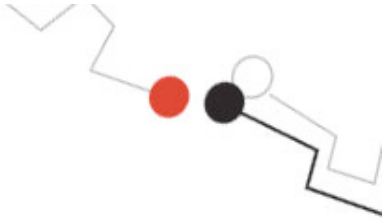
PARIS-
AMSTERDAM
LOGIC MEETINGS
OF YOUNG
RESEARCHERS



Note: This file doesn't contain the extended abstract for Paul Egré's talk. Please download it separately.

Towards a 'sophisticated' model of belief dynamics

Brian Hill
(IHPST)

It has been often noted that classical models of belief do not accurately represent the limits of human doxastic capacity, at least if belief is understood in a sufficiently *internal* sense. This is equally true of models of belief revision (Hansson, 2003). Furthermore, Rott (2004) has recently raised doubts about several of the most important of Gärdenfors postulates relating to belief revision; he called for a "more sophisticated model of belief formation". This paper is attempt to tackle some of these challenges. It proposes a model of belief states which (1) is faithful to a certain *finiteness* of human doxastic capabilities, whilst at the same time (2) permitting a general model for *iterated* belief revision which (2a) satisfies the basic G¨ardenfors postulates for belief revision and (2b) reproduces several of the suggested notions of iterated belief revision as 'special cases'. Finally (3), it provides a natural analysis of Rott's supposed counterexample, in terms of a 'framing effect' which is easily modeled in the proposed framework.

The fundamental observation behind the proposal is the importance of the notion of a sentence or an issue which is 'in play' at a particular moment. This observation seems pertinent not only to certain cases of failure of logical omniscience, but equally to cases of 'overlooked' beliefs. For example, if the agent forgets to go to his meeting at 10.00, it is not that at 10.00 he believes there is no meeting, nor that at 10.00 he neither believes that there is a meeting nor that there is no meeting, but rather that the subject of the meeting *doesn't 'enter into his mind'*, or it *doesn't enter 'into play'*.

A simple way of accommodating the notion of being (or not) 'in play' is to explicitly specify the sentences involved at a given moment, and permit this set of 'pertinent' sentences to change with time. A '*local* ' language, or at least a '*local fragment*' of language, with its own 'local' logical structure (notion of logical equivalence between its sentences, and so on), will be 'operational' or 'relevant' at a given moment; the only beliefs which are explicit or active at that moment are those towards the sentences of this local language.

However, we will not follow the proposals of Fagin and Halpern (1988), which effectively consist in taking the ordinary possible worlds belief framework and adding appropriate restrictions on the belief operator relating to a set of sentences, since, for one thing, this suggestion proves a little clumsy and inoperable when it comes to studying belief revision. Rather, we replace the whole possible worlds structure by a set of "small possible worlds", if you like, where only the sentences of the 'local language' are assigned truth values. Technically (in the simple propositional case considered here) a model of the *local* logical structure at a given moment will comprise of a local language consisting of Boolean combinations of a set of atomic sentences, and an interpretation of this language, that is, an appropriate function from the language into a set of 'states' or '*small* worlds' (these worlds are '*small* ' in the sense that *only* sentences of the local language receive an interpretation in them).

The belief state at a given moment is modeled by a transitive, connected, finitarily stoppered order on the states of the logical structure at this moment (the sentences which are believed are those which are true in all states which are minimal with respect to this order). Such an order provides a semantics for belief revision which validates the Gärdenfors postulates (Gärdenfors, 1988; Grove, 1988). Hence the Gärdenfors postulates are satisfied *with respect to the local logical structure*: for example, the only logical consequences whose belief is implied by the belief in a sentence *A* are the consequences of *A* in the local logical structure, and the only revisions of belief which are accounted for are revisions by sentences in the local language. Note furthermore that the sort of structure described is algebraic: the set of sentences of the local language and the set of sets of states each form a Boolean algebra, and the interpretation function is a quotient homomorphism between the two algebras. We shall call this sort of algebraic structure equipped with the appropriate order *ordered algebra*.

The question of belief dynamics is how new information – and possibly sentences not figuring in the previous local language – come into play. New information shall be modelled by an ordered algebra: the information 'learnt' consists of the sentences true in all the states which are minimal according to the order; the order on the rest of the states represents anticipated revisions of the new information in the light of possible subsequent information. This permits the representation not only of the new information, but equally of the details of how it was learnt, and of certain conditions under which it may be weakened or overturned. The 'standard' case, where no account is given of possible revision of the new information, is a special case of this type of model.

Once one represents the new information as an ordered algebra, belief revision by this new information corresponds to a sort of 'fusion' of two ordered algebra. To model this 'fusion', one may call upon familiar algebraic operations (especially given that the structures involved are algebraic), namely product and quotient operations on algebras and order relations. We shall define an operation which provides a *general model of iterated belief revision*. That is, firstly, it satisfies the G¨ardenfors postulates for 'one shot' belief revision. Secondly, since the result of a belief revision has the form of a *belief state* (it is an ordered algebra), which can undergo further revisions, this is a model of *iterated* belief revision. Thirdly, other proposed models of iterated belied revision, such as those of Segerberg (1998) and Konieczny and Pérez (2000) are reproduced as special cases, corresponding to restrictions on the ordered algebra representing the new information.

Finally, using this formalism, an analysis shall be given of Rott's (2004) supposed counterexample to several of the G¨ardenfors postulates, identifying the apparent problem with a "framing effect" accounted for by an appropriate modeling of the new information with which the belief is to be updated. Under this analysis, Rott's example does not invalidate the fact that the Gärdenfors postulates are satisfied by this model, but only underlines the fact that they are satisfied only in a particular sense, or in rather special cases. Ordered algebras are thus a good candidate for the more 'sophisticated' or 'realistic' models of belief revision which have recently been called for.

References

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

Grove, A. (1988). Two modelings for theory change. *Journal of Philosophical Logic*, 17:157–170.

Gärdenfors, P. (1988). *Knowledge in Flux : Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA.

Hansson, S. O. (2003). Ten philosophical problems in belief revision. *Journal of Logic and Computution*, 13:37–49.

Konieczny, S. and P´erez, R. P. (2000). A framework for iterated revision. *Journal of Applied Non-Classical Logics*, 10(3-4):339–367.

Rott, H. (2004). A counterexample to six fundamental principles of belief formation. *Synthese*, 139:225–240.

Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39:287–306.

Revision of Description Logic ABoxes

Meghyn Bienvenu
(IRIT – Université Paul Sabatier)

It is generally accepted that any reasonable real-world knowledge representation system should be able to integrate new information and to deal with the inconsistencies that may result. This problem has been extensively studied in the propositional setting under the name of belief revision. Perhaps the most influential work in this field is that of Alchourron, Gardenfors, and Makinson, who proposed a set of postulates that, according to them, characterize the set of rational revision operators. The AGM postulates stipulate among other things that the new information must be accepted, that the current beliefs should be changed as little as possible, and that the result of the revision should be syntax-independent.

The widespread use of description logics in a variety of applications, most notably as a basis for the semantic web but also in medical informatics, configuration, and natural language processing, makes the revision of description logic knowledge bases a question of great practical interest. It is thus surprising that there has been relatively little work addressing this problem. In fact, the revision of the factual component of description logic knowledge bases (commonly known as the ABox) has, to the authors' knowledge, never been formally addressed in the literature, and it is this problem which is the subject of this talk.

We will begin the talk with a brief introduction to the description logic ALC. We will then consider how the propositional notion of prime implicates can be extended to ALC and will prove that our definition is a faithful generalization of the propositional case. We next propose a weaker notion, which we term relevant implicates, which we then use to define a class of ALC ABox revision operators. We discuss the properties of our operators, showing them to satisfy all basic AGM rationality postulates for belief revision.

Private revision in a multi-agent setting.

Guillaume Aucher
(IRIT – Université Paul Sabatier)

In this talk, we will tackle the issue of how a particular and specified agent *Y* revises his/her beliefs about the world and about the other agents' beliefs when she *privately* receives new (and supposedly truthful) information. The models we consider are supposed to represent the way she personally views the world (which can then be erroneous). The announcement is privately made to the agent *Y* which means that the beliefs of the other agents should not change in reality. Thus there is no real dynamic aspect but everything is static, which makes the revision process very close in nature to the one classically studied in belief revision for a single agent. This process will be described and the technics employed are indeed very close in spirit to the ones of the AGM approach. Finally, in Sect. 4 we provide a constructive way to determine the revised models based on the notion of “canonical model of degree *n*" and show the connection between the expansion operation and the revision operation.

Vagueness and Introspection

Denis Bonnay and Paul Égré
(IJN, CNRS)

One central and debated aspect of the notion of inexact knowledge concerns the non-transitivity of the relation of indiscriminability and how it should be represented. On the epistemic account of vagueness put forward by Williamson, the intransitivity of the relation of indiscriminability is presented as the main source for vagueness ([5]: 237). In [4] and in the appendix to [5], Williamson formulates a *fixed margin for error* semantics for propositional modal logic in which the relation of epistemic uncertainty, based on a metric between worlds, is thus reflexive and symmetric, but non-transitive and non-euclidian. An important consequence of the semantics is that it invalidates the principles of positive introspection (if I know *p*, then I know that I know *p*) as well as negative introspection (if I don't know *p*, then I know that I don't know *p*).

In [1], we argued against Williamson that models of inexact knowledge that preserve the introspection principles can sometimes be desirable, and we presented a non-standard epistemic semantics for the notion of inexact knowledge, in which non-transitive and non-euclidian Kripke models can nevertheless validate positive as well as negative introspection. In [2], Halpern also argued against Williamson that an adequate model of vague knowledge need not invalidate the introspection principles, but following a different route. Instead of taking intransitivity as a primitive, and proving that the introspection principles can be preserved for a logic with one epistemic operator, as we did in [1], Halpern proposes a bimodal account of inexact knowledge that preserves the introspection principles, and he shows that there is a way to derive intransitivity. For Halpern, the intransitivity of vague knowledge is more characteristic of our reports on what we perceive than about our actual perception.

Despite these differences, one can establish a precise correspondence between Halpern's semantics and the semantics presented in [1]. The object of this paper is to spell out the details of this correspondence, and thus to compare two strategies in order to keep together introspection and nontransitivity. Like Halpern, but contra Williamson, we think it does make sense to preserve the introspection principles within a logic of inexact knowledge; unlike Halpern, but in agreement with Williamson, we are ready to see nontransitivity as a property of perceptual knowledge proper.

Dynamic Epistemic Logic - old and new directions

Hans van Ditmarsch
(Otago)

Dynamic epistemic logic is the modal logic of knowledge and belief change. I will discuss various past, present, and (expected) future directions of research in dynamic epistemic logic. Concerning current and future directions, I intend to pay specific attention to (at least one of) (i) the integration of approaches to model factual and epistemic change; (ii) approaches to model different degrees of belief and their interaction with knowledge, conviction, and change; (iii) and a fairly new phenomenon called 'arbitrary announcement', an approach wherein <>phi means "there is a psi such that <psi>phi", where <psi> models 'ordinary' public announcement. My interest typically involves modelling dynamic phenomena in specific multiagent systems: I intend to sprinkle my presentation with examples and suitable logic puzzles.

Metatheory of actions: beyond consistency

Andreas Herzig and Ivan Varzinczak
(IRIT – Université Paul Sabatier)

Traditionally, consistency is the only criterion for the quality of a theory in logic-based approaches to reasoning about actions. This work goes beyond that and contributes to the metatheory of actions by investigating what other properties a good domain description should have. We state some metatheoretical postulates concerning this sore spot. When all postulates are satisfied, we call the action theory modular. We point out the problems that arise when the postulates about modularity are violated, and propose algorithmic checks that can help the designer of an action theory to overcome them. Besides being easier to understand and more elaboration tolerant in McCarthy's sense, we show that modular theories have interesting properties.

From Toronto to Amsterdam

Tiago de Lima
(IRIT – Université Paul Sabatier)

Since Moore [1980], the reasoning about actions community has been interested in epistemic actions and knowledge. Frameworks for reason with incomplete information, where agents are able to perform both physical (ontic) and epistemic actions have been proposed by, e.g., Shapiro et al. [1998], Scherl and Levesque [2003] and Baral and Zhang [2005]. These approaches are based on the solution to the frame problem, in terms of reduction axioms, proposed by Reiter [1991]. This was further extended by Scherl and Levesque [1993] to handle epistemic actions, leading to a reduction method to S5 logic. Combined with S5 theorem proving, it provides a decision procedure for the so-called plan verification problem. In the general case however, the reduced formula is exponentially larger than the original one and, up to now, no efficient reasoning about actions method were known. Here, we present a general framework which is a sum of S5 logic for modelling knowledge, star-free propositional dynamic logic (PDL) for modelling actions together with a perfect recall axiom. Moreover, we define two new operators: public observations and public assignments. Not surprisingly, we show that these two operators correspond to the public announcement from public announcement logic (PAL), firstly proposed by Plaza [1989], and the public assignment proposed by van Ditmarsch et al. [2005]. As showed by Herzig and De Lima [2006], in such formalism every deterministic public action can be decomposed in a sequence of two actions. The first one is a purely epistemic action (i.e., an epistemic actions that does not change the physical state of the world) and the second one is a purely ontic action (i.e., an ontic action that does not increase the knowledge of the agent). We then argue that public purely epistemic actions can be simulated by compound public announcements while public purely ontic actions can be simulated by compound public assignments. Therefore, since there is no need of PDL abstract actions, we restrict our actions to announcements and assignments only. We call the resultant logic epistemic dynamic logic (EDL). We then show a polynomial satisfiability reduction from EDL to S5 based on the polynomial reduction proposed by Lutz [2006]. It follows that validity, and in particular the plan verification problem, for EDL is in coNP for single-agent and in PSPACE for multi-agent environments.

References:

C. Baral and Y. Zhang. Knowledge updates: semantics and complexity issues. Artificial Intelligence, 164(1-2):209–243, 2005.

A. Herzig and T. De Lima. Epistemic actions and ontic actions: A unified logical framework. In J.S. Sichman et al., editors, IBERAMIA-SBIA 2006, LNAI 4140, pages 409–418. Springer-Verlag, 2006. to appear.

C. Lutz. Complexity and succintness of public announcement logic. In P. Stone and G. Weiss, editors, Proceedings of the Fifth AAMAS, pages 137–144, 2006.

R. Moore. Reasoning about knowledge and action. Technical Note 191, SRI International, 1980.

J. Plaza. Logics of public communications. In M. L. Emrich, Pfeifer M. S,
M. Hadzikadic, and Z. W. Ras, editors, Proceedings of the Fourth ISMIS, pages 201–216, 1989.

R. Reiter. The frame problem in the situation calculus: A simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, Papers in Honor of John McCarthy, pages 359–380. Academic Press Professional Inc., 1991.

R. Scherl and H. Levesque. The frame problem and knowledge-producing actions. In Proceedings of the Eleventh AAAI, pages 689–695. The AAAI Press, 1993.

R. Scherl and H. Levesque. Knowledge, action and the frame problem. Artificial Intelligence, 144(1–2):1–39, 2003.

S. Shapiro, Y. Lesp´erance, and H. Levesque. Specifying communicative multiagent systems. In Agents and Multi-Agent Systems - Formalisms, Methodologies, and Applications, volume 1441 of LNAI, pages 1–14. Springer-Verlag, 1998.

H. van Ditmarsch, W. van der Hoek, and B. Kooi. Dynamic epistemic logic with assignment. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. Singh, and M. Wooldridge, editors, Proceedings of the Fourth AAMAS, pages 141–148. ACM, 2005.

Knowing How to Play: Uniform Choices in Logics of Agency

Nicolas Troquard
(IRIT, LOA)

In the last years there has been increasing interest in logics enabling reasoning about strategies of agents and coalitions of agents, and the agents' knowledge about such strategies. Such logics combine two kinds of modal logics:

• logics of knowledge such as S5, and multiagent versions thereof; such logics have modal operators Ka, where Ka' is read agent a knows that ' ;

• logics of agency, including in particular Coalition Logic (CL) and Alternating-time Temporal Logic (ATL) [1]; such logics have constructions such as CL's [A] \phi or ATL's <<A>>**X**\phi, both (roughly) reading group of agents A has an action to ensure that \phi holds (whatever the other agents choose to do).

While each of these logics is by now well-established, the interaction between knowledge and agency is less consensual. A straightforward combination of for example ATL and epistemic logic (called ATEL) was proposed in [8]. In ATEL one can express things such as agent a has an action to ensure that ', but ignores that . It turned out that ATEL is not su cient for modeling sentences like agent a knows how to ensure ' . The problem can be highlighted by the following example.

Example 1 There is a switch, a lamp, and a blind agent a1, which ignores whether the light is on or o . a1 can toggle the switch (and it knows that), and a1 can remain passive.

Clearly, <<{a1}>>**X** *light* holds here, i.e., a1 can ensure that the light is on (viz. by toggling the switch if the light is o , and by doing nothing if the light is already on). We should also be able to conclude that a1 does not know which action to perform in order to do this.

ATEL makes us conclude here that Ka1 <<{a1}>>**X** *light*, i.e. the blind agent a1 knows that it has an action to ensure the light is on. The problem is that this strategy is what has been called non-uniform: it makes a1 choose different actions in possible worlds that are indistinguishable for him. Multiagent variants of our example can also be devised.

Several authors have proposed modified versions of ATEL, trying to accommodate in one way or another the notion of uniform strategy [6, 9, 7]. It seems to be fair to say that all these attempts resulted in rather complex formalisms with heavy notations, and that there is no consensus up to now what the appropriate logic of knowledge and strategies is.

We here take as our starting point a slightly di erent logic of agency that has been developed in philosophical logic. Just as ATL, the logic of Seeing To It That (STIT) [5] is a modal logic enabling us to speak about time and agents' choices of actions. In STIT, CL's and ATL's \forall-\exists-quantification (there is a strategy of group A such that for all actions of the other agents) is split up into two different modal operators:

• an operator of historical possibility ;

• an operator of “seeing to it that” *Sti*t.

In previous work [3] we have shown that STIT is at least as expressive as ATL. We have proved this by translating ATL into STIT. The main clauses of the translation map ATL's <<A>>\phi (group A has a strategy to ensure that \phi) into STIT's <>StitA \phi (it is possible that group A sees to it that \phi).[1]

In this presentation we argue that the STIT framework can easily account for uniform strategies. To support our claim, we first present a straightforward solution in STIT logic augmented by a modal operator of knowledge. Then we offer a simplification, by introducing a modal logic of knowledge-based uniform agency, for choices, alias one-step strategies. Originally presented at AAMAS'06 [4], we shall push the presentation towards the recent perspective of [2].

References

---

[1] The STIT operator used here is the strategic STIT.

[1] R. Alur, T. Henzinger, and O. Kupferman. Alternating-time temporal logic. In Proceedings of the 38th IEEE Symposium on Foundations of Computer Science, Florida, October 1997.
[2] J. Broersen, A. Herzig, and N. Troquard. A STIT-extension of ATL. In Tenth European
Conference on Logics in Arti cial Intelligence (JELIA'06), Liverpool, England, UK, volume 4160 of Lecture Notes in Arti cial Intelligence, pages 69 81. Springer, 2006.
[3] J. Broersen, A. Herzig, and N. Troquard. Embedding Alternating-time Temporal Logic in Strategic STIT Logic of Agency, 2006. To appear in Journal of Logic and Computation.
[4] A. Herzig and N. Troquard. Knowing How to Play: Uniform Choices in Logics of Agency . In G. Weiss and P. Stone, editors, 5th International Joint Conference on Autonomous Agents & Multi Agent Systems (AAMAS-06) , Hakodate, Japan, pages 209 216. ACM Press, 2006.
[5] J. F. Horty. Agency and Deontic Logic. Oxford University Press, Oxford, 2001.
[6] W. Jamroga and W. van der Hoek. Agents that know how to play. Fundamenta Informaticae, 2004.
[7] W. Jamroga and T. Ågotnes. What agents can achieve under incomplete information. In Proceedings of AAMAS'06, 2006.
[8] W. van der Hoek and M. Wooldridge. Tractable multiagent planning for epistemic goals. In AAMAS '02: Proceedings of the rst international joint conference on Autonomous agents and multiagent systems, pages 1167 1174, New York, NY, USA, 2002. ACM Press.
[9] S. van Otterloo and G. Jonker. On epistemic temporal strategic logic. Electronic Notes in Theoretical Computer Science, 85(2), 2004. In Proceedings of the 2nd Int. Workshop on Logic and Communication in Multiagent Systems (LCMAS 2004).

Imaging and Sleeping Beauty - A case for double-halfers

Mikaël Cozic
(ENS Ulm)

Sleeping Beauty's story is well-known. On sunday evening (t0), Sleeping Beauty is put to sleep by an experimental philosopher. She is awaken on monday morning and at this moment (t1), the experimenter doesn't tell her which day it is. Some time later (t2), she is told that it's monday. At this point, what follows depends on the toss of a fair coin. If the result of the toss is heads, then Sleeping Beauty is put to sleep until the end of the week. If the result is tails, then Sleeping Beauty is awaken on tuesday morning. The crucial fact is that the drug that is given to her is such that she cannot distinguish her awaken on monday from her awakening on tuesday: Sleeping Beauty has a kind of memory erasure.

We are interested in the credence that Sleeping Beauty puts on the proposition that the result of the toss is heads (HEADS). More precisely, the two crucial moments are t1 - when Sleeping Beauty is just awaken on tuesday - and t2 - when Sleeping Beauty has learned that it's tuesday. I coin the first question Q1 and the second Q2. I will adopt the following notation:

- P1 is Sleeping Beauty's credence at t1 ie at her awakening on monday morning
- P2 is Sleeping Beauty's credence at t2 ie after having learned that it's monday

What should be the value of P1(HEADS)? There are basically two positions: the halfers and the thirders. The thirders claim (after A. Elga) that P1(HEADS) = 1/3 whereas the halfers claim (after D. Lewis) that P1(HEADS) = 1/2. Now, the answer to Q1 is intimately linked to the answer to Q2. As a consequence, the two positions are best described by giving their answer to both questions. By conditionalization, one obtains P2(HEADS) = 1/2 for the thirders and P2(HEADS) = 2/3 for the halfers.

We can sum up the positions of Lewis and Elga as follows :

	A. Elga	D. Lewis
Q1	1/3	1/2
Q2	1/2	2/3

Both Elga's and Lewis' basic intuitions are appealing. Elga's intuition is that the coin could be tossed on monday night and that in this case, one should endorse the objective probability of HEADS as her or his credence. Lewis' intuition is that on monday morning, there is no new evidence that is relevant to the credence concerning HEADS. Therefore the credence toward HEADS at t1 should remain the same than at t0. The aim of this paper is to propose a case for reconciling these conflicting intuitions. More pr cisely, I will argue that there is a way to vindicate a double-halfer position according to which P1(HEADS) = P2(HEADS) =1/2. My case is based on a recent theoretical exploration of probabilistic change rules (see B. Walliser and D. Zwirn, 2002) that shows that whereas bayesian conditionalization may be justified for revising contexts the much less known rule of imaging (D.K. Lewis, 1976) seems to be the appropriate one for updating contexts. Applying the imaging rule instead of bayesian conditionalization in the Sleeping Beauty story results in a double-halfer position.