Mapping The Mind - A Matter Of Logic?

## 1. Knowledge - a mysterious matter

Perhaps the most impressive feature of man is his ability to accumulate knowledge, to utilise the knowing of his forefathers and contemporaries, and to pass on his own understanding to the next generation. Not that nothing is lost in the process, not that the working of any man's mind is exactly copied in another's, and yet the effect is strong enough to reshape the earth. Knowledge is the most private of possessions, and yet it is possible to represent it, in a material way, so that the representation evokes a very similar state in another man.

How is this possible? Scholars and scientists from many different fields have inquired into the many aspects of this question. Psychologists and neurophysiologists address the internal question about the ongoings in the individual mind, linguists and computer scientists the external one about the nature of the systems which can transmit and represent these ongoings. In philosophy, the branches of epistemology and philosophy of mind face the internal aspect, and those of logic and philosophy of language the external one. The heart of the matter is, of course, how the two aspects interrelate.

How shall one envisage knowledge? The internal and external aspects seem to pull in different directions. Introspectively, most people seem to feel that knowledge derived from personal experience somehow comes in larger units, representing situations rather than single facts. Often the visual sense is evoked - one 'sees' something before one's inner eye. Some psychologists describe memorising, and indirectly also cognising, as mental imagery. On the other hand, it is language which is the prime vehicle for transmitting knowledge from one man to another, and language works by sentences and propositions, which must be represented in a much more abstract way. Are there really two different systems for representing knowledge, or can the one be reduced to the other? What does empirical psychological evidence say? Can logic, being a science of reasoning, contribute anything? Such questions are equally interesting for the purely scientific student of cognition and for those interested in the foundations of the emerging applied field, known as 'knowledge management'.

## 2. Knowledge in philosophy, psychology and computer science

Today, representing knowledge has become the concern of a blooming industry. However, being at present very far from something that could justifiably be called artificial intelligence, that industry's demand for technical applicability does not seem to work superficialising, but rather to encourage more profound research, in particular such that requires cooperation between the different fields of study involved. My concern in this article is to try to establish some basic structural features that any system representing knowledge and thought should possess. I deliberately try to stay clear of any attempt to summarise the state of the art, being convinced that a structuring of the field is much more needed. I will thus emphasise fundamental conceptualisations, using well-known theories and results for illustration. In particular, I wish to focus on the tension between the emphasis on logic in computer science and philosophical epistemology and certain findings in cognitive psychology.

How, then, do different disciplines look upon knowledge? Broadly speaking, philosophers have recently been interested in the origin of knowledge (is it based on what is received through the senses, or on constructions by our own reason?), in the validity and certainty of knowledge (is there any kind of knowledge which it is absolutely impossible to doubt?), or in the nature of knowledge (what distinguishes it from e.g. true belief?), and less concerned about the structure of the mind which produces the knowledge. In each case, it is the knowledge of *a fact* (and only one fact at a time) which is of concern, and logic becomes a useful tool for its analysis. The state of knowledge of a mind tends to be identified with the set of facts known, and processes involving knowledge, such as explanations, tend to be described in terms of logical relations.

In contrast to philosophers, psychologists have focused their interest on other aspects of the mind, such as perception, memory and thought processes. No doubt, much of what is known in these areas is highly relevant to the study of knowledge, but the connections still have to be spelled out and discussed. In particular, it seems to be important to look at the 'dynamic' problem of modelling *processes*, such as thought, *together* with the 'static' problem of modelling memory or knowledge representation. These two aspects of the mind interact in the sense that what we observe is often only their *combined* effect - the classical Quinean problem of indeterminacy, meaning that empirical evidence, no matter how extensive, is insufficient for determining the

true nature of both factors. Anderson (1978) contains an argument to this effect - Hansson (1987) gives a more general argument in a measure-theoretical context. Large parts of this article will be about precisely such issues.

One of the basic problems in computer science is to find efficient ways of storing and retrieving information. The use of logic in this connection is interesting. In classical data base systems certain areas of memory are set aside to hold a predetermined type of information. If the sequence of signs “John Smith” is found in ten successive positions in a certain part of memory, this is interpreted as the name of a person, and if an adjacent cell holds the value 0, this may be interpreted to mean that John Smith is unmarried. For one who knows the purpose to which the different memory cells are set aside, it is possible to infer the fact that John Smith is unmarried. This can be expressed by saying that the system holds the information that John Smith is unmarried, or that the system ‘knows’ this, but it is important to note that the crucial element is an acquaintance with the code which connects memory positions and meanings, and that this is something which resides outside the data base system itself. Certainly, logic is used extensively in this connection, both informally in thinking about the system and formally in e.g. some programming language, but the *representation* of the fact is not logical in character. There is no symbol representing the fact as a whole, and there is no way of introducing a deductive apparatus and infer the logical consequences of Smith’s being unmarried. What is being represented is a collection of traits of a situation, like e.g. name and marital status, but not a proposition as such or something like a logical formula.

However, another type of programming is becoming increasingly popular, in particular when proofs and deductions are to be performed by the computer. Such programming uses a specific representation of propositions, often in the form of sequences of symbols, just like an ordinary logical formula, or in any case in a way which retains the internal logical structure of the proposition. The proposition is thus represented in a two-stage process: first by something which is essentially a formula in a logical language; and then this something is represented in a standard way by the contents of some memory cells.

The two types of programming are not exclusive - logical programming may e.g. be used as a shell around a conventional data base system to facilitate in- and output. But it is not obvious whether the capacities of the two types of programming to represent different types of

propositions are the same. The crucial point is the expressive power of the logic employed.

This alone would make it interesting to take a closer look at logic as a possible tool for an analysis of the structure of mind. In addition, a continuing debate in psychology about the relative merits of representing memories and thoughts as propositions or images points in the same direction. The following section will therefore be devoted to the possible uses and limitations of logic with a special view on the problem of representation.

## 3. The use and limitations of logic

Logic is often regarded as the obvious tool for analysis of proposition-like entities. However, classical definitions of the subject often speak in general terms about the study of the structure of correct reasoning or of sound argument. It may be possible to imagine e.g. a premiss which is a picture and a conclusion which is another picture, and to say that they form an argument, because all information in the second picture was already contained in the first one. In principle, thus, logic should not be biassed towards propositions over images.

But reality is different. Almost all logic presupposes that premisses and conclusions of arguments are propositions, preferably expressible in a formal language. This is one of two basic conceptions of logic which delimits its use for the analysis of the mind.

The second is the view that an argument, or a proof, is essentially a finite chain of propositions. (The alternative to this view is not necessarily that a proof should be an infinite chain, but it could also be a finite chain of something else than propositions.) Together, these two conceptions lead to profound difficulties for the programme of formalising knowledge in logic.

Suppose that we have complete knowledge about one area of reality (in a sense which can be made precise) and that it can be formalised in some logic. Then it can be *proved* that there exist true facts which cannot be proved in the logic, i.e. that the logic's proof concept is too weak in an essential way.

This is *not* Gödel's famous theorem, but in contrast one which can be proved by elementary means (see Hansson (1999)). The crucial point in

the proof is the finiteness of proof chains. What the theorem proves is that there is a *dilemma* between a demand on propositional *representations* - viz. that all the knowledge shall be representable - and a demand on the propositional *process* of thinking - viz. that that process can be modelled as a finite number of operations on propositions. Since both these demands are part of a more general, logical approach to the mind problem, that approach has got an internal problem. This emphasises the importance of widening the perspective to include both representations and processes - both knowledge and thinking.

This result, which is not particularly deep or difficult, seems to have been largely overlooked by both philosophers and psychologists. So does e.g. John R. Anderson, in his widely cited (1978), say that "it seems to be a generally accepted claim that any well-specified set of information can be represented by a set of propositions" (p. 257). Although what he says is true for some well-specified sets of information, it merely transfers the problem to the other horn of the dilemma.

The above result strongly suggests that logic, in its present form, is not directly of use for the analysis of mind. However, metalogical results often crucially depend on exact formulations, and since it is impossible to go into detail here, I have added an appendix to this paper, where this argument is further elaborated.

## 4. Classifying knowledge

In his delightful little book "The problems of philosophy", Bertrand Russell (1912) makes a distinction between *knowledge of things* and *knowledge of truths*, with a further division of the former category into *knowledge by acquaintance* and *knowledge by description*.

These distinctions may serve as a useful starting-point, but a word of warning is needed for those not versed in philosophical usage: 'things' need not refer to material objects - rather it is an ontological category, roughly consisting of such entities as can be referred to by noun phrases, whether or not they be abstract or concrete, mental or material. Russell mentions wishes and facts and concepts such as 'whiteness' and 'brotherhood' as proper objects for knowledge of things. Presumably events would be proper too.

The fundamental thing for Russell is knowledge of things by acquaintance. It is a sort of immediate experience “without the intermediary of any process of inference or any knowledge of truths”. His example is about the colour of his desk:

> The particular shade of colour that I am seeing may have many things said about it - I may say that it is brown, that it is rather dark, and so on. But such statements, though they make me know truths *about* the colour, do not make me know the colour itself any better than I did before: so far as concerns knowledge of the colour itself, as opposed to knowledge of truths about it, I know the colour perfectly and completely when I see it, and no further knowledge of it itself is even theoretically possible (ch. 5, beginning).

It is obvious that knowledge by acquaintance also produces knowledge of truths, but that the two are qualitatively different and that no amount of knowledge of truths can ever be the same as or equivalent to knowledge by acquaintance. There is an ontological definiteness and fullness about knowledge by acquaintance, whereas knowledge of truths is partial and depending on what aspect of the desk one happens to focus on. Although Russell does not explicitly say so, one gets the impression that our way of organising the world by a certain conceptual scheme strongly affects knowledge of truths, but not knowledge by acquaintance.

But it is not Russell’s view that knowledge of things as such is more fundamental than knowledge of truths. Rather, it is *acquaintance* which is fundamental. Acquaintance leads to one sort of knowledge which belongs to the category knowledge of things, but other sorts of knowledge in this category, such as knowledge by description, have no privileged status, but are in fact dependent on knowledge of truths. To know e.g. Bismarck is - for anyone else than Bismarck himself - to know some truths about Bismarck which together suffice for identifying him. In this way, knowledge by description is subordinated knowledge of truths, which in turn is subordinated knowledge by acquaintance.

This leads to an interesting observation with some connection to the concluding paragraph of section 1 above. Russell says that the most important thing about knowledge by description is that it makes it possible for us to transcend the limits of our private experience. So the following picture emerges: we have knowledge by acquaintance of

certain things based on private experience. This is a kernel of highly reliable knowledge, but of limited scope. There is a mechanism for extending this knowledge to knowledge of other things of which we have no private experience, the mediator being knowledge of truths, derived from acquaintance, and giving rise to descriptions.

This should be compared to the role of logic and language in the transfer of knowledge from one person's mind to another. In both cases there is the idea that knowledge in the mind of a person comes in batches, centered around some 'thing', and that it is related to, though qualitatively different from propositional knowledge, which however can serve as an intermediary when knowledge is transferred from one batch to another.

Russell does not give elaborate reasons for his classification of knowledge, but it receives independent support from an analysis of ordinary language usage. We use the following constructions in connection with knowledge:

* Knowledge that (facts or propositions)
* Knowledge of ('things')
* Knowledge how to (activities)
* Knowledge when, where, why, how, what, i.e. knowledge plus interrogative

When we go from words to facts, we first find that the category of 'knowledge how to', i.e. skills or what could be called procedural knowledge, falls outside Russell's classification. This is a useful reminder that such knowledge exists, but it is of a kind which falls somewhat beside the main groove of this article.

'Knowledge that' seems to fit nicely with Russell's knowledge of truths, and 'knowledge of' with his knowledge of things. 'Knowledge plus interrogative' is not an independent category, but usually a special case of 'knowledge that': to know *when* it happened is to know *that* it happened at time *t*. However, in some cases 'how' and 'what' may indicate 'nowledge of': to know what it is to be a woman is knowledge of womanhood. But in no case is there any reason to amend Russell's basic distinction between knowledge of things and knowledge of truths. As for his further distinction between acquaintance and description, this kind of linguistic analysis is without relevance. This latter distinction will only play an insignificant role in the sequel.

Another observation of interest is that many languages make finer distinctions than English about knowledge: German between *wissen, kennen* and *können*, and French between *savoir*, *connaitre* and *pouvoir*, in rough correspondence to 'knowledge that', 'knowledge of' and 'knowledge how to'. Even if the finer details may vary between the languages, it is obvious that it is a widespread impression that we here deal with different mental processes.

## 5. Perception, memory and thought

Knowledge is the philosopher's favourite aspect of the mind. Psychologists pay more attention to perception, memory and thought. Yet, there can be no doubt that all these aspect are closely interrelated. The naive model is that memories are impressions left by perceptions, that knowledge is closely related to memory, and that thought somehow consists in processing elements of either memory or knowledge.

While this naive model is valuable in its emphasis on the interrelations between these mental functions, it is nevertheless important to realise that the exact relationships may take many forms, corresponding to knowledge representations of different types.

To describe the effects of experiences on the mind as 'impressions' is a Humean way of expression which suggests a fairly mechanical relationship between perceptions and consciousness, or perception and memory. Kant saw a much more elaborate structure here, and results from both modern neurophysiology and psychology suggest that he was right. It takes surprisingly long before perceptions reach consciousness, and it seems reasonable to believe that they undergo quite an amount of processing, which suggests that the representation of the conscious product need not resemble the structure of the sense organ or the signal it produces. I will discuss these matters more in detail in section 10 below.

Memory is often regarded as the mental function which is closest to knowledge. Some simply say that knowledge *is* memory, others that memory is a constituent in knowledge. One of the things of which we can have knowledge by acquaintance according to Russell is 'memory-data'. But this does not mean that a good representation of memory is a good representation of knowledge. Knowledge is a more system-oriented concept, involving relations to other things known and to background settings. To know when Descartes died is not only to

remember seeing something like “Descartes, René (1591-1650)” in a dictionary entry - it is also to understand the convention that the figures refer to the years of birth and death, to recognise the genre of dictionaries as opposed to e.g. story-books, to judge the likelihood that dictionary editors are well-informed about this sort of information, etc.

Nor should we presuppose that either memory or knowledge relies on a unique representation format. Apart from the possibility that different categories in a classification require different representations (which will be discussed below), we must notice that memory and knowledge both come in a dormant as well as a vivid form. When a memory is retrieved, it is brought into the work area of the mind and it becomes conscious and vivid, and it is now that people feel that it should properly be described as mental imagery. But the memory must exist in some form also during its long dormant periods. We can think of the retrieval process in either of two ways: either the memory remains as and where it is, but some sort of mental searchlight is directed at it, illuminating it, or it is copied from its store into the work area of consciousness. In the second case it is not necessary that the permanent storage area and the temporary work area operate by the same system. We may think of the computer as an analogy: data are permanently stored as magnetic marks on a disc or a tape, but if you want to take a look at them or process them you fetch them up to your screen, where they appear in a completely different representation format, although they still constitute the same data. In fact, the common assumption that long term memory is neurochemical in character (protein-based) and short term memory neuroelectrical supports such a possibility.

Perhaps thinking can be best understood in terms of such movements of mnemonic or epistemic elements from the background to the fore of the mind. In general terms, thinking is the processing of such elements so that they come to stand in new relations to each other or recombine into new clusters. One way to establish such relations would e.g. be to exhibit a number of intermediary element such that the step from one to the next would follow the rules of some deduction system. But such processing does not take place among dormant elements, but requires consciousness. Perhaps thinking should not be seen as something operating on a single level, with elements of a single category, but as the faculty of efficiently recoding memory elements from long term storage into the format of the work area, and back again.

## 6. Classifying memory; three types of models

It should be obvious that memory shares many structural properties with knowledge, and that a closer look at current psychological research about memory is justified also from a purely cognitive standpoint.

This view is supported i.a. by the fact that memory easily can be classified in a manner similar to that used for knowledge in section 4. We can distinguish between 'memory that' (a clause), 'memory of' (a noun phrase) and 'memory how to' (a verb phrase), and feel convinced that the linguistic distinctions correspond to some sort of real differences between what would naturally be called *propositional, episodic* and *procedural* memory. However, it is more difficult to envisage the further distinction of 'memories of' into memories by acquaintance and memories by description, where the latter conception seems at first sight to be self-contradictory, although the non-authentic memory picture which suggestible persons build up in their mind could perhaps be an example.

Endel Tulving has introduced a distinction between what he calls the 'episodic' and 'semantic' memory systems in Tulving (1972). The distinction is elaborated upon, and experimental evidence for the functional separation of the two memory systems is quoted in Tulving (1983).

His definition of 'episodic' memory is fairly straight-forward and clear. It is "a system that receives and stores information about temporally dated episodes or events, and temporal-spatial relations among them" (Tulving 1983, p. 21). The similarity to 'memory of' is obvious. It must be noted, however, that the episodes are to be personally experienced, i.e. that the autobiographical tag on this type of memory is essential.

However, Tulving admits that the term 'semantic memory' was a less happy choice. In his 1972 chapter he described it as "the memory necessary for the use of language. It is a mental thesaurus, organized knowledge a person possesses about words and other verbal symbols, their meaning and referents, about relations among them, and about rules, formulas, and algorithms for the manipulation of the symbols, concepts, and relations" (p. 386). This encompasses, in standard parlance, both semantics and syntax, and yet it is obvious that Tulving meant an even broader concept: "We know many things about the world

which are neither meaningful nor readily expressible in words or other symbols" (1983, p. 28). An alternative, and in many ways better expression would be "knowledge of the world", says Tulving in his 1983 book, where he even considers it an open question whether lexical memory should be regarded as 'semantic'. This alternative expression suggests something like 'memory that'.

A further source of uncertainty is Tulving's reference to Bergson (1896) and Russell (1921) as precursors of his own distinction. These philosophers make a distinction between habit-memory and recollections of unique events, i.e. between procedural and episodic memory, with no obvious counterpart to Tulving's 'semantic' category. (It is clear from Tulving (1983) that he intends procedural memory to be a further category, on par with episodic and 'semantic' memory.) It seems clear that the concept of 'semantic' memory needs more elaboration before it attains the same definiteness as its episodic counterpart. The reason may be that we have not yet found the right definition, or, in my opinion more plausibly, that the term does not refer to a unitary idea, but has simply come to be used for whatever is not clearly episodic.

It is an important point for Tulving that his distinction is not merely one of mnemonic content, but between different memory *systems*. This admits and even suggests that the two systems may operate with different *representations*. A natural thought would be to connect on to the debate about mental imagery versus propositional representation, and hypothesise that the natural units with which episodic memory operates are best represented by images, and that the more detached content of a 'semantic' memory is better seen as a proposition.

One must admit that it is often tempting to dissolve a problem, rather than to solve it, by admitting that both parties are right in a way, but there are also obvious difficulties in this case. From the memory of a particular episode, there originates a large number of 'memories that' of singular facts, which were present in the episode. Are these memories still in the episodic system (i.e. do there exist 'memories that' in each of the two systems), or have they been transferred to the 'semantic' system (which then would contain elements both with and without personal anchoring)? Either way, it seems that we would be multiplying entities beyond necessity.

Since Tulving refers to several types of experimental results in support of his views, this may be the proper point to proceed from general

analysis to a more detailed examination of the empirical evidence. I will do so in the remaining sections, but I will prepare that examination by saying something about what distinguishes propositional and image-like representations in principle and about my own preliminary position.

An image-like representation always has a mark of personal presence tagged onto itself. But images differ from propositional representations also with respect to pure information content. First, images are richer. Language is a coarse net in which to catch reality. Many parts of an image my be such that no accurate description is possible. Logic has taught us that a complete theory need not be categorical, i.e. that knowledge of the truth value of all sentences in a language need not determine the structure of the reality it describes.

But images are poorer too, in a sense. Not in their expressive strength, but in their ability to vary the degree of specificity. They are conceived as wholes, and some traits can hardly be left out of a whole, whereas a proposition can abstract from even salient features of an object and select to convey only partial information. It can say that he had a black eye, or that the colour of the car was bright, but an image must specify which eye was black and what colour the car had, if the person or the car were in any way in the centre of interest.

There are other differences too, but those mentioned suffice to bestow different logical properties to representations based on images and propositions. But one should not conclude that these are the only two possibilities. One can e.g. employ the notion of a *mental model*, which shares some but not all properties of a mental image. This notion, which admits considerable latitude in technical details, has been used in an interesting way in Johnson-Laird (1983).

My own position is very preliminary, but could well be described as some sort of mental model. It concerns only non-procedural knowledge. The reason I state it, despite its inchoate state, is simply that one has to have some idea in the back of one's head when one sets out to interpret experimental data, and I thought it most fair to indicate what it is.
So my purpose is not to argue for my position - I am certain that it will change in many respects as I dig further into this area. But I can perhaps hope to prove that it is compatible with several known results, and that some of these results can be interpreted in other ways than they have been.

So, in short, and without justification: I think of memory as unitary - in fact, I think that knowledge is more basic, and that memory is just a special part of a unitary epistemic system. Knowledge is fundamentally *episodic* knowledge or 'knowledge of things', i.e. organised in clusters around some central theme, which may be e.g. an event. Memory is knowledge which is based on acquaintance (in a somewhat extended, un-Russellian sense, in which some inferences are admitted). This should correspond fairly well to Tulving's episodic memory.

*Propositional* knowledge or 'knowledge that' (and 'memory that' and parts of Tulving's 'semantic' memory) has no independent existence (at least not in the long-term store). It emanates from episodic knowledge and is used as a vehicle when memories are retrieved and brought into the fore of consciousness, or when knowledge is to be dressed up in linguistic clothes. It may thus be an important element in *thinking*.

Epistemic elements are not freely combined in general, but have a tendency to group together in *stereotypes*, representing common, typical events, objects, processes, etc. Stereotypes vary greatly in complexity. The acquisition of a sufficiently large repertoire of stereotypes is called education and is a prerequisite for a smoothly functioning epistemic system. The word is modelled on the printer's technical term and any psychiatric connotations should be suppressed. Stereotypes are organised in an associative network.

The *representation* of a thing known consists of one or several stereotypes plus a number of distinctive traits, both additions and corrections. The encoding of a perception consists of a search process for a suitable stereotype plus the creation of the distinctive traits. The depth of an encoding (this metaphoric expression should not be taken too literally) is related to the number of associative links of the stereotype and to the saliency of the distinctive traits. The retention and retrievability of a piece of knowledge depend on the depth of encoding.

## 7. Empirical evidence: the dissociation of performance

One of the main arguments that Tulving puts forward for his view that episodic and semantic memory are functionally different systems is that people tend to perform differently on episodic and semantic tasks. One of the experiments to which he refers is the one reported by Jacoby and Dallas (1981). It is similar in structure to several previous experiments - see e.g. Craik and Tulving (1975).

The experiment is divided into two phases. In the first phase, a question was presented on a screen for one second, followed by a target word (a five-letter noun), which remained in view until the question had been answered. This was repeated for 60 words for each subject. The questions were all of the yes-or-no type and were of three different kinds: twenty were about constituent letters (does the word contain the letter L?), twenty about rhyme (does the word rhyme with 'train'?), and twenty about meaning (does the word refer to the centre of the nervous system?). The idea was to force three different level of mental processing in the subjects. Yes and no answers were required equally often in each category.
In the second phase, the subjects were divided into two groups. The first group - called A for later reference - were put to a recognition test, in which they were presented with 80 words (the 60 words from the first phase plus 20 new ones), one at a time, and required to say 'yes' if the word had occurred during phase 1, and 'no' otherwise.

Group B were presented with the same words, but they were flashed on the screen for only 35 milliseconds, and the subjects were required to say what word it was.

The frequencies of correct responses were the following:

| | Question type | | | | | | |
|---|---|---|---|---|---|---|---|
| | Letter | | Rhyme | | Meaning | | New |
| | Yes | No | Yes | No | Yes | No | word |
| Group A | 0.51 | 0.49 | 0.72 | 0.54 | 0.95 | 0.78 | 0.85 |
| Group B | 0.78 | 0.81 | 0.82 | 0.80 | 0.80 | 0.83 | 0.65 |

The table is to be interpreted thus: frequencies are separated depending on the type of question asked about the word (constituent letters, rhyme, and meaning), and within each such category also depending on whether the correct answer was in the positive or in the negative (i.e. on whether the word in fact contained the letter L, etc.). It was found

that the hypothesised increase in mental processing as we go from left to right (keeping the yes/no variable constant) corresponded to increased performance for group A, but not for group B. It was also found that questions requiring a yes answer more frequently produced subsequent correct recognition than questions requiring a no answer.

Mean reaction times were also measured and found to be shorter for high response frequencies and *vice versa*.

Tulving interpreted the tasks of groups A and B to require episodic and semantic memory respectively and concluded that the experiment "reveals a strong dissociation of performance on episodic and semantic tasks, and thus provides evidence in support of the distinction of episodic and semantic memory as functionally different systems" (p. 89).

An evaluation of this experiment requires first an analysis of the character of the tasks of the two groups, and secondly a discussion of whether the hypothesis about distinct systems best explains the observed response pattern.

The characterisation of the task of group A is comparatively unproblematic. Jacoby and Dallas call it a test of recognition memory, and Tulving prefers the expression 'episodic recognition'. Since the subjects are in fact required to retrieve a minor recent episode in their lives, the episodic character of the task is evident, even if it may be doubtful whether proper recognition is involved (see below).

The task of group B is, however, more confusing. Jacoby and Dallas call it 'perceptual recognition', and Tulving declares that he will use 'perceptual identification' instead, but when he prints the response frequencies, he uses the label 'semantic identification'. In view of his explicit denial that lexical memory is a clear case of semantic memory (1983, p. 69-70), the readiness to classify B as 'semantic' is a bit surprising. Perhaps 'lexical identification' would be the least biassed label.
But even if we grant Tulving his classification of the A and B tasks, does it really follow that there are two distinct memory *systems*? To show that it is not so, I will sketch three alternative interpretations.

For the first one, I wish to direct attention to the distinction between merely remembering or recognising something, and remembering something *about* something (or recognising something *as* something).

The distinction is related to ‘knowledge of’ versus ‘knowledge that’, or to acquaintance with an object versus knowledge that a fact is true. To recognise something properly is to realise that you are acquainted with it, i.e. to acknowledge the existence of that immediate tie which is characteristic of ‘knowledge of things’ and which constitutes proof of some previous encounter. If you then, in addition, remember something *about* this ‘thing’ (or recognise it *as* belonging to a special kind), then what you do is the further act of extracting ‘knowledge that’ about that thing with which you are acquainted. This will take longer, and your degree of success will depend on the depth of the encoding.

Seen in this way, the task of group B is pure recognition (“which word is it?”), while that of group A is more complex, consisting of an identical first element plus a second element of analysing the representation of that word in order to see if it is a fact about it that it was mentioned in the first phase of the experiment. The ease with which this second element is performed will of course depend on the depth of encoding, which is lowest for words processed at the graphemic level and highest for those at the semantic level.

This idea can be empirically tested. Suppose that those in group A were asked something about each word which reflected ‘knowledge of the world’, but did *not* relate to phase 1 (e.g. “is this a part of a human body?”). It is conceivable that the responses would show the same pattern (measured, perhaps, by reaction times), although this task would clearly have to be classified as ‘semantic’ by Tulving.

My second alternative interpretation consists simply in noting that the subject’s control process is not a factor which is controlled in this experiment. Control processes, which have been discussed by Atkinson and Shiffrin (1977), are voluntary strategies adopted by a subject. Examples are coding procedures, rehearsal operations, and search strategies, and they are influenced by the nature of the instructions, the meaningfulness of the material, and the individual subject’s whim. In the present case, the different instructions to group A and B may well induce the subjects to adopt different control processes, determining what parts of memory they search and what parts they discard.

Finally, the result could also be regarded as a time effect. The short time allowed the subjects in group B to recognise the word may prevent all but the most elementary processing of the perception, thereby making it irrelevant whatever traces phase 1 has left in memory.

## 8. Contradictions in memory and knowledge

It is clear from everyday experience that inconsistent pieces of information may coexist in memory, if the contradiction is not too obvious, but it is equally clear that certain blatant contradictions ("the car is green all over and it is also blue all over") will never be present in a sound mind.

This is a recurrent problem for logical models of mind. Usually, knowledge is supposed to be closed under logical consequences, but then no contradictions can exist. And if *some* conclusions are automatically drawn, but not all, which is the non-arbitrary criterion which separates the admissible consequences from the non-admissible?

This problem is shared by all *propositional* models. Propositions have few natural relations other than logical ones, and it therefore becomes difficult to define some such criterion. An *episodic* model is better equipped in this respect: a natural start is to say that contradictions are detected and removed if they are wholly *within* one episodic representation (cluster, stereotype plus distinctive traits), but not necessarily otherwise.

The idea is that a stereotypical representation of an event or an object implies that certain attributes are proper. A person has a size, sex, colour of complexion, hair colour, state of clothing, etc. There are, so to speak, slots in the stereotype, which may be empty (for vaguely conceived objects) or filled with a definite attribute, but which cannot hold several (contradictory) attributes. This provides a mechanism for blocking (at least some) contradictions within one representation, but it admits freely of contradictions between representations.

Elizabeth F. Loftus has described a number of studies of hers, which are relevant to this question (1979). She is interested in cases where a witness has received contradictory information and she puts the question whether all received pieces coexist in memory, in which case a resolution has to be made at the time of retrieval, or if memory has been altered and kept self-consistent.

They all have to do with suggestibility. In one experiment the subjects saw a series of colour slides depicting a wallet snatching. The next day they read a description of the event, allegedly written by a psychology professor who had studied the slides more in depth. The description for

some of the subjects contained four subtle errors (a green notebook carried by a minor character was e.g. described as blue), but for the others it contained a blatant mistake as well (the red wallet taken by the thief was referred to as brown). This functioned as an alarm and made the latter group resist the suggestions better - also the four subtle ones.

In a follow-up experiment the blatantly erroneous information was not presented together with the subtle suggestions, but was delayed for two days. If everything received is stored in memory and contradictions are resolved only at the time of retrieval, then the alarm would function in this case too and make the subject realise the contradiction between the four subtle suggestions and the still available original memories. This did not happen - the subjects were as suggestible as if no blatant error had existed. This suggests that memory has actually been altered, which fits better with the idea of an episodic cluster than with a non-structured set of propositional representations.

In another experiment, it was suggested to the subjects that a car passed a stop sign when in fact it passed a yield sign. When tested for their recollection of the event, the subjects showed no longer response time than subjects who had not received any suggestive information, which again implies that any conflict had been resolved beforehand.

In yet another experiment, a so-called second-guess technique was used. Subjects were tested for their recollection of colours of objects they had seen on slides. Their task was not only to indicate which colour best represented their recollection of the object, but also which was their second choice, assuming that the first choice was incorrect. There are some tricky problems in such an experiment, but when they were kept under control, it turned out that those who had been suggested to give a wrong answer as their first choice were not right more often than chance in their second choice either. The correct memory did not lurk beneath the surface, ready to take its proper place when the intruder was revealed, but seems simply to have disappeared. It is hard to see why this should have to be so if memories are represented by propositions which can coexist, but it is more natural if we have a stereotype with a slot for colour, and this slot has already been filled.

## 9. Thinking in syllogisms

So far, I have been concerned mostly with the more permanent representation of memory and knowledge. I have also indicated that *thinking* may be different, that it has to do with placing information in the fore of consciousness, and that it may use other forms of representation. It is not unreasonable to think that individual control processes are of major importance here, and that we can expect considerable individual differences, depending on habit and education, as well as intra-individual differences, depending on mood and the nature of the task.

The closest we have come so far to a model of thinking is the discussion of natural deduction systems in section 3. It is perhaps not surprising that logical deduction should be one of the first models offered for thinking in general, because it is such a thoroughly researched field. But psychologists have not devoted much attention to deduction in general, although some special problems of deduction have received considerable interest.

One of these is the classical, but narrow and artificially delimited field of *syllogisms*. A syllogism is, for the present purposes, a deduction from two premisses of the form “All A are B”, “No A are B”, “Some A are B”, or “Some A are not B” to a conclusion of the same form, with some additional restrictions on the A’s and B’s. Many theories about syllogistic reasoning have been advanced, but they have had little applicability to deductive thinking in general. A recent attempt to build a model with somewhat better prospects of generality can be found in Johnson-Laird and Steedman (1978), where the model is also put to test.

Their paper also gives information about what conclusions people actually draw from syllogistic premisses, independent of theoretical assumptions, and it is therefore useful also for those who wish to test other models. I will not discuss Johnson-Laird’s and Steedman’s model in detail, but only use it for comparison. My aim is instead to illustrate to role of individual control processes by exhibiting two sketches of models, both based on my own introspective observations, but discussed with enough logicians to ensure that they are not completely idiosyncratic.

The point of comparing two models is that they correspond to different levels of ambition. The first one is intended to reflect every-day

thought and it turns out to correspond well with how people actually think, according to Johnson-Laird and Steedman, although the conclusions are far from always correct. The other one focuses on correctness and accuracy, and is to be used when one approaches a problem as a professional logician. I often feel that I switch between the two models, and that I do so at will.

I prefer to present the models in a visual language, but it is important to note that the images are not to be interpreted as complete pictures, but may well be mere visual symbols.

The places of the A's and B's in a typical syllogism are filled by category nouns or class names, like 'philosophers', 'Greeks', 'bee-keepers', etc. In my first model each such class is thought of in terms of a single representative, equipped with some attribute which reflects the class. The different types of premisses are then envisaged as follows:

| | |
|---|---|
| All A are B | one representative with double attributes |
| No A are B | two representatives with a barrier between |
| Some A are B | two representatives with a hazy link between |
| Some A are not B | two separate representatives |

The first premiss is about A and B, the second one about B and C, and the conclusion is to be about A and C. One then looks at the representatives for A and C and sees if they stand in any discernible relationship to each other. If A and C have the same representative, then we have "All A are C" or "All C are A" depending on whether A or C comes first in one of the premisses. If the representative for B is also the representative for A or C (say A), then the relation between A and C is the same as that between B and C (barrier, link, or nothing), and our conclusion is formulated accordingly. If, finally, all three terms have different representatives, then we are too uncertain about what we see, and our verdict becomes 'no conclusion'.

Evidently, the pictorial language is not essential. We can simply speak about identification of classes and the substitution of one class for another in a completely abstract way. But a visual language is more true to my introspection and it may provide guidance to those who want to amend the model to something more specific in the last case with three representatives.

This simple, rough and ready model is surprisingly accurate if we use it for predictions about what conclusions people will in fact draw. In 51 of the 64 syllogisms it predicts the same conclusion as the majority chose in Johnson-Laird's and Steedman's experiment. A direct comparison with Johnson-Laird's and Steedman's model is complicated by the fact that their model works in two steps: first an initial conclusion is drawn, which is then subjected to a logical test and modified or abandoned if found incorrect. A second complication is that their model only predicts that the answer will be one of two, or sometimes even one of three possible conclusions. The following table compares the predictive accuracy of the models:

Number of cases in which model predicts majority's conclusion

| | A | B | C | D | E | |
|---|---|---|---|---|---|---|
| 1st figure | 14 | 7 | 7 | 12 | 13 | |
| 2nd figure | 15 | 7 | 7 | 12 | 12 | |
| 3rd figure | 10 | 1 | 3 | 12 | 16 | |
| 4th figure | 12 | 2 | 4 | 12 | 16 | |
| Sum | 51 | 17 | 21 | 48 | 57 | (maximum = 64) |

A: prediction by present model
B: first-mentioned initial prediction by JLS model
C: any initial prediction by JLS model
D: first-mentioned final prediction by JLS model
E: any final prediction by JLS model

But majority is often wrong. Many syllogisms prove to be tricky, and even Johnson-Laird and Steedman make mistakes (or at least rely on tacit additional premisses) when they classify their subjects' answers as correct or incorrect. It is of some interest for the sequel to look a little closer at one of the trickier cases.

Suppose you are given the following premisses:

* All bankers are athletes
* No councillors are bankers

Chances are great that you will not be able to draw any conclusion in syllogistic form from these premisses if you work intuitively (see Johnson-Laird (1983), p. 66). If you set out to work more systematically, you may use set-theoretic diagrams and represent

bankers by a small circle inside a bigger circle, which represents athletes. This will illustrate the first premiss. Then you wish to draw a third circle for the councillors. By the second premiss, it has to be wholly outside the bankers' circle, but you can draw it either wholly outside, wholly inside, or intersecting the athletes' circle, but *not* wholly including it, for then it would not be outside the bankers' circle. So it *cannot* be the case that all athletes are councillors, i.e. some athletes are not councillors.

If you can come up with this or a similar piece of reasoning without help or training, then your logical talent is great indeed, and yet the reasoning contains a logical flaw. The impossibility to draw the third circle around the athletes' circle depended on the bankers' circle not being void! If there were no bankers in that circle, then there are no restrictions for the third circle. An empty set is very difficult to represent visually, for it is a subset of all other sets, including disjoint ones, at the same time, and therefore it cannot be represented by a circle at any particular location. So our conclusion would only be valid if we had a third premiss to the effect that there were in fact bankers present. As it stands, the problem admits no solution, just as you thought in the first place.

Logicians are of course well aware of such traps and device safety measures to avoid them. Such measures may take the form of elaborations of set-theoretic diagrams, and one such elaboration will constitute my second model for syllogistic reasoning.

Think, like before, of the classes as being represented by areas on a map, enclosed by boundary curves. But we must not only survey the country, but also try to spot all the people in it. When we spot a person, we mark off that area as inhabited. That all bankers are athletes now means that the inhabited parts of the bankers' land (the land where any banker and no-one else lives) are included in the inhabited parts of the athletes' land, or, equivalently, that those parts of the bankers' land which are outside the athletes' land are uninhabited. Similarly, that no councillors are bankers means that there is no common inhabited area of the bankers' and councillors' land. Since neither of our premisses actually requires any area to be inhabited, we cannot hope for conclusions which establish habitation, so-called particular conclusions, such as "Some A are C" or "Some A are not C". The other type of conclusion, exemplified by "All A are C" or "No A are C", is called universal and requires *un*inhabitation of certain areas (of the "A-but-not-C" and "A-and-C" areas respectively). There is nothing in

our mental landscape which prevents habitation of such areas, and we therefore fail to draw any such conclusion too.

In contrast to the previous one, this is a model which always gives correct answers, and usually all the correct answers. But it requires greater mental concentration and perhaps some training before it can be used efficiently. I use it when I need to be sure, but in my lazy moods I fall back on the simple model with representatives. Both models have been dressed up beyond necessity - the latter is in fact equivalent to a certain way of employing Venn diagrams. This suggests that the visual appearance is not functionally essential, although it seems to be important to some people, perhaps because it easier to scan a visual field for certain objects than to search an auditory or an abstract representation.

## 10. The processing of perceptual input

I have referred above to the dispute between those psychologists who claim that mental representations are best thought of as images, and those who prefer more abstract entities, like propositions. I have shown, towards the end of section 6, that the two positions do not only differ with respect to their appeal to every-day introspection, but also in some of their logical properties.

How, then, does the idea about representation by stereotypes and distinctive features, which I sketched in section 6, compare to the two alternatives in this dispute? I have found the propositional approach wanting in many respects, and stereotypes and imagery have one important aspect in common, namely clustering of features, so it would be natural to put stereotypes close to imagery. However, there are also several points of difference. The most important one is the relation between the representation and the perceptual input.

The theory of mental imagery, at least in its cruder forms, sees the mnemonic representation as a more or less detailed copy of the sense input. Apart from the enormous demands this makes on the capacity of memory, it is of little help in explaining why memory is selective, how it changes with time, how associative links are built up, and many other problems. The idea of stereotypes, however, presupposes extensive processing before a representation is formed. I have only vague ideas about the precise nature of this processing, but it must include a search procedure to find a suitable stereotype and the

establishment of some new associative links. This also means that stereotypes need not resemble sensual inputs, but may be regarded as abstract entities.

Some empirical findings in neurophysiology support this idea of extensive processing. I am thinking of Benjamin Libet's comparisons of ordinary sensory sensations and those elicited by cortical stimulation (Libet 1981).

If a certain part of the surface of the cortex (the postcentral gyrus) is stimulated electrically, the subject will feel a sensation, located in the opposite side of the body. If the electric pulses are at liminal intensity, they will have to be continued for at least 500 milliseconds (ms) if this is to occur. Stimulations of shorter duration will not elicit any sensory experience at all. By contrast, even a very short (15 ms) electrical pulse applied to the skin in the relevant region of the body will cause a conscious sensory experience, but even in this case neuronal activities will continue to develop for more than 500 ms.

This latter fact by itself need mean no more than that neuronal activity dies out slowly after it has done its work, but the fact that no sensation reached consciousness until after about 500 ms of cortical stimulation suggests that the prolonged neuronal activity is a necessary prerequisite for conscious experience also in the skin case. If this is so, we never experience the present, but only the recent past. This is not to say that nothing happens in the mind during the 500 ms between stimulation and experience. We may react on startling stimuli in considerably shorter time than that, and there may be formed permanent records in the mind, which may influence us in the future, but which are outside our conscious reach.
To test this idea, Libet showed that the sensation induced by skin stimulation could be suppressed if the cortex was stimulated somewhat later. In some cases the skin sensation was not suppressed, but enhanced instead, probably depending on the exact location of the cortical stimulation. Either effect usually occurred if the onset of the cortical stimulation was delayed about 200 ms, although in some cases a delay of 500 ms was found to be effective. It should be noted that the first nerve impulses reach the cortex in considerably shorter time than that. The cortical stimulation had to go on for at least 100 ms to be effective, so in fact events happening 300-600 ms after the skin stimulation could prevent it from reaching consciousness.

These findings tell us that many things happen in our minds before they reach consciousness and presumably before they reach our memory. This speaks against simple imagery models and in favour of more elaborate representations. However, it tells us little about *what* goes on. Another finding by Libet suggests that one of the things which happen is that sensations become temporally marked in the process. Suppose that cortex is stimulated as before between times *t* and *t*+500 ms, and that the skin is stimulated at e.g. *t*+300 ms. We would not expect the subject to experience the two sensations at the time they occurred, but with some delay, which however would be equal for the two sensations, so that they were experienced at *t*+500 and *t*+800 ms respectively. But subjects experience that the skin stimulation occurs *first*. This result is notoriously difficult to interpret, but Libet's suggestion is that the sensation has received a time tag in the process, and that such time tags determine the experienced time order, irrespective of when the sensations enter consciousness. Kant would have been interested!

If there are time tags, there may be many other sorts of tags too. In computer language, this suggests that memory could function as an index-sequential register, where events are stored in one place, and their location is stored elsewhere in a number of index registers, each specialising in some 'dimension' of life, such as time, place, smell, sound, children, food, or any X about which it is possible to start a conversation by saying "Speaking of X, I remember when I....". Associative linking and searching would then take place in these index registers, and not in the main store of memory. This is all speculation, of course, but interesting speculation.

## 11. Conclusion

There is a wealth of observations about every-day cognitive life in classical philosophical epistemology and there is quite a number of elaborate and general theories for understanding them in a systematic way. There is a wealth of precise empirical data about not-so-every-day situations in psychology, but there are surprisingly few general theories. Not that psychologists don't theorise, but that they, in comparison to philosophers, theorise about very narrow fields. You will find elaborate theories for the game of 'gomuku', but not for games in general, for syllogisms, but not for deduction in general or inference in general. Those studying concept formation share little theory with those interested in hypotheses, and problem solving is studied with few outlooks at internal representation.

This reflects a situation of increasing specialisation and increasing technical skill in experimental design and there is no doubt that the detailed knowledge of the human mind is growing rapidly. But the situation is not favourable for someone coming from another field, such as philosophy or computer science, wondering to what extent psychological results may bear on his own problem. In this article I have tried, from a philosopher's point of view, to find out what limits experimental evidence sets to general theories about the nature of the human mind. I don't claim to have found any definite answers, but I do claim that I have found some evidence that it might be a fruitful strategy to analyse empirical results from different psychological fields in a philosophical manner with an eye on their relevance for more general theories than is customary in psychology.

## APPENDIX ON THE RELEVANCE OF LOGIC

In the broadest sense, logic can be defined as the study of correct reasoning. Understood thus, logic is at the same time of unlimited relevance for the study of the mind, since there is no limitation to the kind of reasoning to which it applies, nor to the type of representation one may use, and of doubtful relevance, since the methods for studying correct and actual reasoning respectively might differ considerably.

As to the latter point, its importance crucially depends on the purpose of our study. If we are interested in designing computer-based aids in decision-making or else in improving human abilities, then correctness will undoubtedly be a main concern, but certainly not the only one, since only a system with some structural similarity to the real processes of the human mind will be able to supplement these. But if we are driven by a pure interest to understand these processes, we should not jump to premature conclusions about the usefulness of logical methods.

But in common parlance logic is something much more specific. It is a system for formally representing the contents of thoughts and judgements (i.e. propositions) and a method of explaining why and how some such contents are logically true and some logically follow from others. A logic in this sense consists of three parts:

a) a logical *language*, i.e. a system of representation of propositions, which is essentially language-like,

b) a concept of *model*, by which the ideas of logical truth (truth in all models) and logical consequence are made precise, and

c) a *deductive apparatus*, which clarifies the ideas of proof and derivation.

In order to study the usefulness of logic for an analysis of mind I will concentrate on two issues: Can it form the basis for a system of representation of that which is known (or memorised, or perceived)? Can it shed any light on the process of reasoning which leads from one such representation to another? The first question is about the internal *structure of an entity*; the second one about a *process*.

If a particular logic is to be used, it is imperative that the *language* of that logic has an expressive capacity which is large enough to encompass all that can be known or thought. It is well known that ordinary (first order) predicate logic has limited such capacity, and that e.g. modal attributes (such as necessity), generalised quantifiers (like 'most of'), branching quantifiers, quantification over predicates and predicate modifiers (functioning like adverbs) fall outside.

However, in each of these cases it has been possible to extend the language to include also the new feature. But this does not make it plausible that there exists a 'universal' logic, which includes *all* such extensions, or even that it is meaningful to think in terms of an exhaustive list of necessary extensions.
In addition, we know that predicate logic has some nice properties, which are not in general preserved by extensions. One such property is *completeness*, i.e. that all logical truths are provable. This result, which can be said to reflect the appropriateness of the model theory and deductive apparatus for one another, is one without which the deductive apparatus would lose much of its interest as a model for the process of reasoning, since it would then not be powerful enough to prove all truths. Completeness is not necessarily preserved by extensions.

The *model theory* for a logic defines logical truth as truth in all possible models, where of course the definition of a 'model' depends on the logic in question. This is a non-constructive definition, and we have in general no efficient procedure to determine whether something is a logical truth. Model theory can thus not tell us anything about the process of reasoning.

As for the questions of representations, it should be noted that it is a consequence of the definition of logical truth that two propositions are logically equivalent if and only if they are true in exactly the same set of models. In other words: if logically equivalent propositions are thought of as reflecting the same knowledge (as is generally assumed), then that knowledge can in principle be represented by the corresponding set of models. However, this representation is essentially infinitistic for all but trivial logics, and even if its intelligibility depends on how 'model' is defined in the particular case, it is in all likelihood a poor candidate for a realistic view of the mind.

The *deductive apparatus* usually consists of axioms and inference rules. A *proof* or a *deduction* is defined as a sequence of formulas, each of which is either an axiom or an immediate consequence of previous formulas according to one of the inference rules. The steps of the process, and the degree to which these steps may resemble parts of the natural thought process, are thus determined by the formulation of the inference rules.

Early deductive systems were very parsimonious about their inference rules, and usually employed only one. As a consequence, proofs became opaque and long-winded, and it required quite some craftiness even to prove rather obvious truths. So-called *natural deduction systems* try to remedy this by employing a much larger set of inference rules, which is built up in a systematic way from one introduction rule and one elimination rule for each logical constant involved, each rule being a very obvious reflection of the meaning of the logical constant. In this sense, the inference rules deserve the epithet 'natural', although one should be aware that no attempt has been made to argue that they are also 'natural' in the sense of reflecting real-life thought processes.

In fact, by slightly changing the concept of an inference rule, natural deduction systems are able to do away with axioms altogether. It is also possible to prove, for common logics, that any proof can be transformed into a standard form, and that theorem-proving therefore becomes a more strategic enterprise, with less need for *ad hoc* technicalities. *Proof theory*, as this study is also called, is thus perhaps that part of logic which has the greatest chance of being able to contribute to our understanding of the working of the mind, although it can, of course, only serve as a platform, on which more articulated models can be built and tested. Very little has so far been done towards putting this theory in contact with empirical facts about the mind. Two obstacles should be noted:

The first one is simply that there is no evidence that human thought proceeds in a sequence of minimal steps. Rather the impression is that it leaps ahead and sometimes misses the target. But it is conceivable that such leaps could themselves be sequences of minimal steps, i.e. that some such sequences are more natural than others. Any theory build upon proof theory should explain why this is so, and also why incorrect sequences are sometimes chosen.

The other obstacle is more fundamental. Even if a natural deduction system is an improvement on classical deduction systems, it is still a deduction system *for the same logic*. Even if the structure of a proof is better understood, no more and no fewer formulas are provable than before. So we must ask the question: is the deduction system, extensionally speaking, appropriate?

I have discussed this question in my (1999). It has some similarity to Gödel's famous result, but the relevant theorem is provable by elementary means. The basic idea is to look at some well-defined area of knowledge and see to what extent it can be formalised.

As an example of such an area I will choose the theory of natural numbers. There is no very deep reason for this particular choice - what I need is an actual infinity which is as well understood as possible. This theory was given an axiomatic formulation by Peano in 1889 (although Dedekind and, perhaps, Peirce had formulated the axioms even earlier), based on the idea of a first number ('zero') and the operation of forming a 'next' number to any given number. If these ideas are clear in one's mind, and if one can understand the totality of natural numbers as those entities which can be reached from zero by repeated use of the next number-rule, then one has a firm idea of the natural numbers. Mathematicians are satisfied that this is about how well one can understand an infinite set.

The theory of natural numbers can be formalised in second order predicate logic with identity. The mathematician's trust in the theory is supported by the fact that it is easily shown that the formalisation is *categorical*, i.e. that all models are isomorfic. This means that all structural properties of the natural numbers are uniquely determined by the axioms. The theory of natural numbers is thus an example of a well understood area of knowledge which is formalisable in a particular logic. Let us now see whether the deductive apparatus of this formalisation is appropriate.

(The theory of natural numbers *cannot* be formalised in ordinary, *first order* predicate logic, since the crucial axiom is essentially second order. One sometimes sees a fragment of the theory thus formalised, even under Peano's name, but that fragment altogether lacks the intuitive simplicity of the original theory.)

It is proved, by elementary means, in my (1999), that *there is no logic in which a categorical theory about an infinite set of objects can be formalised and in which the deductive apparatus is strong enough to prove all logical truths*. This result is much simpler than Gödel's famous theorem (the crucial point being the finiteness of proofs), but seems to be at least as relevant from an epistemological point of view.

It follows that *no* deductive system is appropriate for the theory of natural numbers, be it natural or classical. Modern logic can thus offer no *immediate* model of the mind. Whether it can offer a *basis* for such a model is not so much a question of logical results, but a matter of empirical fitness.

This somewhat discouraging result should not worry us in the long run. It pertains only to the formal view of logic as the science of relations between propositions, and not necessarily to the broad idea of logic as the study of correct reasoning, viz. if correct reasoning can be modelled by other means than finite sequences of propositions.
It may strike some readers that the above discussion is more similar in tone to what was customary in the first few decades of this century than to modern logic. In those days, logic was seen as a tool for elucidating fundamental epistemological problems in the foundations of mathematics, and the tool was adapted to the problem, which meant i.a. that higher-order theories were freely used. But it seems that with the publication of several important metalogical results around 1930, there appeared a cleft between logicians and mathematicians, which has been widening ever since, with logicians more and more preferring the well-behaved world of first order logic, even if only scattered fragment of the interesting theories can be rendered in it, to the real problem of analysing the mental world that mathematicians have built. Perhaps it is the logic of Frege, Peano, Russell, Zermelo, Sierpinski and the early Tarski that will prove to be of importance to the cognitive sciences, rather than that of the post-gödelians?

*Bengt Hansson*
*Department of Philosophy*
*Lund University*

## References


Anderson, John R. (1978). “Arguments concerning representations for mental imagery”, *Psychological Review* vol. 85 (1978), p. 249-277.

Atkinson, R.C. and Shiffrin, Richard M. (1977). “Human memory: a proposed system and its control processes”, in *Human memory. Basic processes* (ed. by Gordon Bower), New York 1977.

Bergson, Henri (1896). *Matière et mémoire*, Paris 1896.

Craik, Fergus I.M. and Tulving, Endel (1975). “Depth of processing and the retention of words in episodic meory”, *Journal of experimental psychology: general* vol. 104 (1975), p. 268-294.

Hansson, Bengt (1987). “Risk aversion as a problem of conjoint measurement”, in *Decision, probability and utility* (ed. by Peter Gärdenfors and Nils-Eric Sahlin), Cambridge 1987.

Hansson, Bengt (1999). “Are there fundamental limits to human knowledge. The importance of logic to epistemology”, forthcoming 1999.

Jacoby, Larry L. and Dallas, Mark (1981). “On the relationship between autobiographical memory and perceptual learning”, *Journal of experimental psychology: general* vol. 110 (1981), p. 306-340.

Johnson-Laird, Philip N. (1983). Mental models, Cambridge 1983.

Johnson-Laird, Philip N. and Steedman, Mark (1978). “The psychology of syllogisms”, *Cognitive psychology* vol. 10 (1978), p. 64-99.

Libet, Benjamin (1981). “The experimental evidence for subjective referral of a sensory experience backwards in time: reply to P.S. Churchland”, *Philosophy of science* vol 48 (1981), p. 182-197.

Loftus, Elizabeth F. (1979). *Eyewitness testimony*, Cambridge Mass. 1979.

Russell, Bertrand (1912). *The problems of philosophy*. London 1912.

Russell, Bertrand (1921). *The analysis of mind*. London 1921.

Tulving, Endel (1972). *Organization of memory* (ed. by Endel Tulving and Wayne Donaldson), New York 1972.

Tulving, Endel (1983). *Elements of episodic memory*, Oxford 1983.