

V.F. HENDRICKS. *Mainstream and Formal Epistemology*. Cambridge University Press, 2006, xii + 188 pp.

Vincent Hendricks' book is an interesting and original attempt to bring together different traditions in epistemology. The term "mainstream epistemology" is used by the author to refer to the philosophical tradition of conceptual analysis of knowledge, which the author characterizes as the seeking of "necessary and sufficient conditions for the possession of knowledge using largely common-sense considerations and folksy examples and counterexamples" (p. 4). Under the head of "formal epistemology", Hendricks subsumes first and foremost work by epistemic and modal logicians on knowledge representation, information processing and the mathematical modelling of uncertainty (including, although more briefly in the book, work in Bayesian epistemology).

**Main argument.** The book rests on two observations. The first is that analytic epistemologists and epistemic logicians seem largely to pursue different goals: while analytic epistemologists are interested mostly in the precise *definition* of the concept of knowledge, with a concern to distinguish it from near approximations such as true belief or justified true belief (E. Gettier, "Is Justified True Belief Knowledge?", *Analysis* 23:6, 1963), epistemic logicians focus rather on how belief and knowledge are *processed*, and usually set aside foundational issues concerning the precise definition of the concept of knowledge. Hintikka thus argues that "epistemic logic and its applications (...) are completely independent of the issues concerning the definition (precise criteria) of knowledge" ("Knowledge acknowledged: Knowledge of Propositions vs Knowledge of Objects" *Philosophy and Phenomenological Research* 61, 1996, p. 254 – a quote not from Hendricks' book), by pointing out that the concepts of information and belief are more fundamental to epistemic logic than the concept of knowledge itself.

Despite this – and this is Hendricks' second observation – a common perspective may still be taken on both enterprises if one considers the way both analytic epistemologists and epistemic logicians assign truth conditions to knowledge. Since Hintikka's seminal work – discussed in Chapter 6 of the book, and undeniably a main bridge between the two traditions – knowledge and belief are characterized in semantic terms as *elimination of uncertainty*: a subject  $S$  knows that  $P$  if and only if  $P$  is true in all scenarios or possible worlds compatible with the subject's information. Equivalent definitions are at the heart of much of the philosophical literature dealing with the problem of skepticism, but with a different emphasis. For instance, D. Lewis, in his influential paper "Elusive Knowledge" (*Australasian Journal of Philosophy* 74:4, 1996) – discussed in Chapter 5 of the book – takes as his starting point the idea that  $S$  knows that  $P$  if " $S$ 's evidence eliminates every possibility in which not- $P$ ". The skeptical challenge to philosophers, however, concerns the grounds that allow a given subject to be sure that she has taken all possibilities of error into consideration, namely what epistemic alternatives should count as relevant or irrelevant for knowledge and attributions thereof.

As Hintikka emphasized, the problem of skepticism is usually not of interest to epistemic logicians, because "most of the questions concerning

the logic of knowledge are independent of how the dichotomy between admitted and excluded alternatives concerning the world is set up. Only the dichotomy itself matters” (*ibid.*). Belief therefore is the primitive category, and knowledge is defined as true belief, leaving issues of external justification implicit. The account of epistemic logic given in Chapter 6 adequately underlines this internalist perspective prevailing on knowledge in epistemic logic. More precisely, Hendricks stresses the link of Hintikka’s work to G.E. Moore and N. Malcolm’s tradition of “autoepistemology”, and puts the right emphasis on the fact that most of the philosophical issues connected to epistemic logic have to do with self-knowledge and rationality. Such is the case in static epistemic logic, with issues concerning introspection or closure of belief under logical consequence, but also in dynamic epistemic logic, where mechanisms of belief revision and informational updates do not depend on whether the information received is true, as much as on whether it is consistent with previous information.

One of the main originalities of the book, in this context, is the attempt made to relate the internalist perspective of epistemic logicians to the externalist concern of philosophers addressing the problem of skepticism. More precisely, Hendricks proposes to subsume under a common concept, which he calls *epistemological forcing*, the definitions and characterizations of knowledge proposed by different authors within the two traditions. The second main originality is the room given to the paradigm of *computational epistemology*, a branch of formal epistemology which Hendricks proposes to regiment in a modal logic framework.

**Organization of the book.** The book consists of nine chapters which can be grouped into four main parts. The first part (Chapters 1 and 2) delimits the two paradigms of mainstream and formal epistemology and introduces the notion of epistemological forcing. The idea is that knowledge can always be forced to hold over some space of relevant possibilities, and therefore be *consistently* attributed to an agent, even if the agent “commits many errors, even grave ones, in other but irrelevant possibilities” (Preface, p. ix). The notion is introduced by way of an analogy with Cohen’s set-theoretical notion of forcing (forcing an axiom to hold in a generic extension of a model of a given theory, in order to show that the axiom is consistent with the theory), even though Hendricks denies any deeper substantial import from the analogy in a footnote, and presents it as a metaphor in Chapter 9. Indeed, from a model-theoretic point of view, Hendricks’ idea of epistemological forcing generally corresponds to evaluating knowledge over a *restricted* subset of the set of all possibilities (as in the anti-skeptical strategies discussed in Chapters 3-5), rather than considering something like an extension. The analogy becomes more substantial in Chapters 7 and 8, however, where a notion of limiting convergence to knowledge is formally defined.

The second part of the book (Chapters 3-5) gives a survey and critical examination of the epistemological theories of three “mainstream” epistemologists, namely A. Goldman (Chap.3), R. Nozick (Chap. 4) and D. Lewis (Chap. 5), and of their respective attempts to defeat skepticism. Chapter 3 discusses the successive versions of Goldman’s epistemic reliabilism, in particular his idea that a belief-forming method is reliable if it

produces a high ratio of true beliefs in “normal” worlds. Chapter 4 examines Nozick’s characterization of knowledge in terms of true belief that is counterfactually sensitive to truth. Chapter 5 focusses on Lewis’s version of contextualism in epistemology, in which knowledge depends on which possibilities or error can be properly ignored depending on the context.

The third part of the book, on formal epistemology (Chapters 6-8), can itself be divided into two blocks. Chapter 6 (“Logical Epistemology”) is a self-contained introduction to modal epistemic logic and some of its recent developments and applications, including in game theory. The emphasis of the chapter is on canonical systems of propositional epistemic logic of the S4 and S5 type, and on discussions about the plausibility of the respective axioms. Hendricks also stresses the contrast between static and dynamic epistemic logics, namely combined logics of time and knowledge.

The temporal perspective on epistemic logic is the basis of Chapters 7 and 8, which together make a separate block and are generally more technical than the rest of the book. Chapter 7 introduces the paradigm of computational epistemology developed in particular by K. Kelly (*The Logic of Reliable Inquiry*, Oxford UP, 1996), E. Martin and D. Osherson (*Elements of Scientific Inquiry*, MIT Press, 1998), and Hendricks himself in several papers and in an earlier book (*The Convergence of Scientific Knowledge, a View from the Limit*, Trends in Logic: Studia Logica Library Series, Kluwer, 2001). With this chapter, Hendricks turns to yet another epistemological framework, in which the notions of inductive knowledge and knowledge acquisition become central. Chapter 7 gives some useful background on computational epistemology and its origins, in particular in the work of E.M. Gold on formal language learning and the identification of formal grammars on the basis of inductive evidence. The cornerstone is Gold’s notion of *identification in the limit*, namely the existence of a finite time after which all guesses made on the basis of partial evidence about the identity of a language are the same and correct.

In Chapter 8 (“Modal Operator Epistemology”), Hendricks presents a possible worlds model of time and evidence developed with S.A. Pedersen, which he uses to define a semantics of propositional epistemic logic with temporal and alethic modalities, based on the same idea. The semantics is original in that it incorporates notions of evidence and heuristics, traditionally wanting in epistemic logic, via the concepts of *discovery method* (function from finite evidence sequences to propositions) and *assessment method* (fonctions from finite evidence sequences and propositions into truth values). By definition, a hypothesis  $h$  is *knowable in the limit* relative to a discovery method  $\delta$ , given some sequence of evidence and time, if there is some time such that at all later times and for all further allowable inputs,  $\delta$  entails  $h$  and  $h$  is true. The formula  $K_\delta h$  is interpreted as “ $\delta$  knows  $h$  in the limit”. The rest of the chapter examines the validity of the axioms of S5, in particular axioms 4 and 5, when  $K_\delta$  is thus interpreted, depending on the structural constraints imposed on  $\delta$ . Further results are mentioned about *knowledge transmissibility*, namely the interaction of epistemic operators indexed by different methods.

Chapter 9, entitled “ ‘Plethoric’ Epistemology”, concludes on methodological considerations about mainstream and formal epistemology. To

some extent, Hendricks converges with the criticisms found in the writings of naturalized epistemologists against the excessive use of thought experiments (a recent illustration, not from Hendricks' book, is M. Bishop and J.D. Trout's polemical "The Pathologies of Standard Analytic Epistemology", *Noûs* 39:4, 2005, whose delineation of "Standard Analytic Epistemology" happens to overlap Hendricks' characterization of "Mainstream Epistemology"). Hendricks explicitly distances himself from naturalized epistemologists, however, by presenting formal methods as a way of controlling and giving structure to normative intuitions, better than simply dispensing with them (in which I side entirely with him). More generally, Hendricks puts forward a collaborative program of "plethoric" epistemology, in which conceptual analysis and formal methods are encouraged to regulate each other. The notion of "plethora" is quite sensibly opposed to the ecumenical fiction of a unity or complete merge between the two traditions.

**Critical comments.** Hendricks' book appears in a context of numerous and active publications in different areas of epistemology, and the first question the book raises concerns the adequacy of the opposition drawn between "mainstream" and "formal" epistemology. The distinction is not meant to be exclusive, as Hendricks points out (p. 159), and in some cases it is a matter of stylistic or discursive polarity. Indeed, prominent philosophers like D. Lewis, R. Stalnaker or T. Williamson all contributed both formal logical work *and* analytic argumentative pieces in epistemology. Nevertheless, the distinction made by Hendricks makes very good sense in my opinion, notably from a sociological point of view: the community of formal epistemologists generally include logicians, computer scientists and economists who are not professional philosophers and who ignore the concerns of philosophers; conversely, epistemologists in the analytic tradition are often unaware of advances in epistemic logic.

The book covers a lot of useful material in a concise, efficient and very clearly structured manner; it proposes an original synthesis between the two traditions, and as such it should be brought to the attention of both communities. The chapters are written with a palpable intention to reduce technicalities and make the formal results accessible to non-technically minded readers. The book could easily be used as an introduction to epistemology, especially Chapters 1 to 7. Chapter 8, the most technical, contains several diagrams which help to understand the formal definitions. All technical results are explained, but formal proofs are deferred to relevant papers and to Hendricks' earlier book: a technical appendix would have been useful to make the chapter entirely self-contained, however, especially for the section concerning the status of axioms 4 and 5 (pp. 144-145), which is quite dense.

One aspect the book does not cover, in particular in Chapter 6 on epistemic logic, concerns first-order modal logics of belief and issues of intensionality, which played a significant part in Hintikka's original treatment. While the focus on propositional logics is legitimate, a brief mention would have been welcome (for instance of *de re* vs *de dicto* beliefs, a distinction relevant to one of Gettier's original puzzles). The selection of "mainstream" epistemologists given in Chapters 3-5, although representative of three main trends in contemporary analytic epistemology, is

not supposed to be exhaustive either, a point Hendricks should have emphasized and motivated more clearly at the start. Thus some readers will probably regret the lack of discussion of more recent theories and ongoing debates in “mainstream” epistemology, in particular on the topic of contextualism.

Chapter 3 on Goldman’s epistemology is entitled “Mainstream Epistemology” and would better be named “Reliabilist Epistemology” to avoid any specific identification of “mainstream” with a specific theory. Two slight historical inaccuracies should be corrected in the next edition of the book: on p. 20, Russell is said to have anticipated Gettier on the difference between knowledge and true belief in the late 1940s; it is right that Russell anticipated Gettier, but he actually did so as early as 1912 in *The Problems of Philosophy*, Chap. 13; on p. 3 and in Chapter 4, F. Dretske’s paper “Epistemic Operators” is referred to as the anticipation of Nozick’s counterfactual conditions on knowledge, but Dretske’s relevant paper is actually “Conclusive Reasons”, published 1971. Conversely, more emphasis might have been placed on Dretske’s “Epistemic Operators” directly within Chapter 5 on contextualist epistemology, especially for the discussion of Lewis’s views about epistemic closure.

The part of the book on computational epistemology is the most personal and stimulating in my opinion, in particular because this is where the concept of epistemological forcing gets its bite, both formally and conceptually. As is the tendency with metaphilosophical concepts, Hendricks’ notion of epistemological forcing strikes me as too unifying rather than too narrow in the book (a criticism Hendricks lucidly entertains at the end, p. 165: “Perhaps the forcing metaphor is too narrow and too unifying at the same time”), although the idea of finding a common strategy behind all replies to skepticism remains original and forceful. I do not think, in particular, that Hendricks’ logic will be the right logic to analyze and solve the original Gettier examples, which do not involve learning, as much as justification (even though the concept of limiting knowledge may be *immune* to structurally analogous puzzles, as Hendricks suggests, somewhat elliptically, on p. 140).

Despite this, the concept of *knowledge in the limit* introduced in Chapters 7 and 8 appears as the clearest illustration of Hendricks’ forcing idea, and justifies his remark that computational epistemology “balances delicately between pessimistic skepticism and optimistic epistemology” (p. 128): on the one hand, discovery methods are always fallible (below the limit); on the other hand, categorical knowledge can always be ascribed to learning agents in principle (namely in the limit), even if the agents themselves are not necessarily in a position to know that they know (due to the threat of negative evidence). This distinction between first-person and third-person perspective, brought up in Chapter 2, is another useful clarification made in the book, notably for the discussion of logical omniscience given in Chapter 6.

One can foresee several interesting perspectives to the notion of knowability in the limit and the way it is regimented in Hendricks’ system, in particular with regard to two other current areas of research that are at the cross-roads of formal and mainstream epistemology. The first concerns the issue of knowability in principle, as raised by the Fitch Paradox

(“if all truths are knowable, then all truths are known”). The paradox is not discussed in the book, but it has been a topic of growing interest to both formal and mainstream epistemologists recently. It raises the problem of the combination of epistemic modalities with temporal and alethic modalities, and as such it should belong in an extension of what Hendricks calls “modal operator epistemology” (Hendricks’ logic contains special modalities of *empirical* possibility and necessity, but no *metaphysical* modalities). Another area with which one suspects there might be interesting connections is the emerging paradigm of evidence-based logics of knowledge, namely epistemic logics with explicit justifications, as developed by S. Artemov and his school. In Hendricks’ logic, epistemic agents are identified with discovery methods, which relate received evidence to propositions. Evidence segments come close to epistemic justifications therefore, even though they cannot be referred to in the object-language, and remain implicit in that sense. The logic may be enriched, however, to allow explicit reference to evidence.

I mention these further links not to detract from the main contribution of the book, but on the contrary, and in line with what Hendricks calls “mergers” between traditions, to emphasize the wealth of perspectives brought up by Vincent Hendricks’ stimulating and valuable book.

Paul Égré, Institut Jean-Nicod, CNRS, Paris