Annals of Pure and Applied Logic 165 (2014) 106–133

Contents lists available at ScienceDirect

Annals of Pure and Applied Logic

www.elsevier.com/locate/apal

Evidence and plausibility in neighborhood structures

Johan van Benthem [a,b], David Fernández-Duque [c], Eric Pacuit [d,*]

[a] *University of Amsterdam, The Netherlands*
[b] *Stanford University, United States*
[c] *University of Seville, Spain*
[d] *University of Maryland, United States*

ARTICLE INFO

Article history:
Available online 27 August 2013

MSC:
03C35
03B44
03B45
03B62

Keywords:
Modal logic
Neighborhood semantics
Plausibility ordering
Evidence-based reasoning
Doxastic logic

ABSTRACT

The intuitive notion of evidence has both semantic and syntactic features. In this paper, we develop an *evidence logic* for epistemic agents faced with possibly contradictory evidence from different sources. The logic is based on a neighborhood semantics, where a neighborhood N indicates that the agent has reason to believe that the true state of the world lies in N. Further notions of relative plausibility between worlds and beliefs based on the latter ordering are then defined in terms of this evidence structure, yielding our intended models for evidence-based beliefs. In addition, we also consider a second more general flavor, where belief and plausibility are modeled using additional primitive relations, and we prove a representation theorem showing that each such general model is a p-morphic image of an intended one. This semantics invites a number of natural special cases, depending on how uniform we make the evidence sets, and how coherent their total structure. We give a structural study of the resulting 'uniform' and 'flat' models. Our main result are sound and complete axiomatizations for the logics of all four major model classes with respect to the modal language of evidence, belief and safe belief. We conclude with an outlook toward logics for the dynamics of changing evidence, and the resulting language extensions and connections with logics of plausibility change.

1. Introduction

It has become standard practice in Artificial Intelligence and Game Theory to use possible-worlds models to describe the knowledge and beliefs of a group of agents. In such models, the agents' knowledge is based on what is true throughout the set of epistemically accessible worlds, the current information range. Following a similar pattern, the agents' beliefs are based on what is true in the set of most "plausible" worlds. Now, it is often implicitly assumed that the agents arrived at these structures through some process of investigation, but these details are no longer present in the models.

In a number of areas, ranging from epistemology to computer science and decision theory, the need has been recognized for models that keep track of the "reasons", or the *evidence* for beliefs and other

* Corresponding author at: University of Maryland, United States.
E-mail address: epacuit@umd.edu (E. Pacuit).

0168-0072/$ – see front matter
http://dx.doi.org/10.1016/j.apal.2013.07.007

informational attitudes (cf. [23,20,3,34]). Encoding evidence as the current range of worlds the agent considers possible ignores how the agent arrived at this epistemic state. This also ignores the fine-structure of evidence that allows us to consider or modify just parts of it. One extreme for recording this additional structure are models with complete syntactic details of what the agent has learned so far (including the precise formulation and sources for each piece of evidence) (cf. [39]). In this paper, we will explore a middle ground in between ranges of possible worlds and syntactic fine-structure, viz. neighborhood models, where the available evidence is recorded as a family of sets of worlds. Our models are not unlike some earlier proposals in the study of conditionals and belief revision (cf. [21,40,13,28]), but we quickly strike out in other directions, and provide a more in-depth logical treatment.

This paper is a continuation of our earlier work in [8,38] on a new evidence interpretation of neighborhood models.[1] In a neighborhood model, each state is assigned a collection of subsets of the set of states. We view these different collections of subsets as the evidence that the agent has acquired – allowing the agent to have different evidence at different states. Given such an explicit description of the evidence that the agent has acquired, one can explore different notions of beliefs and related cognitive attitudes over neighborhood models. The logical systems that arise on natural model classes of this sort with modalities for evidence and belief have been axiomatized completely in [38]. In this paper, we go one step further, and add some further crucial structure to the neighborhood structures.

In general, there are two ways we can enrich neighborhood structures with descriptions of the agent's beliefs. The first approach is to add new accessibility relations corresponding to each epistemic or doxastic attitude in the neighborhood structure. We then impose constraints on these new relations to ensure that they are "appropriately grounded" on the available evidence. The second approach is to define the agent's beliefs and related cognitive attitudes directly using no more than the given evidence structure in the neighborhood structures. The latter "intended models" may be considered as a special case of the former "general models". This paper offers a careful study of these two approaches.

Our second new contribution is to elaborate on the relationship between our neighborhood models and another general framework for belief change in the modal tradition. Originally used as a semantics for conditionals (cf. [22]), *plausibility models* are wide-spread in modal logics of belief [35,36,6,15]. The main idea is to endow epistemic ranges with an ordering $w \preccurlyeq v$ of relative *plausibility* on worlds (usually uniform across epistemic equivalence classes): "(according to the agent) world v is at least as plausible as w".[2] Plausibility orders are typically assumed to be reflexive and transitive, and often even *connected*, making every two worlds comparable. Connections between evidence structure and plausibility order will occur throughout this paper, and their reflection in logical axioms will be determined.

In all, we shall consider four variants of the logic of evidence-based belief, which depend on some fundamental assumptions one may make about evidence models, in terms of uniformity of evidence across worlds, and total coherence of maximally consistent sets of evidence. For each of the resulting logics, we prove two main results. The first is a characterization theorem for general evidence models as being *p-morphic images* of intended models. The second result determines a complete deductive calculus for each logic. Here our representation using extended evidence models is crucial, since it permits us to employ familiar techniques from modal logic.

Anyone familiar with Sergei Artemov's work will have seen a similarity by now. It is very natural to attach to every believed proposition a "justification" for that proposition. This idea was first studied in Artemov's seminal paper [2] and applied to epistemic logic in [4] (see [3] for an overview and pointers to the relevant literature). In particular, $t : \varphi$ is intended to mean that the agent believes φ and that t is the justification for this belief. Here t is a proof term, and sophisticated logical systems have been developed that

[1] Neighborhood models have been used to provide a semantics for both normal and non-normal modal logics. See [29] for an early discussion of neighborhood models and their logics, and [18,25,19] for modern motivations and mathematical details.

[2] In conditional semantics, such plausibility or 'similarity' orders are typically world-dependent.