Inference, Promotion, and the Dynamics of Awareness

Johan van Benthem (J.vanBenthem@uva.nl)
Institute for Logic, Language and Computation, Universiteit van Amsterdam.
Department of Philosophy, Stanford University.

Fernando R. Velázquez-Quesada (F.R.VelazquezQuesada@uva.nl)
Institute for Logic, Language and Computation, Universiteit van Amsterdam.

Abstract. Classical epistemic logic describes implicit knowledge of agents about facts and knowledge of other agents, based on semantic information. The latter is produced by acts of observation or communication, that are described well by dynamic epistemic logics. What these logics do not describe, however, is how significant information is also produced by acts of inference – and key axioms of the system merely postulate "deductive closure". In this paper, we take the view that all information is produced by acts, and hence we also need a dynamic logic of inference steps showing what effort on the part of the agent makes a conclusion explicit knowledge. Strong omniscience properties of agents should be seen not as static idealizations, but as the result of dynamic processes that agents engage in.

This raises two questions: (a) how to define suitable information states of agents and matching notions of explicit knowledge, (b) how to define natural processes over these states that generate new explicit knowledge. To this end, we extend earlier epistemic "awareness models" into a dynamic system that includes acts of public observation, but also adding and dropping formulas from the currently 'entertained' set, we give a completeness theorem, and we show how this dynamics updates explicit knowledge. Similar ideas have been proposed before, but they were restricted to update with factual propositions; our new dynamic system applies to arbitrary formulas.

We also extend our approach to multi-agent scenarios where awareness changes may happen privately. Finally, we mention further directions and related approaches.

1. The problem of omniscience: what is 'missing in action'

The usual discussions of the problem of omniscience in epistemic logic revolve around the distribution axiom $K(\varphi \to \psi) \to (K\varphi \to K\psi)$. Is knowledge closed under drawing logical inferences? If it is, so the story goes, then we have idealized our knowing agents too much.

But this story is misleading on two accounts. First, with the usual semantics of epistemic logic, the K operator really just describes implicit semantic information of the agent, which definitely has the preceding closure property. The point is rather that closure need not hold for a related, but different intuitive notion, viz. explicit "aware-that" knowledge $\mathrm{Ex}\,\varphi$, in some suitable sense to be defined. So, what we really need is not "epistemic logic bashing", but a richer account of agents' attitudes. Our first question, then, is how to define explicit knowledge.

But there is more. The interesting issue is not whether explicit knowledge has deductive closure. It is rather: "*what do agents have to do* to make their implicit knowledge explicit?" Consider the premises $\mathrm{Ex}\,(\varphi \rightarrow \psi), \mathrm{Ex}\,\varphi$ of the distribution axiom, saying that the agent explicitly knows both $\varphi \rightarrow \psi$ and φ. These do imply $K\psi$, that is, the agent knows ψ implicitly. But crucially, in order to make this information explicit, the agent has to do some *work*, namely, perform an act of "awareness raising" that leads to $\mathrm{Ex}\,\psi$. Stated more syntactically, the usual implication $\mathrm{Ex}\,(\varphi \rightarrow \psi) \rightarrow (\mathrm{Ex}\,\varphi \rightarrow \mathrm{Ex}\,\psi)$ contains a *gap* [ ]:

$$\mathrm{Ex}\,(\varphi \rightarrow \psi) \rightarrow (\mathrm{Ex}\,\varphi \rightarrow [\;]\mathrm{Ex}\,\psi)$$

and in that gap, we should place an action. Note that, then, the agent is no longer omniscient, but she is not defective either: with the right repertoire of actions available, she can do awareness raising as needed.

This paper explores these ideas. We first introduce simple epistemic awareness models, with a specific interpretation of the syntactic component as the formulas 'entertained' by the agent – and recall their standard axiomatization. Then, we explore some proposals for defining explicit knowledge, picking one that we will work with. Next, we define basic dynamic actions that modify our models, and provide examples of their behaviour, alone and in combination. Representing the actions in the language yields a sound and complete logic that clarifies our initial issues. We also develop some formal properties and raise some open problems. Then we move to the multi-agent case, developing tools for private and even unconscious versions of our actions, that were public so far. Finally, we relate our proposal to earlier ones, and end with conclusions and further directions, in particular, toward agents with beliefs that are modified by default inferences.

2. A static system for different agent attitudes

We assume that the reader is familiar with classical epistemic logic (*EL* for short). We have already stated our motivation for working with this: even though this system fails for its intended interpretation of 'full-blooded knowledge', like so many logical systems, it has turned out quite adequate for other, perhaps originally non-intended interpretations. In particular, it deals well with implicit knowledge of the semantic range kind (cf. van Benthem and Martínez (2008)).

Now we extend the base language of *EL*. We add an operator $E\varphi$ saying that the agent "is aware of φ" (Fagin and Halpern, 1988) or, in less psychological terms, that she "entertains φ" as a matter of attention. Notice that this does not imply any attitude pro or con: the agent may believe φ, but also reject it. Stated in other, but related terms, "awareness of" does not imply "awareness that".

Definition 2.1 (Language $\mathcal{L}$)**.** Let P be a set of atomic propositions. Formulas φ of the *epistemic awareness language* $\mathcal{L}$ are given by

$$\varphi ::= p \mid E\varphi \mid \neg\varphi \mid \varphi \wedge \psi \mid K\varphi$$

with $p \in \mathsf{P}$. Other Boolean connectives $\vee, \rightarrow, \leftrightarrow$, as well the existential modal operator $(\widehat{K})$ are defined as usual.

We will read formulas $E\varphi$ as *"the agent entertains φ"*, and formulas $K\varphi$ as *"the agent knows φ implicitly"*. The language is interpreted in epistemic models assigning to each agent in each world a set of formulas, representing the information she entertains.

Definition 2.2 (Semantic model)**.** An *epistemic awareness model* is a tuple $M = \langle W, R, \mathsf{E}, V\rangle$ where

- $\langle W, R, V\rangle$ is a standard epistemic model: a set or worlds W, an accessibility relation $R \subseteq (W \times W)$, and a valuation $V : \mathsf{P} \rightarrow \wp(W)$.
- $\mathsf{E} : W \rightarrow \wp(\mathcal{L})$ is the *"entertain" function* giving the formulas that the agent 'has in mind'. $\mathsf{E}(w)$ is the *entertained set* at w.

As usual, a *pointed model* (M, w) also has a distinguished world w.

The semantic interpretation of formulas in $\mathcal{L}$ is entirely as expected:

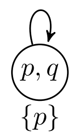
Definition 2.3. Let (M, w) be a pointed semantic model with $M = \langle W, R, \mathsf{E}, V\rangle$. Atomic propositions and boolean connectives are interpreted as usual; for $E\varphi$ and $K\varphi$ we have:

$$(M, w) \models E\varphi \quad \text{iff} \quad \varphi \in \mathsf{E}(w)$$
$$(M, w) \models K\varphi \quad \text{iff} \quad \text{for all } u \in W,\ Rwu \text{ implies } (M, u) \models \varphi.$$

On these models we can impose standard epistemic assumptions about the accessibility relation, such as reflexivity, transitivity, and symmetry. But these are orthogonal to our main concerns in this paper.

It is easy to visualize how our mixed models work:

Example 1.



In this one-world model, the agent knows implicitly that p and also that q. But while she is explicitly aware that p holds, she is not aware that q holds, so her explicit knowledge about p, q differs.

In models with more than one world, a genuine issue of interplay arises: is awareness introspective? We will not assume strong introspection in the sense that entertained sets are closed under the operator E. But in our intended class of models $\mathbf{M}$, we will assume the weaker introspection principle that being entertained is implicitly known:

Weak introspection: Entertained formulas are preserved under epistemic accessibility: if $\varphi \in \mathsf{E}(w)$ and Rwu, then $\varphi \in \mathsf{E}(u)$.

The calculus of reasoning with these notions is standard. The formulas of $\mathcal{L}$ valid in models in $\mathbf{M}$ are those provable in the minimal modal logic K for implicit knowledge plus the axiom of *'weak introspection'*:

$$E\varphi \to KE\varphi$$

Soundness and completeness are proved by standard techniques.

3. Defining explicit knowledge

Explicit knowledge as a defined notion. Combining *implicit* and *entertained* information produces several kinds of *explicit* information. Options start with the $K\varphi \wedge E\varphi$ of Fagin and Halpern (1988), which says that φ is implicitly known and explicitly considered. But there are others. It is not so clear to us which notion is most suited, but there are some obvious desiderata. Explicit knowledge should imply implicit knowledge, and it should be subject to implicit, though not to explicit introspection. This suggests the alternative version

$$\mathrm{Ex}\,\varphi := K(\varphi \wedge E\varphi)$$

or equivalently, $K\varphi \wedge KE\varphi$. In general, this definition is not equivalent to that of Fagin and Halpern (1988), but it is if we assume weak introspection like above, plus reflexivity of accessibility. Then we get

$$\models_{\mathbf{M}} K(\varphi \wedge E\varphi) \leftrightarrow (K\varphi \wedge E\varphi)$$

It is also easy to see that both notions satisfy our earlier desiderata, modulo some mild conditions on the epistemic base logic. More generally, Grossi and Velázquez-Quesada (2009) discuss how assumptions on the underlying base logic of K and E determine properties of explicit information and relations between options. In particular, they discuss positive and negative implicit introspection about explicit knowledge:

$$\mathrm{Ex}\,\varphi \to K\mathrm{Ex}\,\varphi \qquad \neg\mathrm{Ex}\,\varphi \to K\neg\mathrm{Ex}\,\varphi$$

We will assume the definition $\mathrm{Ex}\,\varphi := K(\varphi \wedge E\varphi)$ henceforth for explicit knowledge. Among other candidates, one that appeals to us is: $K\varphi \wedge EK\varphi$, where the implicit knowledge itself has the agent's explicit attention. We will not pursue this option – but it does show the need for a framework that can describe the dynamics of many proposals.

Explicit knowledge as a primitive notion. It may be useful at this stage to explain how we are deviating here from our earlier work (van Benthem, 2008; Velázquez-Quesada, 2009). There we assumed a *primitive notion* of explicit knowledge ("awareness that"), treated as a map assigning formulas to worlds, but not identified with any combination of operators K and E. We only assumed that all explicitly known formulas were implicitly known, and hence in particular, that they were true. But then, updates that can change truth values of epistemic assertions: I might be aware now that I do not know that p, but learning that p will invalidate this explicit knowledge. And other natural actions may change truth values of formulas involving E. Based on these observations, we found the need for dropping all persistence of explicit knowledge under update, except for purely factual assertions. Our current approach of defining explicit knowledge will circumvent this difficulty, since the definition automatically 'recomputes' what is explicit knowledge following an update with epistemic side-effects.

On the other hand, and we acknowledge that this is a drawback of the approach in this paper, there may be something sui generis to the notion of explicit knowledge that φ, not reducible to just implicit knowledge and awareness of φ. I might be thinking of φ, and also know it implicitly, and still fail to see directly that it is true. Think of a conclusion that I am pondering, and that in fact follows from some premises whose truth I am aware of. I could still fail to see how it follows explicitly. We will return to this discussion later.

For now, we move to our second main issue, viz. how agents can 'improve' their current brand of knowledge about a proposition. This is not a matter of static implications between brands of implicit and explicit knowledge. As we have said before, the correct question to ask here is a dynamic one: *what does an agent have to* do *to upgrade her implicit knowledge?* To some readers, introducing the explicit actions that lead to more omniscient states may demystify them. In Conan Doyle's detective stories, the explanation offered at the end turns Holmes' 'magical powers' into a sequence of simple observations and deductive acts, making the procedure "elementary, my dear Watson". While this is true, it also underscores the power of successive small steps.

4. Operations on epistemic awareness models

Our epistemic awareness models suggest a natural and simple dynamics. Though the agent is not logically omniscient anymore, she can get new information by various acts, including observation and inference. But we want to dig deeper. In line with our definition for explicit knowledge, it also makes sense to look for simplest actions transforming models that can be put together to analyze more complex informational acts. We will see later on how these transform explicit knowledge.

Defining the basic actions. Our models have two separate components for representing information: the accessibility relation and the entertained sets. The following operations modify these components in a simple way, allowing us to define complex epistemic actions later on.

The *consider* operation represents an "awareness raising" action:

Definition 4.1 (Public *consider* operation)**.** Let $M = \langle W, R, \mathsf{E}, V\rangle$ be a model and χ any formula in $\mathfrak{L}$. The model $M_{+\chi} = \langle W, R, \mathsf{E}', V\rangle$ is M with its entertained sets extended with χ, that is,

$$\mathsf{E}'(w) := \mathsf{E}(w) \cup \{\chi\} \qquad\qquad \text{for every } w \in W$$

We must check at once that this stays inside our intended models:

Proposition 1. *The* consider *operation preserves the assumptions on* ***M****: if M satisfies weak introspection, then so does $M_{+\chi}$.*

Proof. Take a world w in M and any $\varphi \in \mathsf{E}'(w)$. Suppose Rwu. If φ is already in $\mathsf{E}(w)$, then $\varphi \in \mathsf{E}(u)$ because M satisfies the property, and then $\varphi \in \mathsf{E}'(u)$ by the definition of E'. If φ is not in $\mathsf{E}(w)$, then it should be χ itself, which by the above definition is also in $\mathsf{E}'(u)$. □

'Considering' extends the information an agent entertains, but we can also define a *neglecting* operation with the opposite effect: reducing entertained sets. This fits with the operational idea that agents can expand and shrink the set of issues having their current attention.

Definition 4.2 (Public *neglecting* operation)**.** Let $M = \langle W, R, \mathsf{E}, V\rangle$ be a model and χ a formula in $\mathfrak{L}$. The model $M_{-\chi} = \langle W, R, \mathsf{E}', V\rangle$ reduces M's entertained sets by removing χ, that is,

$$\mathsf{E}'(w) := \mathsf{E}(w) \setminus \{\chi\} \qquad\qquad \text{for every } w \in W$$

Proposition 2. Neglecting*, too, preserves weak introspection.*

This operation can be seen as a form of "forgetting", an action usually disregarded in dynamic-epistemic logic (but see van Ditmarsch et al. (2009) and van Ditmarsch and French (2009) for proposals).

The preceding actions affect what an agent entertains. The next, known from dynamic-epistemic logic, modifies her implicit knowledge:

Definition 4.3 (Public *implicit observation*)**.** Let $M = \langle W, R, \mathsf{E}, V \rangle$ be a model and χ a formula in $\mathfrak{L}$. The model $M_{!\chi} = \langle W', R', \mathsf{E}', V' \rangle$ has

– $W' := \{\, w \in W \mid (M, w) \models \chi \,\}$
– $R' := R \cap (W' \times W')$
– $\mathsf{E}'(w) := \mathsf{E}(w)$ for all $w \in W'$
– $V'(p) := V(p) \cap W'$

Proposition 3. *Implicit observation preserves weak introspection for entertained propositions.*

Proof. The sub-model $M_{!\chi}$ has the same entertained sets at its worlds as M, and its epistemic accessibility is a subrelation of that for M. $\square$

Building complex actions. Complex actions can now be built by combining basic ones. As an example, it seems natural to think that a public observation of some fact is in fact done consciously, generating awareness. The corresponding operation of "explicit seeing" (van Benthem, 2008; Velázquez-Quesada, 2009) can be defined as an implicit observation followed by an act of consideration:

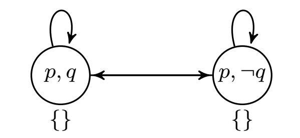
$$M_{\mathrm{PA}(\chi)} := (M_{!\chi})_{+\chi}$$

The definition would also work in the opposite order, given that we are transforming two independent components of our models. Of course, the earlier-mentioned issue of acts invalidating explicit knowledge returns here. We might observe an epistemic fact φ, consider its content, and yet, through the very update, find that we no longer know that φ is true. But this seems right: and indeed, the earlier definition of awareness that in terms of implicit knowledge and entertainment will tell us which propositions we explicitly know after the act.

Still, one might argue that implicit observation and considering take place *simultaneously*. While this makes sense, we will not pursue it here.

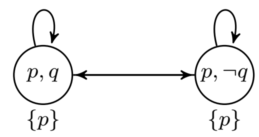
5. The actions in action

Consider the following model:

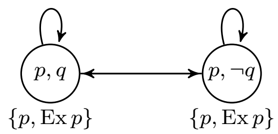


In the leftmost world the agent does not even know implicitly that q. But she knows implicitly that p, though not explicitly.

After the agent *considers* p, we get the model on the right: in its leftmost world, the agent now knows explicitly that p.

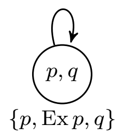


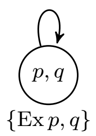
Lifting a restriction in van Benthem, Velázquez-Quesada (2008, 2009), our agent can also get explicit knowledge about her own entertained, implicit and explicit information. Here is how this can happen:



When she *considers* $\mathrm{Ex}\, p$, we get the model on the left. The agent knows explicitly that she has explicit knowledge of p. By acting, she has achieved positive introspection.

Next, consider the above *explicit public announcement* of q: an *implicit observation* followed by *consideration* of q. This yields the model on the right where q is now explicitly known by the agent ($\mathrm{Ex}\, q$).





Finally, *neglecting* p makes the agent lose earlier explicit knowledge about it (in our case, we get $\neg\mathrm{Ex}\, p$). Moreover, she no longer has explicit knowledge that $\mathrm{Ex}\, p$, since the latter formula is no longer true, and therefore, it is no longer implicitly known.

There are many further scenarios with complex many-world patterns, but the above will suffice to show the interest of our setting.

6. A complete dynamic logic

In order to express how our dynamic operations affect implicit knowledge, entertainment, and explicit knowledge, we extend the static epistemic awareness language with modalities representing each basic operation. If χ and φ are formulas in the resulting extended language (still called $\mathcal{L}$ in this section), then so are

$[+\chi]\varphi$	*after the agent considers χ, φ is the case.*
$[-\chi]\varphi$	*after the agent neglects χ, φ is the case.*
$[!\chi]\varphi$	*after the agent implicit observes χ, φ is the case.*

Definition 6.1. Let (M, w) be a pointed semantic model with $M = \langle W, R, \mathsf{E}, V\rangle$, and let χ, φ be formulas in the extended language $\mathcal{L}$:

$$\begin{array}{lll}(M,w) \models [+\chi]\varphi & \text{iff} & (M_{+\chi}, w) \models \varphi \\ (M,w) \models [-\chi]\varphi & \text{iff} & (M_{-\chi}, w) \models \varphi \\ (M,w) \models [!\chi]\varphi & \text{iff} & (M,w) \models \varphi \text{ implies } (M_{!\chi}, w) \models \varphi\end{array}$$

The main difference among the new modalities is the precondition. The agent can consider or neglect a formula χ without any further requirement, but for her to implicitly observe χ, χ needs to be *true*.

6.1. Dynamic completeness theorem

We now formulate a sound and complete logic for the semantic validities in the extended language $\mathcal{L}$:

Theorem 1. *The valid formulas of the dynamic-epistemic awareness language $\mathcal{L}$ (in models of **M**) are just those provable by the axioms and rules for the static base language (see Section 2) plus the reduction axioms and modal inference rules listed in Table I.*

Proof. We use standard techniques from dynamic-epistemic logic (cf. van Benthem et al. (2006)). Our axioms run through all cases needed to reduce a innermost occurrence of a dynamic operator. Iterating this produces an equivalent formula in the static base language. □

Table I. Sound and complete logic for dynamic epistemic awareness logic.

$\vdash [+\chi]p$	$\leftrightarrow\ p$	$\vdash [-\chi]p$	$\leftrightarrow\ p$
$\vdash [+\chi]E\chi$	$\leftrightarrow\ \top$	$\vdash [-\chi]E\chi$	$\leftrightarrow\ \bot$
$\vdash [+\chi]E\varphi$	$\leftrightarrow\ E\varphi$ for $\varphi \neq \chi$	$\vdash [-\chi]E\varphi$	$\leftrightarrow\ E\varphi$ for $\varphi \neq \chi$
$\vdash [+\chi]\neg\varphi$	$\leftrightarrow\ \neg[+\chi]\varphi$	$\vdash [-\chi]\neg\varphi$	$\leftrightarrow\ \neg[-\chi]\varphi$
$\vdash [+\chi](\varphi \wedge \psi)$	$\leftrightarrow\ [+\chi]\varphi \wedge [+\chi]\psi$	$\vdash [-\chi](\varphi \wedge \psi)$	$\leftrightarrow\ [-\chi]\varphi \wedge [-\chi]\psi$
$\vdash [+\chi]K\varphi$	$\leftrightarrow\ K[+\chi]\varphi$	$\vdash [-\chi]K\varphi$	$\leftrightarrow\ K[-\chi]\varphi$
From $\vdash \varphi$ infer $\vdash [+\chi]\varphi$		From $\vdash \varphi$ infer $\vdash [-\chi]\varphi$	
$\vdash [!\chi]p$	$\leftrightarrow\ \chi \to p$		
$\vdash [!\chi]E\varphi$	$\leftrightarrow\ \chi \to E\varphi$		
$\vdash [!\chi]\neg\varphi$	$\leftrightarrow\ \chi \to \neg[!\chi]\varphi$		
$\vdash [!\chi](\varphi \wedge \psi)$	$\leftrightarrow\ [!\chi]\varphi \wedge [!\chi]\psi$		
$\vdash [!\chi]K\varphi$	$\leftrightarrow\ \chi \to K[!\chi]\varphi$		
From $\vdash \varphi$ infer $\vdash [!\chi]\varphi$			

These principles express the syntactic basics of considering and neglecting, merged with the axioms of public announcement logic *PAL*. For instance, how does entertained information change when the agent *considers* χ? We have two cases. After considering χ, the agent entertains a $\varphi \neq \chi$ iff she entertained φ before; but also, considering χ always makes the agent entertain χ. The effect of *neglecting* χ is analogous. The additional commutation clauses express the independence of modifying the domain of worlds and sets of entertained formulas.

6.2. How the logic describes our major issues

Our logic states how each basic operator of the language is affected by our three actions. By combining effects and unfolding definitions, the logic also explains how derived notions of *explicit knowledge* fare under these actions. We discuss a few cases, using our earlier definition $K(\varphi \wedge E\varphi)$, and suppressing detailed calculations:

Explicit knowledge. For the action of *considering* χ and explicit knowledge about a different formula φ we get the valid principle

$$[+\chi]\mathrm{Ex}\,\varphi \leftrightarrow K([+\chi]\varphi \wedge E\varphi) \qquad (\text{for } \varphi \neq \chi).$$

We leave it to the reader to put this analysis into words. One might have expected a simpler direct reduction principle $[+\chi]\mathrm{Ex}\,\varphi \leftrightarrow \mathrm{Ex}\,\varphi$, but this formula is not valid, since the *considering* action may have changed truth values for sub-formulas of φ.

In the particular case of explicit knowledge about χ itself, however, unfolding via the reduction axiom yields

$$[+\chi]\mathrm{Ex}\,\chi \leftrightarrow K[+\chi]\chi$$

This shows how the *considering* action makes implicit knowledge explicit: we will look at this in more detail below. Going back to the initial discussion about the distribution axiom, the formula

$$\mathrm{Ex}\,(\varphi \to \psi) \to (\mathrm{Ex}\,\varphi \to \mathrm{K}\,\psi)$$

is valid, since the antecedent implies the premises of the distribution law for K. Then, *considering* is the action that 'fills the gap':

$$\mathrm{Ex}\,(\varphi \to \psi) \to (\mathrm{Ex}\,\varphi \to [+\psi]\mathrm{Ex}\,\psi) \quad \text{is valid}$$

One might think that the real act here is a richer one of *drawing the inference*, but in our analysis is the explicit consideration of the conclusion that 'gives the last little push' toward explicit knowledge. Behind this observation lies a general fact that reflects the analysis of actions of epistemic "promotion" in van Benthem (2008):

Fact. The formula $K\varphi \to [+\varphi]\mathrm{Ex}\,\varphi$ is valid.

Proof. Using our reduction axioms as above, we get that

$$\begin{aligned} [+\varphi]\mathrm{Ex}\,\varphi &\leftrightarrow [+\varphi]K(\varphi \wedge E\varphi) \\ &\leftrightarrow K[+\varphi]\varphi \wedge K[+\varphi]E\varphi \\ &\leftrightarrow K[+\varphi]\varphi \end{aligned}$$

But the latter follows from $K\varphi$, since we have quite generally that

the formula $\varphi \to [+\varphi]\varphi$ is valid.

The reason is that, given our semantics, and act of considering φ can only change truth values for $E\varphi$ and formulas containing it. But the formula φ itself is too short to be affected by this. □

Thus, our current proposal realizes the intuitive expectations in our Introduction. But it can describe more, including the behaviour of explicit knowledge under the *neglecting* operation. Here is what happens with formulas φ that differ from the neglected χ:

$$[-\chi]\mathrm{Ex}\,\varphi \leftrightarrow K([-\chi]\varphi \wedge E\varphi) \qquad (\text{for } \varphi \neq \chi)$$

With explicit knowledge about the formula χ itself, *considering* produces a trivialization to 'end-points':

$$[-\chi]\mathrm{Ex}\,\chi \leftrightarrow K\bot$$

If the agent's implicit information is true, however (say, epistemic accessibility is reflexive), we get $\neg[-\chi]\mathrm{Ex}\,\chi$: one never has explicit knowledge about χ after neglecting it. Still, the agent does keep χ as implicit knowledge, witness the valid law

$$\mathrm{Ex}\,\chi \to [-\chi]K\chi$$

A precise justification runs like that for the preceding Fact.

Finally, we analyze the effect of an implicit observation over explicit knowledge. For any φ and χ, unfolding the definition of explicit knowledge via our axioms (we suppress intermediate steps here) gives

$$[!\chi]\mathrm{Ex}\,\varphi \leftrightarrow (\chi \to K([!\chi]\varphi \wedge (\chi \to E\varphi)))$$

Again, we leave it to the reader to state this fact informally. This outcome is our solution to the earlier-mentioned problem of update making explicit knowledge 'out of synch' with implicit knowledge. (Recall that this was the reason for the restriction to purely factual assertions in van Benthem, Velázquez-Quesada (2008, 2009).) Explicit knowledge is now a defined notion, so it automatically re-adjusts to whatever happens to the modalities K and E, and our logic tells us precisely how.

We have extracted the effect of our basic epistemic actions over explicit knowledge defined as $K(\varphi \wedge E\varphi)$. Thus, we replace discussion whether agents *have* epistemic closure by a much richer picture of what they can *do* to update and "upgrade" their knowledge. Moreover, this style of analysis works not only for the stated notion of explicit knowledge; it can also provide us with validities expressing the way other possible notions are affected, like the already mentioned $K\varphi \wedge EK\varphi$.

6.3. Schematic validities and algebra of actions

While all this seems straightforward dynamic epistemic technique, there is a catch. In deriving the principles of the previous section, we have used more than the reduction axioms of our logic per se. Several important 'schematic' principles did not follow from our reduction axioms. In particular, we have used the two principles

$$[+\chi]\chi \leftrightarrow \chi \qquad \text{and} \qquad [-\chi]\chi \leftrightarrow \chi$$

whose validity involved additional considerations. Of course, each specific instance of such a formula can be derived, given our completeness theorem. But that does not mean there is any illuminating uniform derivation of an "algebraic" sort. Indeed, it is a well-known (though not much-publicized) open problem, even for public announcement logic, to characterize its schematic validities, not dependent on the special treatment of atomic formulas in *PAL*. Given the importance of such principles here, that problem becomes even more urgent.

Algebra of actions. We end with one further source of schematic validities. As important as it is to understand how actions affect our information, their own structure is of interest, too. We briefly mention some relevant validities, to show that we have the beginnings of an interesting *"algebra of actions"*:

- In general, *neglecting* does not neutralize *considering* ($[+\chi][-\chi]\varphi \leftrightarrow \varphi$ is not valid): if the agent initially entertains χ, *considering* makes no change, but *neglecting* does, yielding a model where χ is not entertained. The actual validity is the qualified

$$\neg E\chi \rightarrow ([+\chi][-\chi]\varphi \leftrightarrow \varphi)$$

- The dual case behaves in the same way: *considering* does not neutralize *neglecting* in general – but we do have:

$$E\chi \rightarrow ([-\chi][+\chi]\varphi \leftrightarrow \varphi)$$

As for unqualified algebraic laws, we do have idempotence:

– A sequence of *considering* has the same effect as a single one, and the *neglecting* operation behaves in a similar way:

$$[+\chi]\varphi \leftrightarrow [+\chi][+\chi]\varphi \qquad \text{and} \qquad [-\chi]\varphi \leftrightarrow [-\chi][-\chi]\varphi$$

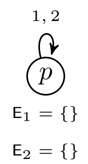
Given the dynamics of the system, we do not expect strong commutation laws between considering and neglecting, but we do with implicit observation, that modified an independent component of our models.

We will not pursue the resulting action algebra here (it will get even nicer when we restrict attention to just factual assertions). But it does demonstrate the same desideratum beyond standard axiomatizations in dynamic-epistemic logic. Even public announcement logic PAL has a hidden algebra of successive announcements – but it tends to go unnoticed, as two successive announcements can be compressed into one. But this compression disappears with extensions to relation change for belief or preference, as has been noted in the literature. The same is true here, and hence the additional axiomatization issues.

7. From single to multi-agent scenarios

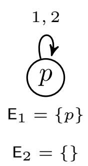
So far, we have considered activities of single agents, including their observations, but also their acts of inference. Now the latter are typically private, and hence it makes sense to look at scenarios with privacy. A bit paradoxically, privacy only becomes visible in a multi-agent setting. Here is an initial illustration, with two agents:

Example 2. Consider the following model M, generalizing the single-agent framework to a multi-agent setting in a straightforward way:



In the only world of the model, each agent knows implicitly that p, but no agent entertains p ($\neg E_1 p \wedge \neg E_2 p$). Moreover, agents have implicit higher-order knowledge about each other. E.g., agent 2 knows implicitly that agent 1 does not entertain p ($K_2 \neg E_1 p$).

Now let an event take place: agent 1 considers p: M_{+p} is given by



In the new situation, agent 1 entertains p ($E_1 p$), and now has explicit knowledge about it. But there is more: agent 2 now knows implicitly that agent 1 entertains p ($K_2 E_1 p$), but without knowing this explicitly.

Is this a realistic scenario? It seems strange that an action of 1 alone can affect the information of agent 2. To get clearer on this, we need a more detailed analysis of how epistemic awareness models should change, in a setting allowing privacy.

7.1. Multi-agent static framework

The extension of the static epistemic awareness framework to a setting with many agents in a group $\mathtt{A}$ is straightforward. In the language of multi-agent $\mathcal{L}$, we just add agent indexes to the E and the K modalities (E_i and K_i, respectively). In the semantic models, R becomes a *function* from $\mathtt{A}$ to $\wp(W \times W)$ returning an *accessibility relation* R_i for each agent $i \in \mathtt{A}$, and E becomes a function from $\mathtt{A} \times W$ to $\wp(\mathcal{L})$ returning the *entertained* set $\mathsf{E}_i(w)$ of each agent i at each possible world w. The semantic interpretation of formulas is then as before, using E_i and R_i to interpret modalities $E_i\varphi$ and $K_i\varphi$, respectively.

In this multi-agent case we will not impose special semantic constraints, such as the earlier weak introspection principle. In private settings with single-agent actions of *considering* and *neglecting*, it makes more sense to reinterpret the K modalities as *beliefs* that can be mistaken, as is also done in standard dynamic-epistemic logic.

7.2. Multi-agent actions: the general case

To make our actions *private*, we need a mechanism that lets actions affect agents in different ways. The *action models* of Baltag et al. (1999) will do this, provided we extend them in the manner of van Benthem et al. (2006). That is, events can now really change the world, coming not just with *preconditions* on their executability, but also with *postconditions* describing what changes they bring about:

Definition 7.1 (Multi-agent action model)**.** With $\mathtt{P}$ the set of atomic propositions and $\mathtt{A}$ the finite set of agents, a *multi-agent action model* is a tuple $A = \langle S, T, \mathrm{Pre}, \mathrm{Pos}\rangle$ where:

- $\langle S, T, \mathrm{Pre}\rangle$ is an action model (Baltag et al., 1999) with S a finite non-empty set of *events*, $T : \mathtt{A} \to \wp(W \times W)$ a function returning an *accessibility relation* T_i for each agent $i \in \mathtt{A}$ and $\mathrm{Pre} : S \to \mathcal{L}$ the *precondition* function indicating where each event can be executed.

- $\mathrm{Pos} : (\mathtt{A} \times S \times \wp(\mathcal{L})) \to \wp(\mathcal{L})$ is the *postcondition* function, assigning a new set of formulas in $\mathcal{L}$ to every tuple of an agent, event, and (old) set of formulas in $\mathcal{L}$.

A *pointed action model* (A, s) has a distinguished world s.

Recall that we want to model private versions of our operations that *modify* entertained sets. This is exactly the role of the Pos function, a variation on the *substitution function* defined in van Benthem et al. (2006) for representing factual change. The following update rule describes how our action models modify epistemic awareness models.

Definition 7.2 (Product update)**.** Let $M = \langle W, R, \mathsf{E}, V\rangle$ be a multi-agent semantic model and let $A = \langle S, T, \mathrm{Pre}, \mathrm{Pos}\rangle$ be a multi-agent action model. The *product model* $M \otimes A = \langle W', R', \mathsf{E}', V'\rangle$ is given by

- $W' := \{ (w, s) \mid (M, w) \models \mathrm{Pre}(s) \}$
- $R'_i(w, s)(w', s')$ iff $R_i ww'$ & $T_i ss'$
- $V'(p) := \{ (w, s) \in W' \mid w \in V(p) \}$
- $Y'_i(w, s) := \mathrm{Pos}_i(s, Y_i(w))$

Note how the Pos function works: for each agent i and each event s, Pos takes agent i's entertained set at w in M, and returns her entertained set at (w, s) in $M \otimes A$. Later on, we will look at restrictions on the syntactic format of definition for the postcondition function.

In order to express how product updates affect the agent's information, the *extended* multi-agent language $\mathfrak{L}$ has extra modalities:

> if (A, s) is a pointed action model and φ is a formula in the extended multi-agent $\mathfrak{L}$, then so is $[(A, s)]\varphi$.

The semantic interpretation of these new formulas is provided here:

Definition 7.3. Let (M, w) be a pointed multi-agent semantic model and let (A, s) be a pointed action model with $A = \langle S, T, \mathrm{Pre}, \mathrm{Pos}\rangle$.

$$(M, w) \models [(A, s)]\varphi \quad \text{iff} \quad (M, w) \models \mathrm{Pre}(s) \ \text{ implies } \ (M \otimes A, (w, s)) \models \varphi$$

7.3. Private *considering* and *neglecting*

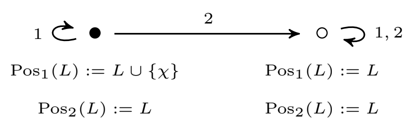
Now we can define *private* actions more precisely. Here are simple versions of the earlier *considering* and *neglecting*. As usual, these encode what takes place, but also how different agents 'view' this:

Definition 7.4 (Private *considering* action)**.** Let χ be any formula in multi-agent $\mathfrak{L}$. The *private considering action of agent j* is the pointed action model $(\mathrm{Pri}^j_{+\chi}, \bullet)$ with $\mathrm{Pri}^j_{+\chi} = \langle S, T, \mathrm{Pre}, \mathrm{Pos}\rangle$ as

$$- S := \{\bullet, \circ\} \quad - T_i := \begin{cases} \{(\bullet, \bullet), (\circ, \circ)\} & \text{if } i = j \\ \{(\bullet, \circ), (\circ, \circ)\} & \text{otherwise} \end{cases} \quad - \mathrm{Pre}(\bullet) = \mathrm{Pre}(\circ) := \top$$

$$\begin{array}{lll} - \mathrm{Pos}_j(\bullet, L) := L \cup \{\chi\}, & \mathrm{Pos}_j(\circ, L) := L & \\ - \mathrm{Pos}_i(\bullet, L) := L, & \mathrm{Pos}_i(\circ, L) := L & \text{for } i \neq j \end{array}$$

The diagram on the right shows the model $\mathrm{Pri}^1_{+\chi}$ in the 2-agent case (preconditions omitted).

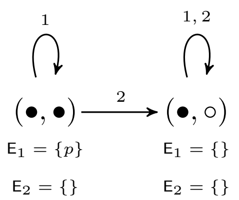


Definition 7.5 (Private *neglecting* action)**.** Let χ be any formula in multi-agent $\mathfrak{L}$. The *private neglecting action of agent j*, given by the pointed action model $(\mathrm{Pri}^j_{-\chi}, \bullet)$, differs from *private considering* only in its postcondition function for agent j in $\bullet$:

$$\mathrm{Pos}_j(\bullet, L) := L \setminus \{\chi\}$$

These actions transform our initial epistemic awareness model:

Example 3. Recall the model M from Example 2. After agent 1 considers p privately (i.e., after applying $(\mathrm{Pri}^1_{+p}, \bullet)$), we get the model:



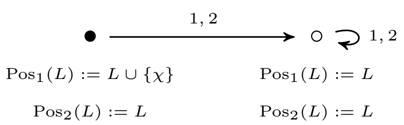
In the evaluation point, the leftmost world, agent 1 entertains p ($E_1 p$), just like she does after publicly considering p. But this time, agent 2's implicit knowledge does not change: she still believes implicitly that agent 1 does not entertain p ($K_2 \neg E_1 p$).

7.4. Unconscious versions

The flexibility of the postcondition mechanisms is great. We can represent many further scenarios, even *unconscious* actions, hidden from all agents, including the one that 'performs it! We just give an illustration:

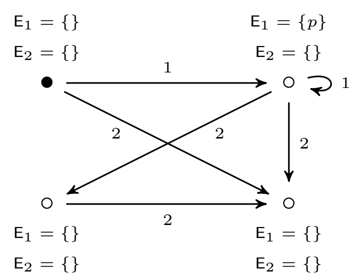
Definition 7.6 (Unconscious *neglecting* action)**.** Let χ be a formula in the multi-agent $\mathfrak{L}$. The *unconscious neglecting action* of agent j, given by the pointed action model $(\mathrm{Unc}^j_{-\chi}, \bullet)$, differs from its *private* counterpart only in the definition of the accessibility relation:

$$T_i := \{(\bullet, \circ), (\circ, \circ)\} \qquad \text{for all agents } i$$



The diagram on the left depicts $\mathrm{Unc}^1_{-\chi}$ in a 2-agent case.

Example 4. Consider the pointed model $(M \otimes \mathrm{Pri}^1_{+p}, (\bullet, \bullet))$ of Example 3. If agent 1 now unconsciously neglects p, we get the model



In $\bullet$, the agent does not entertain p ($\neg E_1 p$), but she implicitly believes that she entertains it ($K_1 E_1 p$).

Much more can be said about this scenario, and we feel that we have a promising take here on unconscious actions such as *forgetting*. But our purpose here was just to demonstrate the flexibility of the framework.

7.5. Complete dynamic logic for the multi-agent version

A complete dynamic logic for this system looks like our earlier single-agent logic, with indices attached. Its principles for atomic formulas, Boolean operations, and epistemic implicit knowledge are the usual *DEL* versions. As an illustration, we have the valid equivalence

$$\vdash [(A,s)]K_i\varphi \quad \leftrightarrow \quad \mathrm{Pre}(s) \rightarrow \bigwedge_{Tsr} K_i[(A,r)]\varphi$$

But the crucial issue is the reduction axiom for entertainment given the *postconditions*. Consider the axioms that we gave for our basic operations of *considering* and *neglecting*. These described the postconditions inside the language, exploiting the simple format of their effects. For instance, we had $E\varphi$ after an act $+\chi$ if either we had it before, or the two formulas were the same, and then it was true automatically. This case distinction in the reduction axiom directly reflects the simple disjunctive definition of the postcondition for the action $+\chi$: being the union of the old entertained set with χ. The same is true in our more general setting: *simple uniform definitions of postconditions in our action models will automatically generate matching reduction axioms.*

We will not pursue the technicalities of this mechanism here, but there should be a definition scheme $\delta(\phi)$ within our language for postconditions of awareness-set changing events, stating when a formula ϕ belongs to the new set. The syntax of this scheme can then be used to derive reduction axioms. The system for epistemic and factual change in (van Benthem et al., 2006) is an illustration, exploiting the fact that postconditions of its events were definable as simple syntactic substitutions.

Even at this preliminary stage, our private awareness analysis suggests interesting extensions of standard dynamic epistemic logic.

8. Other approaches

Many approaches deal with the problem of logical omniscience, and we mention only a few. Our paper relates to these as follows.

First, we have provided a dynamic account for implicit and explicit information in terms of epistemic actions of *considering*, *neglecting* and *implicitly observing*. This extends the original approach in Fagin and Halpern (1988), where no explicit dynamics is presented except for temporal transitions without internal structure.

Second, unlike in recent approaches to inferential dynamics (Jago (2009), van Benthem (2008), Velázquez-Quesada (2009)), our agents' explicit knowledge is not only about factual formulas. It also includes 'higher' information about their own information.

Third, our multi-agent framework extends Grossi and Velázquez-Quesada (2009) by allowing us to handle public, private and even unconscious versions of the *considering* and *neglecting* actions.

Finally, we do not claim that our proposals make these earlier proposals obsolete. In particular, we still feel that our definition-based "reductionist" approach to explicit knowledge also has its drawbacks. In an intuitive sense, explicit knowledge may be *sui generis*, and, for instance, non-refutational uses of inference may also be considered as 'single events' that just increase this explicit knowledge. Even so, we hope to have shown the potential of the reductionist approach.

9. Conclusions and further directions

Our paper shows how a significant informational dynamics can take place over existing epistemic awareness models, generalizing acts of observation and inference. We also show how this leads to technical systems and results about these, in the spirit of dynamic epistemic logic. We have by no means explored all aspects of our proposal: our multi-agent setting can describe many more agent activities than what we have shown. Also, many technical issues remain unresolved: in particular, the issue of schematic validities and action algebra, and the precise logic of postconditions, dependent on their format of definition.

A most urgent desideratum in our view concerns the clear interpretation shift in our multi-agent section. It was *beliefs* of agents that made more sense there than *knowledge*. But once we take beliefs seriously, we should redo our analysis in the setting of dynamic logics for acts of *belief revision* (cf. Baltag and Smets, van Benthem (2008, 2007)) that work over epistemic plausibility models. Indeed, it makes much sense to relate belief revision, not just to new observations, but also to new inferences. This inferential dynamics should then also include special mechanisms that affect belief, rather than knowledge, that is: *default rules*. It is a significant issue how all this should be done. van Benthem (2009) is a first exploration, including some major changes in the plausibility relation, that now does not just order worlds, but worlds plus partial syntactic descriptions.

But our main point is not technical results, or concrete directions. It is rather the general picture of agency arising from our analysis. We replaced agents with "supernatural" abilities like omniscience by human ones that must, and can work to improve their information. The resulting mathematical model is rich, and much more attractive than the usual ones. We mentioned Sherlock Holmes at some point in our story, famous for combining observation and deduction. Dynamic logics are about what makes this tick. As visitors to Reichenbach Falls in Switzerland can see, our hero died a (fictional) death at the hands of the evil mathematician Professor Moriarty. But though these two minds were enemies, their fields were not. We hope to have shown how delicate philosophical issues can profit from mathematical modeling.

References

Baltag, A., L. S. Moss, and S. Solecki: 1999, 'The Logic of Public Announcements, Common Knowledge and Private Suspicious'. Technical Report SEN-R9922, CWI, Amsterdam.

Baltag, A. and S. Smets: 2008, 'A Qualitative Theory of Dynamic Interactive Belief Revision'. In: G. Bonanno, W. van der Hoek, and M. Wooldridge (eds.): *Logic and the Foundations of Game and Decision Theory (LOFT7)*, Vol. 3 of *Texts in Logic and Games*. Amsterdam University Press, pp. 13–60.

Fagin, R. and J. Y. Halpern: 1988, 'Belief, awareness, and limited reasoning'. *Artificial Intelligence* **34**(1), 39–76.

Grossi, D. and F. R. Velázquez-Quesada: 2009, '*Twelve Angry Men*: A Study on the Fine-Grain of Announcements'. In: X. He, J. F. Horty, and E. Pacuit (eds.): *LORI*, Vol. 5834 of *Lecture Notes in Computer Science*. pp. 147–160, Springer.

Jago, M.: 2009, 'Epistemic Logic for Rule-Based Agents'. *Journal of Logic, Language and Information* **18**(1), 131–158.

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

van Benthem, J.: 2008, 'Merging Observation And Access In Dynamic Logic'. *Journal of Logic Studies* **1**(1), 1–17.

van Benthem, J.: 2009, 'Logic, Mathematics, and General Agency'. In: P. Bour, M. Rebuschi, and L. Rollet (eds.): *Festschrift for Gerhard Heinzmann*. Nancy: Laboratoire d'histoire des ceinces et de la philosophie. To appear.

van Benthem, J. and M. Martínez: 2008, 'The Stories of Logic and Information'. In: P. Adriaans and J. van Benthem (eds.): *Philosophy Of Information*, Handbook of the Philosophy of Science. Amsterdam: North-Holland, pp. 217–280.

van Benthem, J., J. van Eijck, and B. Kooi: 2006, 'Logics of communication and change'. *Information and Computation* **204**(11), 1620–1662.

van Ditmarsch, H. and T. French: 2009, 'Awareness and forgetting of facts and agents'. In: *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technologies (WI-IAT 2009)*. Milan. To appear.

van Ditmarsch, H., A. Herzig, J. Lang, and P. Marquis: 2009, 'Introspective forgetting'. *Synthese (Knowledge, Rationality and Action)* **169**(2), 405–423.

Velázquez-Quesada, F. R.: 2009, 'Inference and Update'. *Synthese (Knowledge, Rationality and Action)* **169**(2), 283–300.