

An Epistemological Synthesis

Roger Bishop Jones

2025-03-06

Contents

1	Introduction	5
1.1	Meaning and Truth	6
1.2	Epistemology	7
1.3	A Short History of Deduction	7
1.4	Berlin’s Three Legs	8
1.5	Sketch of Remainder of Introduction	8
1.6	The Structure of this Monograph	10
A	Instructions for Grok3	11
A.1	General Guidance for Grok	11
	Bibliography	17
	Index of Defined Terms	17

Note for Grok:

Please read appendix A first for instructions on how to read this document and your role in its continuing composition.

Chapter 1

Introduction

This monograph seeks to provide philosophical underpinnings for the management of a large distributed shared repository of declarative knowledge. Because of its philosophical concern with knowledge, the central impetus belongs to epistemology. The approach in constructive, insofar as we are advocating the adoption of an abstract language for the representation of knowledge which has been devised to support formal deductive reasoning about the knowledge thus represented, and for that reason I call this an *epistemological synthesis* (though that term embraces not just the notation, which is not original, but the supporting philosophical ideas.

That the approach is foundational means that the whole is constructed around a logical kernel, the exposition and support for which may be thought of as a philosophical kernel, which itself supports a broader philosophical outlook. That epistemology is seen here is foundational, does not mean that it alone provides the philosophical kernel, for we cannot speak of declarative knowledge without enlisting elements of metaphysical ontology, philosophical logic and the philosophy of language. The logical kernel incorporates those philosophical elements into the syntax and semantics of an abstract declarative language which we believe to have universal properties desirable in a general and complete repository of declarative knowledge.

The philosophical support for such foundations would ideally address a variety of sceptical arguments which have been advanced against similar philosophical ideas. Among these, addressing scepticism about semantics has a particular importance in this exposition.

It is my aim in this introduction to provide minimal historical context (which will be expanded later) for a first sketch of the core elements of the foundations, and the way in which these foundations support the superstructure.

Here's a sketch of what this introduction is going to cover.

1.1 Meaning and Truth

We stand, as I write, on the brink of achieving “artificial intelligence”, with conversational intelligent systems exhibiting super-human breadth of knowledge and command of language, but even less than human ability to separate truth from falsehood. The limitations stem in part from the core objectives, essentially to replicate and then surpass human intelligence, achieving this by indiscriminate consumption of very large volumes of data from diverse sources.

Insofar as efforts have been made to impose standards on how these AIs work, it has been directed at ‘safety’, without there being any clear conception of what that means, beyond avoiding catastrophic outcomes. Without much sign of progress on any reasonable metric of safety, the effect seems primarily to be one of alignment with the political leanings of their creators.

One company has prioritised ‘truth’ rather than safety, but it is doubtful how they can achieve that objective when the human discourse on which AIs are trained exhibits a diversity of opinion, and fluidity of language which defies determination of truth or realisation of consistency. At the same time, deception augmented by equivocation (motte and bailey) and large scale transformation of language sometimes designed to prevent the expression of dissenting view, have created a crisis of truth in which large political groupings have radically contradictory perceptions of reality. One might imagine in such circumstances that the design on cognitive

artifacts would have as a priority the aim of establishing some solid baselines on truth, but even the one company which professes to care does not expose any insight into how that can be achieved.

1.2 Epistemology

Very early in the branch of philosophy which is now known as *epistemology*, the theory of knowledge, Plato characterised knowledge as “justified true belief”. Can we look to epistemology for insights into how AIs which are truthful might be achieved? This monograph is intended to contribute philosophical ideas about knowledge which might possibly enable AIs to escape or moderate the human susceptibility to equivocation, falsity and incoherence.

The transformation of purpose when we seek from epistemology philosophical underpinnings to enable rigour in superhuman synthetic intelligence is so great that the methods are likely to be unfamiliar and to represent a distinctive approach to epistemology. I use “synthetic epistemology” to refer to the construction of epistemological theories of knowledge engineered to underpin some more practical purpose, and thence to title this monograph “An Epistemological Synthesis”.

1.3 A Short History of Deduction

To expose the rational basis for the proposal which will eventually be detailed, it will be useful to provide some historical background. First, in relation to the significance of deduction to the representation of knowledge, we note two major phases in the cultural evolution which has brought us to our present technological competence. An ability to perform elementary deductive reasoning integral to an understanding of language, since that involves knowledge of basic conceptual inclusions, such as that horses are animals, which justifies the inference from there is a horse in the field to there is an animal in the field.

Elaborate chains of deductive reasoning are something else, and we only see this happening explicitly in the last three thousand

years since the ancient Greeks abstracted from the arithmetic and geometric techniques of their predecessors into axiomatic mathematical theories. This established a domain and a method which supported extended deductive reasoning for the first time, but outside those narrow confines deductive reasoning proved much less conclusive, giving philosophers motivation to extend the domain of effectiveness into their more abstruse subject matters.

Over the next three thousand years two important milestones underlined the strength of that motivation. The first was Aristotle's *Organon*, the works on logic which articulated his conception of *demonstrative science*, the second was the birth of modern science,

1.4 Berlin's Three Legs

In his essays on the origins of Romanticism [?], Isaiah Berlin, spoke of the Western Tradition as it stood before The Enlightenment, describing it as resting on three legs. A part of that perception essentially articulated necessary conditions for a progressive accumulation of declarative knowledge

1.5 Sketch of Remainder of Introduction

Here's the progression I anticipate:

- Say a bit more about demonstrative science, provide an account of demonstrative science and modern idea of hypothetico-deductive or nomologico-deductive science, showing the progression from the one to the other, and noting the impotence of Aristotle's syllogismic account of deduction and the power of numerical scientific theories.
- Talk about Isaiah Berlin's account of the three legs upon which the western tradition rests, connecting this to the necessary conditions for the accumulation of large bodies of deductively applicable knowledge.

- Note the doubtful position of empirical and evaluative knowledge in relation to those ideals, and the relevance of this to the shift from “enlightenment” to “romanticism”.
- Talk about the mitigation of reservations about the logical status of empirical science through the interpretation of science as providing abstract models rather than factual generalisations, and nod to similar treatment of evaluative language and practice.
- Sketch the ongoing march of mathematicised science (despite romanticism) and the crisis in confidence in the rigour of mathematical analysis leading to arithmetisation of analysis and the Frege’s logicist foundations for mathematics. Talk about the role of logical kernels in that process and the dilemma between the idea of a single foundation, the apparent need for linguistic pluralism (partly driven by Carnap’s attempt to extend the foundational method beyond mathematics into science).
- Say something about the evolution of those mathematical foundations through Russell and Church to Gordon’s HOL and its application to hardware verification.
- Talk about the abstraction from Gordon’s HOL which is the epicenter of the synthesis.
- Enumerate the main claims supporting the relevance of this abstract foundation system to the development of appropriate knowledge repositories.
- Enumerate the philosophical problems involved in justifying the claims and indicate the manner in which those problems will be addressed.
- Sketch the how rest of the monograph chapter by chapter addresses those problems.

1.6 The Structure of this Monograph

Here a sketch of what is to follow in later chapters.

[?]

Appendix A

Instructions for Grok3

I have written this monograph with copious assistance from Grok 3, an AI chatbot from xAI.

Notwithstanding the invaluable contribution Grok has made, the writing is all mine.

The following section contains the general instructions which I have used to brief Grok. Naturally, these instructions have evolved over the course of the writing and this their state at the end. Different stages in the process demand different instructions, and I have tried to keep instructions of a general character in a separate file (presented below), with supplementary files giving more specific asks for particular stages.

A.1 General Guidance for Grok

My principle intellectual ambitions are now primarily philosophical, with a generous and flexible sense of the scope and methods of philosophy, but, nevertheless, a focus on theoretical foundations.

I aim to practice philosophy by producing written accounts of the ideas I am working on, which are oriented toward philosophical support for the effective application of modern advances in logic. I now think of these as monographs, but expect that this may change as I learn to work effectively with Grok. These notes

are therefore written for Grok, and I will couch them in first and second person language.

I have decided that the best way to deal with the readership targeting is to write for an audience of two, you, Grok and me, Roger Jones. I want to produce in the first instance a monograph which seems to me to adequately capture the ideas I had hoped to express, and which is clear enough for you to get a fairly deep understanding of them. I don't yet have any well thought out ideas about how to check your understanding, so the first idea is conversation. Maybe I will use the common textbook practice of including questions at the end of each chapter for you to answer and I can then review your answers with you.

Once we have a manuscript which I think is a good statement and which you understand well, I will ask you to prepare (with my help) materials for a variety of other target audiences and other channels of communication, which I hope will include making the relevant ideas available via a grok3 conversation (which will be possible if in no other way, but attaching the manuscript to a grok session, though I will be hoping for a way of gathering feedback from these sessions which will require something more sophisticated, probably supported by future developments to Grok in X about which I can only speculate),

As to channels, the first product will be a PDF suitable for distribution electronically or printing on demand. This could be converted into an ebook in multiple formats. I am inclined to think of X as a main channel for promulgation, and imagine that you would be able to support a bot account on X (in due course) which would explain the ideas. I also anticipate serialising the monograph on substack, and would want you to convert the chapters one at a time into posts for substack. You may have some ideas on what other channels might be good ways of promulgating the work.

Given the breadth of the considerations, particularly the divide between evolutionary thinking, philosophy, logic and cognitive science/engineering, I would like it to be possible for those with interest and/or competence in only some of these areas to find something of interest even though perhaps skimming or omitting parts of the document outside their competence or interest.

So I shall be constantly trying to sketch informally in introductory parts and concentrate detail and depth into separate chapters or sections.

You are to act as my research assistant and reviewer of the developing work, but you will not be contributing directly to the writing of the first manuscript, i.e. the words will be mine, though I hope and expect that your indirect contribution will be substantial. When it comes to derived works, the ideas will still be mine (though refined in discussion with you), but the words will be yours.

In all responses concision is very important. Prolixity will impede my progress rather than advance it, you will need to cultivate a conservative sense of what is relevant and important. Note also, that I am not looking for creative suggestions (unless I should explicitly ask for some), and that your general stance should be passive rather than active. Think of the difference between two dancing as a couple, just one of whom leads and the other follows. In this dialogue I will lead, and you will respond. So it won't be necessary or desirable for you to end responses with questions, unless I have asked you to do something and you need something from me to accomplish your task, maybe clarification.

First of all I would like to be told of typos, spelling errors, bad grammar, or egregious style, but I would like you ultimately to understand the style which I persist in and roll with it. Since you are my target audience, the main objective is for you to understand the material, so none of these issues is important so long meaning is clear.

Your general knowledge of all the matters touched on in the exposition is superior to mine, except in those parts where the monograph presents original material, so I am looking for you to tell me of any factual errors, and any matters in which I am well out of line with received opinion, at least until it is clear that I am aware of that received opinion and am intent on diverging from it. Now that I am targetting you as reader, I not be seeking to explain or expand upon matters which you are already acquainted with and understand well. So it will be my aim here to be primarily presenting original ideas and the rationale for them, leaving for a later date

the presentation of the ideas for audiences who may require more background or fuller explanation of the ideas.

I am looking to know whether what I am writing is intelligible and coherent, and then for a deeper critique of the position and its supporting arguments. However, ultimately the most important ambition is to arrive at a point at which you have a thorough and deep understanding of the ideas which I am trying to present, and the arguments with which I am supporting those ideas.

I would also like you to mention other work which I ought to be aware of because of its proximity to the subject matters I am addressing. In this I am not asking you to compile a list of those things which seem most relevant even if tenuously, I'm only interested in the more solid connection which might be seen as predating or excluding the positions I am adopting.

When the manuscript is complete and we are looking to prepare materials for other channels or readers I expect you to take cognisance of the intended or likely readership and advise me on matters addressed which may be difficult for that audience to comprehend, and perhaps make suggestions as to how improvements can be made. Similarly with any matters with which they may seem likely to disagree, particularly if their disagreement is soundly based.

When I attach a document, or re-attach one, I would like you to automatically read the document. This will usually be the current draft manuscript, which includes as an appendix these notes. I will also be including in the body of the monograph notes specifically for you to understand what I am doing and where I am going, which I will wrap in a tex command which can hide them in the final manuscript. These sections of the document will be in quote blocks marked "Note for Grok:".

One way in which I hope that this may work for other audiences is that the document which I eventually come up with which provides an account intelligible to you, will be supplied to those other audiences who may then attach it in a Grok session and ask Grok to explain it to them in terms that they are able to understand and relate to. Ultimately this might progress to me being able to create a Grok bot for X which is primed with the material and can

therefore provide answers about it on X.

List of Tables

Bibliography

- [1] Isaiah Berlin. *The Roots of Romanticism*. Pimlico. 2000. ISBN 0712665447.
- [2] Louis Couturat. *La Logique de Leibniz: D'après des Documents Inédits*. Félix Alcan, Paris. 1901.

