota_i(w'') = \iota_i(w')$ and that $w''(i) = w'(i)$. Furthermore, because $\sim_i$ is an equivalence relation we know that $[w'']_i = [w']_i$. This means that $w'' \in ||IR_i^*||$. This gives us that $\iota_{i|IR_i^*}(w') = \iota_i(w') \cap ||IR_i^*||$ and also that $w'' \in \iota_{i|IR_i^*}(w') \cap [w']_{i|IR_i^*}$, as required.

These two facts show that epistemic cleaning does indeed compensate for uncertainty about intentions, something that is beyond the reach of non-epistemic cleaning. One announcement of intention-rationality can make some agents intention-irrational when there is uncertainty about what the others intend, but this cannot happen when knowledge-consistency is announced.

Announcing intention-rationality is self-fulfilling if it is repeated often enough, though, and this is a more or less direct consequence of the simple intention-revision policy that I defined above. To see this requires two prefatory facts: one which states that the intention revision policy indeed always preserves consistency, and the other that repeated announcement of intention-consistency stabilizes after a finite number of steps.

Fact 6 $\mathbb{M} \models \bigwedge_{i\in\mathcal{A}}[\phi!]\mathsf{i}_i\top$ for all models for game structure $\mathbb{M}$.

Proof We show the validity of $\bigwedge_{i\in\mathcal{A}}[\phi!]\mathsf{i}_i\top$ via a syntactic derivation and Theorem 2. See Appendix 4.4 for details.

Definition 14 (*Announcement stabilization*) Given a pointed game model $\mathbb{M}, w$, let $\mathbb{M}^k_{|\phi}, w$ be the pointed model that results after announcing k times ϕ at w. The announcement of ϕ *stabilizes* at k for $\mathbb{M}, w$ whenever $\mathbb{M}^k_{|\phi}, w = \mathbb{M}^{k+1}_{|\phi}, w$.

The announcement of intention-rationality does indeed stabilize, in this technical sense. The details of the proof of this fact are given in Appendix 4.5.

Fact 7 (*Stabilization of* $[\bigvee_{i\in\mathcal{A}} IR_i!]$) For any pointed model $\mathbb{M}, w$, the announcement of $\bigvee_{i\in\mathcal{A}} IR_i$ stabilizes at some k.

Putting these two together, we directly obtain the following.

Corollary 2 *At any k where the announcement of intention-rationality stabilizes for a given pointed model $\mathbb{M}, w$:*

$$\mathbb{M}^k_{|\bigvee_{j\in\mathcal{A}} IR_j}, w \models \bigwedge_{i\in\mathcal{A}} \mathsf{i}_i\top$$

With this in hand, it is straightforward to get the following.

Fact 8 (*Self-fulfilling of* $[\bigvee_{i\in\mathcal{A}} IR_i!]$ *at the stabilization point*) At any k where $[\bigvee_{i\in\mathcal{A}} IR_i!]$ stabilizes, $\mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i}, w \models A \bigvee_{i\in\mathcal{A}} IR_i$.

Table 6 The intentions of the agents in counterexample for Fact 10

	$\iota_1(w)$	$\iota_2(w)$
w_1	w_2	w_1
w_2	w_1	w_4
w_3	w_3	w_3
w_4	w_4	w_4

Proof Assume the contrary, then there is a $w' \in W^k_{|\bigvee_{i\in\mathcal{A}} IR_i}$ such that $\mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i}, w \models \neg \bigvee_{i\in\mathcal{A}} IR_i$. But then $w \notin \mathbb{M}^{k+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$, against the assumption that the announcement of $\bigvee_{i\in\mathcal{A}} IR_i$ stabilizes at k.

This means that, even though non-epistemic cleaning is not necessarily self-fulfilling after one announcement, it is in the long run. The route to a stable contracted epistemic model with intentions is much quicker with epistemic cleaning, though.

Fact 9 For any pointed model $\mathbb{M}, w$, the announcement of $\bigvee_{i\in\mathcal{A}} IR^*_i$ stabilizes after one announcement.

Proof By definition, $W_{|\bigvee_{i\in\mathcal{A}} IR^*_i} = ||\bigvee_{i\in\mathcal{A}} IR^*_i||$. But we also know from Fact 5 that for all w' in $W_{|\bigvee_{i\in\mathcal{A}} IR^*_i}$, $\mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR^*_i}, w' \models \bigvee_{i\in\mathcal{A}} IR^*_i$. This means that $\mathbb{M}^2_{|\bigvee_{i\in\mathcal{A}} IR^*_i} = \mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR^*_i}$.

Moreover, as the example in the proof of Fact 4 suggests, these stabilization points can be different.

Fact 10 (*Fixed points divergence*) There exist models $\mathbb{M}$ where the announcement of intention-rationality stabilizes at k such that:

$$\mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR^*_i} \neq \mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i}$$

Proof Take a model $\mathbb{M}$ with two agents and four states, w_1 to w_4, where the strategies are as in Fact 4, and $[w]_i = \{w\}$ for all states. Fix the intentions as in Table 6. It should be clear that in all states, $\mathbb{M}, w \models \bigvee_{i\in\mathcal{A}} IR_i$. This means that for all states, $\mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR_i}, w = \mathbb{M}, w$, i.e. this announcement does not remove any states, and so that $\mathbb{M}$ is its own stabilization point. Observe, on the other hand, that $\mathbb{M}, w_2 \not\models \bigvee_{i\in\mathcal{A}} IR^*_i$. But since $\iota_1(w_1) = \{w_2\}$, we get $\iota_{1,|\bigvee_{i\in\mathcal{A}} IR^*_i}(w_1) = \{w_1, w_3, w_4\}$ after the announcement of knowledge-consistency at w_1. But then it is clear that $\iota_{1,|\bigvee_{i\in\mathcal{A}} IR^*_i}(w_1) \neq \iota_{1,|\bigvee_{i\in\mathcal{A}} IR_i}(w_1)$, and since in this case the announcement of $\bigvee_{i\in\mathcal{A}} IR_i$ "stabilizes" at $k = 0$, we get that $\mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR^*_i} \neq \mathbb{M}^k_{\bigvee_{i\in\mathcal{A}} IR_i}$

This last result shows that the iterated behaviour of the two forms of cleaning does indeed differ, as the intention revision policy preserves consistency of intentions, but it sometimes forces agents to adjust their intentions in the face of epistemic cleaning in a way that would not have been necessary for non-epistemic cleaning. Observe that this would also be the case if I had used a more sophisticated revision policy. The different behaviour of non-epistemic and epistemic cleaning comes from the fact that in this case the latter, and not the former, forces the agents to revise their intentions, and not from the way they proceed with this revision.

This difference between the iterated behaviour of epistemic and non-epistemic cleaning is, however, the only one that can occur: knowledge-consistency is robust to any number of altruistic cleanings. This is a direct consequence of the fact that knowledge-consistency implies intention-rationality, and that they are both introspective.

Fact 11 For all pointed models $\mathbb{M}, w$, if $\mathbb{M}, w \models \bigvee_{i\in\mathcal{A}} IR_i^*$ then $\mathbb{M}, w \models [\bigvee_{i\in\mathcal{A}} IR_i!] \bigvee_{i\in\mathcal{A}} IR_i^*$.

Proof Assume that $\mathbb{M}, w \models \bigvee_{i\in\mathcal{A}} IR_i^*$, i.e. that there is an i and a $w' \sim_i w$ such that $w' \in \iota_i(w)$. Because IR_i^* is introspective, this means that $\mathbb{M}, w' \models \bigvee_{i\in\mathcal{A}} IR_i^*$. But then $\mathbb{M}, w' \models \bigvee_{i\in\mathcal{A}} IR_i$, which means that both w' and w are in $W_{|\bigvee_{i\in\mathcal{A}} IR_i}$, and also that $w' \in \iota_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)$. But then $\mathbb{M}, w \models [\bigvee_{i\in\mathcal{A}} IR_i!] \bigvee_{i\in\mathcal{A}} IR_i^*$.

From this we also know that the model that would result from epistemic cleaning is always a sub-model of the one which would result at the stabilization point of the non-epistemic variant of this operation.

Corollary 3 *Suppose that for a pointed model* $\mathbb{M}, w$ *the announcement of intention-rationality stabilizes at k, then*

$$\Big\| \bigvee_{i\in\mathcal{A}} IR_i^* \Big\|_{|\bigvee_{i\in\mathcal{A}} IR_i^*} \subseteq \Big\| \bigvee_{i\in\mathcal{A}} IR_i \Big\|^k_{|\bigvee_{i\in\mathcal{A}} IR_i}$$

Proof Follows directly from Fact 8, 9 and 11.

Let us wrap up the results from this section. I have answered four questions left open by van Hees and Roy (2008) and by Roy (2009). I first showed that cleaning corresponds to a public announcement of "intention-rationality" under common knowledge of intentions. Granting that this is quite an strong assumption, I moved on to an analysis of public announcements of intention-rationality in situations where agents are uncertain about the intentions of others, and compared it with the announcement of "knowledge-consistency" of intentions. I showed that only this second type of announcement is "self-fulfilling" when there is uncertainty about intentions, but that they both are otherwise. I then analyzed the iterated behaviour of these operations, and showed that they both reach a stabilization point, and studied how these relate. This allowed us finally to look at the interplay between these two operations, which revealed that knowledge-consistency is indeed "robust" to announcements of intention-rationality.

3 Conclusion

In this paper I have proposed a logical system to describe and study reasoning about information and intentions in interactive situations. I have completely axiomatized this system and used it to further the work on intention-based transformation of decision problem initiated by van Hees and Roy (2008) and Roy (2009), by answering four questions left open by their approach. All in all, I have provided a framework that merges insights from and contributes to three strong traditions in the study of rational agency: dynamic epistemic logic, epistemic game theory and philosophy of action.

Three issues need most urgent investigation: the connection between intentions and preferences, the relation between the present framework and the BDI (Belief-Desire-Intention) architectures for multi-agent systems, e.g. Cohen and Levesque (1990); Georgeff et al. (1999) and Wooldridge (2000), and extensions to more subtle policies of intention revision. In this paper I have completely left aside the question of how intentions should relate to preferences, and whether the dynamic processes that I studied could, or maybe should, be reduced to known preference-based transformation of games in strategic form, for instance iterated elimination of strictly dominated strategies. This is an important question, which would help to understand how the approach studied here relates to mainstream game theory. Although very similar in method and aims, the BDI models have not been developed for direct application to games in strategic forms, which makes it at the least non-trivial to see how they relate to the logical system presented here. The investigation into more subtle intention revision policies will require the extension of the already quite loaded framework with even more technical machinery, and this is the main reason why I have not taken up this task here. This paper provides a good understanding of the basic dynamics of epistemic models with intentions, though, and with this to hand I believe it would be timely to move to such richer models.

4 Appendix

4.1 Model-theoretic definition of common knowledge

Given a group of agents $G \subseteq \mathcal{A}$ and a model $\mathbb{M}$, let $\sim^*_G$ denote the reflexive-transitive closure of the union of the relations $\sim_i$ for all $i \in G$. Denote $[w]^*_G$ the set of state w' such that $w \sim^*_G w'$. The formula ϕ is *common knowledge among* G at state w whenever $\mathbb{M}, w' \models \phi$ for all $w' \in [w]^*_G$.

This definition is well known to correspond, at the syntactic level, to the infinite conjunction of “everybody knows that...” mentioned in Sect. 1.2.1. See e.g. (Fagin et al. 1995). The reader can also check that the notion of uniform intentions defined in Sect. 2.1 does indeed imply that intentions are common knowledge in the present sense.

4.2 Proof of theorem 1

Definition 15 (*Named and pasted MCS*) Let Γ be a maximally consistent set (MCS) of $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$. We say that Γ is *named* by σ if $\sigma \in \Gamma$. If σ names some MSC(s) Γ we denote it (them) Γ_σ. Γ is *pasted* whenever $E(\sigma \wedge <> \phi) \in \Gamma$ implies that $E(\sigma \wedge <> \sigma') \wedge E(\sigma' \wedge \phi)$ is also in Γ.

Lemma 2 (Extended Lindenbaum lemma) (Blackburn et al. 2001, *p. 441*) *Let S' be a countable collection of nominals disjoint from S, and let $\mathcal{L}_{\mathcal{SG}}{}'$ be $\mathcal{L}_{\mathcal{SG}} \cup S'$. Then every $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$ consistent set of formulas can be extended to a named and pasted $\Lambda_{\mathcal{L}_{\mathcal{SG}}{}'}$-MCS.*

Proof **Naming** Enumerate S', and let σ be the first new nominal in that enumeration. For a given consistent set Γ^*, fix $\Gamma_\sigma = \Gamma \cup \{\sigma\}$. By (Name) Γ_σ is consistent. **Pasting** Enumerate the formulas of $\mathcal{L}_{\mathcal{SG}}{}'$ and take $\Gamma_0 = \Gamma_\sigma$. Assume Γ_n is defined, and let ϕ_{n+1} be the $n^{th} + 1$ formula in the enumeration. Define Γ_{n+1} as Γ_n if $\Gamma_n \cup \{\phi_{n+1}\}$ is inconsistent. Otherwise form Γ_{n+1} by adding ϕ_{n+1} to Γ_n if ϕ_{n+1} is not of the form $E(\sigma' \wedge \psi)$. If ϕ_{n+1} is of form $E(\sigma' \wedge \psi)$, then we paste with the first new nominal σ'' in the enumeration of S'. I.e. $\Gamma_{n+1} = \Gamma_n \cup \{\phi_{n+1}\} \cup \{E(\sigma' \wedge <> \sigma'') \wedge E(\sigma'' \wedge \phi)\}$. By (Paste), Γ_{n+1} is also consistent. Finally, set $\Gamma = \bigcup_{n \leq \omega} \Gamma_n$. This is clearly a named and pasted MCS.

Definition 16 (*Yielded MCS*) The sets *yielded* by a $\Lambda_{\mathcal{L}_{\mathcal{SG}}{}'}$-MCS Γ are the sets Δ_σ such that $\Delta_\sigma = \{\phi : E(\sigma \wedge \phi) \in \Gamma\}$.

Lemma 3 (Properties of yielded sets) (Blackburn et al. 2001, *p. 439*) *Let Δ_σ and $\Delta_{\sigma'}$ be any yielded sets of a $\Lambda_{\mathcal{L}_{\mathcal{SG}}{}'}$-MCS Γ, for arbitrary nominals σ and σ' in $\mathcal{L}_{\mathcal{SG}}{}'$.*

1. *Both Δ_σ and $\Delta_{\sigma'}$ are named $\Lambda'_{\mathcal{L}_{\mathcal{SG}}}$-MCS.*
2. *If $\sigma' \in \Delta_\sigma$ then $\Delta_\sigma = \Delta_{\sigma'}$.*
3. *$E(\sigma \wedge \phi) \in \Delta_{\sigma'}$ iff $E(\sigma \wedge \phi) \in \Gamma$.*
4. *If σ'' names Γ then Γ is itself the yielded set $\Delta_{\sigma''}$.*

Proof

1. By (Exists$_\sigma$), $E\sigma \in \Gamma$, and thus Δ_σ is named. Assume now it is not consistent. This means that there are $\xi_1 \wedge \cdots \wedge \xi_n \in \Delta_\sigma$ such that one can prove $\neg(\xi_1 \wedge \cdots \wedge \xi_n)$ in $\Lambda_{\mathcal{L}_{\mathcal{SG}}}$. But this means that $A\neg(\xi_1 \wedge \cdots \wedge \xi_n) \in \Gamma$, by (Nec). This, in turns, means that $\neg E(\xi_1 \wedge \cdots \wedge \xi_n) \in \Gamma$. But that cannot hold. Recall that $(\xi_1 \wedge \cdots \wedge \xi_n) \in \Delta_\Gamma$ iff $E(\sigma \wedge \xi_1 \wedge \cdots \wedge \xi_n)$ is also in Γ. But then by (K) for E, we get that $E(\xi_1 \wedge \cdots \wedge \xi_n) \in \Gamma$. For maximality, observe that a formula ϕ and its negation are not in Δ_σ iff neither $E(\sigma \wedge \phi)$ nor $E(\sigma \wedge \neg\phi)$ are in Γ. But because the latter is a MCS, this means that both $\neg E(\sigma \wedge \phi)$ and $\neg E(\sigma \wedge \neg\phi)$ are in Γ. The first formula implies $A(\sigma \to \neg\phi) \in \Gamma$, but then, given that $E\sigma \in \Gamma$, by standard modal logic reasoning we get that $E(\sigma \wedge \neg\phi)$, contradicting the consistency of Γ.
2. Assume $\sigma' \in \Delta_\sigma$. This means that $E(\sigma \wedge \sigma') \in \Gamma$. By (Inc$_{E-\sigma}$) we get that both $A(\sigma \to \sigma')$ and $A(\sigma' \to \sigma)$ are in Γ, and so by K for E, we get $A(\sigma \leftrightarrow \sigma') \in \Gamma$. Assume now that $\phi \in \Delta_\sigma$. This means that $E(\sigma \wedge \phi) \in \Gamma$. But by standard K reasoning we get that $E(\sigma' \wedge \phi) \in \Gamma$, which means that ϕ is also in $\Delta_{\sigma'}$. The argument is symmetric for $\phi \in \Delta_{\sigma'}$, and so $\Delta_\sigma = \Delta_{\sigma'}$.
3. We first show the left-to-right direction. Assume that $E(\sigma' \wedge \phi) \in \Delta_\sigma$. This means that $E(\sigma \wedge E(\sigma' \wedge \phi)) \in \Gamma$. But then this implies, by K for E that $EE(\sigma' \wedge \phi) \in \Gamma$, which in turn, because of axiom 4 for E, implies $E(\sigma' \wedge \phi) \in \Gamma$. For the converse, assume that $E(\sigma' \wedge \phi) \in \Gamma$. By axiom 5 for E, we get that $AE(\sigma' \wedge \phi) \in \Gamma$. But we also know by (Exists$_\sigma$) that $E\sigma \in \Gamma$, from which

we get by standard K reasoning that $E(\sigma \wedge E(\sigma' \wedge \phi)) \in \Gamma$. This means that $E(\sigma' \wedge \phi) \in \Delta_\sigma$.

4. Assume that $\sigma \in \Gamma$. For left to right, assume that $\phi \in \Gamma$. This means that $\sigma \wedge \phi \in \Gamma$, which implies by axiom T that $E(\sigma \wedge \phi)$ and so that $\phi \in \Delta_\sigma$. Now assume that $\phi \in \Delta_\sigma$. This means that $E(\sigma \wedge \phi) \in \Gamma$, which in turn implies that $A(\sigma \rightarrow \phi)$ by ($\text{Inc}_{E-\sigma}$). But then by axiom T again we get that $\sigma \rightarrow \phi \in \Gamma$, and ϕ itself because $\sigma \in \Gamma$.

Definition 17 (*Epistemic model for completeness*) Let Γ be any named and pasted $\Lambda_{\mathcal{L}_{\mathcal{SG}}'}$-MCS. The named game model $\mathbb{M}^\Gamma$ yielded by Γ is a tuple $\langle W^\Gamma, \mathcal{A}, \sim_i^\Gamma, \iota_i^\Gamma, V^\Gamma \rangle$ such that:

– W^Γ is the set of sets yielded by Γ.
– $\mathcal{A}$, defined as $\{i : \text{there is a} <>_i \phi \text{ in } \mathcal{L}_{\mathcal{SG}}\}$, is the set of agents.
– $\Delta_\sigma \sim_i^\Gamma \Delta_{\sigma'}$ iff for all $\phi \in \Delta_{\sigma'}$, $\Diamond_i \phi \in \Delta_\sigma$.
– $\iota_i^\Gamma(\Delta_\sigma) = \{\Delta_{\sigma'} : \text{for all } \phi \in \Delta_{\sigma'}, \text{i}_i\phi \in \Delta_\sigma\}$.
– For all $x \in \text{PROP} \cup (S \cup S')$, $V^\Gamma(x) = \{\Delta_\sigma : x \in \Delta_\sigma\}$.

Lemma 4 (Existence Lemma for $E\phi$, I_i and K_i) *If $\Diamond_i \phi \in \Delta_\sigma$ then there is a $\Delta_{\sigma'} \in W$ such that $\phi \in \Delta_{\sigma'}$ and $\Delta_\sigma \sim_i^\Gamma \Delta_{\sigma'}$. Similarly for $I_i\phi$ and $E\phi$. Furthermore, if $\phi \in \Delta_\sigma$ then for all $\Delta_{\sigma'}$, $E\phi \in \Delta_{\sigma'}$.*

Proof (Blackburn et al. 2001, p. 442) for K_i and I_i. The argument for $E\phi$, including the "furthermore" part, is a direct application of Lemma 3.

Lemma 5 (Truth Lemma) *For all $\phi \in \Gamma$, $\mathbb{M}^\Gamma, \Delta_\sigma \models \phi$ iff $\phi \in \Delta_\sigma$.*

Proof By induction on ϕ. The basic cases, including the nominals, are obvious. The inductive cases is a standard modal logic argument from Lemma 4.

All that remains to be shown is that $\mathbb{M}^\Gamma$ is indeed an epistemic model with intentions. We start by looking at the epistemic relation $\sim_i^\Gamma$.

Lemma 6 (Adequacy of $\sim_i^\Gamma$ - Part I) *The relation $\sim_i^\Gamma$ is an equivalence relation.*

Proof All $S5$ axioms are canonical (Blackburn et al. 2001, p. 203).

This means that $\{[\Delta_\sigma]_i : \Delta_\sigma \in W^\Gamma\}$ partitions the set W^Γ, for each agent. We use these partitions directly to define the strategic choices at each Δ_σ. That is, for each "state" Δ_σ, we set $f(\Delta_\sigma)(i) = [\Delta'_\sigma]_i$ such that $\Delta_\sigma \in [\Delta'_\sigma]_i$. By the previous lemma we automatically get that this function is well-defined. The rest of the adequacy lemma for $\sim_i^\Gamma$ is then easy.

Lemma 7 (Adequacy of $\sim_i^\Gamma$ - Part II) *For all Δ_σ and Δ'_σ, if $\Delta_\sigma \sim_i^\Gamma \Delta'_\sigma$ then $\Delta_\sigma(i) = \Delta_{\sigma'}(i)$ and $\iota_i^\Gamma(\Delta_\sigma) = \iota_i^\Gamma(\Delta'_\sigma)$.*

Proof The first part is a trivial consequence of the way we set up $\Delta_\sigma(i)$. For the second part, observe that by the definition of ι_i^Γ all we need to show is that for all $|\phi| \in \iota_i^\Gamma(\Delta_\sigma)$, $|\phi|$ is also in $\iota_i^\Gamma(\Delta_{\sigma'})$, with $|\phi| = \{\Delta_\sigma \in W^\Gamma : \phi \in \Delta_\sigma\}$. So assume the first. This means that $I_i\phi \in \Delta_\sigma$, which means by (K-I) that $K_i I_i \phi$ is also in Δ_σ. But then, because $\Delta_\sigma \sim_i^\Gamma \Delta_{\sigma'}$, we obtain by a routine modal logic argument that $I_i\phi \in \Delta_{\sigma'}$, which is just to say, $|\phi|$ is also in $\iota_i^\Gamma(\Delta_{\sigma'})$.

Lemma 8 (Adequacy of ι_i^{Γ}) *For all Δ_σ, $\iota_i^{\Gamma}(\Delta_\sigma)$ does not contain the empty set.*

Proof Follows directly from the canonicity of (Ser).

4.3 Proof of Theorem 2

We define $\Lambda_{\mathcal{DLGF}}$ as $\Lambda_{\mathcal{LSG}}$ together with the formulas in Table 4. Showing completeness boils down to showing soundness for these new axioms (van Ditmarsch et al. 2007).

Proof Soundness of the first four axioms is well known, except for the nominal case. For this case take an arbitrary pointed model $\mathbb{M}, w$, and assume that $\mathbb{M}, w \models [!\phi]\sigma$. This happens iff $\mathbb{M}_{|\phi}, w \models \sigma$ whenever $\mathbb{M}, w \models \phi$. Now if the antecedent holds, then $w \in W_{|\phi}$ and so $V_{|\phi}(\sigma)$ is well defined, and so that $\mathbb{M}_{|\phi}, w \models \sigma$ iff $V_{|\phi}(\sigma)$ iff $w \in V(\sigma)$ iff $\mathbb{M}, w \models \sigma$, as required. The other direction is completely analogous.

It remains to be shown that the fifth axiom is valid. Take an arbitrary pointed model for game a structure $\mathbb{M}, w$, and assume that $\mathbb{M}, w \models \phi$ (otherwise we are done) and that $\mathbb{M}, w \models [\phi!]I_i\psi$. This means that for all $w' \in \iota_{i|\phi}(w)$, $\mathbb{M}_{|\phi}, w' \models \psi$. There are two cases to consider.

1. $||\phi|| \cap \iota_i(w) \neq \emptyset$. This means that $\mathbb{M}, w \models \mathsf{i}_i\phi$. We have to show that $\mathbb{M}, w \models I_i(\phi \rightarrow [\phi!]\psi)$ as well. Take $w' \in \iota_i(w)$ and assume that $\mathbb{M}, w' \models \phi$, for otherwise the implication is trivially satisfied. Because of this, we know that $w' \in W_{|\phi}$, and by the definition of ι_i we know that $w' \in \iota_{i|\phi}(w)$ as well. But then by assumption we know that $\mathbb{M}_{|\phi}, w' \models \psi$, which means that $\mathbb{M}, w' \models [\phi!]\psi$.
2. $||\phi|| \cap \iota_i(w) = \emptyset$. In this case we have that $\mathbb{M}, w \models \neg\mathsf{i}_i\phi$, so it remains to show that $\mathbb{M}, w \models [\phi!]A\psi$, that is that $\mathbb{M}_{|\phi}, w' \models \psi$ for all $w' \in W_{|\phi}$. Take a $w' \in \iota_{i|\phi}(w)$. By assumption we know that $\mathbb{M}_{|\phi}, w' \models \psi$, but because $||\phi|| \cap \downarrow \iota_i(w) = \emptyset$ we also know that $\iota_{i|\phi}(w) = W_{|\phi}$, and so we are done.

4.4 Proof of Fact 6

In this proof the numbers refer to Table 4.

We start with $[\phi!]\neg I_i\bot$, which is the same as $[\phi!]\mathsf{i}_i\top$. By (2), this is equivalent to

$$\phi \rightarrow \neg[\phi!]I_i\bot$$

Now, by (5), the consequent expands into two parts $\Phi = \mathsf{i}_i\phi \wedge I_i(\phi \rightarrow [\phi!]\bot)$ and $\Psi = \neg\mathsf{i}_i\phi \wedge [\phi!]A\bot$, which we treat separately to keep the formulas readable.

$$\phi \rightarrow \neg(\phi \rightarrow (\Phi \vee \Psi)).$$

Before looking at each disjunct, some redundancy can be eliminated by propositional reasoning, to get:

$$\phi \rightarrow \neg(\Phi \vee \Psi)$$

Now let us look first at $\Phi = \mathsf{i}_i\phi \wedge I_i(\phi \rightarrow [\phi!]\bot)$. By (1) we get

$$\mathsf{i}_i\phi \wedge I_i(\phi \rightarrow (\phi \rightarrow \bot))$$

because $\bot$ can be treated as a propositional atom. This is equivalent in propositional logic to:

$$\mathsf{i}_i\phi \wedge I_i(\neg\phi).$$

But the second conjunct is just the negation of the first, which means that Φ is just equivalent to $\bot$. We are thus left with:

$$\phi \rightarrow \neg(\bot \vee \Psi)$$

Which is just the same as:

$$\phi \rightarrow \neg\Psi$$

Now, recall that ψ is the following:

$$\neg\mathsf{i}_i\phi \wedge [\phi!]A\bot$$

By (4), this expands to:

$$\neg\mathsf{i}_i\phi \wedge A(\phi \rightarrow [\phi!]\bot)$$

By (1) again, we thus get:

$$\neg\mathsf{i}_i\phi \wedge A(\phi \rightarrow (\phi \rightarrow \bot)).$$

This again reduces to:

$$\neg\mathsf{i}_i\phi \wedge A(\neg\phi).$$

Putting this back in the main formula, we get:

$$\phi \rightarrow \neg(\neg\mathsf{i}_i\phi \wedge A(\neg\phi)).$$

But then propositional reasoning gets us:

$$(\phi \wedge A\neg\phi) \rightarrow \mathsf{i}_i\phi.$$

But the antecedent is just a contradiction of the T axiom for E, $\phi \rightarrow E\phi$, and so we get:

$$\bot \rightarrow \mathsf{i}_i\phi$$

which is propositionally equivalent to $\top$. Since we took an arbitrary i, we can conclude that this is also the case for $\bigwedge_{i\in\mathcal{A}}[\phi!]\mathrm{i}_i\top$.

4.5 Proof of Fact 7

We provide here a direct proof of stabilization, because the announcement we analyze is not a monotone map. Had that been the case, i.e. if it were the case that $\mathbb{M}_{|\bigvee_{i\in\mathcal{A}} IR_i} \subseteq \mathbb{M}'_{|\bigvee_{i\in\mathcal{A}} IR_i}$ provided that $\mathbb{M} \subseteq \mathbb{M}'$, then the existence of a stabilization point would have been ensured by Tarski's fixed point theorem. We indeed have that $W_{|\bigvee_{i\in\mathcal{A}} IR_i} \subseteq W'_{|\bigvee_{i\in\mathcal{A}} IR_i}$ if $\mathbb{M} \subseteq \mathbb{M}'$. The non-monotonicity of this announcement lies in the update rule for the intention set. One can easily devise an example where $\mathbb{M} \subseteq \mathbb{M}'$ but in which there is a $w \in W$ and an $i \in \mathcal{A}$ such that $\iota_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w) \not\subseteq \iota'_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)$.

Proof Assume that there is no such k. This means that there is no k such that $\mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i}, w = \mathbb{M}^{k+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}, w$. Since we are working with finite models, this means that there is a finite n-step loop where $\mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i} = \mathbb{M}^{k+n+1}_{|\bigvee_{i\in I} IR_i}$ such that

$$\mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i}, w \neq \mathbb{M}^{k+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}, w \neq \cdots \neq \mathbb{M}^{k+n}_{|\bigvee_{i\in\mathcal{A}} IR_i} \neq \mathbb{M}^{k+n+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$$

Now, observe that by Definition 8:

$$W^k_{|\bigvee_{i\in\mathcal{A}} IR_i} \supseteq W^{k+1}_{|\bigvee_{i\in\mathcal{A}} IR_i} \supseteq \cdots \supseteq W^{k+n}_{|\bigvee_{i\in\mathcal{A}} IR_i} \supseteq W^{k+n+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$$

But since $\mathbb{M}^k_{|\bigvee_{i\in\mathcal{A}} IR_i} = \mathbb{M}^{k+n+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$, all these inclusions are in fact equalities.

$$W^k_{|\bigvee_{i\in\mathcal{A}} IR_i} = W^{k+1}_{|\bigvee_{i\in\mathcal{A}} IR_i} = \cdots = W^{k+n}_{|\bigvee_{i\in\mathcal{A}} IR_i} = W^{k+n+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$$

Given the definition of the relation $\sim_i$, it must then be that for all $0 \leq \ell \leq n$, there is a $i \in \mathcal{A}$ and a $w \in W^{k+\ell}_{|\bigvee_{i\in\mathcal{A}} IR_i}$ such that $\iota^{k+\ell}_{i,|\bigvee_{i\in\mathcal{A}} IR_i}(w) \neq \iota^{k+\ell+1}_{i\,|\bigvee_{i\in\mathcal{A}} IR_i}(w)$. But this cannot be, as the following two cases show, and so there cannot be such a loop.

1. Assume that:

$$\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w) \cap \left\| \bigvee_{i\in\mathcal{A}} IR_i \right\|^{k+\ell} \neq \emptyset \tag{1}$$

   This means that:

$$\iota^{k+\ell+1}_{i\,|\bigvee_{i\in\mathcal{A}} IR_i}(w) = \left(\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w) \cap \left\| \bigvee_{i\in\mathcal{A}} IR_i \right\|^{k+\ell}\right)$$

   But observe that, because $W^{k+\ell+1}_{|\bigvee_{i\in\mathcal{A}} IR_i} = W^{k+\ell}_{|\bigvee_{i\in\mathcal{A}} IR_i}$ we have:

$$\left\| \bigvee_{i\in\mathcal{A}} IR_i \right\|^{k+\ell} = W^{k+\ell+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$$

This means that:

$$\iota^{k+\ell+1}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)=(\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)\cap W^{k+\ell}_{|\bigvee_{i\in\mathcal{A}} IR_i})=\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)=\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)$$

So (1) cannot hold while:

$$\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)\neq\iota^{k+\ell+1}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)$$

2. Assume then that:

$$\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)\cap\left\|\bigvee_{i\in\mathcal{A}} IR_i\right\|^{k+\ell}=\emptyset \tag{2}$$

In this case $\iota^{k+\ell+1}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)$ just becomes $W^{k+\ell+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$. But recall that by definition, $W^{k+\ell+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}$ is just $||\bigvee_{i\in\mathcal{A}} IR_i||^{k+\ell}$. But since we know that $W^{k+\ell+1}_{|\bigvee_{i\in\mathcal{A}} IR_i}=W^{k+\ell}_{|\bigvee_{i\in\mathcal{A}} IR_i}$, this means that $||\bigvee_{i\in\mathcal{A}} IR_i||^{k+\ell}=W^{k+\ell}_{|\bigvee_{i\in\mathcal{A}} IR_i}$. But that would mean:

$$\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)\cap W^{k+\ell}_{|\bigvee_{i\in\mathcal{A}} IR_i}=\emptyset$$

which is just to say that

$$\iota^{k+\ell}_{i|\bigvee_{i\in\mathcal{A}} IR_i}(w)=\emptyset$$

which is impossible by Fact 6.

Acknowledgements Many thanks to Johan van Benthem, Jens Ulrik Hansen, Barteld Kooi, Alexandru Baltag, Martin van Hees and the two anonymous referees of *Synthese* for their remarks and comments. The paper also greatly profited from discussions during the Seminar on Logics for Dynamics of Information and Preferences at the Institute for Logic, Language and Computation in Amsterdam, and at the LOCCOL Workshop in September 2008 in Prague. Financial support from the Conseil de Recherches en Sciences Humaines du Canada, Scholarship scholarship #752-2006-0345 is gratefully acknowledged.

References

Alchourron, C.E., Gardenfors, P., & Makinson, D. (1985). On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic, 50*(2), 510–530.

Areces, C., & ten Cate, B. (2007). Hybrid logic. In P. Blackburn, J. van Benthem, & F. Wolter (Eds.), *Handbook of modal logic*, pp. 821–868. Amsterdam: Elsevier.

Aumann, R. (1999). Interactive epistemology I: Knowledge. *International Journal of Game Theory, 28*, 263–300.

Aumann, R., & Dreze, J. (2008). Rational expectations in games. *American Economic Review, 98*(1), 72–86.

Baltag, A., Smets, S., & Zvesper, J. (2009). Keep ‘hoping’ for rationality: A solution to the backward induction paradox. *Synthese*, *169*(2), 301–333.

Blackburn, P., de Rijke, M., & Venema, Y. (2001). *Modal logic*. Cambridge, MA: Cambridge University Press.

Blackburn, P., van Benthem, J., & Wolter, F. (Eds.) (2006). *Handbook of modal logic*. Amsterdam: Elsevier.

Brandenburger, A. (2007). The power of paradox: Some recent developments in interactive epistemology. *International Journal of Game Theory, 35*, 465–492.

Bratman, M. (1987). *Intention, plans and practical reason*. London: Harvard University Press.

Bratman, M. (1999). *Faces of intention; selected essays on intention and agency*. Cambridge: Cambridge University Press.

Bratman, M. (2009). Intention, belief, practical, theoretical. In J. Timmerman, J. Skorupski, & S. Robertson (Eds.), *Spheres of reason*. Oxford: Oxford University Press.

Bratman, M., Israel, D., & Pollack, M. (1988). Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4, 349–355

Cohen, P., & Levesque, H. (1990). Intention is choice with commitment. *Artificial Intelligence, 42*(2–3), 213–261.

Fagin, R., Halpern, J., Moses, Y., & Vardi, M. (1995). *Reasoning about knowledge*. Cambridge: MIT Press.

Gargov, G., & Goranko, V. (1993). Modal logic with names. *Journal of Philosophical Logic, 22*(3), 607–636.

Georgeff, M., Pell, B., Pollack, M., Tambe, M., & Wooldridge, M. (1999). The belief-desire-intention model of agency. In *Intelligent agents V: Agent Thories, Architectures, and Languages. 5th International Workshop, ATAL’98. Proceedings*. pp. 1–10. Springer–Verlag.

Gerbrandy, J. (1999). *Bisimulations on planet kripke*. Ph.D. thesis, University of Amsterdam, ILLC. Dissertation Series DS-1999-01.

Girard, P. (2008). *Modal logics for belief and preference change*. Ph.D. thesis, Stanford University.

Gul, F., & Pesendorfer, W. (2001). Temptation and self-control. *Econometrica, 69*(6), 1403–1435.

Hansen, J. (2009). *Hybrid public announcement logic* (Manuscript).

Harman, G. (1976). Practical reasoning. *Review of Metaphysics, 29*(3), 431–463.

Liu, F. (2008). *Changing for the better: Preference dynamics and agent diversity*. Ph.D. thesis, University of Amsterdam, ILLC Dissertation Series DS-2008-02.

McClennen, E. (1990). *Rationality and dynamic choice: Foundational explorations*. New York: Cambridge University Press.

Osborne, M., & Rubinstein, A. (1994). *A course in game theory*. Cambridge: MIT Press.

Plaza, J. (1989). Logics of Public Communications. In M. Emrich, M. Pfeifer, M. Hadzikadic, & Z. Ras (Eds.), *Proceedings of the fourth international symposium on methodologies for intelligent systems: Poster session program, Oak Ridge National Laboratory*, pp. 201–216.

Rott, H. (2001). *Change, choice and inference: A study of belief revision and nonmonotonic reasoning*. Oxford: Ford University Press.

Roy, O. (2009). Intentions and interactive transformations of decision problems. *Synthese, 169*(2), 335–349.

ten Cate, B. (2002). Internalizing epistemic actions. In M. Martinez (Ed.), *Proceedings of the NASSLLI 2002 student session*, Stanford University, pp. 103–123.

ten Cate, B. (2005). *Model theory for extended modal languages*. Ph.D. thesis, University of Amsterdam, ILLC Dissertation Series DS-2005-01.

van Benthem, J. (2007a). Dynamic logic for belief revision. *Journal of Applied Non-classical Logics, 17*(2), 129–155.

van Benthem, J. (2007b). Rational dynamic and epistemic logic in games. *International Game Theory Review, 9*(1), 13–45.

van Benthem, J. (to appear). *Logical dynamics of information flow*. Cambridge: Cambridge University Press.

van Benthem, J., Girard, P., & Roy, O. (2008). Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic, 38*(1), 83–125.

van Benthem, J., van Otterloo, S., & Roy, O. (2006). Preference logic, conditionals, and solution concepts in games. In H. Lagerlund, S. Lindström, & R. Sliwinski (Eds.), *Modality matters: Twenty-five essays in honour of Krister Segerberg*, Vol. 53. Uppsala: University of Uppsala.

van der Hoek, W., Jamroga, W., & Wooldrige, M. (2007). Towards a theory of intention revision. *Synthese, 155*(2), 265–290.

van Ditmarsch, H., van de Hoek, W., & Kooi, B. (2007). *Dynamic epistemic logic*, Vol. 337 of *Synthese Library Series*. Springer.

van Hees, M. & Roy, O. (2008). Intentions and plans in decision and game theory. In B. Verbeek (Ed.), *Reasons and intentions*, pp. 207–226. Aldershot, England: Ashgate Publishers.

Velleman, J. (1997). How to share an intention. *Philosophy and Phenomenological Research, 57*(1), 29–50.

von Neumann, J., & Morgenstern, O. (1994). *A theory of games and economic behaviour*. Princeton, NJ: Princeton University Press.

Wooldridge, M. (2000). *Reasoning about rational agents*. Cambirdge: MIT Press.