‘Tell It Like It Is’: Information Flow in Logic

Johan van Benthem, Amsterdam & Stanford, http://staff.science.uva.nl/~johan

1 Logic and information flow

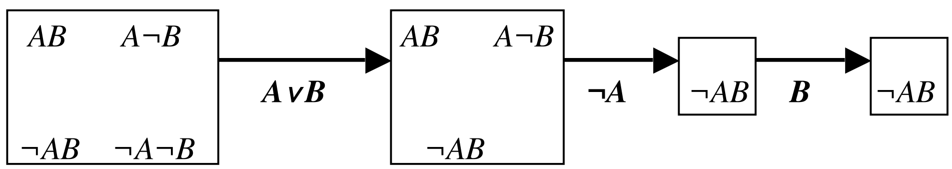
Logic is traditionally taken to be the systematic study of *inference*, drawing new conclusions as a means of elucidating or ‘unpacking’ information that is implicit in the given premises. This is the sense in which, for instance, the conclusion B of a typical propositional inference ‘from $A \vee B$, $\neg A$ to B’ tells us more about the situations satisfying the two premises. Inferences like this drive mathematical proof, automated deduction in computers, and the minds of the many people on buses and trains today doing their daily Sudoku puzzles, which involve long strings of propositional inferences.

But there is much more to information flow than deduction inside one’s own mind! The following list occurs with many authors, going back far into the Western – and the Indian – logical tradition. Suppose I want to know if a proposition P is true. Perhaps the easiest way of finding out is through *observation*: I just look at the relevant situation and see if P is the case. Only if this is not feasible for me, since P is about some remote, or otherwise inaccessible situation, I might try the option of *inference*, trying to deduce P or its negation from the stock of propositions already at my disposal. Finally, if this second road does not help either, I could use a third method, step out of my individual setting, and ask some reliable person, turning to multi-agent *communication*. Thus, information flow involves ‘social dynamics’: actions by different agents are crucial. The point of this paper is to show how all these phenomena fall squarely within the scope of modern logic, viewed as a general account of information flow. To emphasize the point, asking a question and giving an answer is just as ‘logical’ as drawing a conclusion! We will give examples of recent developments in this area of ‘logical dynamics’, and point at relevant parts of the literature.

In particular, we will show how mixtures of information are well within the scope of logical theory. In particular, one simple dynamic mechanism: giving new information through *public announcement*, gives fresh perspectives on many existing logical topics.

2 Epistemic logic: 'truth to the best of my information'

Information as range The best-known logical paradigm incorporating a natural semantic notion of information is *epistemic logic* – an approach from philosophy, but nowadays also found in computer science and economics. Models represent agents' current state of uncertainty by means of 'information ranges' listing all ways the actual world might be. New information will shrink these ranges. E.g., take the inference 'from $A \vee B$, $\neg A$ to B' once more, but now imagine that the premises come in one by one, starting from an initial situation where the agent has no knowledge whatsoever about A and B. Here is how the 'updates' work out for the two premises, ruling out 3 of the 4 options – while we can also see that updating with the statement B would not change the information state any more:



There are two crucial aspects to this scenario, viz. (a) *Statics*: what the agent knows or does not know at each stage of this process, and (b) *Dynamics*: the events themselves that trigger information flow. We start with task (a), the original focus in logical systems so far.

Logical basics The basic epistemic language describes knowledge in terms of "to the best of an agent's information". The syntax has proposition letters $p, q, \ldots,$ Boolean connectives $\neg, \vee$, and modal operators $K_i\phi$. The latter express that agent *i knows that* ϕ, while the dual $\langle i \rangle\phi = \neg K_i\neg\phi$ says that *i considers* ϕ *possible*. Models for the language are triples

$$\boldsymbol{M} = (W, \{\sim_i \mid i \in G\}, V),$$

where W is a set of worlds, the $\sim_i$ are binary accessibility relations between worlds, while V is a propositional valuation. The worlds ('states', 'situations', or whatever is relevant to the setting) in the set W represent the options for how the actual situation might be, while the relations $\sim_i$ encode the uncertainty, or alternatively, the current information of the agents:

$x \sim_i y$ says that, at world x, i considers y an option for being the actual world.

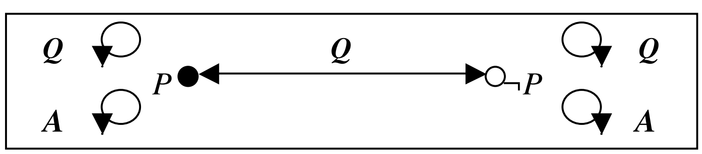
These accessibility relations may be different for different agents, who need not all have the same information. One often takes the $\sim_i$ to be equivalence relations, but this is not crucial, and it does not matter to what follows. In this setting, the semantic truth condition for the knowledge operator uses a universal quantifier over the current range:

Agents know what is true throughout their current range of uncertainty:

$\boldsymbol{M}, s \models K_i\phi$ iff for all t with $s \sim_i t$: $\boldsymbol{M}, t \models \phi$

The dual $<i>\phi$ is then the existential quantifier 'in some currently accessible world'. We will use the 'knowledge' terminology for the operator $K_i\phi$ in of what follows. Even so, the more neutral term '*to the best of i's information*' for $K_i\phi$ states better what the above universal quantification over the current information range really achieves.

Factual and higher-order information Bringing in more than one agent, the simple epistemic language can formulate complex scenarios. Consider a model for two agents ***Q***, ***A*** ('questioner' and 'answerer') with one world where the atomic fact P holds, and another where P fails. We assume that the real world (there always *is* one!) is the one indicated by the black dot in the picture – though this is an outsider's annotation, rather than the agents' own information. Labeled lines linking worlds indicate *uncertainties*. In particular, ***Q*** does not know which world is the actual one, while ***A*** is better informed: if the actual world has P then she knows that is the case, and if it does not, then she know that, too:



M

This diagram also encodes 'higher' information agents have about their own information and that of others. E.g., ***Q*** knows that she does not know if P, while ***A*** does. Hence it would be a good idea for ***Q*** to *ask a question* to ***A***, and find out. But before getting to that dynamic feature, let us take stock of the model ***M***, i.e., the current state of the informational process. In formulas of the epistemic language, the following are true in the world to the left:

$P, K_AP, \neg K_QP, K_Q\neg K_QP, K_Q(\neg K_QP \wedge \neg K_Q\neg P)$ (***Q*** knows that she does not know that P), $K_Q(K_AP \vee K_A\neg P)$ (***Q*** knows that ***A*** knows whether P)

Thus, epistemic logic can express complicated multi-agent patterns. In particular, *iterated* knowledge about oneself and others reveal something essential about the notion of information as used by humans, keeping our actions grounded in mutual expectations.

Group-level information In addition to describing individual agents and their interactions, epistemic logic also has further notions of *group knowledge*, doing justice to the fact that, when individual agents meet, they can form larger groups with new properties of their own. In particular, facts stated in public tend to create common knowledge: we all know the facts, but also that the others know it, and that they know that we know, and so on, up to any level of iteration. Technically, this emergent form of group knowledge $C_G\phi$ ('ϕ is common knowledge in group G') is read in the above information models as follows:

> $\boldsymbol{M}, s \models C_G\phi$ iff for all worlds t reachable from s by
> some finite sequence of $\sim_i$ steps ($i \in G$): $\boldsymbol{M}, t \models \phi$.

Another important epistemic group notion is *distributed knowledge*, which describes roughly what the group could come to know through internal communication:

> $\boldsymbol{M}, s \models D_G\phi$ iff for all t with $s \bigcap_{i \in G} \sim_i t$: $\boldsymbol{M}, t \models \phi$

Logic as information calculus Next, the valid laws of epistemic logic provide a 'calculus of information'. There are many axiomatic systems here, but we just give an illustration. Assuming that the accessibility relations are equivalence relations, the complete logic is axiomatized by the modal logic *S5*, on top of any complete classical propositional logic:

$Kj(\phi \to \psi) \to (Kj\phi \to Kj\psi)$	*Knowledge Distribution*
$Kj\phi \to \phi$	*Veridicality*
$Kj\phi \to KjKj\phi$	*Positive Introspection*
$\neg Kj\phi \to Kj\neg Kj\phi$	*Negative Introspection*

The complete system describes agents' own reasoning, but also our reasoning as theorists about what they know. Here are the required additional axioms for *common knowledge*:

$$C_G\,\phi \leftrightarrow \phi \,\&\, E_G\,C_G\,\phi \qquad \textit{Equilibrium Axiom}$$

$$(\phi \,\&\, C_G\,(\phi\rightarrow E_G\,\phi)) \rightarrow C_G\,\phi \qquad \textit{Induction Axiom}$$

Here $E_G\phi$ says that everyone in the group G knows that ϕ. The complete logic is decidable – be it more complex than propositional logic (describing interaction takes its toll). These laws can be applied in every concrete scenario, just like the principles of probability theory.

There are many more aspects epistemic logic viewed as a theory of information, including the *expressive power* of epistemic languages, and the proper *structural equivalences* between information models. Cf. van Benthem 2006A, van der Hoek & Pauly 2006.

3 Dynamic epistemic logic: learning from observation and communication

Structure and process It is hard to think of information in isolation from the processes which create, modify, and convey it. This combination is natural in many disciplines. In computer science, one designs data structures in tandem with the processes that manipulate them for specific tasks. The same duality occurs in philosophy and linguistics, witness the saying ‘Meaning Is What Meaning Does'. We can only give good representations of linguistic expressions when we state how these are going to be *used*: in communication, inference, and so on. Thus, *structure should always come in tandem with a process!* And so, which processes drive the notion of information? There is a great variety of candidates: computation, inference, update, revision, correction, question answering, communication, interaction, games, learning, and so on. Some of these processes concern activities of single agents, while others are multi-agent 'social' phenomena – with interesting connections between the two, as in the passage from multi-agent argumentation to single-agent formal proof. Some of these processes will be mentioned in the sections of this paper.

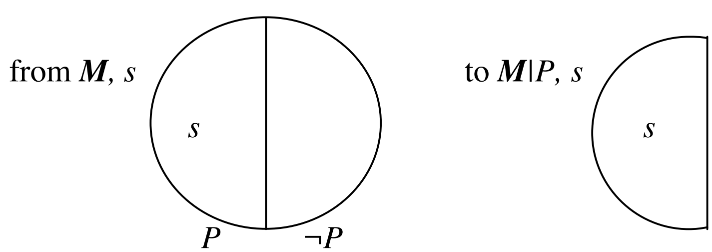
Questions and answers For a start, as a simplest epistemic process – repeated thousands of times throughout our lives –, consider the following conversational scenario:

(a) ***Q*** asks ***A*** the question "*P?*",

(b) ***A*** gives the true answer "Yes".

Here the situation depicted in the current model $\boldsymbol{M}$ changes, since information will start to flow. If this episode is simple cooperative Gricean communication, the question (a) itself already conveys new information, viz. that $\boldsymbol{Q}$ does not know the answer, but also, that she thinks $\boldsymbol{A}$ might know. In general, this can be highly informative and useful to the answerer. Next, the answering event (b) conveys that $\boldsymbol{A}$ knows that P, and its public announcement in the group $\{\boldsymbol{Q}, \boldsymbol{A}\}$ makes sure that $\boldsymbol{Q}$ now also knows that P, that both agents know this about each other, and so on to the earlier-mentioned common knowledge. In particular, we see that learning both factual and higher epistemic information can be important in multi-agent communication: an earlier point, but now returning in a dynamic setting.

Information update as model change Our question–answer scenario involves the simplest form of information flow, triggered by *public announcement* $!P$ of some fact P. This covers both communicative and single-agent settings, as it applies to both the *Observation* and *Communication* items on our original list. One can think of $!P$ as a real statement coming from some authoritative source in conversation – or more generally, as a totally reliable observation based on one's senses. (Of course, observation might also be cast as a conversation between 'Me' and 'Nature'.) If I see that the Ace of Spades is played on the table, I now come to know that no one of us holds it any more. Of course, we can be mistaken about what we hear or see, but logic starts with first understanding the idealized setting. These events $!P$ of incoming new *hard information* change what I know. Formally, they trigger changes in the current epistemic model. More precisely, $!P$ eliminates all worlds incompatible with P, thereby zooming in on the actual situation. Thus the current model $(\boldsymbol{M}, s)$ with actual world s changes into its sub-model $(\boldsymbol{M}|P, s)$, whose domain is the set $\{t \in \boldsymbol{M} \mid \boldsymbol{M}, t \models P\}$. In a picture, one goes



Public announcement logic: compositional analysis of communication Typically, truth values of formulas change in an update step: agents who did not know P now do after $!P$. This leads to subtle phenomena, but one can keep track of them in *public announcement logic PAL*, extending epistemic logic with action expressions:

Formulas	P:	$p \mid \neg\phi \mid \phi \vee \psi \mid K_i\phi \mid C_G\phi \mid [A]\phi$
Action expressions	A:	$!P$

The fundamental semantic clause for the dynamic action modality is as follows:

$$\boldsymbol{M}, s \models [!P]\, \phi \quad \text{iff} \quad \text{if } \boldsymbol{M}, s \models P, \text{ then } \boldsymbol{M}|P, s \models \phi$$

Here is the complete calculus of information flow under public announcement:

Theorem 1 *PAL* is axiomatized completely by the usual laws of

Static epistemic logic plus the following *reduction axioms*:

$[!P]q$	$\leftrightarrow$	$P \rightarrow q$	*for atomic facts* q
$[!P]\neg\phi$	$\leftrightarrow$	$P \rightarrow \neg[!P]\phi$	
$[!P]\phi\wedge\psi$	$\leftrightarrow$	$[!P]\phi \wedge [!P]\psi$	
$[!P]K_i\phi$	$\leftrightarrow$	$P \rightarrow K_i(P \rightarrow [!P]\phi)$	

This logical calculus highlights basic issues of communication in a very compact format. In particular, the reduction axioms perform a step-by-step compositional analysis of the effects of arbitrary assertions, reducing them down to plain epistemic ones. Here, the central axiom (d) relates the knowledge agents have following new information to knowledge they had before. This highlights an important feature of logics of information flow, viz. *pre-encoding*. The current information state $\boldsymbol{M}$ already has the right information about knowledge after events $!P$ through *relativized knowledge* of the form $K_i(P \rightarrow$.

Group knowledge Finally, *PAL* can be extended to deal with common knowledge, but to analyze $[!P]C_G\phi$ we must enrich epistemic logic with a new notion of 'conditional common knowledge' $C_G(\phi, \psi)$ which says that ϕ is true in every world reachable with steps staying inside the ψ–worlds. Then we have the following valid reduction axiom:

$$[!P]C_G(\phi, \psi) \leftrightarrow C_G(P \wedge [!P]\phi,\ [!P]\psi).$$

There is much more to be said about this logical system – for which we refer to the technical literature. In particular, *PAL* has a modal bisimulation-based model theory, and it raises many new issues of expressive power and computational complexity. We refer to van Benthem 2006B, van Benthem, van Eijck & Kooi 2006 for more details about all this.

In what follows, we show how adding a *PAL*-style component to existing logics leads to many new issues and increased modeling power for logical systems.

4 Communication

PAL itself is the result of adding events of public announcement to an existing system, viz. epistemic logic. The resulting calculus of communication and learning raises many issues.

Agent ability and diversity There is still more to axiom (d) than what we said so far. It also has a remarkable *interchange* between two operators: the dynamic $[!P]$ and the epistemic K_i. Such a commutativity law clearly fails for knowledge and action in general. For instance, you know right now that after drinking you are terribly boring. But the tragedy of life is that, after drinking, you do not know that you are boring... The reason is that drinking, unlike mere observation or talk, impairs agents' epistemic abilities. More precisely, knowledge action interchanges presuppose unimpaired memory, i.e., *Perfect Recall*. Thus, studying information flow immediately raises issues of *agent ability*. We can also describe agents with less idealized properties, including forgetfulness, and even groups of agents with different abilities, a reality in real life (van Benthem & Liu 2004, Liu 2006).

Keep talking Other noteworthy features of *PAL* reside in its further validities. Here is one:

$$[!P][!Q]\phi \quad \leftrightarrow \quad [!(P \wedge [!P]Q)]\phi \qquad \#$$

What this says is that any effect of two consecutive assertions $!P$, $!Q$ could also have been produced by making *only one assertion*. You may think this is just the conjunction $!(P \wedge Q)$ – but you would be wrong. $!P$ may have changed the current model, so after that, Q may have changed its truth value. The more complex equivalence # really gets things right here.

Paradoxes of coming to know What do we learn from a public announcement $!P$? Is not its sole point to produce common knowledge of P? Or, in *PAL* terms, is the following valid:

$[P!]C_G P?$

This was not on our list of axioms. And indeed, though intuitively plausible at first sight, this principle of epistemic action founders on *Moore-type* statements. Consider this case:

$\neg K_{you}P \ \& \ P$ "you do not know it, but *P*"

This proposition may well be true right now. But once uttered, it will make you know that *P*, and indeed, *P* becomes common knowledge, invalidating $\neg K_{you}P \ \& \ P$ as a whole. Thus, announcing a true Moore sentence makes it false in the new information state! Natural self-refuting cases are ignorance assertions in well-known epistemic puzzles such as 'Muddy Children'. Even so, many assertions do produce common knowledge when announced, but it is not known yet precisely which syntactic forms *P* in public announcements *!P* produce common knowledge – an open logical problem known as the 'Learning Problem'. These matters take us right into contemporary epistemology. Consider the simple but exasperating *Fitch Paradox* concerning the Verificationist Thesis "*What is true can be known",* or in epistemic terms, plus some unspecified modality for the "can" (van Benthem 2004):

$$\phi \rightarrow \Diamond K\phi.$$

Fitch gave the following simple but surprising argument showing that the Verificationist Thesis is inconsistent. The heart of the problem is again a Moore-style assertion:

$$P\wedge\neg KP \ \rightarrow \ \Diamond K(P\wedge\neg KP) \rightarrow \Diamond(KP\wedge K\neg KP) \ \rightarrow \ \Diamond(KP\wedge\neg KP) \rightarrow \bot$$

Therefore, we may conclude that $P \rightarrow KP$, i.e., knowledge and truth would collapse!

Iterated announcement, learnability, and conversation We have seen that no public announcement can make us know a Moore-type proposition. So, what *can* be learnt in principle through single or *repeated* public announcements of true propositions? There is an incipient literature on this. E.g., Miller & Moss 20005 have shown that with repeated announcements, complexity of the logic of communication increases considerably. *PAL* with added iteration operators $!P^*$ for arbitrary finite sequences of repetitions of *P* (as long as it is true) is *undecidable,* non-axiomatizable, and even much worse in complexity. On the other hand, Balbiani et al. 2007 have just axiomatized a logic *APAL* with an added operator $<>\phi$ standing for 'after some public announcement, ϕ is the case' which describes effects of future information though without specifying which form it will take. Finally we mention another recent development along these lines. *PAL* assumes that any true proposition P can be publicly announced by an action $!P$. But in actual conversation, there may be restrictions on what can be said, and in what order. We can represent this in temporal *conversation models* specifying the admissible sequences of true announcements, starting form some initial informational situation. But once we do this, the axioms of *PAL* as they stand will no longer be true. In particular, an existential base principle like

$$<!P>T \leftrightarrow P$$

no longer holds – since even when P is true, its announcement may be prohibited by the scenario of the model. (The direction from left to right will still hold in conversation models.) Thus, simple assertions $<!P>T$ now convey procedural information about available assertions in the conversation. Therefore, the correct base axiom becomes:

$$<!P>q \leftrightarrow <!P>T \;\&\; q$$

Using this and similar observations for other reduction axioms, van Benthem, Gerbrandy & Pacuit 2007 have shown that the logic of epistemic conversations remains manageable:

Theorem 2 The logic of public announcement over conversation models is decidable.

5 Belief revision

Doxastic logic In general, there is a conceptual distinction between information coming from some source, and agents' attitudes and responses to it. Agents can have other attitudes toward propositions than knowledge, including belief or doubt. In particular, 'doxastic logics' analyze assertions $B_i\varphi$ for '*agent i believes that* φ'. The semantics adds a new idea to the information ranges in our epistemic modeling so far. We assume further gradations, in the form of a *plausibility ordering* of worlds as seen from some vantage point:

$\leq_{i,s} xy$ in world s, agent i considers y at least as plausible as x.

Thus, while the earlier ranges of epistemic alternatives corresponded to the strict information that we have, the same ranges ordered by plausibility give finer gradations. In particular, we now define belief semantically as '*truth in the most plausible options*':

$$\boldsymbol{M}, s \models B_i\phi \text{ iff } \boldsymbol{M}, t \models \phi \text{ for all } t \text{ which are maximal in the ordering } \lambda xy. \leq_{i,s} xy.$$

Here is an elementary example. Consider a model with two worlds that are epistemically accessible, but the one with ¬P considered more plausible than the other:

$$\boldsymbol{P} \underset{\leq}{\longrightarrow} \neg P$$

In this model, at the actual world where $\boldsymbol{P}$ holds, the agent does not know whether P, but she does (mistakenly!) believe that $\neg P$. It is crucial that our beliefs can be false. There are some technical complications in making this work in infinite models, but this is the idea. As with epistemic logic, there are complete doxastic logics and a whole theory around them (Fagin et al. 1995). In general the resulting logics also analyze the *interplay* between knowledge and belief, with information models having two relations $\sim_i$, $\leq_j$ which can be entangled in various ways, reflecting a stand on such issues as whether knowledge implies belief, or whether one knows one's beliefs. While inter-relations between attitudes toward information are an important topic, we focus on belief in what follows.

Pre-encoding and conditional belief In doxastic logic, one soon finds that mere beliefs are not sufficient for explaining agents' behaviour. We want to know what they would do in certain scenarios where they receive new information. This requires *conditional belief*:

$$\boldsymbol{M}, s \models B_i(\phi \mid \psi) \text{ iff } \boldsymbol{M}, t \models \phi \text{ for all worlds } t \text{ which are } \textit{maximal} \text{ for } \lambda xy. \leq_{i,s} xy \text{ in the set } \{u \mid \boldsymbol{M}, u \models \psi\}.$$

Conditional beliefs $B_i(\phi \mid \psi)$ are like logical conditionals in general, in that they express what might happen under different circumstances from where we are now. In particular, they *pre-encode* beliefs in ϕ that we would have if we were to learn new things ψ. The analogy is so close that conditional belief on reflexive transitive plausibility models and satisfies exactly the laws of the minimal conditional logic given by Burgess and Veltman. We will return to this setting later on when discussing the mechanics of belief revision.

Hard and soft information Combined epistemic-doxastic-conditional logics also suggest a richer picture of incoming information. There is *hard information* in the current range of epistemically accessible worlds. Next, fine-structure may be present on these ranges, through plausibility orderings. These can be viewed either subjectively as representing agents' attitudes, or more objectively, as the result of receiving what might be called *soft information* making some proposition P more plausible, but not necessarily ruling out $\neg P$-worlds. It is the total interplay of all these attitudes which describes our stances toward information, and the way they are affected by new incoming information. Thus, we get a mixture of information per se, and the ways in which agents take it – and this logical entanglement is such it is hard to say where one notion ends and the other begins.

Changing beliefs The dynamic process perspective on information change underlying *PAL* also applies to our *beliefs*, and how to *revise* these on the basis of incoming information. This involves changes, not in the range of available worlds or epistemic accessibility, but rather in the *plausibility orderings* $\leq_{i,\ s}\ xy$ among worlds. How this works precisely, depends on the incoming signal. When we receive hard information $!P$, update will proceed by world elimination as before. We then get new beliefs related to our earlier conditional beliefs, and the counterpart to the above reduction axiom (d) are laws saying which new beliefs – and indeed, conditional beliefs – are acquired (van Benthem 2007B):

Theorem 3 The logic of conditional belief under public announcements is axiomatized completely by (a) any complete base logic of $B_i(\phi / \psi)$ for one's chosen models, (b) *PAL* reduction axioms, plus (c) a reduction axiom for conditional beliefs:

$$[!P]\, B_i(\phi / \psi) \quad \leftrightarrow \quad P \rightarrow B_i(\, [!P]\phi \mid P \wedge [!P]\psi)$$

Hard information involves some interesting phenomena with belief. True information can be misleading! For instance, consider this model, where all worlds are epistemically accessible, with a plausibility ordering $1 \leq 2 \leq 3$. Here the agent believes that p in the actual world 1, but for the wrong reason, as she considers world 3 most plausible:

$$\boxed{\mathbf{1}\ p, q \quad \leq \quad 2\ \ r \quad \leq \quad 3\ p, s}$$

Now suppose that a true public announcement $!\neg s$ is made. This eliminates world *3*, but in the remaining model with domain *{1, 2:*

1 *p, q*	$\leq$	*2 r*

the agent now believes, incorrectly, that $\neg p$. Following Stalnaker, Baltag & Smets 2006 have emphasized that there is room for another natural agent attitude here, viz. of *safe belief*, being those beliefs which cannot be changed by new true information. This provides an additional robustness to these beliefs, moving them closer to knowledge on an epistemic scale. Technically, safe beliefs are about those formulas ϕ which hold *in all worlds that are least as plausible as the current one*. Thus, a dynamic perspective on information change can also suggest new static epistemic-doxastic operators.

Genuine belief revision Still, we are not yet dealing with genuine belief revision. The latter rather occurs with *soft information* concerning a proposition *P*, as mentioned above. This just increases our 'preference' for *P*-worlds, without totally abandoning the others. Soft information leads to plausibility change, not world elimination. This can come in various sorts. A quite typical 'belief revision policy' in this spirit is *lexicographic upgrade* $\Uparrow P$ (Rott 2005) which replaces the current ordering relation ≤ between worlds by the following: *all P-worlds become better than all ¬P-worlds, and within those two zones, the old ordering remains*. Belief changes under such policies can be axiomatized completely in *PAL*-style. E.g., here is the logic for $\Uparrow P$ revision (van Benthem 2007B). It looks complex – but after all, we are now describing a more subtle informational process than epistemic update:

Theorem 4 The dynamic logic of lexicographic upgrade is axiomatized completely by the logic of conditional belief + compositional analysis of effects of revision:

$$[\Uparrow P]\, q \leftrightarrow q, \qquad [\Uparrow P]\, \neg\phi \leftrightarrow \neg[\Uparrow P]\phi, \qquad [\Uparrow P]\, (\phi\wedge\psi) \leftrightarrow [\Uparrow P]\phi \wedge [\Uparrow P]\psi$$

$$[\Uparrow P]\, B(\phi|\psi) \leftrightarrow (\langle\rangle(P \wedge [\Uparrow P]\psi\,) \wedge B([\Uparrow P]\phi \mid (P \wedge [\Uparrow P]\psi\,))$$

$$\vee\, (\neg(\langle\rangle(P \wedge [\Uparrow P]\psi\,) \wedge B([\Uparrow P]\phi \mid [\Uparrow P]\psi\,))$$

Here <> is the existential epistemic modality saying 'in some world of the current range'. Richer dynamic doxastic logics handle many further policies (cf. Baltag & Smets 2006, Baltag, van Ditmarsch & Moss 2007). There are also strong analogies between plausibility reordering and dynamic logics of *preference change* (van Benthem & Liu 2006).

6 Inference and observation:

Entanglement Let us now go back tot the original examples in Section 1. Actual planning and problem solving often involves an entanglement of two logical processes: *inference* and *update*. E.g., in solving a puzzle or playing a game, I update my current representation of the facts with new information (say, an observation of a card being played) and read off the new diagram, but I also quickly infer things from information already at my disposal. Likewise, philosophical accounts of intentional behaviour (Paul 2007) have stressed the interplay of internal inference making us better aware of what the intention involves, and relevant observations keeping us on track while following a plan of action. The dynamics of the two processes is quite different. While we have said a lot about update, we have said nothing at all about how to represent growth of information through inference steps.

Information flow in inference Indeed, there is an issue here which has been labeled the 'scandal of deduction'. We all agree that inference can be informative – but in just which sense? Clearly the information states have to be more fine-grained now than sets of worlds. As we saw in Section 1, the conclusion in the inference 'from $A \vee B$, $\neg A$ to B' added nothing in terms of update to what the premises had already established. There is no consensus about how to model the finer grain size, though interesting proposals exist by various authors cf. Allo 2007. In this section, we will take an extremely simple example, inspired by Jago 2006, and through his work, by logic programming. Consider information states consisting of sets of propositional literals $(\neg)p$. In this setting, information generating events are applications of Horn-clause rules of the form p_1 & …, & $p_n \to q$. Thus, an initial information state might grow step-by-step as in the following example:

$$\boxed{p, \neg q, r} \xrightarrow{(p \,\&\, q) \to s} \boxed{p, \neg q, r, s} \xrightarrow{(t \,\&\, r) \to q} \boxed{p, \neg q, r, s, \neg t}$$

A modal logic for this would have proof rules as action expressions, and we can then describe the modal theory of these sets of literals connected by labeled inference steps with operators of the form $<rule>\phi$. Starting from some initial set, the information states obtained in this way may be viewed as partial descriptions of the set of all models in our earlier semantic sense satisfying the initial literals as well as all the Horn clause rules used in generating the further inferential information states.

Two-level dynamic logics Combining this perspective with the *PAL* paradigm, we can now have a two-stage procedure integrating inferential information flow with update through observation. For simplicity, take a propositional language: we omit subtleties of epistemic dynamic operators in this section. Also, we do not provide precise formal definitions, but merely sketch the main ideas. A semantic model is now built using valuations explicitly accepting or rejecting each proposition letter, i.e., complete worlds in the earlier sense for our simple language. But the model also specifies a family of inference rules ***R*** whose corresponding conditionals are valid in the given set of valuations. Finally, it has partial sets of literals X which do not yet decide every proposition letter. Each such set X stands in a compatibility relation with all total valuations V containing it. More precisely then,

> A model ***M*** is a *set of pairs (V, X)* where X is a set of literals compatible with V, together with a family R of Horn clause inference rules.

One can think of this as an agent's current semantic range (in the earlier sense) about the empirical facts in the actual world *plus* its current inferential approximations to the worlds in it. Of course, in a matching logical language, we will need to have suitable static operators accessing both components. Here a knowledge operator $K\varphi$ might operate on V components only, just as before. Next, one could access the X component by an operator $I\varphi$ saying that the agent 'realizes that φ' if the literal φ is actually present in X. In this setting, there are several natural kinds of informative action. First, consider 'internal elucidation':

> *Inferential steps* do not change the V components of the current state,
> but merely take some available rule and add literals to the current X
> components where it applies (its premises are available).

The second kind of action is the familiar running thread in this paper:

> *PAL*-style *public announcements !P*, removing all states *(V, X)*
> from the current one where V fails to satisfy P.

But there may be other 'intermediate' actions with something of both. In particular, think of the act of seeing in some conscious sense. We can model this with

> *Explicit observations* $+q$ of literals q, which operate like $!q$, but then also place q directly into all X components of pairs (V, X) in the remaining model.

Putting all this into a language with dynamic modalities, we can make combined assertions which describe the fine-structure of these informational processes. E.g.,

$$Ip \rightarrow [+q]\langle (p \& q) \rightarrow r\rangle r$$

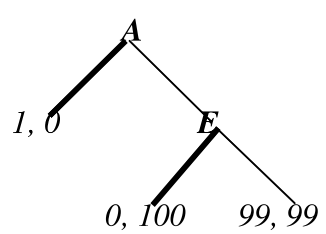
says that, if the agent realizes that p in the current information state, and she then explicitly observes that q, then she can come to realize that r by explicitly performing an inference step invoking the Horn clause rule $(p \& q) \rightarrow r$. This is just a simple example. But it does show that dynamic logics of both internal inferential elucidation and external observation make sense. A full-fledged logic with a language containing the above two static modalities $K\varphi$ and $I\varphi$ plus dynamic modalities for the above three informational actions remains to be developed in formal detail. But we trust the reader will have seen by now that this sort of mixed informational calculus is both a feasible and a worth-while endeavour.

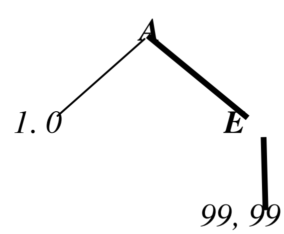
Cognitive architecture Two-level models seem closer to the actual working of the human mind, where different processes run in parallel, as has been pointed out by cognitive psychologists. In particular, it has been suggested that belief formation involves some sort of inferential component (Velazquez 2007 proposes Peircean abduction for this purpose). Cf. the two-level view of pure information processing combined with belief formation on the basis of this information, where belief changes occur in a sort of 'reflective dynamics' in Rott 2007, tracking models at one level in terms of 'belief bases' at another.

7 Games

From logic to games As was noted in Section 1, informational processes are *dynamic*, involving actions and events on a par with static information structures. They are also *social*, as basic scenarios involve more than one agent, with notions of group knowledge crucial to understanding information flow. This dynamic social character of information comes together in the concept of *intelligent interaction* between agents (LogiCCC 2006). And interaction, of course, is the main topic investigated in game theory. There are many current interfaces between logic and game theory (de Bruin 2004, van Benthem 2005, van der Hoek & Pauly 2006), having to do with epistemic analysis of equilibrium solutions in terms of rational action, or logical analysis of games as a paradigm for generalized interactive computation (Parikh 1985, Abramsky 2007). Here we merely mention a few appealing interfaces which are related to introducing public announcements. What happens in dynamic scenarios of information flow where games can *change*?

Promises and intentions The famous game-theoretic solution procedure of Backward Induction solution (*BI*) for extensive games often produces 'bad equilibria' representing some 'socially' undesirable outcome. An example is given to the left in the following picture, where the bad equilibrium *(1, 0)* predicted by reasoning about players' 'rationality' makes both hugely worse of than the cooperative outcome *(99, 99)*. One way of doing something about this, following van Benthem 2007C, is by making *promises*. These may be viewed as public announcements of intentions. E.g., ***E*** might promise that she will not go left, changing the game to the new one depicted on the right– and the new equilibrium *(99, 99)* results, making both players better off by restricting the freedom of one of them!





But one can also announce other things. Van Otterloo 2005 has a more general logic of players' strategic powers and preferences, where models change through announcement of players *intentions*. Complete logics for such systems arise from intertwining *PAL* as above with modal logics of actions and preferences, in a relatively straightforward manner.

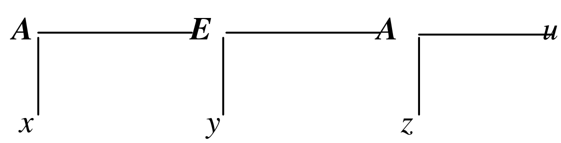
'Rational Dynamics' Van Benthem 2003 uses public announcement to analyze further solution procedures. *Strategic games* induce epistemic models ***M*** of strategy profiles with preferences and uncertainty relations for players who know their own strategy, but not that of the others – and they have their own solution procedures, such as Iterated Removal of Strictly Dominated Strategies. Then a combined modal-preference language can formulate statements of Weak Rationality ("no player chooses a move which she knows to be worse than some other available one") and Strong Rationality ("every player chooses a move which she thinks may be the best possible one"). When announced, these eliminate worlds, and iterating this, one finds a smallest sub-model where announcements have no further effect: *WR* or *SR* are now common knowledge.

Theorem 5 The result of iterated announcement of *WR* is the usual solution concept of Iterated Removal of Strictly Dominated Strategies; and it is definable inside ***M*** by means of a formula of a modal μ–calculus with inflationary fixed-points. The same for iterated announcement of *SR* and game-theoretic Rationalizability.

If *A* has ‘existential-positive’ syntax (*SR* does), the definition is in a standard μ–calculus. In this scenario of 'internal deliberation' players keep recalling their rationality. But one can announce many other statements. A similar analysis applies to extensive games:

Theorem 6 The Backward Induction solution for extensive games is obtained through repeated announcement of the assertion "no player chooses a move all of whose further histories end worse than all histories after some other available move".

Here is how this works out for a Centipede game, with branches indicated by name:



Stage 0: rule out u, Stage 1: rule out z, Stage 3: rule out y.

The procedure ends in largest sub-models where players have common belief of rationality.

Again, there are other announcement options than the ruthless egotism of *BI*. Van Benthem 2003 discusses history-oriented announcements, where players steer future actions by reminding themselves of the *legitimate rights of others*, because of ‘past favours received’.

Excursion: strategies Games are not just about single moves, but also about longer-term interactive strategies. If we use propositional dynamic logic *PDL* for strategies and moves in games (cf. van Benthem 2002), this leads to the obvious logic *PDL+PAL* adding public announcements $[!A]$. It is easy to show that *PDL* is closed under relativization to definable sub-models, both in its propositional and its program parts, using an operation $\pi|A$ for programs π which basically wraps them in tests $?A$.

Theorem 7 *PDL+ PAL* is axiomatized by merging their separate laws while adding the following reduction axiom: $[!A]\{\sigma\}\,\phi \leftrightarrow (A \rightarrow \{\sigma|A\}[!A]\phi)$.

Of course, there is more structure to games than just moves and strategies, and *PAL*-style announcement scenarios also make sense with combined epistemic preference languages.

8 Further examples and further directions

More public dynamics Our discussion has by no means exhausted the potential for adding public announcement operators to existing logics, making their underlying mechanisms of information flow explicit. Without much further elaboration, we mention a final example which seems well-worth exploring. Inspired by earlier work by Yamada, van Benthem & Liu 2006 consider preference changes due to commands or suggestions, and interleave this with new information via public announcement. One setting where this combination occurs is in *social choice theory*, when analyzing voting patterns. Mathematical frameworks in this area usually consider just preference aggregation functions, proving constraints on what criteria these can satisfy. But in reality, there is a stepwise dynamics to this process. Prior to voting, there is deliberation, with incoming new information. And this information can affect agents' preferences in various ways. E.g., if my preferences are for things which I consider most plausible in terms of my beliefs, then new *factual information* can change my beliefs and hence also my preferences. I wanted to vote for a party promising Golden Mountains, but when I learn that no such geological formations exist, I change my vote. And there are even more subtle phenomena. Learning about *other people's preferences* can also be relevant, e.g., when I want to align my votes with theirs. This is why in deliberation, so-called 'straw votes' are useful, telling us in a preliminary manner how the preferences of the other participants lie. Deliberation and voting seems to be a rich source of experience, which invites logical formalization involving both static epistemic/doxastic preference languages, and mechanisms which input new information about the empirical facts, and about other agents involved. Adding public announcement to social choice theory seems an excellent initial step toward combining 'top-down' analysis of voting procedures with 'bottom-up' accounts of procedural justice and rational deliberation.

Broader frameworks There are many further strands in the current literature extending public announcement as a mechanism of information flow. E.g., often agents have only *partial powers of observation*. General dynamic-epistemic logics (Baltag, Moss & Solecki 1998, van Benthem, van Eijck & Kooi 2006) work with 'event models' ***A*** describing complex scenarios where agents have different observational access. So-called 'product update' then turns a current model ***M*** into a model ***MxA*** recording information of different agents. Product update redefines the universe of relevant possible worlds, and the epistemic accessibility relations between them. Conversations, games, internet transactions, and other real activities are like this. Other structures arise when we pay attention to computational *program structures* in conversation, which involve sequential composition, guarded choice, iteration, and even parallel composition. Van Benthem 2007A discusses how one can combine all this with explicit logic to give an account of intelligent interaction between rational agents. Next, as for the longer term, there are *epistemic-temporal* logics describing universes of all possible temporal evolutions of informational processes, used in philosophical logic, computational logics, game logics, learning theory, situation calculus, belief revision, and other areas. Van Benthem & Pacuit 2006 is a survey of the vast literature drawing many strands together.

9 References

S. Abramsky, 2007, 'Information, Processes and Games', Computing Lab, Oxford University. In P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier, Amsterdam, to appear.

P. Allo, 2007, *On Logics and Being Informative*, Dissertation, Philosophical Institute, Free University Brussels.

P. Balbiani, A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, T. de Lima, 2007, 'What can we achieve by arbitrary announcements?', *Proceedings TARK* 2007, University of Namur.

A. Baltag, H. van Ditmarsch & L. Moss, 2007, 'Epistemic Logic and Information Update', in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier, Amsterdam, to appear.

A. Baltag, L. Moss & S. Solecki, 1998, 'The Logic of Public Announcements, Common Knowledge and Private Suspicions', *Proceedings TARK 1998*, 43 – 56, Morgan Kaufmann Publishers, Los Altos.

A. Baltag & S. Smets, 2006, 'A Qualitative Theory of Dynamic Interactive Belief Revision', University of Oxford and Free University Brussels. To appear in *Proceedings LOFT 2006*, University of Amsterdam Press, 2007.

J. van Benthem, 2002, 'Extensive Games as Process Models', *Journal of Logic, Language and Information* 11, 289–313.

J. van Benthem, 2003, Rational Dynamics and Epistemic Logic in Games, in S. Vannucci, ed., *Logic, Game Theory and Social Choice III*, University of Siena, Department of political economy, 19–23.Also in *International Game Theory Review* 9:1, June 2007, World Scientific, Singapore, 13 – 45.

J. van Benthem, 2004, 'What One May Come to Know', *Analysis* 64 (282), 95 – 105.

J. van Benthem, 2005, 'Open Problems in Logic and Games', in S. Artemov, H. Barringer, A. d'Avila Garcez, L. Lamb & J. Woods, eds., *Essays in Honour of Dov Gabbay*, King's College Publications, London, 229–264.

J. van Benthem, 2006A, Epistemic Logic and Epistemology: the state of their affairs, *Philosophical Studies* 128, 2006, 49 - 76.

J. van Benthem, 2006B, 'One is a Lonely Number: on the logic of communication', in Z. Chatzidakis, P. Koepke & W. Pohlers, eds., *Logic Colloquium '02,* ASL & A.K. Peters, Wellesley MA, 96 – 129.

J. van Benthem, 2007A, 'Computation as Conversation'. To appear in *Proceedings Computation in Europe 2005*.

J. van Benthem, 2007B, 'Dynamic Logic of Belief Change', ILLC Amsterdam. To appear in *Journal of Applied Non-Classical Logics*.

J. van Benthem, 2007C, 'Rationalizations and Promises in Games, *Philosophical Trends*, special issue on logic, Chinese Academy of Social Sciences, Beijing.

J. van Benthem, J. van Eijck & B. Kooi, 2006, 'Logics of Communication and Change', *Information and Computation* 204(11), 1620 – 1662.

J. van Benthem, J. Gerbrandy & E. Pacuit, 2007, 'Merging Frameworks for Interaction: *DEL* and *ETL*', ILLC Amsterdam & Informatics Torino. *Proceedings TARK* 2007, University of Namur.

J. van Benthem & E. Pacuit, 2006, 'The Tree of Knowledge in Action', *Proceedings Advances in Modal Logic,* ANU Melbourne.

J. van Benthem & F. Liu, 'Diversity of Logical Agents in Games', *Philosophia Scientiae 8:2,* 163 – 178.

J. van Benthem & F. Liu, 2006, 'Dynamic Logic of Preference Upgrade', ILLC Amsterdam. To appear in *Journal of Applied Non-Classical Logics*.

B. de Bruin, 2004, *Explaining Games*, Dissertation. ILLC, University of Amsterdam.

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

W. van der Hoek & M. Pauly, 2006, 'Modal Logic for Games and Information', in P. Blackburn, J. van Benthem & F. Wolter, eds., *Handbook of Modal Logic*, Elsevier, Amsterdam, 1077 – 1148.

M. Jago, 2006, *Logics for Resource-Bounded Agents*, Dissertation, Department of Computer Science, University of Nottingham.

F. Liu, 2006, 'Diversity of Agents', Paper presented at ESSLLI Workshop on Resource-Bounded Agents, Malaga. Report 2006-37, ILLC, University of Amsterdam.

LogiCCC, 2007, *Modeling Intelligent Interaction: Logic in the Humanities, Computational and Social Sciences*, European Science Foundation, Strassburg.

J. Miller & L. Moss, 2005, 'The Undecidability of Iterated Modal Relativization', *Studia Logica* 79(3): 373 – 407.

S. van Otterloo, 2005, *A Strategic Analysis of Multi-Agent Protocols*, Dissertation DS-2005-05, ILLC, University of Amsterdam & University of Liverpool.

R. Parikh, 1985, 'The Logic of Games and its Applications', *Annals of Discrete Mathematics* 24 (1985), 111 – 140.

S. Paul, 2007, 'From Intention to Knowledge', Department of Philosophy, Stanford University.

H. Rott, 2007, 'Information Structures in Belief Revision', Philosophical Institute, Regensburg. To appear in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier, Amsterdam.

F. Velazquez, 2007, 'Belief Revision as Abduction', Working paper, ILLC, University of Amsterdam.