# M. D' Agostino, S. Modgil and C. Larese. *Depth-Bounded Reasoning. Volume I: Classical Propositional Logic. College Publications, London. 2024, xvii + 225, ISBN 978-1-84890-442-2*

:: Alejandro Solares-Rojas

**Abstract**
D'Agostino et al. recently launched book on the College Publication series on Logic and Bounded Rationality is reviewed. Applications to human-oriented AI are emphasized.

**Keywords**
Depth-bounded Boolean logics; bounded rationality; book review

**How to Cite**
Solares-Rojas, A. *M. D' Agostino, S. Modgil and C. Larese. Depth-Bounded Reasoning. Volume I: Classical Propositional Logic. College Publications, London. 2024.* The Reasoner, 19(2). http://doi.org/10.54103/1757-0522/28422

Can logic-based 'slow thinking' models complement machine learning 'fast thinking' methods in meaningful and practical ways? According to Neuro-Symbolic AI, logic-based models constitute a promising interface between opaque machine-oriented methods and human-oriented design and audit, as these models rely on widely developed theories of inference and argumentation.

http://doi.org/10.54103/1757-0522/28422

However, standard theories are highly idealized in that they model omniscient agents who can recognize all logical consequences of their assumptions. Approaches like the one proposed in this book aim to provide realistic models, accounting for the cost that inferences imply to resource-bounded agents. These approaches recognize that making inferences often exceeds cognitive resources and is generally computationally hard, as evidenced by the likely intractability of Classical Propositional Logic (CPL). In particular, the book's approach approximates classical-logic reasoning by defining a hierarchy of increasingly stronger, yet tractable, sub-logics that converge to CPL. These sub-logics can be intuitively associated with resource-bounded agents who approximate ideal omniscient ones. The conceptual basis of the approach is the distinction between actual and virtual information, namely, operational information that is practically available to the agent, versus hypothetical information that the agent does not actually hold but temporarily assumes as if she did.

The Prequel explains this distinction using the sudoku puzzle. Reasoning with actual information corresponds to steps performed using an ink pen, while reasoning with virtual information corresponds to steps requiring a pencil and eraser. A typical reasoning pattern based on actual information is the Single Candidate Principle (SCP): using only available information and known constraints, a single candidate is determined by excluding all other options. By contrast, some reasoning patterns essentially require the introduction of alternate hypotheses and keeping track of their consequences, i.e., the introduction of virtual information that is not even implicitly contained in the information held by the agent. The more nested use of virtual information required, the harder the deduction. The maximum number of these nested uses yields a sensible measure of the difficulty or depth of the deductions. Unfortunately, standard models are structurally inadequate to account for a notion of depth that is semantically well-founded, in the sense that the meaning of the logical operators remains the same throughout the corresponding hierarchy of approximations. Therefore, the approach resorts to non-standard semantics and proof theory. *Chapter 1* focuses on the basic, 0-depth approximation related to easy deduction steps that depend only on understanding the operators' meaning and applying the SCP accordingly. An informational semantics for the operators is given, based on the notions of informational truth and falsity. These notions satisfy a corresponding version of Non-contradiction but not of Bivalence, under penalty of omniscience. CPL's standard semantics is thus not suitable, so two equivalent alternative semantics are explored: constraint-based and non-deterministic. Both fix the meaning of the operators solely in terms of actual information, with no use of virtual information at all, and yield a notion of implicit-information extraction that is easy and 'local'. The 0-depth approximation is Tarskian, has no tautologies, and there is no functionally complete set of operators for it. So, different operators' choices define actually different 0-depth logics. Under any choice, however, the induced logic can be decided in quadratic time. This is shown via a proof-theoretic characterization that is a non-standard Natural Deduction system, where the introduction and elimination rules have a linear format, involve only actual information, and correspond to typical deduction patterns. These rules are taken to fix the meaning of the operators and are indeed sound and complete with respect to any of the two equivalent semantics. Quadratic-time tractability follows from the system's satisfaction of the subformula property, meaning the rules stand not only for easy but also 'local' steps. Moreover, the proof system enjoys an inversion principle, and derivations with the subformula property are uniformly shorter than those without it.

*Chapter 2* studies two alternative, albeit not equivalent, ways of characterizing the approximations of greater depth, yielding the weak or strong version, respectively. Both ways are characterized semantically and proof-theoretically, sharing the same 0-depth basis and overall conceptual framework. In both, a deduction's depth is identified with the maximum nested use of a single rule that implements Bivalence and controls the introduction of virtual information. However, the rule format and the specific induced measure of depth distinguish between weak and strong approximations. The strong format can represent non-nested applications within the same derivation, while the weak format does not. Thus, the depth of the 'same' deduction may well be lesser in the former format than in the latter. Their semantics vary according to the format, but they are essentially recursive extensions of the 0-depth approximation semantics. The main point is that depth-increase corresponds to the indispensable introduction of information that cannot otherwise be obtained by the operators' meaning and SCP applications. This intuitively involves more costly reasoning steps whose cost increases proportionally to their nesting. Thereby, hierarchies of approximations are defined, where up to $k$ nested uses of virtual information are allowed, and whose tractability is guaranteed whenever $k$ is fixed and the subset of formulas that can be conclusions of the respective Bivalence-rule or the introduction rules is suitably restricted. Fewer restrictions yield deductively stronger approximations, and their suitability depends on the intended application. Remarkably, the approach provides a logical measure of the difficulty of single deductions, their tractability being a by-product. The $k$-depth approximations, $k > 0$, may be Tarskian depending on the mentioned subset restrictions. Moreover, tautologies increase with $k$ and derivations can be normalized. Furthermore, non-refutational normal proofs enjoy the non-contamination property, which is a sort of variable-sharing property that bans irrelevant applications of *ex-contradictione quodlibet*.

This last property is useful for applications in argumentation theory, which is the topic of *Chapter 3*. Argumentation provides a unifying and promising setting for a variety of non-monotonic logics. Specifically, it allows for dialogues between agents, where they reason together by exchanging information, resolving conflicts, and finding joint deliberations. However, standard models impose counterintuitive and highly idealized requirements on agents. First, they usually leave implicit the proof-theoretic means by which arguments are constructed and thus also their persuasive force. Second, they imply omniscience by assuming that all arguments defined by a base can be constructed and included in the corresponding framework, and that the legitimacy of each argument is verified by checking, prior to inclusion, that its premises are consistent and non-redundant. These assumptions depart from real-world argumentation and are intractable.

The book's approach facilitates models that are suitable for practical desiderata and rational with respect to resource-bounded agents. Specifically, a notion of argument is given that distinguishes between premises that the agent commits to and those 'supposed for the sake of argument'. This allows for realistic models of premises' inconsistency demonstration via dialectical inter-agent argumentation. Then, intractable checks for arguments' legitimacy are dropped in favor of frameworks that include only the arguments within the agents' construction capabilities. The approach takes normal proofs as explications of arguments, and the resource-boundedness notion is exploited in that agents may still be credited as rational when tractably constructing arguments up to a given *k*-depth. Accordingly, the non-contamination property of *k*-depth normal proofs stops the generation of obviously redundant proofs by tractable means.

*Chapter 4* discusses how the approach can help solve philosophical problems arising from the view that CPL is informationally trivial. Orthodoxy holds that, in any valid deduction, the information of the conclusion is implicitly contained in the information of the assumptions. Valid deductions are said to be analytic in the semantic sense that their validity depends merely on the meaning of the operators. However, CPL's probable intractability strongly suggests that the conclusion of certain complex inferences may convey information that is not contained in the assumptions, in the objective sense that there is probably no feasible procedure for extracting it. Therefore, these inferences should be regarded as synthetic. According to the approach, CPL is 'trivial' only for omniscient agents, and not for realistic agents who consume resources when reasoning. Only 0-depth deductions are analytic, whereas deductions of greater depth are increasingly synthetic, in that their validity does not depend solely on the operators' meaning and their conclusion conveys information that is not even implicitly contained in the assumptions.

The *Conclusion* discusses ideas and methods closely related to the approach's non-standard semantics and proof theory. A brief overview of the state-of-the-art of an emerging research program is also given, which spans from covering logics other than CPL to applications in Probability.

In summary, the book provides logic-based models of resource-bounded agents within a robust and well-motivated conceptual framework. These models are useful in a range of practical and multidisciplinary applications, particularly in human-oriented AI, which is of current importance. I would have liked to find more pointers to alternative approaches for designing these models in the book. However, scattered references were perhaps avoided, and a robust survey definitely deserves independent treatment. I believe that the book's contents are of interest and accessible to a wide audience, having a clear multidisciplinary appeal at the intersection of AI, Economics, Philosophy, and Cognitive Science, to name a few. Except for some easily recognizable typos, the book is generally well-written and strikes a balance between technicalities and the intuitions underlying them. The content difficulty is kept to a minimum, requiring basic to medium technical training from the readers. Still, given that the book is well-organized and generally self-contained, readers can easily select content more suited to their background and interests, relying on pointers to more basic or complementary material. The book constitutes an excellent start for the series on *Logic and Bounded Rationality*, and I am sure that it will be a key reference for years to come.

ALEJANDRO SOLARES-ROJAS

Universidad de Buenos Aires