Annals of Pure and Applied Logic 165 (2014) 49–81

Contents lists available at SciVerse ScienceDirect

Annals of Pure and Applied Logic

www.elsevier.com/locate/apal

The logic of justified belief, explicit knowledge, and conclusive evidence

Alexandru Baltag, Bryan Renne [*,1], Sonja Smets [2]

University of Amsterdam, Institute for Logic, Language, Information and Computation (ILLC), Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Available online 19 August 2013

This paper is dedicated to Professor Sergei Artemov on the occasion of his 60th birthday.

MSC:
03B42
03B70
68T27
68T30

Keywords:
Dynamic epistemic logic
Justification Logic
Belief revision
Logical omniscience
Gettier
Evidence

ABSTRACT

We present a complete, decidable logic for reasoning about a notion of *completely trustworthy* ("*conclusive*") *evidence* and its relations to *justifiable* (implicit) belief and knowledge, as well as to their *explicit justifications*. This logic makes use of a number of evidence-related notions such as availability, admissibility, and "goodness" of a piece of evidence, and is based on an innovative modification of the Fitting semantics for Artemov's Justification Logic designed to preempt Gettier-type counterexamples. We combine this with ideas from belief revision and awareness logics to provide an account for explicitly justified (defeasible) knowledge based on conclusive evidence that addresses the problem of (logical) omniscience.

1. Introduction

Justification Logic, due to Sergei Artemov and originally conceived as a solution to a long-standing open problem concerning the intended semantics of Gödel's provability logic [3], has since developed into a wide-ranging study of the notions of *evidence* and *justification*; see, e.g., [1,2,4–10,13,18–21,23–25,29–31,33,42–45,50,53,54,59–62]. By making explicit the "evidence" supporting a given assertion, this formalism can capture one of the main ingredients in the epistemological analysis of knowledge: the *epistemic justification*

* Corresponding author.
E-mail address: bryan@renne.org (B. Renne).
[1] Funded by an Innovational Research Incentives Scheme Veni grant from the Netherlands Organisation for Scientific Research (NWO).
[2] Funded in part by an Innovational Research Incentives Scheme Vidi grant from the Netherlands Organisation for Scientific Research (NWO) and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC Grant agreement No. 283963.

0168-0072/$ – see front matter
http://dx.doi.org/10.1016/j.apal.2013.07.005

underlying the knowledge or belief possessed by an agent endowed with only limited logical resources and bounded rationality. As a consequence, this approach can be used to address the problem of "logical omniscience" [11,12,29,31,59,60] that affects other formalizations of doxastic/epistemic logic; indeed, Justification Logic is perhaps the most philosophically far-reaching and computationally sophisticated approach to this well-known problem.

Melvin Fitting's semantics [10,28] for this logic relates in an elegant way to Kripke's well-known relational semantics for knowledge and belief: essentially, in the Fitting semantics, a piece of evidence t justifies the belief in (or knowledge of) an assertion φ iff t is "admissible" for φ (i.e., it has the syntactic shape of a well-formed argument having φ as one of its conclusions) and in addition φ is implicitly believed/known (in the sense of Kripke's relational semantics). However, as we argued in a previous paper [15], the Fitting semantics is "Gettierizable," by which we mean that it is vulnerable to counterexamples of the type given by Gettier in his celebrated paper [32]. Moreover, we argued in [15] that this semantics is prone to other more basic problems: it leads to the counterintuitive conclusion that an agent must accept (believe in the legitimacy of) a body of evidence t supporting an assertion φ only because she believes the conclusion φ, even if in fact her belief in φ has nothing to do with the evidence t.[3] In the same paper, we sketched a proposal for an alternative semantics that could address this problem.

In this paper, we build on the work in [15] by proposing a new solution to the problem of the "Gettierizability" of Fitting semantics, a solution based on introducing a notion of *actual availability of "conclusive" (or "good") evidence*. A conclusive body of evidence is one that is fully reliable, in the sense that it is truthful whenever it is available, and moreover all of its component pieces of evidence are similarly reliable. Intuitively, we say that the body of evidence t is "explicit good evidence" (actually available to the agent) at a world w (and write that Et holds at w) if the agent has "constructed" (i.e., computed, observed, etc.) t and if in addition t is indeed conclusive evidence at world w. Despite what the name may suggest, an agent can still be mislead about whether Et holds: she might believe some constructed evidence to be conclusive, while in fact it is not. And conversely, t can be explicit good evidence without being actually accepted by the agent: although the evidence t is available to her (say, because she has observed or constructed t), she does not believe it to be fully reliable (though in fact it is). So whether evidence is accepted (i.e., believed, or "known," to be conclusive/legitimate) is independent of whether that evidence is in fact legitimate.

We present here a static logical account for reasoning about the notions of conclusive evidence, justifiable belief (and conditional belief), defeasible knowledge, and ("hard") information. In particular, while an agent may have an implicit belief, which may be justifiable in principle by some legitimate piece of evidence that is implicitly accepted by the agent, it need not be the case that this belief is explicit: the required evidence might not be currently available to the agent (due to her computational limits or to a lack of time). Only evidence that is available to the agent (via conscious observation, explicit logical construction, or actual computation) can serve as explicit justification for the agent's beliefs. When evidence is both available to the agent and accepted as legitimate by her, then any assertion supported by it is *explicitly believed*: the evidence provides the agent with an *explicit justification* for this belief. If in addition the evidence is also known to be conclusive, then any assertion supported by it is *explicitly known*: the evidence provides the agent with an *explicit conclusive justification* for this piece of knowledge.

It will therefore be useful to distinguish between the *implicit* notions of belief, conditional belief, defeasible knowledge, and ("hard") information and the *explicit* statements that are supported by specific pieces of evidence that say why the statement in question holds. As we will see, the implicit notions are closed under logical consequence and therefore suffer from the *problem of logical omniscience*, wherein the notion

[3] More precisely, if the agent implicitly believes φ in the sense of Kripke's semantics, then *any* evidence t that is admissible for φ justifies the belief. In particular, the semantics does not allow us to have two different pieces of evidence t and t', both admissible for a believed formula φ, and yet just one of these pieces of evidence is the justification. Further, an admissible t for a believed formula φ always provides a justification, even if in fact the agent does not believe some of the other assertions supported either by t or by one of its constituent pieces of evidence!