Exhuming the criticism of the logicist

Randy Goebel

Technical Report CS-86-33
September 1986

Exhuming the criticism of the logicist

Randy Goebel
Logic Programming and Artificial Intelligence Group
Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
rggoebel@waterloo.csnet

September 14, 1986

Abstract

McDermott has recently explained his fundamental philosophical shift on the methodology of Artificial Intelligence, and has further suggested that this shift is both necessary and inevitable. This shift results from a perception that a trend towards over formalisation has detached the real problems from the research results. McDermott's criticism is an enlightened exhumation of the criticisms of the seventies, and explains new ways in which the logical methodology can be abused. I argue that McDermott's criticism should not discourage the use of logic, but force a timely re-examination it's fundamental role in AI.

1 Introduction

Artificial intelligence is as a methodology for constructing rational machinery and is therefore concerned with machines that can represent and use knowledge. Philosophers, psychologists and computer scientists continue to debate the form and substance of knowledge, but an advantage befalls the latter, as they can experiment as both analysts and *synthesists.*

This potential for the synthesis and the subsequent re-analysis of knowledge as information structures is fundamental to the AI scientist's methodology. Heated debates over issues like procedures versus declarations [Win75] or predicate calculus versus associative networks [IB81] focus on the old distinction of "knowing that" versus "knowing how" (e.g., see [Den69, p. 184]), and have contributed insight into the fundatmental problems of representation and reasoning. The question remains, however, as to which alternative is more appropriate for designing intelligent machinery.

AI currently leans toward the use of logic for representation and reasoning. McDermott's strong criticism attempts to rebalance current trends in favour of a more procedural approach [McD86]. In examining his criticism, I argue that the picture is not so bleak as to convince us to abandon the logical approach.

2 What the logical approach should be

Many have eloquently explained what logic is and isn't (e.g., [Hay77,IB81]). Still McDermott claims to present what "most" AI people believe, so it is important to again review the fundamental issues regarding logic and representation in order to argue that his concept of "most" is misleading.

2.1 Semantics

The most important feature of logic is the correspondence between symbols of a formal language and objects in a domain. The correspondence provides the basis to pose the question "What does this expression mean?" about expressions of the formal language. Without it, *any* rational conclusion about the result of manipulating such symbols is impossible.

In some symbol storage and manipulation systems the correspondence between symbols and domain objects is never officially specified. But it exists, if only in the eye of the system's implementors and users. This need for *some* correspondence theory is what is accepted by most AI researchers, rather than McDermott's *fait au complet* assumptions regarding how much can be learned and what must be *a priori* represented, (e.g., "Even a program that learns must start out knowing may more facts than it will ever learn." [McD86, p. 1]).

The requirement for some correspondence between symbols and problem domain does not entail the use of any particular correspondence theory (e.g., Tarskian semantics), nor does it entail that a correspondence theory be absolute and unalterable. Tarskian semantics serves to establish the truth value of an assertion with respect to a world conceived in terms of relations on individuals. We can ask "what does the expression Φ mean?" and receive the answer that "It means that Φ has truth value α, in the Tarksian sense of truth."

But similarly, other basic properties of sets of symbols are of interest. For example, the *identity*[1] of symbols is another issue; one could ask "What does this sequence of symbols refer to?" The naive use of Tarskian semantics to identify "Fred" with Fred the tadpole may fail when Fred matures; many more or less elaborate manoeuvers will circumvent the required static correspondence (e.g., introducing time or suitable transformation functions) However, an executable semantics might preserve the simplicity of the original naming

[1] I.e., the thing to which a symbol refers, as opposed to those things claimed to be equivalent.

scheme by providing a procedure for re-establishing correspondences with the world. One may require a basis for determining truth when identities are dynamic, and so might select some combination of static and dynamic correspondence theories. The point is, of course, that logic requires that there be a correspondence theory regardless of *how* or *when* it is provided.

2.2 Reasoning

The wide acceptance of deduction does not imply that it is the only form of logical reasoning. For example, despite McDermott's critique of abduction and induction, Kunifuji et al. [KST*86] suggest that these styles of reasoning need not be non-logical.

To consider the possible role of other forms of inference, first recall that the correspondence theory of a particular logic justifies inference, e.g., that *modus ponens* is *truth* preserving. Similar justifications based on other kinds of correspondence theories are possible for abduction and induction. For example, one might required that an abductive inference rule preserve the already established identities. Intuitively, we might ask "what hypothesis can be abduced to explain an observation *without* introducing new objects?" This is related to the intuition behind circumscription, and the notion of a "circumscription policy" [Lif86]. Alternatively, one might want to abduce a hypothesis that minimally disturbs established identities, as in reasoning by analogy where we might explain someone's eating habits by hypothesising their equivalence with a previously identified swine.

Abduction and its role in automatic theory formation is of particular interest, as it requires the concept of preferred explanation or theory. That is precisely what the *Theorist project* at the University of Waterloo is investigating [PGA86,GFP86,Poo86]. Similar projects are in progress elsewhere, including those of Genesereth et al. [Gen82], Reggia et al. [RNW83], Reiter [Rei85], and de Kleer et al. [dW86b,dW86a].

McDermott is correct in noting that the focus of concern in this research on deductive nomological theory is the issue of preferred explanation or preferred theory[Hem65]. It is also the major issue in existing approaches to non-monotonic reasoning, especially as proposed by McDermott and Reiter [MD80,Rei80]. In our theory formation investigations, we have found that theory preference criteria depend on the problem. For diagnostic reasoning, for example, Poole argues for *least presumptive* theories [Poo86]. In planning, Goodwin and I advocate heuristic selection of *most persistent* theories [GG86]. In reasoning by analogy, where Greiner [Gre86] has shown how theory extension provides a model for learning by using analogies, Jackson and I are investigating a preference heuristic based on *simplest most relevant* theories [JG86]. The point is that the theory formation framework is so utterly simple and productive, it begs further investigation.

Induction might similarly be supported as a rational method of reasoning. For example, Bryan Magee explains that Popper's insight into induction and scientific reasoning is expressed in the statement "...although no number of observation statements reporting observations of white swans allows us logically to derive the universal statement 'All swans are white', one single observation statement, reporting one single observation of a black swan, allows us logically to derive the statement 'Not all swans are white'." [Mag73, p. 22] This doesn't seem to be an irrational conclusion. Furthermore, the theory formation work shows how induction is modelled by that rational selection of potentially refutable hypotheses. Israel has already argued that this kind of rationality is most important in the pursuit of a framework for common sense reasoning [Isr80].

3 The meta-methodological concern: rational machines

The concern with logic is not with deduction but with rationality. Logic's primary contribution is a method for describing the mechanisms of rational behaviour. By assuming that "rational" and "logical" are equivalent, we are further relieved from McDermott's despair with logic. Of course our study of reasoning must be founded on some practical and objective way of measuring, or at least detecting, rational behaviour. We otherwise succumb to a lack of discipline where that assumes rationality on the basis of a kind of shallow machine-behaviourism.

What might this measure of rational objectivity be? As philosophically elusive as it might seem, computer scientists are daily confronted with examples of rational and irrational machines. Systems software that reports failure because a needed file is missing is clearly more rational than software that ungraciously runs in a tight loop.

Other conceptions of machine rationality are possible. One is proposed by Donald Michie, in a paper entitled "Advice programming and the human window." [Mic80]. Michie's point is that somewhere on the continuum between complex procedural reasoning and direct table lookup lie programs that operate within the "human window." Such programs are distinguished by their ability to provide humanly understandable explanations of their behaviour. This point is well known to expert system developers — a system isn't *expert* unless it can explain its conclusions.

An ability to explain suggests rationality, but at least two further requirements remain. The first is that explanations must be at the appropriate level of detail; the second is that explanations for preferences be available. For example, consider a diagnostic system working in the theory formation framework. We view observations as theorems of an unspecified theory, so that diagnosis amounts to solving for values of Θ in the expression

$$facts \cup \Theta \models observations$$

where Θ takes on possible diseases as values. If potential diseases are specified in terms of their symptoms we can hypothesize all combinations of diseases, and test to see which entail the observations (de Kleer and Smith's work on diagnosing multiple faults provides a program that need not consider all possible combinations of such diseases[dW86a]).

Regardless of the mechanism used to solve the expression, two kinds of question arise for any particular value of Θ. The first is "why is Θ is a solution?" e.g., "Why does rheumatoid arthritis explain the observation 'aching knee'?" The second is about preferred values for Θ, e.g., "why did you diagnose rheumatoid arthritis instead of scarlet fever?"

The answer to the first question involves the procedure that verified the $\models$ relation between $facts \cup \Theta$ and the *observations*—the most likely candidate is $\vdash$, or ordinary deduction. The terse explanation "$facts \cup \Theta \vdash$ *observations*" might do for some, but more detail would be required by most. The required elaboration, typically given by existing expert systems, traces the application of inference rules to the axioms that participated in drawing the conclusion.

The second question asks how one theory is preferred over another. We might offer only a heuristic explanation like "rheumatoid arthritis is more common," or "scarlet fever is associated with a necessary but missing symptom." Contrary to McDermott's criticism, the explanation required may not *yet* be logically formalised, but it is certainly not illogical—there are plenty of rational reasons for preferring one theory over another. The question of formalising such explanations is really the question of formalising the process of scientific reasoning—the answer remains unknown.

A simpler but related question is whether such explanations should be expressed in terms of computation (i.e., procedures on data structures) or reasoning (i.e. inference rules on axioms). This brings us to the final section, where we examine McDermott's criticism of "deducto-technology."

4 Abstract computation: from Turing machines to theorem-provers, and beyond

McDermott's analysis and criticism of meta-reasoning and "deducto-technology" should be unraveled in a framework of multiple levels of abstract computation. Intuitively, one requires a mental picture where equivalent theories of computation are stacked upon one another, according to some informal "level of abstraction" ordering (this ordering is, in itself, controversial). For example, we might place Turing machines at the bottom, followed by Post's correspondence systems, followed by LISP (λ calculus) systems, followed by Horn clause systems.[2]

Here arises a most important question about the appropriate level for providing explanations. For example, how can we determine when explanations are more appropriately provided in the language of λ-conversion or as Horn clause proof theory? The answer is not always the same, but there must certainly be some criteria for choosing. The proper

[2] See [Tar77] or [vSe82] for results on Horn clause computability that provide a precise characterisation of computation as deduction.

choice of description language is also considered by cognitive scientists like Stich [Sti83], Pylyshyn [Pyl84] and Churchland [Chu84] in their analysis of "folk psychology" as an appropriate vocabulary for analysing cognitive behaviour. McDermott fails to acknowledge the importance of this choice when he criticises logic programmers for referring to "...something as trivial as list-processing operations..." [McD86, p. 9] as logical inference. It is true, for example, that one can analyze the time complexity of a Prolog append program in terms of list computations (e.g., cons, car, cdr) or inferences. The choice depends on the motivation: the former is appropriate when comparing LISP and Prolog systems, but the latter is appropriate for comparing two different Prolog axiomatisations of the append relation.

Another example can be given in terms of the above diagnosis scenario. In naive implementations, theory preference might only be explained in terms of the order in which a procedure generates candidate hypotheses (e.g., lexicographic). A more rational implementation could describe a theory preference heuristic in terms of the inferences that led to the preference, e.g., inferring one disease over several others on the basis of experimental observations or the frequency analysis of case data. This strategy need not be deductive, but it is obviously rational.

To infer an ordering of explanatory theories requires meta reasoning. This too, McDermott criticises. For example, he correctly points out that meta-theoretical manipulations are not constrained to retain soundness at the object level. However, the diagnosis example clearly shows how the meta reasoning required to order theories can be decoupled, in terms of soundness, from the object level theory. The reasoning that orders theories has no impact on whether or not any particular theory deductively implies the observed symptoms.

In this case McDermott's criticism is more appropriately aimed at meta reasoning systems that merely encode efficiency heuristics that "short-circuit" object level derivations. The inference system of the two levels could be identical; the only difference is that the meta level has "compiled" search strategies while the object level uses interpretative search strategies.

5 Summary and conclusions

Logic is an appropriate tool for the study of machine representation and reasoning. McDermott's despair of the logical methodology for artificial intelligence is actually a criticism of the abuse of logic, rather than despair with logic. By taking an unnecessarily narrow view of the logical methodology, McDermott criticises the "logicists" for problems not entailed by the most general concept of logic. The antidote is to equate logical machines with rational machines, and argue that rational behaviour can only be interpreted in terms of operational descriptions at the appropriate level of abstraction.

In support of McDermott's position, it is clear that we know too little about how to pursue the problem of providing a machine with a theory of rationality. The appropriate combination of logical inference and procedure execution is yet unknown, but it is clear that both are involved. McDermott's reaction is reasonable, given the current tendency to overformalise and loose the motivation of constructing a rational machine. However, the major contribution of McDermott's paper is the further research it will stimulate.

Acknowledgements

Romas Aleliunas, Scott Goodwin, Eric Neufeld and David Poole contributed valuable comments on an earlier draft of this paper. This research was supported by National Sciences and Engineering Research Council of Canada grant A0894.

References

[Chu84] P. Churchland. *Matter and consciousness: A contemporary introduction to the philosophy of mind.* MIT Press, Cambridge, Massachusetts, 1984.

[Den69] D.C. Dennett. *Content and consciousness.* Routledge and Kegan Paul, London, England, 1969.

[dW86a] J. de Kleer and B.C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 1986. [to appear].

[dW86b] J. de Kleer and B.C. Williams. Reasoning about multiple defaults. In *Proceedings of the AAAI-86 Conference*, pages 132–139, University of Pennsylvania, Philadelphia, Pennsylvania, August 11-15 1986.

[Gen82] M.R. Genesereth. The use of design descriptions in automated diagnosis. In D.G. Bobrow, editor, *Qualitative reasoning about physical systems*, pages 411–436, MIT Press, Cambridge, Massachusetts, 1982.

[GFP86] R. Goebel, K. Furukawa, and D. Poole. Using definite clauses and integrity constraints as the basis for a theory formation approach to diagnostic reasoning. In *Proceedings of the Third International Conference on Logic Programming*, pages 211–222, Imperial College, London, England, July 14-18 1986.

[GG86] S.G. Goodwin and R. Goebel. *Theory preference based on persistence.* Research report CS-86-34, Department of Computer Science, University of Waterloo, Waterloo, Ontario, September 1986.

[Gre86] R. Greiner. *Learning by understanding analogies.* Technical Report, Department of Computer Science, University of Toronto, Toronto, Ontario, August 1986.

[Hay77] P.J. Hayes. In defence of logic. In *Proceedings of the Fifth IJCAI*, pages 559–565, MIT, Cambridge, Massachusetts, August 22-25 1977.

[Hem65] C.G. Hempel. *Aspects of scientific explanation and other essays in the philosophy of science.* The Free Press, New York, 1965.

[IB81] D.J. Israel and R.J. Brachman. Distinctions and confusions: A catalogue raisonne. In *Proceedings of the Seventh IJCAI*, pages 452–459, The University of British Columbia, Vancouver, British Columbia, August 24-28 1981.

[Isr80] D.J. Israel. What's wrong with non-monotonic logic? In *Proceedings of the AAAI-80 Conference*, pages 99–101, Stanford University, Stanford, California, August 18-21 1980.

[JG86] W.K. Jackson and R. Goebel. *Using theory formation for reasoning by analogy.* Technical Report, Department of Computer Science, University of Waterloo, Waterloo, Ontario, 1986. [in preparation].

[KST*86] S. Kunifuji, H. Seki, T. Takewaki, K. Furukawa, and K. Tsurumaki. *Toward mechanization of deductive, inductive and abductive inference.* Technical Report, Institute for New Generation Computer Technology, Tokyo, Japan, April 1986.

[Lif86] V. Lifschitz. Pointwise circumscription: Preliminary results. In *Proceedings of the AAAI-86 Conference*, pages 406–410, University of Pennsylvania, Philadelphia, Pennsylvania, August 11-15 1986.

[Mag73] B. Magee. *Popper.* Fontana Press, London, England, 1973.

[McD86] D.V. McDermott. *A critique of pure reason.* Technical Report, Yale University, New Haven, Connecticut, June 1986.

[MD80] D.V. McDermott and J. Doyle. Non-monotonic logic I. *Artificial Intelligence*, 13(1 & 2):41–72, 1980.

[Mic80] D. Michie. *Advice programming and the 'human window'.* Technical Report, Machine Intelligence Unit, University of Edinburgh, Edinburgh, Scotland, 1980.

[PGA86] D. Poole, R. Goebel, and R. Aleliunas. *Theorist: A logical reasoning system for defaults and diagnosis.* Research report CS-86-06, University of Waterloo, Waterloo, Ontario, February 1986. [to appear in *Knowledge Representation*, N. Cercone and G. McCalla (eds.), Springer-Verlag].

[Poo86] D. Poole. *Default reasoning and diagnosis as theory formation.* Research report CS-86-08, Department of Computer Science, University of Waterloo, Waterloo, Ontario, March 1986.

[Pyl84] Z. Pylyshyn. *Computation and cognition: Toward a foundation for Cognitive Science.* MIT Press, Cambridge, Massachusetts, 1984.

[Rei80] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13(1 & 2):81–132, 1980.

[Rei85] R. Reiter. *A theory of diagnosis from first principles.* Technical Report TR-187/85, Department of Computer Science, University of Toronto, Toronto, Ontario, December 1985.

[RNW83] J. Reggia, D. Nau, and P. Wang. Diagnostic expert systems based on a set covering model. *International Journal of Man-Machine Studies*, 19(4):437–460, 1983.

[Sti83] S. Stich. *From Folk Psychology to Cognitive Science: The case against belief.* MIT Press, Cambridge, Massachusetts, 1983.

[Tar77] S-Å. Tärnlund. Horn clause computability. *BIT*, 17:215–226, 1977.

[vSe82] Šebelék, J. and Štěpánek, P. Horn clause programs for recursive functions. In K.L. Clark and S-Å. Tärnlund, editors, *Logic Programming*, pages 325–340, Academic Press, New York, 1982.

[Win75] T. Winograd. Frame representations and the declarative/procedural controversy. In D.G. Bobrow and A. Collins, editors, *Representation and understanding*, pages 185–210, Academic Press, New York, 1975.