

Logic of change, change of logic

Hans van Ditmarsch · Brian Hill · Ondrej Majer

Received: 17 July 2009 / Accepted: 17 July 2009 / Published online: 14 August 2009
© Springer Science+Business Media B.V. 2009

Understanding human behaviour, and indeed human beings more generally, requires an understanding of their attitudes: their beliefs, desires and preferences of course, but also a plethora of other attitudes or attitudinal factors, such as their presuppositions, their assumptions, their intentions, their attention, their awareness and their emotional states. But people's attitudes change over time: to take the least controversial example, their beliefs change on learning new facts. It is not sufficient to understand the role human attitudes play at any particular moment; an understanding of how different attitudes change is also required.

Formal models—be they from logic or mathematics—have historically played an important role in the modelling and study of some of these attitudes. Notable examples include the Bayesian model of (partial) beliefs as probability functions ([de Finetti 1937](#); [Ramsey 1931](#)), the modal logician's representation of belief and knowledge in terms of accessibility relations or partitions on sets of possible worlds ([Hintikka 1962](#); [Aumann 1976](#)), typical models of preferences as order relations over sets of

H. van Ditmarsch (✉)
Universidad de Sevilla, Filosofía, Lógica, CI Camilo J Cela, 41018 Sevilla, Spain
e-mail: hans@cs.otago.ac.nz

H. van Ditmarsch
University of Otago, Computer Science, P.O. Box 56, Dunedin 9054, New Zealand

B. Hill
HEC Paris, 1 rue de la Libération, 78351 Jouy-en-Josas, France

B. Hill
IHPST (CNRS/Paris 1/ENS), 13 rue du Four, 75006 Paris, France

O. Majer
Institute of Philosophy, Academy of Sciences of the Czech Republic, Jilská 1, 110 00 Prague 1,
Czech Republic

alternatives (Debreu 1959) and the use of utility or desirability functions to represent what the agent values (Ramsey 1931; von Neumann and Morgenstern 1947; Jeffrey 1972). Although accounts of attitude change couched in terms of some of these models have been around for a long time, recently there has been intensive development of accounts of change in other modelling paradigms. In the realm of belief change, for example, to the traditional accounts of belief change formulated in the Bayesian model (de Finetti 1937; Savage 1954) and variants (Levi 1974; Jeffrey 1972), have been added models of belief change which rely simply on machinery from ordinary propositional logic (Alchourron et al. 1985; Gärdenfors 1988) and more recently, on machinery from modal logic (Seegerberg 1998; van Ditmarsch 2005; Baltag and Smets 2006; van Benthem 2007) building on previous work on knowledge change after announcements dating back to Plaza (1989).

Faced with this plethora of competing models of attitudes and attitude change—or, perhaps it is more accurate to speak of *frameworks* for modelling attitudes and attitude change—it is perhaps the moment to take a step back and ask: what do we want from a theory of change?

This question, as philosophical and methodological as it is technical, was at the heart of the colloquium *Logic of Change, Change of Logic* organised by the Institute of Philosophy of the Academy of Sciences of the Czech Republic in Prague in September 2008. It is a question of interest to all those whose research involves formal models of human attitudes and attitude change—philosophers, applied logicians, computer scientists, psychologists, economists and social scientists. Indeed, we were fortunate enough to have present at the colloquium representatives from many of these fields.

This special issue of *Knowledge, Rationality & Action*, a subjournal of *Synthese*, contains selected papers from the colloquium, which provide an overview of some current issues concerning or related to change in human attitude. Roughly, the papers can be organised along two dimensions: that of the attitude or attitudes studied, and that of the modelling framework adopted. Of course, these two dimensions are not strictly independent, to the extent that several frameworks better lend themselves to modelling change of some attitudes rather than others; nevertheless, they serve to organise the following résumé of the field.

Concerning attitudes, the traditional distinction between beliefs and desires has, in some form or another, largely dominated the literature. It is fair to say that the motor of much work on change has been an interest in change in belief, largely because it is controversial whether there are changes in “fundamental” desires or preferences: that is, changes which cannot be reduced to a change in belief (Grüne-Yanoff and Hansson 2009). However, it is becoming increasingly clear to some that a study of attitudes which do not immediately and simply reduce to the classic pair—acceptance, presupposition, or awareness, to take but several examples—are required. Some of the contributions in this volume will concentrate on the “traditional” attitudes, whereas others will be concerned to extend the techniques developed to other attitudes which have to date received less attention.

Concerning models or modelling frameworks, perhaps the simplest distinction which can be drawn is that between those which are cardinal and those which are ordinal. Cardinal models normally involve the “degree” to which one can hold an

attitude as measured by a real number. For example, under the classic Bayesian view, rational agents' beliefs can be represented by probability functions: the probability of the proposition "it is raining now in Prague" captures one's degree of belief that it is currently raining in Prague. (We shall assume here that the object of attitudes are propositions, although this is far from uncontroversial. See for example [Jeffrey \(1972\)](#); [Hansson \(1999\)](#) for discussion.) By contrast, in ordinal models real numbers cannot be meaningfully attributed to propositions: a weaker structure is used to represent the agent's attitudes. For example, in traditional representations of beliefs used in many of the traditional models proposed by philosophical logicians and computer scientists, any particular proposition is either believed to be true, believed to be false, or neither: there is no notion of "degree" of belief. Even in developments of these models, involving say an epistemic entrenchment on propositions believed ([Rott 1999](#)) or a plausibility relation on worlds ([Grove 1988](#)), the structure representing beliefs is entirely captured by an ordering which does not determine unique real values for propositions.

It should be emphasised that this distinction does not precisely correspond to the more subtle distinction that one could draw between quantitative models—ones which use numbers—and qualitative ones—ones which do not. This is for two reasons. Firstly, because models which use natural numbers, rather than real numbers, count as ordinal, rather than cardinal, though they may be thought of as quantitative. For example, the models of beliefs by probabilities are cardinal, but the models using natural numbers, such as the ranking functions introduced by [Spohn \(1988\)](#), are ordinal. Secondly, because the qualitative–quantitative distinction is itself not always straightforward. In particular, it is often possible to *derive* a quantitative representation (in terms of numbers) from a qualitative one (in terms of an ordering). To take a famous example, suppose that the agent has preferences over a certain number of potentially uncertain options, such as going to the beach tomorrow (uncertainty: whether it is raining or not). Then a classic representation theorem, due to [Savage \(1954\)](#) (see also [Ramsey \(1931\)](#)), gives conditions on the preference relation which are necessary and sufficient for the existence of a unique probability function and a utility function which is unique up to positive linear transformation, such that the agent prefers option A to option B if and only if the expected utility of A (calculated with that probability and utility) is greater to that of B. The probability function is often interpreted as the agent's beliefs, and the utility function as a representation of his desires. So this result relates a qualitative representation of preferences (as an order) to quantitative representations of belief and desire. Similarly, from a representation of beliefs by a plausibility relation on possible worlds which has some basic properties, it is possible to assign natural numbers to propositions which can be thought of as "degrees of belief" ([van Ditmarsch 2008](#)). To this extent, many models which are first presented as qualitative may be also considered to be quantitative (see in particular [van Ditmarsch 2008](#) and [Baltag and Smets 2006](#)). However, the distinction between cardinal and ordinal representations is quite clear: it corresponds to the possibility of meaningfully using real numbers in the representation, and this can be cashed out in the *uniqueness* of the numbers used in the model ([Krantz et al. 1971](#)). The probability function featuring in Savage's theorem is unique: it is this uniqueness which licenses the qualification as cardinal. By contrast, the numbers got from a plausibility order on worlds are not unique

real numbers (there are many transformations which change the “degrees of belief” without changing the plausibility relation): this model is ordinal. Of course, in practice, it is not always possible to obtain the degree of uniqueness one would want (Jeffrey’s decision theory, in its original form, does not enjoy a representation result in which the probabilities are as unique as in Savage’s; see Jeffrey (1972)); nevertheless we can say that a model is cardinal if, by using it, the modeler is (implicitly) committing himself to the required uniqueness of the real numbers. It is interesting to note that the worry about the uniqueness of the representation used, although central to decision theory, is almost entirely absent among logicians.

Among the cardinal models of belief and desires, there are two general frameworks: the Savage framework, which dominates economics, and the Jeffrey framework, which is the most popular among philosophers. A crucial difference between the two lies in the objects of belief and desire. Whereas in the Savage setup the objects are distinct—probabilities are defined on a set of states, and utilities on a (distinct) set of consequences—in the Jeffrey framework, they are the same—there is a domain of propositions, and any proposition receives a probability value and a desirability value. This has made the Jeffrey framework particularly tractable for treating the interaction between beliefs and desires. In particular, it provides an attractive treatment of changes in belief and desire, and the effects of changes in one on changes in the other (Bradley 2007).

Among the ordinal models developed in the logical tradition, there is a distinction between two “schools”, which, as in the case of cardinal models, corresponds to a difference in the object of the attitude. The difference can be seen as arising from a difference in the languages in which the objects of the agent’s attitudes are couched. On the one hand, the so-called AGM paradigm of belief revision uses a propositional language—or more precisely, does not assume any structure on the language beyond the structure encapsulated in the classical Boolean connectives. It is naturally interpreted as considering propositional beliefs—that is, beliefs about propositional sentences such as “it is raining”. By contrast, the models in the Dynamic Epistemic Logic or DEL paradigm use a modal language, which can speak not only about the current beliefs (preferences etc.) of the agent, and perhaps of other agents, but also what the agent would come to believe (prefer etc.) on accepting a particular input (command, announcement, information, etc.). This framework emphasises the question of the change in the agent’s beliefs in propositions about his beliefs or those of others, such as “I don’t believe it is raining”.

In ‘Revising Incomplete Attitudes’, Richard Bradley, a major proponent of Jeffrey’s decision-theoretic framework, considers a weakening of that framework that brings out an interesting relationship to the AGM framework for belief revision. Jeffrey, along with Savage and much of the tradition, assume that people are ‘maximally opinionated’: that is, for any pair of options available, they can state whether they prefer one to the other or whether they are indifferent. Formally, this amounts to the assumption that the preference relation is complete. As Bradley notes, this is one of the least attractive assumptions in decision theory; he proposes a model which does not rely on the assumption. The basic idea is to model a non-opinionated agent’s state of mind, not by a single pair of probability and desirability functions as in Jeffrey’s original model, but by a set of such pairs. The central part of the paper contains a discussion of different

sorts of changes which can occur in such non-opinionated states of mind. On the one hand, generalised conditioning—the change operation which imposes particular probability and desirability values on a specific partition on the space of propositions (Bradley 2007)—can be defined on non-opinionated states of mind: it models changes where the agent adopts specific beliefs or desires with respect to particular propositions. On the other hand, the fact that non-opinionated agent’s states of mind are represented by *sets of* probability and desirability pairs allows for other change operations, such as changes where pairs are removed from the set—wherein the agent becomes more opinionated about certain issues—and changes where new pairs are added to the set—wherein the agent withdraws some of the opinions he previously held. Bradley shows how such changes could be characterised in his model, considers the relationship to the AGM operations of expansion and contraction, and discusses how these operations could be involved in deliberation.

Sven Ove Hansson’s contribution, ‘Preference-Based Choice Functions—a Generalized Approach’, also draws part of its motivation from the possibility of incomplete preferences, which, as he too notes, are important in understanding the dynamics of preferences. However, rather than the relationship between preference and the attitudes of belief and desire, Hansson’s paper deals with the relationship between preference and choice. Classical results in choice theory have related choice to preferences under the assumption that preferences are complete and transitive: formally, they have given necessary and sufficient conditions on choice functions—functions associating to each possible set of available alternatives the subset containing those which the agent may allow himself to choose—for there to exist a transitive and complete preference relation generating them. While these results have been extended to preferences which are complete and non-cyclic, it is not even clear how one would generate a choice function from a preference relation which was incomplete or cyclic. Hansson’s paper is a contribution to the recent literature on this subject. He discusses and compares the formal properties of several methods of constructing choice functions from preference relations which need not be complete nor non-cyclic, and proposes a new construction, which he characterises.

‘The Game Of Inquiry’ by Emanuel Genot presents the game theoretical approach to belief revision. Jaakko Hintikka, the founder of this approach, argues that we need to study belief change as sequences of choices, and that game theory provides the best framework for studying such sequential decision processes. The idea is to consider the traditional operations of belief change in the AGM paradigm—expansion, contraction and revision—as the result of a special kind of game of information-seeking by questioning, called interrogative games. This kind of game is played by the Inquirer against Nature, it includes both deductive and (non-monotonic) interrogative steps (‘putting questions to sources’), and a play of the game involves a step-by-step procedure, the outcome of which is actual belief change. An application of a given rule in a particular interrogative game should be justified from an epistemological viewpoint (by reference to what the author calls ‘inquiry-relevant factors’). Genot introduces the framework of Interrogative inquiry games based on a modified version of Beth Tableaux with additional rules for interrogative steps. He gives a representation theorem for the basic AGM operations, and provides an extension of the system which deals with multiple and possibly conflicting sources. He demonstrates the behaviour

of his system on a counterexample to one of the controversial rules in Belief Revision—the rule of Recovery: he shows that in his system possible failure or success of the rule is related to the reliability of the sources of answers.

Olivier Roy's paper, 'A Dynamic-Epistemic Hybrid Logic for Intentions and Information Changes in Strategic Games', focuses on an attitude which, although it has been recognised as important by philosophers of action, has received relatively little attention in the formal literature: intention. Indeed, the aim of the paper is to undertake a formal study of intention which is informed by the existing philosophical literature. Roy uses the same tools to model intention as are used in the classical modal models of belief and knowledge in games—an accessibility relations on an appropriate set of states—hence incorporating intention into models of strategic interaction in a group of agents. He then considers the dynamics of these models, using a dynamic-epistemic hybrid logic including the standard epistemic operator for knowledge as well as a special operator for intention (agent i intends that ϕ) and a dynamic epistemic operator for public announcement. This framework allows him to model situations when agents change their intentions in the light of a new information (he speaks of 'intention revision'). One of the central notions of the article is 'intention rationality': the agent does not choose strategies which exclude the achievement of her intentions. Roy studies some of the properties of this notion, and in particular the relationship to the stronger notion of 'knowledge consistency' of intentions: the agent only intends what he considers achievable. Finally, he discusses the dynamic properties of repeated public announcements of intention rationality and knowledge consistency of individual agents, and in particular the long-term consequences on the information and intentions of the group.

'On the dynamics of institutional agreements' by Andreas Herzig, Tiago de Lima, and Emiliano Lorini considers the attitude of acceptance, focusing on the importance of the context in which agents receive new information. The statements you wish to accept as a Labour member of British Parliament, disciplined by the party whip ('the war in Irak is just') may be quite different from the statements you wish to accept privately, or as a member of Amnesty International. The authors present a logic with explicit operators to represent that a set of agents G (that may be a subset of the set of all agents) accepts (the truth of) a formula ϕ in a given institutional context x . They also introduce two dynamic modal operators to express change of acceptance. The first of these operators expresses the fact that all the agents learn that a formula ϕ is true in context x . This operator is like a public announcement and consists of model restriction (to the announcement formula ϕ), except that the model restriction is for the given institutional context x only. The second of these operators expresses that the agents in G accept formula ψ in the institutional context x , and the semantics involves the update of their acceptances, as members of x , in order to effect that. One might say that the second operator has more a feel of *belief revision*, including the aspect of resolving possible inconsistencies, whereas the first operator is more like *belief expansion*.

The idea to devote a special issue of *Knowledge, Rationality and Action*, a sub-journal of *Synthese*, to the logic of change has its origin on the colloquium *Logic of Change, Change of Logic*, which took place in Prague in September 2008. We would like to thank all the participants for a wonderful creative and inspiring atmosphere:

Guillaume Aucher, Alexandru Baltag, Richard Bradley, Peter Brössel, Mathijs de Boer, Tiago de Lima, Cédric Dégremon, Franz Dietrich, Sebastian Enqvist, David Etlin, André Fuhrmann, Emmanuel Genot, Konstantinos Georgatos, Sven Ove Hansson, Conrad Heilmann, Brian Hill, Jingde Cheng, Timothy Childers, Namjoong Kim, Piotr Kulicki, Emiliano Lorini, Ondrej Majer, David Makinson, Michal Peliš, Francesca Poggiolesi, Hans Rott, Olivier Roy, Fabien Schang, Sonja Smets, Ken Turner, Leon van der Torre, Hans van Ditmarsch, Arnis Vilks, and Bernard Walliser. We benefited from the stunning surroundings and the admirable hospitality of the Villa Lana, the beautiful villa near the centre of Prague which the Czech Academy of Sciences allowed us to use for the conference; it proved the perfect environment for stimulating and interesting discussion. We would also like to thank to the institutions which supported this colloquium: Institute of Philosophy of the Academy of Sciences of the Czech Republic, the grant agency of the ASCR (Grant No. GA401/07/0904 “Logical analysis—theories and applications”), the Grant Agency of the Czech Republic (Grant No. IAA900090703 “Dynamic formal systems”) and the Czech Society of Computer Science. Finally, we would like to thank several anonymous referees who helped with this special issues: their advice was invaluable both to us as editors and to the authors.

References

- Alchourron, C. E., Gärdenfors, P., & Makinson, D. (1985). The logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50, 510–530.
- Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, 4, 1236–1239.
- Baltag, A., & Smets, S. (2006). Dynamic belief revision over multi-agent plausibility models. In Bonanno, G., van der Hoek, W., & Wooldridge, M. (Eds.), *Proceedings of the 6th conference on logic and the foundations of game and decision theory (LOFT'06)*, pp. 11–24
- Bradley, R. (2007). The kinematics of belief and desire. *Synthese*, 56, 513–535.
- de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*
- Debreu, G. (1959). *Theory of value. An axiomatic analysis of economic equilibrium*. Wiley, New York
- Gärdenfors, P. (1988). *Knowledge in flux: Modeling the dynamics of epistemic states*. Cambridge, MA: MIT Press.
- Grove, A. (1988). Two modelings for theory change. *Journal of Philosophical Logic*, 17, 157–170.
- Grüne-Yanoff, T., & Hansson, S. O. (Eds.). (2009). *Preference change. Approaches from philosophy, economics and psychology, volume 42 of Theory and Decision Library A*. Dordrecht: Springer.
- Hansson, S. O. (1999). *A textbook of belief dynamics, theory change and database updating*. Dordrecht: Kluwer.
- Hintikka, J. (1962). *Knowledge and belief: An introduction to the logic of the two notions*. London: College Publications. Nouvelle édition. 2005.
- Jeffrey, R. C. (1972). *The logic of decision* (2nd ed.). Chicago: University of Chicago Press.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement* (Vol. 1). San Diego: Academic Press.
- Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy*, 71, 391–418.
- Plaza J. (1989). Logics of public communications. In *Proceedings of the 4th international symposium on methodologies for intelligent systems*, pp. 201–216.
- Ramsey, F. P. (1931) Truth and probability. In *The foundations of mathematics and other logical essays*. Harcourt, Brace and Co., New York.
- Rott, H. (1999). Coherence and conservatism in the dynamics of belief. Part I: Finding the right framework. *Erkenntnis*, 50, 387–412.
- Savage, L. J. (1954). *The foundations of statistics* (2nd ed.). New York: Dover. 1971.
- Segerberg, K. (1998). Irrevocable belief revision in dynamic doxastic logic. *Notre Dame Journal of Formal Logic*, 39, 287–306.

- Spohn, W. (1988). Ordinal conditional functions: A dynamic theory of epistemic states. In W. Harper & B. Skyrms (Eds.), *Causation in decision, belief change, and statistics* (2nd ed.). Dordrecht: Kluwer.
- van Benthem, J. (2007). Dynamic logic for belief revision. *Journal of Applied Non-Classical Logics*, 17, 129–155.
- van Ditmarsch, H. (2008). Comments on the logic of conditional doxastic actions. In K. Apt & R. van Rooij (Eds.), *New perspectives on games and interaction, texts in logic and games* (Vol. 4, pp. 33–44). Amsterdam: Amsterdam University Press.
- van Ditmarsch, H. P. (2005). Prolegomena to dynamic logic of belief revision. *Synthese*, 147, 229–275.
- von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton: Princeton University Press.