Folk Psychology and Naïve Physics

Murray Shanahan

Imperial College
Department of Computing,
180 Queen's Gate,
London. SW7 2BZ.
England

Abstract

This essay concerns the relationship between folk psychology and the prospective products of the so-called logicist approach to artificial intelligence. It is suggested that the products of the logicist research programme may demystify the terms of folk psychology, such as "belief", since the everyday belief ascriptions of the folk psychologist could be translated into the more precise, but fundamentally similar, language of the artificial intelligence scientist. The purpose of this essay is to emphasise this similarity through descriptions of both folk psychology and logicist artificial intelligence.

Introduction

The subject of this essay is a particular sort of activity, namely *everyday problem solving*, an aptitude for which is displayed by each human being in his ability to attain simple goals, such as navigating an unfamiliar city, or making a cup of tea, or driving a car. Even cats and dogs display such an aptitude, to a lesser degree than human beings but to a greater degree than house flies, in their capacity to find food and shelter and sexual solace.

Folk psychology is the day-to-day "theory" used to explain and predict such activity. As a folk psychologist, I have a degree of understanding of everyday problem solving, but I seek a deeper understanding, an understanding to which folk psychology, as it stands, is inadequate, an understanding sufficient for the fulfilment of the vision of artificial intelligence (AI) research. Beginning with a portrait of contemporary folk psychology, I will proceed to a discussion of the goals of AI research, and will go on to explore the relationship between the two, emphasising the importance of the capacity to translate between the language of the folk psychologist and the language of the AI scientist.

1. Folk Psychology

Folk psychology is the day-to-day theory we use to explain and predict exveryday problem solving behaviour. We can imagine a languageless folk psychologist, who could not explain but who could predict, and whose capacity to do so would be reflected in an ability to influence the actions of her subject according to her own desires, and to adapt her own actions to what she expects her subject to do. This is what is meant by the *practices* of folk psychology. Such practices are not linguistic, although it is hard to imagine someone acquiring them without the aid of language. In contrast, the *customs* of folk psychology are reflected in folk psychological talk, in the verbal *attitude* ascriptions involved in everyday explanations of action. With the prevailing folk psychology, the nature of predictive practice is revealed in explanatory custom. By investigating the language of the folk psychologist we discover the state of the art in the folk understanding of everyday problem solving.

Suppose that the folk psychologist is sitting in a pub and is asked to explain the behaviour of someone who goes to the bar and buys a drink. She might say something like this: "He saw that the bar was open, he knew that the bar sold drinks and he was

thirsty so he approached the bar and bought a drink". A folk-psychological explanation involves the ascription of various propositional attitudes to a subject — beliefs, desires, hopes, fears, intentions, suspicions, etc (in this essay I shall concentrate on belief and desire), which are characterised by their use of *embedded sentences*. It also involves the assumption that certain "causal" relationships obtain between a subject's perceptions, propositional attitudes and actions — perceptions give rise to attitudes, attitudes interact to form further attitudes, and attitudes give rise to actions. The folk psychologist is able to make (reasonably) accurate predictions and to give (fairly) satisfactory explanations of a subject's behaviour. So folk psychology is a sort of theory, albeit an unwritten one, manifest only in practices and customs.[1]

It is sometimes useful to categorise the beliefs ascribed by the folk psychologist — they can be about particular states of affairs, such as this room at this instant, or they can be part the subject's grasp of some concept, such as water or London or the number five. If the folk psychologist says of someone: "He believes that the sink is full of water", then she is crediting him with possessing the concept of water, the concept of a sink and with an understanding of what it is for something to be full of a liquid. Let us consider what it is to possess the concept of water. The subject is able to recognise water — he knows what it looks like, that it is clear and sparkles and shines, he knows what it sounds like, gushing, trickling or splashing, he knows what it feels like and tastes like. He knows how it behaves, how it falls and splashes, how it spreads across surfaces or soaks into them, how it runs downhill, how it fills containers and how they overflow. He knows the many uses of water, for drinking, for swimming in, for washing with.[2] The concept of water (in some form or other) is common to (almost) all human beings, whilst other concepts are less common, such as the concept of my pet budgie or of transcendental idealism.

The folk psychologist displays an understanding of her subject's perceptual apparatus when she says: "He saw that the sink was full of water". She knows the circumstances in which he can see such a thing; he must be looking in the right direction, his eyes must be open, there must be enough light to see by, etc. When she describes

---

[1] See P.M.Churchland, Eliminative Materialism and the Propositional Attitudes, Journal of Philosophy, vol LXXVIII (1981), no 2, p 67 and A.Clark, From Folk Psychology to Naïve Psychology, Cognitive Science, vol 11 (1987), p 139 for opposing views on the status of folk psychology as a theory.

[2] I am not, of course, claiming that all of the beliefs mentioned are necessary for a grasp of the concept of water. I am simply suggesting some beliefs which seem to us to be characteristic of such a grasp.

what her subject perceives, the folk psychologist employs the terms of the web of conceptual beliefs she has attributed to him.

The folk psychologist attributes to her subject certain dispositions to form and revise beliefs about particular states of affairs according to what he perceives, and to reason from one such belief to another. Suppose it is close to last orders in the pub and that someone goes to the bar to buy a drink. When he arrives he sees that there is no one to serve him, so he concludes that the bar is closed, and walks away. But then the barman (who has been changing the barrel) comes rushing in to take last orders, and the customer returns. The folk psychologist says: “He saw that the barman had gone, so he thought that the bar was closed. But then he realised that the barman was only changing the barrel. So he went back and bought a drink”.

The folk psychologist only ascribes *dispositions* to reason from one belief to another. If someone asked her whether the barman knew that “Tigers don’t have pink stripes”, she would consider it a most peculiar question. She might say something like this: “In a sense he knew it, because everyone knows that tigers only have black stripes. But he surely didn’t need to know it to change the barrel”.

If the subject does not possess a particular concept, then the folk psychologist’s predictions and explanations can cope, although the more exotic the subject seems to her, the less intelligible she finds his behaviour and the more work is involved in understanding it. Suppose the folk psychologist is dealing with a particularly peculiar subject, who doesn’t know that water is drinkable. Then she might still say of him: “He believes the sink is full of water”, and she might also add “But he doesn’t know that water is drinkable (poor chap)”. His lack of grasp of the concept will be manifest in his behaviour — he refuses to drink water, even when he is thirsty and has access to it. Of course, a subject’s lack of grasp of a concept admits of degree, depending on what beliefs are missing and how important they are.

It may happen that the subject will learn that water is drinkable. He may observe somebody drinking it, or he may accidentally ingest some himself and find it agreeable. The folk psychologist can cope with such changes in conceptual belief. She says of her subject: “He saw someone drinking water and realised that it was drinkable”. Similarly, the folk psychologist has an understanding of concept acquisition. Consider a young child who has not yet learnt how to use a knife and fork. The folk psychologist sees him playing with them and says: “He is learning about knives and forks”. She can tell when

he makes progress, when he begins to display an ability to wield his cutlery properly. She even knows how to speed up the child's progress, through suitable instruction and demonstration.

So the folk psychologist conceives her subject's web of conceptual beliefs to be in a continuous state of flux. New beliefs can be acquired and old ones revised and, as Quine points out,[1] the revision of one belief may in turn bring about the revision of other beliefs which are logically related. There are some *important* conceptual beliefs whose revision would cause major disruption to a subject's web of conceptual beliefs, such as the belief that nothing can be in two places at once. There are others, less important, whose revision would cause only minor disruption, such as the belief that the pub is always busy on Fridays. And there are, of course, beliefs of every intermediate shade of doxastic importance. The folk psychologist tends to project her own important conceptual beliefs, her own *conceptual framework*, onto her subject.

Characterisations like that above barely scratch the surface of our folk-psychological understanding. They are necessarily imprecise — the customs and practices in question do not admit of precise characterisation, they manifest themselves differently in different individuals, they vary from culture to culture, and are subject to evolution. These customs and practices are embedded in a linguistic culture which encourages the habit of philosophical enquiry. In particular, it allows the Socratic question: "What is belief", and it admits discourse on the nature of belief even though this discourse consistently fails to produce satisfactory answers and generates the illusion of a puzzle. The field of AI hopes to foster a very different sort of understanding.

2. Artificial Intelligence

The term "artificial intelligence" is applied to many kinds of research, ranging from the study of search algorithms, through the construction of theorem provers and the design of certain programming languages, to the study of computational models of cognition. Whilst much of this research produces tools with immediate application outside the sphere of AI itself, there is a clearly discernible vision motivating AI research, and each of these tools is a prospective contribution towards its realisation.

[1] W.V.O.Quine, Two Dogmas of Empiricism, in From a Logical Point of View.

> There is no reason to think that artificial intelligence, unlike other disciplines, has a unique goal. It is inspired, however, by a unique vision — of fully autonomous, flexibly intelligent, rational (though artificial) agents.[1]

But it is not enough merely to be able to build such machines. What is sought is a thorough understanding of everyday problem solving. What does it mean to have such an understanding? In some sense, a spider understands webs. This understanding is manifest in an ability to spin them and repair them in a variety of differently shaped niches. But a spider has a very meagre understanding of tension and stress and structures. It could not apply its understanding to the construction of bridges, nor could it communicate its understanding to other spiders. There are different degrees and different kinds of understanding.

Central to the realisation of the AI vision is a formal study of methods for problem solving in the everyday world, because a great deal of intelligent behaviour just is problem solving in the everyday world, and the development of a proper understanding of everyday problem solving demands a rigorous, mathematically founded investigation of its underlying principles. So, both folk psychology and AI are concerned, amongst other things, with everyday problem solving, but their approaches are very different in style. By restricting her domain of enquiry to everyday problem solving, the AI scientist avoids serious philosophical issues, like subjectivity and privacy and first-personal perspective, but through the rigour and precision of her language helps to dispel illusory ones, like the nature of belief.

Let us be quite clear about the scope of this enquiry. Quine remarks that,

> Different persons growing up in the same language are like different bushes trimmed and trained to take the shape of identical elephants. The anatomical details of twigs and branches will fulfill the elephantine form differently from bush to bush, but the overall outward results are alike.[2]

A certain sort of enquiry would be interested in the shape of the bush, another sort might be interested in particular anatomical structures which realise this shape. Note that a description of the shape of the bush captures the space of possible anatomical structures which could realise that shape. Everyday problem solvers are also like appropriately trimmed and trained bushes. The kind of AI research which is the subject of this essay is

---

[1] D.Israel, A Short Companion to the Naïve Physics Maifesto, in Formal Theories of the Commonsense World, ed J.Hobbs and R.C.Moore, Ablex (1985), p 427.

[2] W.V.O.Quine, Word and Object, MIT Press (1960), p 8.

interested in the shape of the bushes — the nature of the activity not the mechanisms underlying that activity.[1]

It may turn out that in order to usefully describe this activity, we require certain constructions of language — such as the propositional attitudes of folk psychology. But in using this sort of language we are not saying anything about mechanism. The domain of folk psychology is not restricted to humans. We tend also to use it to explain and predict the activity of certain animals and machines, and would probably use it for Martians if we ever happened to meet any. Even the actions of a wooden golem which worked by magic would be within the domain of folk psychology. As Dennett points out, the behaviour of a chess computer could be explained in terms of the algorithms it employs or even in terms of its physical construction. But it is easier to adopt the "intentional stance" and employ the language of folk psychology, invoking the machine's beliefs and desires (such as a desire to "get its queen out early").[2] If somebody *forced* us to employ attitude talk only with respect to humans, then it would be necessary to invent new linguistic constructions which performed the same function but which had wider scope.

Furthermore, the products of AI research are insensitive to the *particular* world (or simulated world) to which they happen to be connected. It is a matter of indifference to the AI scientist whether her system is attached to a simulated environment (the programmer taking the rôle of a Cartesian demon), to our Earth or to Twin Earth. The AI scientist, then, is a kind of "methodological solipsist".[3] Suppose the AI scientist is asked to consider the beliefs of a system connected to our Earth and those of a system connected to a simulated environment. She might say of both that they believe the sink is full of water. If we then pointed out to her that the beliefs are, in a sense, different, she might say: "Of course they are different. The system connected to the simulated environment doesn't have a belief about a real sink, whilst the other system does. However, this doesn't affect my research programme".[4]

---

[1] This kind of AI is sometimes identified with the "McCarthy school", as opposed to the "Minsky school". See D.Israel, A Short Companion to the Naïve Physics Maifesto, in Formal Theories of the Commonsense World, ed J.Hobbs and R.C.Moore, Ablex (1985), p 427.

[2] D.Dennett, Intentional Systems, in Brainstorms, Harvester Press (1981), p 6.

[3] See H.Putnam, The Meaning of Meaning, in Essays on Mind, Language and Reality, Cambridge University Press (1975), p 215.

[4] The so-called wide/narrow debate is explored in P.Pettit and J.McDowell (eds), Subject, Thought and Context, Oxford University Press (1986).

3. Naïve Physics

One approach to the realisation of the AI vision involves an attempt at a deep analysis of basic folk-theoretical concepts in the hope that this will yield a sufficiently formal theory — a theory which will illuminate the concept's rôle in the production of behaviour. The analysis of folk-theoretical concepts involves making public a number of sentences about those concepts, instituting a convention which will more firmly fix their meaning. But every sentence made public employs more folk-theoretical terms, which may require further conceptual analysis. According to the *logicist* thesis, this process will converge on a body of sentences whose interpretation is universally agreed, and which are written in a formal language. This body of sentences will capture the commonsense knowledge which is brought to bear in everyday problem solving. I shall not rehearse the arguments for the logicist position, which can be consulted elsewhere.[1] My concern here is with the philosophical status of prospective products of the logicist research programme.

Someone whose everyday world is, say, present day London, has a grasp of a great many culturally specific concepts — things like buses, tube trains, shops and restaurants. Even the shallowest conceptual analysis of such things soon exposes a collection of underlying naïve physical and metaphysical concepts — of objects, arrangements of objects, the stuff that objects can be made from, the ways objects behave, when they are lifted, pushed, dropped, hit or simply left alone — of spatial and temporal location, of up and down, far away and near, in front and behind, before and after. Naïve physical and metaphysical concepts are the components out of which complex conceptual frameworks are built, and their analysis is the first step towards an understanding of such frameworks. The project of performing such an analysis deeply enough to yield a formal theory is described by Hayes.[2]

> I propose the construction of a formalization of a sizable portion of common-sense knowledge about the everyday physical world: about objects, shape, space, movement, substances (solids and liquids), time, etc.[3]

---

[1] P.J.Hayes, In Defence of Logic, Proceedings IJCAI 77, p559. R.C.Moore, The Rôle of Logic in Knowledge Representation and Commonsense Reasoning, Proceedings AAAI 82, p 428. See also D.McDermott, A Critique of Pure Reason, Computational Intelligence, vol 3, no 3, p 151., and the many commentaries in the same volume.

[2] P.J.Hayes, The Second Naïve Physics Manifesto, in Formal Theories of the Commonsense World, ed J.Hobbs and R.C.Moore, Ablex (1985), p 1.

[3] Ibid, p 2.

Hayes's motivation for this project is partly to get away from the "toy domains" which have been the traditional concern of AI research, such as the Blocks World, and to provide a richer domain for the study of problem solving. He suggests that the project should not initially be concerned with the inference mechanisms that will be used on the resulting formalisation, and he recommends the first-order predicate calculus as its "reference language".[1] The resulting theory is expected to be very large, comprising perhaps a hundred thousand axioms, and it seems unlikely that it will contain any isolated sub-theories. It may, however, be structured into *clusters*. A cluster is a densely connected (though not isolated) collection of axioms, which fix the meanings of a number of closely related concepts. For instance, there seems to be a family of concepts associated with places and positions, whose analysis will yield a cluster.

> Consider the following collection of words: inside, outside, door, portal, window, gate, way in, way out, wall, boundary, container, obstacle, barrier, way past, way through, at, in.
>
> I think these words hint at a cluster of related concepts which are of fundamental importance to naïve physics. This cluster concerns the dividing up of three-dimensional space in pieces which have physical boundaries, and the ways in which these pieces of space can be connected to eachother, and how objects, people, events and liquids can get from one such place to another.[2]

Besides places and positions, Hayes discusses our everyday concepts of spaces and objects, qualities, quantities and measurement, change, time and histories, energy, effort and motion. Writing about the composition of objects, Hayes says,

> As far as I can judge, all naïve-physical objects are either a single piece of homogenous stuff, or are made up as a composite out of parts which are themselves objects. The essence of a composite is that its component parts *are* themselves objects, and that it can (conceptually if not in practice) be taken apart and reassembled, being then the same object. Examples of composites include a car, a cup of coffee, a house, four bricks making a platform. Examples of homogeous objects are a bronze statue, a plank of wood, the Mississippi, a brick. Homogenous objects have no parts, and can only be taken apart by being broken or divided in some way, resulting in *pieces*.[3]

---

[1] By "reference language", Hayes means a single language into which more exotic representation languages can be translated. He does not object to the use of such exotic languages if it is convenient.

[2] Ibid, p 19.

[3] Ibid, p 27.

Hayes goes on to attempt a formalisation of our everyday concept of liquids — the containment of liquid, its behaviour and the individuation of liquid objects.[1] The seventy-four axioms he provides are powerful enough to permit the prediction of the behaviour of liquids in various circumstances. For instance, they can be used to predict the behaviour of a glass of milk as it is poured onto a flat table, spreads out to the sides and spills over the edges. Work has also been done on formalising other everyday concepts, such as shape[2] and substance.[3] Alongside naïve physics, which is a theory of everyday middle-sized objects, the logicist requires other theories, of naïve topography[4] and naïve psychology, for example.[5]

As Hayes comments,[6] (human) naïve physics is pre-Galilean.[7] But the falsity of a theory in naïve physics, human or machine, is no cause for concern. So long as the cases which would falsify the theory do not arise given the precision of the naïve physicists measurements, then the theory continues to be useful. More precisely, a naïve theory can be said to be adequate with respect to a given *granularity*.[8] For certain purposes, a coarse grain of representation is adequate — at a coarse grain, the human body can be represented simply as a cylinder, and this would suffice for tackling the problem of moving about in a crowd (so long as the crowd was not too dense). For some purposes, a finer grain is required — the human body could be represented as a collection of variously sized, connected cylinders (arms, legs, torso and so on), and this would be suitable for problems involving more intimate forms of interaction. Similarly, if a naïve theory displays ontological promiscuity or betrays contentious metaphysical presuppositions, there is no problem so long as the theory serves its purpose.

Whilst a *weak* logicist believes that logic can be used to describe the knowledge required for everyday problem solving, a *strong* logicist believes that logic can also be used to *represent* it in a computer. The fulfilment of the strong logicist research programme demands the development of a whole battery of techniques for the

---

[1] P.J.Hayes, Naïve Physics 1: Ontology for Liquids, in Formal Theories of the Commonsense World, ed J.Hobbs and R.C.Moore, Ablex (1985), p 71.

[2] Y.Shoham, Naïve Kinematics: Two Aspects of Shape, in Commonsense Summer: Final Report, ed J.Hobbs, SRI International, AI Center, 1984.

[3] G.Hager, Naïve Physics of Materials: A Recon Mission, Ibid.

[4] See E.Davies, Representing and Acquiring Geographical Knowledge, Pitman.

[5] Although they have similar domains, it is important to distinguish the AI scientist's formal theory of naïve psychology from the unwritten “theory” of the folk psychologist.

[6] P.J.Hayes, The Naïve Physics Manifesto, in Expert Systems in the Microelectronic Age, ed D.Michie, Edinburgh University Press (1979).

[7] I emphasise *human* naïve physics here because the AI scientist is not necessarily concerned to model human mistakes in reasoning about everyday objects.

[8] See J.R.Hobbs, Granularity, Proceedings AAAI 85, p 432.

construction and use of naïve theories, corresponding to the various capacities familiar to the folk psychologist: mechanisms for theory formation, mechanisms for default reasoning, reason maintenance systems and planners.[1] But whether she has taken the strong or weak logicist approach, the AI scientist can explain her creation's behaviour as if it were the product of an ever-changing set of logical formulae, expressed in naïve-theoretical terms, which mediates between perception and action.

The AI scientist, who is familiar with the construction of naïve theories, is adept at translation between the formal language in which a naïve theory is expressed and the language of folk psychology. Translation is possible because of the close correspondence between the sentences of folk psychology and predicate calculus formulae. Knowing the rôle of a given set of such formulae in the production of behaviour, she can ascribe folk-psychological attitudes to her creations. Conversely, she can generate a set of formulae which correspond (roughly, since folk-psychological language is imprecise) with any given folk-psychological description of a set of attitudes. As well as facilitating communication between AI scientists, the capacity to translate between formal and folk-psychological language affords relief to the sense of puzzlement about the nature of belief, and thus serves an important philosophical purpose.

Concluding Remarks

In sum then, the AI scientist is interested in describing a particular causal surface — the interface between an inner and an outer to whose structure she is indifferent, though, in a sense, the causal surface in question defines the space of possible such structures. The causal surface in question is picked out by the prevailing folk psychological customs and practices. What the AI scientist seeks is an improved set of customs and practices — one which displays a deeper understanding, manifest first in a language which admits of less ambiguity and leaves fewer unanswered (unanswerable) questions, and second in the construction of machines which exhibit a capacity for everyday problem solving. Rather than displacing the old folk-psychological customs and practices, the new language supplements them, and neutralises some of their apparent puzzles via their translation into a purer idiom.

[1] For a detailed inventory of the logicist's toolbox see N.Nilsson and M.Genesereth, Logical Foundations of Artificial Intelligence, Morgan Kaufmann (1987).