ON CAUSAL AND CONSTRUCTIVE MODELLING OF BELIEF CHANGE

A. V. RAVISHANKAR SARMA

Our life in various phases can be construed as involving continuous belief revision activity with a bundle of accepted beliefs, unaccepted beliefs (disbeliefs) and some even misunderstood beliefs. What is evident from such an activity is that we pass through various phases of our life effortlessly without being conscious of the rationale[1] for such changing sets of beliefs. In this process, we give up many mistakenly held beliefs of the earlier phase and replace them by new sets of belief. Ofcourse, in the process one sometimes undertakes the examination of evidence and resorts to reasoning leading finally to a new set of beliefs.[2]. Thus, both our beliefs and knowledge. At this moment we would not make a fine philosophical distinction between "beliefs" and "knowledge" (defined often as"justified true belief") are in a constant state of flux.

The question concerning *why* and *how* we change our mind (represented as a set of beliefs) is quite common in our daily life. The immediate questions we are concerned with are how and why we actually or ought to change our mind (or a set of beliefs). In answering this question, we usually invoke reason, or what we take to be a reason, for changing beliefs. Some times we provide reasons without being questioned about the adequacy of the explanations[3] because the process of belief change is unconventional or puzzling for the observer. Sometimes it is quite apparent that change of beliefs cannot be fully understood unless we invoke causal explanations and causal relevance of some of the factors among a whole range that might be placed before us for consideration.

Thus, a change in our beliefs supported by causal explanation may involve both *reasons* and *causes*. Moreover, while entertaining a set of beliefs we are often confronted with new evidential information obtained from reliable sources. An agent in such a situation might be constrained to readjust her beliefs in such a way that the resultant belief set after accepting the new information remains consistent. Here, we assume that an agent ultimately would be reluctant to entertain an inconsistent set of beliefs, although sometimes she might be accused of holding inconsistent beliefs.

This mundane process of changing beliefs in a rational manner has become the subject matter of a rich and fruitful area of investigation and cognitive modeling, giving rise to various *theories of belief revision*. Most of the research in this area is influenced by the seminal works in philosophical logic, in particular by Alchourron, Gärdenfors and Makinson (here after AGM model) [1, 6], who developed the theory of belief revision. They also proposed a mechanism called the *standard epistemic entrenchment* (SEE) [8] for uniquely determining the revision and contraction.

The problem of *belief revision* is how an agent changes her beliefs when new information obtained from an evidential source causes inconsistencies in the currently held beliefs in

Date: August 9, 2013.

a belief set. BR has been extensively surveyed by Hansson [13], more recently by Hans Rott [15] and Kern- Isberner [12].

This problem of belief revision also arises in a number of areas in computer science and Artificial Intelligence (AI), e.g., in updating logical database, in hypothetical reasoning. All the models of belief revision more or less subscribe to the criterion that information is valuable and unnecessary loss of information should be avoided. It is widely acknowledged that the ability to change beliefs in a rational manner is one of many facets of an intelligent agent. *belief revision* is an ubiquitous process that underlies many forms of intelligent behavior. An analysis of such intelligent behavior exhibits the role of *epistemic states*. *Epistemic states* (which represent an agent's current belief state) are central to the analysis of belief revision. An epistemic state representing the beliefs of an agent also includes the ability of the agent for coherent reasoning including *explanations* of the earlier beliefs the agent entertained. Moreover, the strategies for performing the revision of beliefs by the agent are also encoded in the epistemic states. Besides, the strategies for revision in the form of explanations, especially causal explanations, are represented in terms of conditionals (in the "If.. then" form).

Causality occupies a central position in both philosophical and the Artificial Intelligence (hereafter AI) research. Many epistemological problems and AI processes require an appropriate understanding of the notion of causality [14, 11]. However, in spite of wide interest aroused by this topic, relatively little attention has been given to the essential link between notions of causality, causal explanation and belief revision. A given belief revision system must respect causal consistency apart from logical consistency [5]. In other words, we intend that the belief revision system should work in a causal environment as well. Causal environment in our case includes causal structure, interventions, causal properties, and causal mechanisms.

We find that there is a gap in our understanding of how belief revision would take place in a *causal environment* and the way we represent belief changes as changes of epistemic states in such an environment. The present work is about how an agent changes beliefs in the light of new evidential information guided by the motive of providing causal explanation. The causal explanation in turn would involve structure, intervention, causal properties, and causal mechanism, which would lead to *ordering* of beliefs over a set of sentences in the language.

This ordering would be achieved by imposing constraints in the form of *causal relevance*. In other words, this would demonstrate how the notion of causality influences the belief revision. The dissertation is an attempt to make the underlying causal structure in the process of revising beliefs explicit. Thus, it deals with the dynamics of epistemic states in the light of causal explanations and causal relevance. Since, causality is represented as conditionals, one of the problems with which this dissertation will be concerned is the way the conditionals in question are ordered. The basic idea is that not all conditional beliefs are important for the purpose of revision. Intuitively, we do give emphasis to some conditionals over the other. This idea is captured by a causal and constructive analogue of standard entrenchment ordering called *causal epistemic entrenchment* (CEE).

The purpose of the present study is to investigate the process of belief change when an agent seeks causal explanations and relevant causal factors. And how this belief change leads to an ordering or entrenchment in the framework of CEE. This thesis is about the

modeling of belief change based on causal constraints such as causal relevance. We attempt to show how insights from scientific theory change, guided by such concepts as structure, intervention, causal properties, and causal mechanism, will enable us to construct a theory of belief revision based on the *structural realist notion of causality* [2, 3, 4].

Current belief revision systems are intended to be causal but, none of the existing belief revision models satisfactorily bring out the connection. The other related problems can be found in an inspiring paper by Hansson [10]. In our belief revision theory we show how the selection of the causally relevant part is carried out and demonstrate what role this part plays in the *belief revision* process. Our theory of *belief revision* would require elucidation of two important ideas, namely, the notion of *causal relevance* (CR) and *revision based on causal relevance*. [4]

We formulate an appropriate definition of causal relevance (CR) which takes care of structure (S), intervention (I), causal properties (CP) and causal mechanism (CM). However, the need for formulating causal relevance in the context of belief revision is the following. If a belief state K is revised by a sentence A, then all the sentences in K that are irrelevant (causally irrelevant) to the validity of A should be retained in the revised state of belief. The key idea, following Gärdenfors [7] is that not all beliefs are important for the belief revision and it has been observed that belief revision takes place only in the selective and active part. We present some of the prominent approaches for causal relevance [14] which involves some of the semantic features (not all) structure, intervention, causal mechanism, causal properties. We compare our definition with these approaches while emphasizing the merits of our approach. Our definition of causal relevance includes all the semantic features mentioned above. It can be represented as $CR = f(S, I, CP, CM)$.

Finally, we propose an appropriate entrenchment ordering based on the idea of causal relevance mentioned above. Causal epistemic entrenchment (CEE) is a partial entrenchment ordering with an additional constraint, namely, the constraint of causal relevance. CEE is different from other entrenchment orderings in the sense that it is non-linear and gives importance to the *core beliefs*. CEE is motivated by the need to distinguish equally entrenched beliefs with respect to the causal relevance. In other words, one belief can be more causally relevant than another and hence can have higher entrenchment.

Finally, the distinction between concrete and abstract science, evident in history and philosophy of science [9], would enable to us to view *belief revision* either in a descriptive or explanatory manner. But, explanatory approach for belief revision has an edge over the descriptive belief revision in case of revision due to complex well formed formulae. The dissertation seeks to find an explanatory approach to belief revision. We suggest a Grove style semantic modelling and provide additional postulates in order to obtain a complete characterisation.

The idea of considering explanations in Belief revision is not new. The role of explanation in belief revision is studied by Falappa[5]. But they present explanation in a deductive way. The role of abductive reasoning in belief revision is studied by Pagnuco[6]. It is different from other approaches in *how* we view the ordering of entrenchment of beliefs and *what* we understand to be the *core* of a given knowledge base. In our case a relevant belief is more entrenched than the other beliefs when it has all the features such as structure, intervention, causal properties and causal mechanism. The *core* consists of all the semantic features mentioned above.

Our modeling of belief revision is close to belief revision based on Abduction. The strategy adapted in belief revision based on Abduction is as follows. New information need not necessarily be integrated with the old information. Instead, an agent first seek an explanation or justification for the new information. After doing so, it could incorporate the explanation into its epistemic state together with the new information. In our case the explanation is based on causal description.

Notes

[1]Davidson's Actions, Reasons, and Causes (1963) defends the ostensibly commonsensical view that rationalization is a species of causal explanation. In particular, the thesis that reasons are causes has been embraced by realists who are physicalists about the mental.

[2]Gilber Harman makes a clear distinction between reasoning and inference. Reasoning normally involves change in view, often giving up things previously accepted as well as coming to accept new things, so reasoning is not in general thinking through a proof or argument in logical order, with premises, intermediate steps, and final conclusion.

[3]There are debates as to whether to construe reasons as causes. For instance Davidson construes reasons as causes.
According to Stegmulller's thesis of structural identity of explanation and prediction, consists of two subtheses, namely " (1) that every adequate explanation is potentially a prediction (2) that conversely every adequate prediction is potentially an explanation. We may, for instance, also think that an explanation is needed.

[4]This is attempted in two important chapters (i.e., chapter 5 and chapter 6) in the thesis.

[5]M.A. Falappa and G. Kern-Isberner and G.R. Simari. Explanations, belief revision and defeasible reasoning. In Artificial Intelligence, 141(1-2):1-28, 2002.

[6]Pagnucco, M., The Role of Abductive Reasoning within the Process of Belief Revision PhD Thesis, Department of Computer Science, University of Sydney February 1996.

References

[1] C. A. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530, 1985.
[2] Anjan Chakravartty. The structuralist conception of objects. *Philosophy of Science*, 70:867–878, 2003.
[3] Anjan Chakravartty. Structuralism as a form of scientific realism. *International Studies in the Philosophy of Science*, (18):151–171, 2004.
[4] Anjan Chakravartty. Causal realism: Events and processes. *Erkenntnis*, 63:7–31, 2005.
[5] Adnan Darwiche and Judea Pearl. Symbolic causal networks. In *AAAI*, pages 238–244, 1994.
[6] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Bradford Books, Cambridge, MA, 1988.
[7] Peter Gärdenfors. Belief revision and irrelevance. In *PSA*, pages 349–356, 1990.
[8] Peter Gärdenfors and David Makinson. Revisions of knowledge systems using epistemic entrenchment. In *TARK*, pages 83–95, 1988.
[9] Amitabha Gupta. The concrete and the abstract science: description versus explanation. *The Journal of Indian Council of Philosophical Research*, V1(3):67–88, 1989.
[10] Sven Ove Hansson. Ten philosophical problems in belief revision. *J. Log. Comput.*, 13(1):37–49, 2003.
[11] Daniel Kayser and Farid Nouioua. About norms and causes. *International Journal on Artificial Intelligence Tools*, 14(1-2):7–24, 2005.
[12] Gabriele Kern-Isberner. *Conditionals in nonmonotonic reasoning and belief revision: considering conditionals as agents*. Springer-Verlag New York, Inc., 2001.
[13] Sven ove Hansson. *A Textbook of Belief Dynamics. Theory Change and Database Updating*. The Kluwer Press, Boston, 1999.
[14] Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, MA, 2000.

[15] Hans Rott. *Change, Choice, and Inference*. Oxford University Press, Oxford, 2002.

E-mail address: avrsarma@gmail.com