# COGNITION AS INTERACTION

Johan van Benthem, University of Amsterdam & Stanford University

January 2006

*Abstract*
Many cognitive activities are irreducibly social, involving interaction between several different agents. We look at some examples of this in linguistic communication and games, and show how logical methods provide exact models for the relevant information flow and world change. Finally, we discuss possible connections in this arena between logico-computational approaches and experimental cognitive science.

## 1 Introduction: from lonesome cognitive performance to social skills

When King Pyrrhus of Epirus, one of the foremost well-educated generals of his age, had crossed over to Italy for his famous expedition, the first reconnaissance of a Roman camp near Tarentum dramatically changed his earlier perception of his enemies (Plutarch, "Pyrrhus", Penguin Classics, Harmondsworth, 1973):

> Their discipline, the arrangement of their watches, their orderly movements, and the planning of their camp all impressed and astonished him – and he remarked to the friend nearest him: "These may be barbarians; but there is nothing barbarous about their discipline".

It is intelligent social life which often shows truly human cognitive abilities at their best and most admirable. But textbook chapters in cognitive science mostly emphasize the apparatus that is used by single agents: reasoning, perception, memory, or learning. And this emphasis becomes even stronger under the influence of neuroscience, as the only *obvious* thing that can be studied in a hard scientific manner are the brain processes inside individual bodies. Protagoras famously said that "Man is the measure of all things", and many neuroscientists would even say that it's just her brain. By contrast, this very brief paper makes a plea for the irreducibly social side of cognition, as evidenced in the ways in which people communicate and interact. Even in physics, many bodies in interaction can form one new object, such as a solar system. This is true all the more when we have a meeting of many minds! Perhaps the simplest and yet most striking example of interactive cognitive behaviour is language use in conversation. This will be our key example in what follows.

## 2 Questions, answers, and the spectrum of language use

In modern logic and semantics, the dynamics of language use has come into focus. The individual speech acts that take place can only be understood in terms of mutual information that agents have about each other, and more generally, the resulting *interaction.* Consider a simple conversation like:

***Q*** *"Is the KNAW on this canal?"*
***A*** *"Yes."*

Here the questioner ***Q*** conveys that she does not know the answer, and probably also, that she thinks it possible that the answerer ***A*** does know. After the answer has been given, ***Q*** has not just learnt the fact that this canal is indeed the location of the Royal Academy. She also knows that ***A*** knows that location, and even that ***A*** knows that she knows it now, and so on. Ideally, the two agents have achieved *common knowledge*, to any depth of mutual iteration.

Now, this simple episode of communicative language use lies on a much longer natural chain of inter-locking cognitive abilities. First, before we can process the information in given statements, we need to understand what they say. This is seldom a purely one-agent matter, but it rather involves a process of *interpretation* for reaching an equilibrium between speaker's and hearer's meaning. Bi-directional Optimality Theory describes part of this – but authors like Parikh 2002 and van Rooy 2002 have even full-fledged game-theoretic accounts of the crucial higher-level interactions here. Next, successful *communication* requires an account of the presuppositions and effects of various kinds of speech act. Just which information is passed exactly when we use a particular linguistic construction? This can be much more sophisticated than the above elementary scenario. E.g., in a classroom, if ***Q*** were a teacher, and ***A*** a student, we would not expect ***Q*** to be ignorant of the answer, and we definitely do not expect her to have illusions about the 'knowledgeability' of ***A***. Next, individual questions makes sense only when there is a broader intention behind them. This brings us to conversations with strategies for repeated asking and answering, as well as just the right amount of revealing and hiding information to achieve intended effects and reach goals. Conversations live at the level of *strategic games*, and again they achieve some sort of game-theoretic equilibrium between agents giving and receiving information. But games are again just episodes in a longer stream of linguistic behaviour, which involves accumulated memory over time. For instance, our ideas about the reliability of conversation partners may encode a lot of past experience, as well as the resulting expectations about the future. Thus, we get into the temporal analysis of long-term

*protocols and learning*. And moving beyond the aggregation level and life-span of single agents, we can even look at long-term linguistic practices in societies, or even evolutionary processes of language change. These long-term phenomena transcend the scope of standard logic, and eventually involve the mathematics of *dynamical systems*.

My own current interests lie mostly in the middle of this spectrum, viz. information update, belief revision, and games. Experience in this area has shown two things. First, there is enough substance to create exact theories – but also, such theories need to take their cues from quite diverse disciplines, such as linguistics, philosophy, logic, computer science, economics, and cognitive psychology. The next section provides a concrete example of this confluence, all from a logician's perspective.

## 3 Dynamic epistemic logic for communication

Communication involves update of information. To describe such processes in precise terms, we need to have an account of both the information states of groups of agents, and the basic and complex actions that can change one of these states into another. For this purpose, logical techniques turn out quite useful.

***Epistemic logic*** To analyze the above question/answer episode, we use a well-known system from the philosophical tradition, viz. *epistemic logic* – originally developed as a tool for analyzing epistemological notions and arguments. In a self-explanatory notation, here are some key features of our question-answer episode:

> ***Q*** asked a factual question *"P?"*, ***A*** answered truthfully *"Yes"*.
> For a truthful answer ***A*** must know that *P*: $K_A P$
> A normal cooperative question then has the two presuppositions
> (a) ***Q*** does not know if *P*: $\neg K_Q P \wedge \neg K_Q \neg P$
> (b) ***Q*** thinks it possible that ***A*** knows if *P*: $\langle Q \rangle (K_A P \vee K_A \neg P)$
> After the answer, *P* becomes common knowledge: $C_{(Q, A)} P$

***Information update*** Next, the *updates* that take place when speech acts occur change the relevant *information models*. E.g., here is a simple model for an initial situation where *P* holds, ***Q*** does not know this (Figure 1):

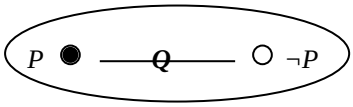


Figure 1

The line between the two worlds indicates the initial uncertainty of agent ***Q***. The absence of such a line for agent ***A*** indicates that the latter knows whether *P* is the case. This absence is transparent for ***Q*** who therefore does know that ***A*** knows if *P*.

Next, the update corresponding to a public assertion *P!* by ***A*** that *P* holds (which is the logical content of the above answer "Yes") removes the ¬*P*-world to the right in this model, leaving us with only the situation depicted in Figure 2:

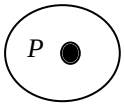


Figure 2

This new model shows pictorially how both agents know that *P* – and indeed, that *P* has become common knowledge among them. As for the general process here, public announcement changes the current information model by *world elimination*.

***Dynamic-epistemic logic*** Epistemic logic describes what agents know at the static intermediate stages of an informational process. But really, it is the state-changing actions *themselves* that seem of primary interest. To bring the latter into focus, one can borrow an idea from computer science, viz. the *dynamic logic* of programs, interpreted as actions that change states of some computing device. A combined *dynamic-epistemic logic* has formulas with both epistemic operators and action modalities. These formulas are interpreted in epistemic models ***M*** (i.e., diagrams as above) at particular worlds *s*, where they describe facts and properties of agents – as seen from the standpoint of *s*. In particular, the following dynamic action modality *[P!]* $\phi$ describes what happens in the updated model after something has been said:

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

The right-hand side of this truth condition says that $\phi$ is true in model ***M*** at world *s* after an action of true announcement of *P* has taken place to produce the model ***M***|*P*. More complex dynamic-epistemic formulas $\phi$ then describe what agents know (or do not know) after an announcement action has taken place:

| | |
|---|---|
| $[P!]K_i\phi$ | agent *i* knows that $\phi$ after *P* has been announced |
| $[P!]C_G\phi$ | after *P* has been announced, $\phi$ is common knowledge |

In this manner, we get a logical system that describes the effects of communication. Its set of valid principles is known to be simply axiomatizable, and it is even decidable by means of a mechanical algorithm that could run – in principle – on a computer.

Important advances in this paradigm of dynamic-epistemic logic have been made in the 1990s by Plaza, Gerbrandy, van Ditmarsch, and others – with Baltag, Moss & Solecki 1998 as a particularly seminal contribution. We refer to van Benthem 2002, 2005 for a survey of the state of the art plus a list of open problems.

For our purpose here, perhaps the main thing to observe is the meeting of academic cultures in just one dynamic-epistemic formula describing effects of communication:

$$[P!]K_i\phi$$

The systematic study of speech acts like *P!* was initiated in the philosophy of language by Austin and Searle. The knowledge operator *K* goes back to Hintikka's work in philosophical logic and epistemology. And the dynamic logic operator *[]* comes from the tradition of Hoare, Pratt, and many others in computer science and mathematics. Finally, the study of communication by such means seems more like a topic in the social sciences. Thus, humanities, natural sciences, and social sciences meet in one single locus. C.P. Snow deplored the chasm between the 'Two Cultures', but they still do meet in unexpected places.

***Benefits: new issues*** Let us explore this interdisciplinary meeting point in a bit more detail. A logical system provides a lense for looking at phenomena that may not have been visible clearly before. For instance, what is the general effect of a speech act like making a public assertion *P*? It seems plausible that this must always produce common knowledge, reflected in the dynamic-epistemic formula

$$[P!]C_G P$$

But the latter principle is problematic as true logical law of communication. Back at the start of analytical philosophy, G.E. Moore formulated his notorious true-but-infelicitous sentences such as

*"P, but I don't believe that P",*

which do not seem appropriate in communication. In particular, given some minimal assumptions in epistemic logic, one can never *know* the analogous logical proposition $P \,\&\, \neg KP$. This phenomenon made its way from philosophy to mathematics in puzzles, such as the famous Muddy Children and its ilk: 'hats', cheating housewives, ... In a given group of muddy and clean children, each child can see the others, but not its own forehead. Now their father tells them publicly that at least one child is dirty. In the ensuing process of questioning which child knows its status, they all say first that they

do not know if they have mud on their foreheads. But as this question and answer process is repeated, a last round occurs where everybody has figured out who are dirty. For instance, in a group with two dirty children and one clean one, the dirty ones can figure out their status after one round:

> If I were clean, the one dirty child I see would have seen only clean children around her, and so she would have known that she was dirty at once. But she did not. So I must be dirty, too!" This reasoning is symmetric for both muddy children – so both know in the second round. The third child knows it is clean one round later, after they announced that.

Thus, the last ignorance assertion in the sequence makes *its own negation* into common knowledge. Fagin et al. 1995 give important applications of the reasoning in such puzzles to understanding computational processes with distributed agents which exchange information. In dynamic-epistemic logic, the point is that any true announcement $(P \,\&\, \neg KP)!$ with a factual assertion $P$ makes the left-hand conjunct $P$ into common knowledge, thereby invalidating the right-hand conjunct $\neg KP$. This phenomenon is ubiquitous. Van Benthem 2004B shows how the same difficulty underlies the 'Fitch Paradox' of verification, which says that the apparently plausible principle that "Every true proposition is learnable" cannot be maintained consistently. In logic itself, this raises the technical issues *just which propositions P* achieve their own common knowledge when truly announced. Van Benthem 2002 has some partial results – but the general question is still open (see also van Ditmarsch et al. 2006 for this, and many other current issues in dynamic-epistemic logic).

***Information update in general*** We have seen how even something as simple as making a public announcement can hide unexpected subtleties. But of course, there are many further types of communication. We seem to be very good in providing *partial* information, making sure that only the right people know what we want them to know, while others do not. This brings in issues of security, hiding, lying, and all those more subtle skills that lubricate civilized social life. Moreover, this information is not just conveyed by speech acts. We also get it by mere observation, or by inference from what we already know. All these sources can occur intertwined, with switches from one to the other. If I want to know if paracetamol is sold close to the Royal Academy, I can either try to reason it out inside my head, using the *KNAW* documentation at my disposal, or I can perceptually inspect all adjoining houses on the Kloveniersburgwal, or I can resort to communication, and ask some reliable-looking person in the vicinity. Richer systems of dynamic-epistemic logic exist which can handle communication and perception at the same time, including partiality and hiding, by using update with *event models*. Indeed, this was the main innovation due to Baltag, Moss & Solecki 1998. And once we have

such more subtle systems in place, we can start looking at the total area of human communicative behaviour, asking which tasks are harder than others – and where are the natural boundaries of complexity separating one practice form another.

## 4 General logical dynamics of interaction

Epistemic logic started in philosophy as the formal analysis of what one single person can be said to know, and which laws govern valid reasoning about knowledge and ignorance. Dynamic epistemic logic shifts this focus in at least two ways:

| | |
|---|---|
| *dynamic* | events themselves become explicit objects of logical study, |
| *social* | events crucially involve what agents know about each other, |

including group phenomena that are *sui generis* such as common knowledge. Both aspects reflect current uses of epistemic logic in computer science and economics, where interactive behaviour of agents (human or artificial) is at centre stage. They also reflect a broader 'Dynamic Turn' which has surfaced in philosophy, linguistics, and computer science since the early1980s (cf. Muskens et al. 1997). Information growth and knowledge are just one of many cognitive phenomena involved in this, others are semantic interpretation, scientific methodology, or epistemology in general.

***Belief revision*** One key source for the Dynamic Turn combining many strands has been the study of *belief revision* (Gärdenfors & Rott 1995) in the face of information which contradicts our current beliefs. Both update and revision are ubiquitous in ordinary life, as we confront current beliefs with new observations, and then modify the former. They also occur on a grander scale in science, when we change theories that contradict the facts (or themselves). Belief revision seems less deterministic than information update: agents have different legitimate options for changing their current beliefs to accommodate some new observation. Accordingly, this phenomenon has been studied mostly in a postulational style, writing down intuitively plausible constraints on changes in beliefs, whatever specific mechanism is chosen. Even so, concrete dynamic logics for belief revision can be designed just as well as for information update: the relevant changes in the models will now affect plausibility orderings of worlds.

The dynamics in revision seems clear. But what would it mean to take a *social* stance here? To see this, an analogy may be helpful with ordinary discourse. As long as you tell me things that are consistent with what I believe (or have committed to for the sake of conversation), I will just go along – whether or not I truly accept what is being said. But once there is an open conflict with what I believe to be true, this will no longer work, and I am forced to perform some sort of action. The latter can range from

'withdrawal' to open contradiction. And conversations often contain episodes where some conflict of opinion, small or large, needs to be resolved to restore consistency. Likewise, the best way of casting belief revision would indeed be social, as a problem of *integrating sources*. There is my old self with its beliefs, there is Nature or some other source coming with a new proposition, and what the new belief state for agents in this setting looks like depends on the priority or reliability assigned to these sources. Some of this integration may be peaceful – but sometimes also, there will be conflict. There has been some work on 'belief merge' in this more social style (cf. Maynard-Reid & Shoham 2001), but so far, a convincing intuitive set of postulates reflecting interactive intuitions still seems lacking. In this connection, the example of scientific communities is also of interest, since these communities also produce collective products (theories and practices) which can change under pressure from outside sources.

***Learning theory*** Evidently, people have various strategies for revising theories, or just our ordinary opinions. In a sense, belief revision theory is not out-and-out dynamics yet, as those processes themselves are not manipulated as first-class citizens in the calculus. An example of the latter move is the explicit theory of *learning mechanisms* in Kelly 1996, merging ideas from the philosophy of science, mathematical topology, and computer science. Update, revision, and learning form a coherent family of issues, going upward from short-term to long-term behaviour.

Again, issues in learning theory have been mainly cast in terms of single learners so far, e.g., computational devices trying to pick up the grammar of some language, when confronted with a potentially infinite evidence stream. But surely, the basic learning scenario is again *social*. A *Student* is learning from a *Teacher* who is presenting material, and it is their interaction which determines the results of this process, whether good or bad. We will return to this social setting for learning in a little while, but first, let us discuss interaction over time in more general terms.

***From single actions to strategic games*** The individual information moves described by dynamic epistemic logics naturally lead to the next step in the ladder of cognitive phenomena. For instance, humans are good at turning any given practice into a sort of *game*. Here is an example. Any topic of discourse, and any conversation space of admissible assertions about it, leads naturally to *information games*, where players must say something non-trivial at every turn, and the goal is to be the first to know the true situation. More complex examples are found in parlour games like "Clue" (van Ditmarsch 2000) which have been designed so that the game remains enjoyable to play, though not too hard to get stuck in sheer complexity.

In this game setting, all our earlier cognitive phenomena return, with new twists:

***Game theory and dynamic epistemic logic*** To describe agents' behaviour in this richer setting where conversation has a goal, and participants are aware of this, we must turn to game theory. The optimal thing to say depends on what we have heard before, plus our current plan. This requires strategies on players' part, and the best outcomes for all are often well-described by the usual Nash equilibria. Though these equilibria may be computed by general techniques (including probabilistic mixed strategies), most of their outcome values are unknown for even simple information games. But even when available, game values are global features at best of the process. Dynamic-epistemic logic adds the *fine-structure* of players' information as they are observing events in the course of the game, while trying to plan their next move.

***Belief revision*** In a game setting, however, update of information is no longer the only relevant process. Players must also be able to *revise their beliefs* when confronted with evidence contradicting their expectations so far. To see this, consider a game of the famous Centipede type. Two players can play either 'Down' or 'Across' (Figure 3):

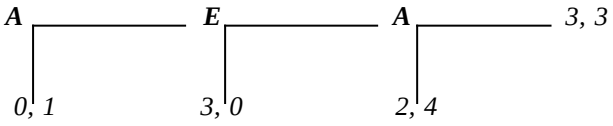


Figure 3

The value of an outcome for ***E*** is indicated first in these pairs, that for ***A*** follows. In this scenario, the standard recommendation of game theory proceeds by a line of reasoning called 'Backward Induction':

> We analyze the behaviour of ***A*** at the end, and then work back toward the initial node. It is to his advantage there to play Down (this gives him 4 instead of 3). But since ***E*** knows this, she will play Down at her move, since this gives her 3 instead of 2. But then, knowing *this* again, at the start of the game, ***A*** should play Down – because this gives him 1 instead of 0.

This reasoning, though it sounds compelling, has the surprising effect that, with this outcome, both players are worse off than they would have been if the game had proceeded toward the far right. In longer Centipede games, the difference in pay-off can be startling. But *will* this initial opt-out happen, and does Backward Induction apply? There is a presupposition involved here. Standard *rational agents* always choose the action that is to their own greatest advantage. But this is not the only possible kind of player, and ***E*** may have other beliefs about the sort of agent she is dealing with, such as

'***A*** is stupid, or generous, or adventurous…'

In particular, even when ***E*** starts with a standard expectation of ruthless 'rationality' – after ***A*** plays across, it seems likely that ***E*** will *change her beliefs* about ***A***. As with information update, the best current logical theories of belief revision come from computer science, e.g. reasoning in AI (Gärdenfors & Rott 1995). But there are also connections with belief revision as studied in cognitive science (Castelfranchi 2004).

***Games and learning*** Games also seem the right setting for studying learning. E.g., the sequential give-and-take, where one's best move depends on preceding ones by others, seems to reflect classroom dynamics. Here is a simple scenario:

> The truth is the situation ♣ in the following diagrams. Your Student, located at the node ○ on the upper left, tries to stay ignorant about this, moving along the lines between points. Your task as a Teacher is to cut successive ignorance lines, forcing him into the right position with no means of escape. You start by cutting a link, then the Student moves along some still available link, and so on. Both of you must perform a move as long as you can.

Whether the Teacher can teach the Student well depends on which player has the winning strategy in this game. (One of them must have such a strategy, by Zermelo's Theorem in game theory.) Figure 4 shows two games with different outcomes:

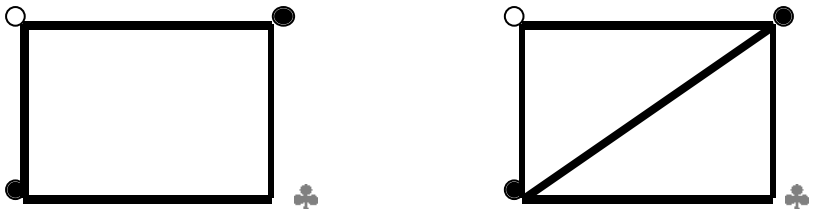

Figure 4

The reader may want to check that the Student has the winning strategy in the first game, and the Teacher in the second! The details of that argument will high-light the interactive dynamics of learning, where opposing forces may be at work. Of course, this particular game is not very realistic yet – though it is certainly more true to life than the idea that teachers just pour knowledge into their grateful students. A more sophisticated game example would be a version more like the popular parlour game "Scotland Yard", where the moves of Students cannot be observed in general, and Teachers can only perform occasional 'tests' on where they are at intermediate stages.

***Diversity of agents*** In the richer game-theoretic setting, many further questions arise for logical analysis of communication and interaction that have a strong cognitive flavour. In particular, agents clearly differ in their powers of reasoning and observation, their propensities in belief revision (eager or conservative), and the sheer cleverness of their strategies. Thus, any society of agents will show diversity along many dimensions, and logical systems should account for this. For a first attempt, cf. van Benthem & Liu 2004, who compare ideal Turing machines with finite automata in their role as information updaters. As one striking instance, when we encounter new agents, we need to determine their type, both as to processing capacities and as to general behaviour. Simple quick conclusions are not always the right ones here. E.g., Axelrod 1984 studied the simple finite-automaton strategy like *Tit-for-Tat*, which' cooperates' if you did so in the preceding round, and 'defects' if that's what you did there. Axelrod showed how, in social encounters, Tit-for-Tat wins out against much more sophisticated forms of punishing and rewarding past behaviour. Likewise, sophisticated analyses of interactive computation using linear logic (Abramsky 1998) uses general *Copy-Cat* strategies which copy another player's move from one game to another. When suitably composed, these create amazingly effective mathematical and computational behaviour in very complex games. These scenarios can be hard to compute, but *movies* provide interesting examples. Understanding 1990s classics like *The Matrix* or *Memento* requires a sort of modeling where game theory and logic have not yet caught up.

***Long-term processes*** Finally, even games as normally understood are just episodes on a larger time-scale. In the total arena of life, finite terminating processes co-exist with ones that are for all practical purposes *infinite*, such as the operating system of my computer, or the general conventions of the society that our behaviour lives in. Understanding this broader setting requires an integration of dynamic epistemic logic and game theory with the mathematical theory of *dynamical systems*. This is happening already in modern evolutionary game theory, and I am sure that logic will follow suit. Cf. the epistemic temporal logic systems of Fagin et al. 1995, Parikh & Ramanujam 2003, Belnap et al. 2001. Thus, serious mathematical models and logical calculi live along the whole spectrum of cognitive phenomena mentioned in Section 2.

## 5 Toward cognitive reality

The preceding account of logics for communication and games may have shown that there are many interesting structures to be discovered in the area of social cognitive phenomena. Nevertheless, it is fair to say, looking at the relevant literature in logic, linguistics, philosophy and computer science, that much of the reference made there to 'real cognition' is mere rhetoric! Even the most innovative authors on the logic side

show little inclination to really go out and be confronted with the experimental evidence. But this is not because such contacts would be useless. In fact, if we were to go out and see whether humans are really 'so good' at communicative subtlety, or even prior to that: how they really do things – we might be in for many creative surprises.

One reason for a persistent armchair syndrom are philosophical preconceptions, such as the celebrated 'anti-psychologism' of Frege and his followers, which has been adopted by mainstream philosophers and logicians ever since. This allows them to theorize, use the attractions of real-life examples (and sell papers by making a few eye-catching claims in that direction), but just at the moment when confrontation with reality threatens: to comfortably retreat to a more normative or theoretical position. Personally, I find this strategy more and more empty, and almost intellectually dishonest. But much worse than dishonesty: it has become *boring*, and the time seems ripe to actually confront all of the above with experimental facts.

What was said in the preceding sections suggest many experimental questions, such as:

> What information do people get out of various types of assertion?
> How do they revise or merge beliefs and resolve conflicts of opinion?
> What are their interactive strategies in conversation or learning?
> How do they cope with diversity of agents?

and on a more technical note, do human actors really feel 'complexity barriers' when jumping from one practice to another (say from honest reporting to lying) as predicted by logico-computational theory? Experimental game theorists have already made their move toward cognitive psychology here: semanticists, philosophers, and logicians might follow. One nice example is the work on interactive variants of "Master Mind" carried our by Rineke Verbrugge and co-workers at the University of Groningen.

Indeed, much of the relevant experimental material lies close at hand. Just watch your students play information games on their *GSM*s (Muddy Children-like puzzles often come as accessories with them), or let people play known games in various degrees of complexity by manipulating rules, such as in the move from Chess to *Kriegsspiel*, where one does not know the positions of the Opponent's pieces. After lectures on dynamic epistemic logic, I often get intriguing responses from members of the audience telling me how they play modified information games, 'complexifying', e.g., "Clue" to make it more interesting. Some variations seem to work there, others do not. E.g., adding a fourth guessing category of Motive in addition to Room, Person, and Murder Weapon is reported as easy – whereas allowing some 'cheating moves' makes the game almost impossible to successfully complete. It would be up to socially minded cognitive scientists to use this spontaneous evidence, and find out…

## 6 A triangle affair: logic, cognition, and computation

There is one more point to be noted to get the full picture of cognitive activities advocated here. In modern research, studying information and communication really involves *three* main ingredients. Logic supplies the theory, cognitive reality supplies the phenomena that we are interested in. But there is always a third party involved, viz. the role of *computation* of some sort. This influence is both theoretical: witness the frequent use of ideas coming originally from so-called 'dynamic logics' of programs – but it is also practical. We communicate with machines perhaps more often nowadays than with fellow humans, and man-created virtual reality is all around us: just as visionary computer scientists had predicted way back when 'single-minded' traditional hardware was still the only game in town (Licklider 1965). And these computational agents are no longer the lonely Turing machines from the 1930s scribbling on their tapes: they live in communities of distributed agents performing Internet transactions, or even playing (robo-)soccer matches. This triangle affair between

*logical theory*, *experimental facts*, and *computational design*

is not an annoying case of diversity. It rather reflects a deep and fruitful insight. The very same phenomenon may manifest itself either in the models of a logical theory, or be embodied in human cognition, or in the design of some computational process running on a machine. There is no more difficulty understanding this diversity, and unity, than there is in grasping the concept of the Holy Trinity.

## 7 Conclusion

This paper has tried to make a case that interactive social cognitive phenomena are important – and also, that they are endowed with a logical-computational structure rich enough to make for interesting scientific analysis. But perhaps, the sheer *social stance* by itself is worth emphasizing. I am always amazed by the 'individualist' bias in cognitive studies. We look at language in terms of individual competence, even though there are usually Speaker and Hearer interacting, and even though the real language skill to be taught is successful communication, not writing grammatically correct sentences in your private diary. And while we are at it, take that teaching itself. We model single agents forming hypotheses about a grammar or some other model for an observed string of phenomena. But again, this is just a one-dimensional projection of the most striking two-agent setting for teaching, being that of a Student with a Teacher. Our models should emphasize the latter, and then specialize to the former. Finally, logical systems emphasize lonesome reasoners, or even crystalline proofs where all

traces of human activity have been washed off with formaline. But surely, the prototypical logical activity is argumentation, which happens in an interactive social setting.

My next main point has been the Trinity of logic, cognition, *and computation*. I really believe this is also the way to go in cognitive studies generally. First, not accidentally, these three areas often undergo the same intellectual movements. Notably, the social perspective emerged in logic and computer science around the same time in the 1980s. But perhaps a stronger argument for the juxtaposition is the emergence of deep and surprising new insights. Here is an example from dynamic epistemic logic again. Puzzles like Muddy Children or complex conversational strategies involve complex behaviour that can be described as standard program constructions *IF THEN ELSE, WHILE DO* telling people to say and ask certain things until certain effects are reached. Now Miller & Moss 2005 have shown that the complexity of this sort of reasoning on top of the logic of public announcement is *undecidable*. This seems a purely negative result, but their method of proof is quite interesting. The authors show how to encode the behaviour of arbitrary Turing machines (and hence, correlated undecidable issues like the Halting Problem) in terms of planning conversational strategies for achieving specified epistemic goals as to who is to know what at the end. Thus, taking a positive view of the result, we see that

> computation and conversation have equal processing power!

Insights like this allow us to look at familiar phenomena in unexpected ways. Computing machines communicate, but communicating humans also engage in complex computational processes. Heaven knows, we might even be able to tap those computational resources! This commonality is reinforced by the earlier observation that we are already living in, and coping with, mixed human-machine societies, involving diversity of agents of many different types. Indeed, our ability to do so constitutes one of the most amazing cognitive skills that we have – and one that should be explained!

One final attractive feature of the social aspect of cognition is that it is all around us. This may be a bit hard to see at first. In science we are used to the idea that some things are *too far* away to see with the naked eye, and astronomers have to send their rockets and other expensive machines out to do the exploring for them. But social cognition may be *too close* to our skins to recognize its importance and structure straightaway. But once we do, we see that our whole world is already a gigantic Cognitive Lab. We do not need multi-billion dollar machines to force elementary particles into accelerating loops that reveal their potential and limitations. Such things happen automatically. Just think of the *Internet*, surely a multi-billion dollar machine, created for free, and all those

agents playing all sorts of differential informative games using their emails, and playing their *cc* and *bcc* buttons for achieving complex epistemic effects! So, the experiments are already running. Now we must go and read off the results…

## 8 References


Abramsky, Samson. 1998. From Computation to Interaction, towards a science of information. BCS/IEE Turing Lecture.

Adriaans, Pieter and Johan van Benthem, eds.. To appear. *Handbook of the Philosophy of Information*, Amsterdam: Elsevier.

Axelrod, Robert. 1984. *The Evolution of Cooperation*, New York: Basic Books.

Baltag, Alexandru, Moss, Lawrence and & Slawomir Solecki. 1998. The Logic of Public Announcements, Common Knowledge and Private Suspicions,. In *Proceedings TARK 1998*, Los Altos: Morgan Kaufman: 43–56.

Belnap, Nuel, Perloff, Michael and & Ming Xu. 2001. *Facing the Future*, Oxford: Oxford University Press.

Benthem, Johan van. 2002. One is a Lonely Number. Tech Report PP-2002-27, ILLC Amsterdam. To appear in P. Koepke and W. Pohlers (eds.), *Colloquium Logicum*, Providence : AMS Publications, 2005.

Benthem, Johan van. 2004A. A Mini-Guide to Logic in Action. Philosophical Researches, Suppl. Beijing: Chinese Academy of Sciences: 21–30.

Benthem, Johan van. 2004B. What One May Come to Know. Analysis 64 (282): 95–105.

Benthem, Johan van. 2005. Open Problems in Update Logics. To appear in T. Rozhkovskaya (ed.) *Mathematical Problems from Applied Logic*, Novosibirsk /New York: Russian Academy of Sciences/Plenum Press.

Benthem, Johan van, and Fenrong Liu. 2004. Diversity of Logical Agents in Games. Philosophia Scientiae *8:2:*163–178.

Castelfranchi, Christiano. 2004. Reasons to Believe: Cognitive Models of Belief Change. Ms. ISTC-CNR, Roma. Invited lecture, Workshop *Changing Minds*, ILLC Amsterdam, October 2004.

Ditmarsch, Hans van. 2000. *Knowledge Games*, Dissertation DS-2000-06, ILLC Amsterdam and University of Groningen.

Ditmarsch, Hans van, Wiebe van der Hoek and Barteld Kooi. 2006. *Dynamic Epistemic Logic*, Cambridge: Cambridge University Press.

Fagin, Ron, Halpern, Joseph, Moses, Yoram, and Moshe Vardi. 1995. *Reasoning about Knowledge*, Cambridge (Mass.): The MIT Press.

Gärdenfors, Peter and Hans Rott. 1995. Belief Revision. In D. M. Gabbay, C. J. Hogger and J. A. Robinson (eds.) *Handbook of Logic in Artificial Intelligence and Logic Programming* 4. Oxford: Oxford University Press.

Kelly, Kevin.1996. *The Logic of Reliable Inquiry.* Oxford: Oxford University Press.

Licklider, John. 1965. Man-Computer Partnership. Palo Alto: Digital Systems Research Center. Also in International Science and Technology.
Maynard-Reid, II , Pedrito, and Yoav Shoham. 2001. Belief Fusion: Aggregating Pedigreed Belief States. Journal of Logic, Language and Information 10:2, 183 – 209.

Miller, Joseph and Lawrence Moss. 2005. The Undecidability of Iterated Modal Relativization. Studia Logica 97:3, 373-407.

Muskens, Reinhard, Johan van Benthem and Albert Visser. 1997. Dynamics. In J. van Benthem & A. ter Meulen (eds.) *Handbook of Logic and Language*, Amsterdam: Elsevier Science Publishers, 587-648.

Parikh, Prashant. 2002.*The Use of Language*, Stanford: CSLI Publications.

Parikh, Rohit and Ram Ramanujam. 2003. A Knowledge Based Semantics of Messages. In J. van Benthem & R. van Rooy (eds.) special issue on Information Theories, Journal of Logic, Language and Information 12:4, 453–467.

Rooy, Robert van. 2002. Signalling Games Select Horn Strategies'. Ms., ILLC, University of Amsterdam. To appear in Linguistics and Philosophy.