ROBERT STALNAKER

THE PROBLEM OF LOGICAL OMNISCIENCE, I

From their beginning,[1] epistemic and doxastic logics – the logics of knowledge and belief have been modeled on modal logic – the logic of necessity and possibility. Knowledge and belief, in such logics, are analogous to necessity. There is a wide variety of modal logics, but all of the normal ones contain certain distribution or deductive closure principles; for example, if '$\Phi \rightarrow \Psi$' is valid, then so is '$\Box\Phi \rightarrow \Box\Psi$'. Most versions of epistemic logic are normal in this sense, accepting analogous principles for knowledge and belief. Developers of such logics invariably remark that the principles of deductive closure are unrealistic, since it is obviously false that knowers in general know all of the deductive consequences of anything that they know. The assumption that knowers do, as a matter of logic, have such knowledge – that they are deductively omniscient – is defended as an idealization. Sometimes the divergence between the assumptions of the ideal theory and the facts about the domain of its intended application is described as a problem for epistemic logic – the problem of logical omniscience. My aim in this paper is to try to get clear about just what kind of idealization such normal epistemic and doxastic logics are making, and what the motivation is for idealizing in this way. If there is a problem of logical omniscience, I want to try to see if I can say what the problem is. My broader aim is to try to get clearer about the concepts of knowledge and belief, and about what work we should expect a logic of these concepts to do.

I shall begin by contrasting two different ways that the divergence between idealization and reality might be explained, and considering several different kinds of reasons that one might have for idealizing in one or the other of the two ways. Then I shall look at the problem of logical omniscience from the perspective of a certain conception of the nature of belief, the sentence storage model. I have little sympathy with this influential conception of belief; I think the problem of logical omniscience helps to bring out its limitations, and to point the way to a more adequate conception. But the problem of logical omniscience is not solved by giving up the sentence storage model. I shall suggest,

Synthese **89**: 425–440, 1991.
© 1991 *Kluwer Academic Publishers. Printed in the Netherlands.*

in conclusion, that it is a symptom of a tension in our ordinary conceptions of knowledge and belief.[2]

1. IDEALIZATION

There are two ways that one might try to reconcile the fact that people do not believe all the logical consequences of their beliefs with a theory that seems to say that they do. On the one hand, one might interpret one's logic to be a logic of belief in the ordinary sense, but restrict its domain of literal application to imaginary believers of a special idealized kind – perhaps to agents who have unlimited memory capacity and infinite computational power and speed. Ordinary people, and even extraordinary real people, can't think of everything, but ideal believers can, and if there were such believers they would believe, in the ordinary sense of 'believe', all the logical consequences of everything that they believe. On the other hand, one might take the domain of literal application of one's logic of belief to be unrestricted, including ordinary agents who have no special computational powers, but interpret the concept of belief that the theory models to be belief in a special sense. The divergence between ideal and real is explained as a difference between belief in the ordinary sense and belief in some special technical sense. For example, one might distinguish belief in the ordinary sense from *implicit* belief: one's implicit beliefs include, by definition, all of the deductive consequences of one's beliefs, whether or not they are or could be recognized as such by the agent. To be logically omniscient with respect to implicit belief is no great feat; not even the most ignorant and unreflective of us can avoid being logically omniscient in this sense.

Both of these stories begin by conceding that ordinary agents do not believe, in the ordinary sense of 'believe', all of the logical consequences of their beliefs. Why, one might ask, should one idealize in either of these ways? Why shouldn't one's logic of belief be a logic of belief in the ordinary sense and at the same time a logic that applies to ordinary believers – to anyone who has beliefs in the ordinary sense? In such a realistic logic a proposition of the form *x believes that P* should entail a proposition *x believes that Q* only if it is impossible, as a matter of logic, for anyone to be correctly described as believing that *P* unless that person also believes that *Q*. Is the problem that a logic of real

belief is too hard to find, and so we have to settle for the simpler logic of a simplified concept, or a simpler domain of application? Or, is there a more positive reason for idealizing: perhaps the concept of real, explicit belief is a concept that picks out uninteresting surface phenomena, while belief in an idealized sense, or the belief we *would* have if we were free of certain limitations we all have, is deeper or more basic or more interesting in some way. I shall look at four different motivations for populating one's theories with idealizations which might provide reasons for idealizing in one or the other of the two ways I have contrasted.

First, one may idealize to get at underlying mechanisms. The complicated behavior of some system may be explained by the interaction of a number of different components, components that can be best understood by seeing how they would work in isolation even if they are, in a realistic context, never found in isolation. A theory may focus on one component, seeing the action of the others as external, interfering factors. In the ideal system these external factors are not there at all. The frictionless planes and weightless pulleys of elementary physics problems are familiar examples of idealization justified in this way. Another is Chomsky's use of the competence-performance distinction to isolate a psychological capacity that is specific to language. Performance is the surface phenomena: what speakers do and don't say, and what expressions they do and don't find odd. But, it is hypothesized, the surface phenomena have different kinds of explanations. Some things are not said, or seem unintelligible to speakers, because they are ungrammatical; others are just too complicated, just too hard to process or remember. One abstracts away from memory and processing limitations by describing the performance of an ideal speaker-listener who has our grammatical competence, but no such limitations. One idealizes in this way in order to better explain the performance of ordinary speakers.

Some theories that idealize for this reason hypothesize that the system the theory is about tends toward some equilibrium state. Various external forces may divert it from its natural state, but when they do its internal dynamics tend to move it back toward the equilibrium. One part of the articulation of such a theory is the description of the equilibrium state – the state that the system *would* be in if it were free from external forces. Even if real systems of the kind one is studying never

reach equilibrium, the description of the ideal, equilibrium state may help to explain their behavior. Economic theories are familiar examples of theories that idealize in this way.

The assumption of deductive omniscience is sometimes conceived, at least implicitly, as an idealization motivated in this way. Failures to know or believe all the consequences of one's knowledge or beliefs are to be accounted for by a kind of cognitive friction impeding a natural process of drawing consequences. It is natural to think of belief sets that are inconsistent or not deductively closed as unstable, tending toward an equilibrium at which they satisfy conditions of perfect rationality, an equilibrium that is never reached because our belief state is constantly perturbed by the receiving of new information from outside. This is an attractive picture, but I shall argue below that it rests on an implausible conception of what belief and knowledge are.

A second reason to idealize is to simplify. Some features of a system that greatly complicate its accurate description and explanation may nevertheless, for some purposes and in some contexts, be negligible. The cost of the distortion that comes from ignoring such features may be less than the benefit of simplification. For example, despite the fact that we know that these things are strictly false, we may, in some contexts, be justified in assuming that mass is concentrated at a point, that the gravitational force between two bodies remains constant as they approach each other, that air offers no resistance, that light always travels in straight lines. It has been suggested that the assumption of deductive omniscience implicit in normal epistemic logics is a simplification of this kind. Robert Moore, for example, writes that logics of knowledge that imply deductive omniscience

> represent idealizations that are reasonable approximations to the truth for many purposes. While no rational agent's knowledge is closed under logical consequence, outside of mathematics there seem to be few cases where this significantly affects an agent's behavior.[3]

But I think this underestimates the extent of the distortion. It is not only mathematicians who need to worry about their failure to know all the consequences of their knowledge. Any context where an agent engages in reasoning is a context that is distorted by the assumption of deductive omniscience, since reasoning (at least deductive reasoning) is an activity that deductively omniscient agents have no use for. Deliberation, to the extent that it is thought of as a rational process of

figuring out what one should do given one's priorities and expectations is an activity that is unnecessary for the deductively omniscient. In fact any kind of information processing or computation is unintelligible as an activity of a deductively omniscient agent. It is hard to see what a logic of knowledge could be for if it were a harmless simplification for it to ignore these activities that are so essential to rationality and cognition.

A third kind of justification for idealization is normative: whatever the inner dynamics of our states of knowledge and belief, and whatever the extent of our divergence from the ideal of deductive omniscience, isn't this ideal at least something that rational agents ought to strive to approximate? Isn't a divergence from deductive omniscience a defect in one's state of knowledge, even if a defect that is unavoidable? This suggestion seems plausible, but a number of philosophers have resisted it, suggesting that one may have good reasons to avoid accepting all the consequences of one's beliefs, even for having beliefs that are inconsistent with each other, and which are recognized to be so. It has been argued, for example, that rationality requires us to recognize our fallibility. Rational people should believe that at least some of their many beliefs are false; if they do, then they will disbelieve some conjunctions of propositions each one of which they believe. A different reason for rationally withholding belief – at least explicit belief – from some consequences of one's beliefs is given by Gilbert Harman: "Many trivial things are implied by one's view which it would be worse than pointless to add to what one believes". Harman proposes a principle of reasoning he calls *clutter avoidance*: "One should not clutter one's mind with trivialities".[4] I don't find either of these reasons for deliberately refraining from believing the consequences of one's beliefs persuasive since I think they presuppose implausible accounts of what belief is,[5] but even if deductive omniscience is a normative ideal, it is not clear that that is a reason to build it into a logic of knowledge.

A fourth motivation for idealizing is more pessimistic. Perhaps the best we can do is to get a logic of the knowledge of an idealized knower, or of knowledge in some special idealized sense. Perhaps we know how to give a clear account of a concept of knowledge from which it follows that the knowers to which it applies are logically omniscient, but that there are insurmountable problems with any account of knowledge we know how to give that lacks this consequence. I shall argue that this fourth kind of motivation comes closest to the reason why we make

this idealization about knowledge. Once we see why this is so, we can be clearer about just what the problem of deductive omniscience is, and how we might solve it.

I have suggested that the nature of the idealization, and the motivation for it, depend on what one takes knowledge and belief to be. I want to look now at a popular picture or model of what belief is, and at the problem of logical omniscience from the perspective of this model. I call it a model rather than a theory because it is not always clear how literally its proponents want it to be taken, but whether meant literally or as a metaphor, it has had a profound influence on the way philosophers and cognitive scientists think about belief, and the logic of belief.

2. THE SENTENCE STORAGE MODEL OF BELIEF

The *sentence storage model* is perhaps common enough to be called the received view.[6] According to this model, one's beliefs[7] are determined by a set of sentences, perhaps of a mental language, perhaps of one's natural language, that one stores in memory. To a first approximation, the idea is that to believe that P is to have a sentence that says that P stored (to use the fashionable idiom) in one's belief box. This is only a first approximation, since no one thinks that *everything* one believes, in the ordinary sense, is explicitly stored. The sentence storage model distinguishes explicit from implicit belief. The explicit beliefs are those in the set of sentences stored in the belief box; other things one believes, for example certain obvious consequences of the sentences one stores, are believed only implicitly. Different proponents of the sentence storage model have different accounts of what implicit belief is, but what is essential to the model is that implicit belief, and belief in general, is determined by the explicitly stored beliefs.

If this is how belief is to be explained, then what should the logic of belief be? First, what kind of logic should belief have if it is given a realistic semantics – one that is intended to apply to the actual beliefs of ordinary believers? Second, what kind of idealization of this model of belief would yield the standard logic of belief, according to which believers are logically omniscient, and what might motivate the idealization? Third, is there a *problem* of logical omniscience on this account? If so, what is it?

One might ask these questions about either explicit or implicit belief.

I shall start with explicit belief, with the logic of the belief set: the set of sentences that is stored in the agent's belief box. One might think that a realistic logic of belief would impose no conditions at all on the contents of belief boxes. How can it be a matter of logic that if some sentence is stored in a belief box, then so is some distinct sentence? And so, one might expect that a realistic logic of explicit belief would have no logical principles at all that depend on the internal structure of belief attributions. But there are two possible sources of principles for a logic of explicit belief, sources that are distinct and important to distinguish. First, even if there are no constraints whatsoever on the sets of mental sentences that may constitute a person's explicit beliefs, there may be some logical relations between different sentences attributing beliefs for the following reason: while it is *sentences* that are the explicit beliefs, according to the sentence storage model, it is what the sentences say, rather than the sentences themselves, that belief attributions refer to. In a sentence of the form "x believes that P", or its formalization in the language of doxastic logic, 'B_XP', the sentence that goes in for 'P' is *used*, and not mentioned. A semantics for the language should tell us what 'B_XP' says – what semantic value it has – in terms of the semantic value of what goes in for 'P'. The semantic value of a sentence, it seems reasonable to assume, is what the sentence says; sentences with the same value are those that say the same thing. Even if to believe (explicitly) is to store a sentence, the sentence stored is not identified in a belief attribution. When we say that x explicitly believes that P, we say (on the storage model) that x stores *some* sentence that says that P. So a sentence of the form "x believes that P" makes an existential claim about x's beliefs: that there exists a sentence in x's belief box that says that P. Now suppose there are distinct sentences of the language of belief attribution (not the believer's language, but the language in which his beliefs are being described) that, as a matter of logic, say the same thing. Suppose, for example, that a sentence of the form '$P \,\&\, Q$' says the same thing as the corresponding sentence, '$Q \,\&\, P$'. Then, '$B_X(P \,\&\, Q)$' will be logically equivalent, on the storage model, to '$B_X(Q \,\&\, P)$'. Any set of sentences at all, and so any belief set, will contain a sentence that says that '$P \,\&\, Q$' if and only if it contains a sentence that says that '$Q \,\&\, P$', on the assumption that any sentence that says the one also says the other.

Just which principles of the logic of belief are validated by this kind of consideration will depend entirely on what is assumed about the

contents of sentences – about the nature of what it is that sentences say. If we individuate contents very finely, then there will be fewer such equivalence principles, while, if we choose a coarse-grained conception, there will be more. Suppose we followed the simplest course, individuating contents by their truth conditions. Then, our semantics for explicit belief will validate the equivalence principle:

$$\text{if } \vdash \Phi \leftrightarrow \Psi \text{ then } \vdash B_x\Phi \leftrightarrow B_x\Psi.$$

This, by itself, is not quite deductive omniscience, but it is a very strong principle, one that would not be plausible for a realistic semantics of belief. It implies, for example, that anyone who believes any necessary truth – for example any trivial tautology – therefore believes all necessary truths. And if Ψ is a consequence of something x believes, say Φ, then while x may not believe Ψ, it will follow from the equivalence principle that x believes the conjunction of Φ and Ψ. Nothing, however, is implied or assumed about the powers of believers to recognize logical equivalences. The choice of a coarse-grained conception of content is a decision about how to describe the sentences that one might find in a belief box, not an assumption about what sets of sentences might be found in one. The decision is, thus, not a decision to idealize but, rather, to describe belief sets only in a very abstract and imprecise way. The decision does bring a problem, since there is a sharp divergence between the way beliefs are ordinarily described and the way the language of doxastic logic, interpreted this way, would describe them. But, assuming the sentence storage model were correct, this would not be a very deep or serious problem. All we need to do to avoid it is to choose a more fine-grained conception of content, one that categorizes sentences in a more revealing way. We don't want contents to be *too* fine-grained, since we want belief attributions to be capable of bringing out similarities between different sentences stored in different belief boxes – sentences that play the same roles in the respective cognitive economies of the agents storing them. But they should be individuated finely enough so that a belief attribution will reveal what is important about the stored sentence that makes the belief attribution true. This decision will to a large extent determine the logic of belief, but it will not have anything directly to say about the nature of the sets of sentences that, according to the sentence storage model, define our beliefs.

The second possible source of principles of a logic of explicit belief does concern the relations between the sentences that may be found in

the belief box. Some proponents of the sentence storage model argue that a belief set must meet certain minimal standards of logical coherence in order to count as a set of beliefs at all. Christopher Cherniak, for example, notes that, "[a] collection of mynah bird utterances or snippets from the *New York Times* are chaos, and so (at most) just a sentence set, not a belief set".[8] The suggestion is that if an alleged belief box were filled with a random collection of sentences, it would not count as a belief box, and the owner of the box would not count as an agent. So, perhaps, a sentence storage theory of belief – even nonidealized belief – should impose some conditions on the relations between the sentences in the set of beliefs that is stored and, perhaps, those conditions will have consequences for the logic of explicit belief.

It is not clear, however, what constraints would be plausible. A conception of minimal rationality could at best justify some consistency requirements – perhaps a requirement that belief boxes not contain sentences that are blatantly contradictory – but it would not justify any closure or inference conditions on explicit belief. Minimal rationality may require that one *believe* certain obvious consequences of the sentences one stores in the belief box, but it cannot justify a requirement that one *store* sentences that are obvious consequences of sentences one stores. The obvious consequences of sentences one stores are just the kind of thing that one does not need to store, since they can always be inferred when needed. Cherniak points out that it may sometimes be unreasonable to make sound inferences from one's beliefs because one's limited time and cognitive resources are better spent on other things, and Harman's principle of clutter avoidance enjoins one not to clutter one's belief box with trivial consequences. It is hard to see why even an idealized conception of explicit belief should have any deductive closure conditions. Would an agent with unlimited memory and computational power and speed have any reason to store the obvious consequences of its beliefs? Such an ideal believer might have less need to avoid clutter in its belief box, but it would also have less positive reason to add any deductive consequences to its store of explicitly represented sentences, since consequences can always be inferred when needed, cost free. Deductive closure, as a constraint on the sentence storage model's notion of *explicit* belief, is plausible neither as a normative ideal nor as an equilibrium state toward which belief sets tend.

A logic of explicit belief motivated by the storage model would reveal nothing about the inferential powers, either real or ideal, of believers,

since it would be a logic of the base from which believers infer, not of the conclusions they draw from this base. Any nontrivial principles of such a logic would derive from decisions about how to describe the sentences stored, and would say nothing about the relations between the sentences themselves. The storage model would provide no motivation to idealize the concept of explicit belief and, so far as I can see, it would give rise to no problem of logical omniscience. Nothing could be easier than to design a realistic logic of explicit belief appropriate to the storage model, but such a logic would have little interest. It would not be a logic of belief in the ordinary sense since, for both real believers and logically perfect ones, most of what is believed in the ordinary sense is believed only implicitly. So let us look at the storage model's concept of implicit belief.

There are at least two different notions of implicit belief, a broad notion and a narrow notion. On the broad notion, the implicit beliefs of a believer include everything the believer is committed to in virtue of having the explicit beliefs he has – all the information that is implicit in what is explicitly stored. This will include all the deductive consequences of the explicit beliefs and, perhaps, also beliefs about what the believer does and does not (implicitly) believe. On the broad notion, implicit beliefs are by definition deductively closed, for ordinary believers as well as for those with extraordinary computational powers. The claim that something is implicitly believed says nothing about whether the believer has access to that belief – whether the believer will assert or assent to it, or act as if he thinks it is true. No one thinks that implicit belief in this broad sense is an analysis of belief in the ordinary sense; all that is claimed for it is that it is a notion of some interest. The logic of implicit belief, in this sense, is simple and unproblematic: it is a normal modal logic, with an assumption of logical omniscience. But implicit belief in this sense tells us no more than explicit belief about the inferential powers of the believer.

Something is an implicit belief in the narrower sense only if it "is *easily* inferable from one's explicit beliefs".[9] Easy inferences will include both inductive and deductive consequences of explicit beliefs that obviously follow from them, and also things that can be easily inferred from the fact that something is explicitly believed. The task of clarifying implicit belief in this sense is the task of characterizing the easy inferences.

Suppose that the storage model were correct – that all of our beliefs

were derivative from a set of sentences in some mental language stored in the belief box – and suppose also that we had a clear explanation of the notion of an easy inference. Would implicit belief in the narrow sense then capture the ordinary notion of belief? Not necessarily. The problem is that there is no reason to assume that even explicit beliefs will be accessible to consciousness. Harman is clear about this: "A belief", he says, "can be explicitly represented in one's mind, written down in Mentalese as it were, without necessarily being available to consciousness".[10] Harman's example of inaccessibility is a Freudian one where the inaccessibility is explained by repression. But explicit beliefs, just like beliefs that are only implicit in the broad sense, may be inaccessible because of the computational limitations of the believer. Search is a kind of computation, and if the belief box is large and full enough, the search may be a long and hard computation. But if an explicit belief is inaccessible because of the computational limitations of the believer, then the fact that something is an easy inference from it will not render it a belief, in the ordinary sense. If a belief is present to consciousness – if the agent is currently judging that it is true – then it will be reasonable to infer that the agent also believes whatever may be easily inferred by that agent from that belief. But being present in a conscious act of judgment is neither necessary nor sufficient for explicitly stored belief. Not only may explicit beliefs be tacit and inaccessible, occurrent judgments – beliefs that are neither tacit nor inaccessible – may also fail to be explicit in the storage model's sense. A belief I am actively entertaining may be something that I inferred from the contents of my belief box, and that I do not, perhaps for reasons of clutter avoidance, add to what is explicitly stored there. The problem is that the distinction between implicit and explicit belief is being used to do two different jobs that one distinction is not suited to do. The manifest fact that we are not logically omniscient is a fact about our computational limitations – the fact that some of the information that is implicit in what we know or believe is, because of computational limitations, not accessible to us. To get at belief and knowledge in the ordinary sense we need a distinction between what is accessible and what is implicit but inaccessible. The explicit-implicit distinction is sometimes tacitly assumed to be this distinction. But, if it is, then it is a completely different distinction from the one that the storage model makes between two different forms in which information is represented, the distinction between propositions expressed by sentences written

down in the belief box and propositions not written down there but somehow implicit in the ones that are. Now I don't see the slightest reason to take seriously this belief box myth as anything more than a highly misleading metaphor; but, even if it were literally correct as a theoretical account of the mechanism by which human beings store the information they use to guide their behavior, it would still not give us the resources to explain the distinction between accessible and inaccessible stored information, and this is the distinction we need in order to give a realistic account of belief and knowledge in the ordinary sense.

Ordinary knowledge is a *capacity*, and ordinary belief a *disposition*. Because of our computational limitations, we may have the capacity constituted by the knowledge that P, or the disposition constituted by the belief that P, while at the same time lacking the capacity or disposition that we would have if we knew or believed some deductive consequence of P. But what is the capacity or disposition as capacity or disposition to do? The storage model has nothing to say about this, and so has little promise of clarifying the problem of logical omniscience. Let me sketch a slightly different picture, one that brings in the uses to which knowledge and belief are put. This will be a very simplified and idealized model that considers only one use to which knowledge and belief are put, but it is a model that helps to bring out some of the obstacles that make it difficult to give a realistic account of knowledge and belief.

3. THE QUESTION-ANSWER MACHINE

Suppose that an agent is a question and answer machine. Its belief and knowledge are to be understood as capacities and dispositions to answer questions. Such a machine will need some mechanism or mechanisms for storing information, and some way of using that information to generate answers to the questions it receives. How it stores the information is not directly relevant to what it knows or believes. It might, for example, have lists of question-answer pairs, pictures or maps and procedural rules for reading them, vast look-up tables, or a small number of axioms and some powerful deductive rules. Or, information might be implicit in the command structure of its programs. Presumably, a necessary condition for the machine to know or believe that P will be that the information (or misinformation) that P be implicit in what is stored or represented in whatever way the machine stores

information. But more will be required: the information must be available – the agent must be able to access the information in order to give answers that express the proposition that *P* when such an answer is appropriate. Propositions that meet the first necessary condition will be implicitly known or believed in a sense that corresponds to the storage model's broad sense of implicit belief – a sense that unproblematically satisfies the logical omniscience condition. But the set of propositions that also meets an appropriate accessibility condition need not be closed under logical consequence.

But how will the accessibility condition be spelled out? There are at least three problems. First, accessibility is clearly a matter of degree. There are questions I can answer quickly with a moment's thought or a minor calculation, and questions that I have the computational resources to answer eventually, but only after a lot of time and effort. For some questions of the latter kind, I may be able to say outright that I have the capacity to produce the answer eventually; for others, I may in fact be able to produce an answer, if I choose the right computational strategy, but may be unable to say whether I can until I actually produce the answer. How easy must the search or computation be in order for the answer to count as something the agent already knows or believes, and not just something it has the capacity to come to know or believe? I assume you know your multiplication tables – you know, for example, that 6 times 4 equals 24. And no doubt you also know certain simple arithmetic truths that it is not plausible to assume are memorized, for example, that 47 times 100 equals 4,700, and that 385 is not a prime number. It is equally clear that you do not know the prime factors of 75,563, even though you know that you could figure it out and might even be able to put a limit on how long it would take you. But there are intermediate cases that are not so clear. There is obviously a continuum here, and no very natural place to draw a line between information that is easily accessible and information that is not. I don't think this is a serious problem. Attribution of knowledge and belief are obviously highly context-dependent, and the line between what we already know and what we could come to know if we made the effort may be one thing determined somewhat arbitrarily in different ways in different situations. A more serious problem is this: on this model, information is accessed in response to a question, and the ease of access will depend on the question. The same proposition may be an answer to different questions, and whatever

one's standards for easy access, the proposition may be easy to access in response to one question, but not to another. For example, it will take you much longer to answer the question, "What are the prime factors of 1591?", than it will the question, "Is it the case that 43 and 37 are the prime factors of 1591?" But the answers to the two questions have the same content, even on a very fine-grained notion of content. Suppose that we fix the threshold of accessibility so that the information that 43 and 37 are the prime factors of 1591 is accessible in response to the second question, but not accessible in response to the first. Do you know what the prime factors of 1591 are or not? The problem is not that the two different questions will affect your knowledge differently – that the second question, but not the first, will bring it about that you know (not just implicitly) what the prime factors of 1591 are. It seems to be a fact, not a problem, that questions – even nonleading questions without presuppositions – can change what we know and believe by bringing out what was previously merely implicit in what we believed. This is the familiar lesson of the Socratic method, made explicit in the *Meno*.[11] Our problem is that we are not just trying to say what an agent would know upon being asked certain questions; rather, we are trying to use the facts about an agent's question answering capacities in order to get at what the agent knows, even if the questions are not asked. But attributions of knowledge and belief are not tied to any particular questions that the knowledge or beliefs might be used to answer. More generally, the problem is that we need to understand knowledge and belief as capacities and dispositions – states that involve the capacity to access information, and not just its storage – in order to distinguish what we actually know and believe, in the ordinary sense, from what we know and believe only implicitly. We can do this only by bringing the uses to which knowledge and belief are put into the concepts of knowledge and belief themselves, but, on the face of it, it does not seem that when we attribute knowledge or belief to someone we are making any claims about what the agent plans to do with that information.

Finally, even if we had a satisfactory account of accessibility for the question and answer model, it would not be clear how to generalize it to an account of knowledge and belief in terms of capacities and dispositions to use information (or misinformation) to guide not just one's question answering behavior, but one's rational actions generally. For we want an account of knowledge and belief, not just for expert systems

and people who staff information booths, but for all kinds of agents. We want a notion that helps to explain why people do what they do, in their nonlinguistic as well as their linguistic behavior. Very roughly, I know whether *P* if I have the capacity to make my actions depend on whether *P*. But, I may have this capacity for some actions, and not for others. Consider, for example, a shrewd but inarticulate chess player who may be able to access information for the purpose of choosing a move even if she is unable to access that same information for the purpose of answering a question, or giving an explanation of why she moved as she did. In the general case, it is even clearer that the accessibility of knowledge and belief can be understood only relative to the actions they are being used to guide.

The problem of logical omniscience, I am suggesting, is the problem of accessibility. The reason we idealize in our logics of knowledge and belief is because we have a much clearer conception of implicit knowledge and belief – the information or informational content that we store – than we do of accessible knowledge and belief – the information and belief that is available to guide behavior. The storage model may yield a logic of explicit belief that avoids logical omniscience, but it does it by avoiding the real problem. We won't have a clear understanding of knowledge and belief, and of an important part of cognition, until we address this question.[12]

NOTES

[1] See Jaakko Hintikka: 1962, *Knowledge and Belief*, Cornell University Press, Ithaca, New York.

[2] Throughout my discussion, I shall talk sometimes about knowledge and sometimes about belief. Different conceptions of knowledge and belief will say different things about the relations between them, but I am assuming that the problem of understanding how and why we do not know all the consequences of our knowledge is essentially the same as the problem of how and why we do not believe all the logical consequences of our belief.

[3] Robert Moore: 1988, 'Is it Rational to be Logical?', in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, Morgan-Kaufman, Los Altos, California, p. 363.

[4] Gilbert Harman: 1986, *Change in View: Principles of Reasoning*, Bradford Books, MIT Press, Cambridge, Massachusetts, p. 12.

[5] The first kind of consideration is discussed in Stalnaker: 1984, *Inquiry*, Bradford Books, MIT Press, Cambridge, Massachusetts, Chap. 5. I shall discuss Harman's principle of clutter avoidance below.

[6] See, for example, Gilbert Harman: 1973, *Thought*, Princeton University Press, Princeton, New Jersey; Jerry Fodor: 1981, 'Propositional Attitudes', in Fodor, *RePresentations*, Bradford Books, MIT Press, Cambridge, Massachusetts. pp. 177–203; Christopher Cherniak: 1986, *Minimal Rationality*, Bradford Books, MIT Press, Cambridge, Massachusetts; Robert Moore and Gary Hendrix: 1982, 'Computational Models of Belief and the Semantics of Belief Sentences', in Stanley Peters and Esa Saarinen (eds.), *Processes, Belief, and Questions: Essays on Formal Semantics of Natural Language and Natural Language Processing*, D. Reidel, Dordrecht, pp. 107–27.

[7] In discussing this account, I shall focus on belief rather than knowledge. If one's beliefs are the sentences stored in one's belief box, then presumably one's knowledge will be explained in terms of a subset of the set of beliefs – beliefs meeting some additional, partly external conditions.

[8] Cherniak: 1986, p. 6.

[9] Harman: 1986, p. 13; my emphasis.

[10] Ibid., p. 14.

[11] Cf. Lawrence H. Powers: 1978, 'Knowledge by Deduction', *Philosophical Review* **87**, 337–71.

[12] A sequel to this paper, 'The Problem of Logical Omniscience, II', will discuss the problem of accessibility in the context of a model of knowledge developed by theoretical computer scientists to describe and explain the behavior of distributed systems.

Dept of Linguistics and Philosophy
MIT, Cambridge, MA 02139
U.S.A.