Psychological Models of Concepts:
Introduction

James Hampton and Danièle Dubois

Hampton, J.A., & Dubois, D. (1993). Psychological models of concepts. *In* Van Mechelen et al, (Eds.) *Categories and concepts: Theoretical views and inductive data analysis (11-34).* London: Academic Press.

The first half of this two-part volume is concerned with our current understanding of how people conceptualise the world around them. In this introductory chapter, we shall present a brief outline of different approaches that have been proposed in answer to the fundamental questions: what are concepts, and how do people understand, represent and use them? The models presented here are mainly derived from cognitive psychology, although the chapter by Michalski offers a machine learning perspective, while that of Sutcliffe offers a more philosophical critique of current psychological work. Within this book, the general questions to be asked are: What is the place of concepts in current theories of knowledge representation, and what theory of concepts best accounts for how people understand the world around them?

The structure of the chapter will be as follows. A first section proposes definitions for some of the terminology presently found in the psychological literature on concepts. The second section then gives a restricted outline and foretaste of the major views on concepts to be presented in the following chapters. This overview leads to a concluding section in which we discuss the more general assumptions implicit in experimental psychological investigations of concepts and in which we point to possible problems arising from these assumptions. In particular, questions arise when the psychological approach is placed within the more general frame of other contemporary cognitive sciences such as linguistics, artificial intelligence (AI) and philosophy. The consequences of adopting these assumptions for the use of inductive data analysis in psychological concept research are briefly sketched. The discussion will lead us to argue that the variety of theoretical approaches enhances the heuristic value of inductive data analysis through placing theories of concepts on solid empirical grounds.

1. Defining terms

The problem in attempting any definition is of course that many different writers in the field have developed their own understanding and use of terms. In any domain where there are rival theories, there will be a degree of incommensurability between the terms of different theories that makes comparison and testing difficult. However, in the interests of clarity, we will attempt to provide some definitions here. These definitions are first approximations, and to some extent informal. They represent our best efforts to guess at a consensual view (a consensus that may be illusory). But where different disciplines use terms in importantly different ways, we will try to make this clear. In particular the reader should note that the chapters by Sutcliffe and Michalski provide their own clear definitions of terms, in accordance with common usage within philosophy and machine intelligence respectively, which will differ from those we propose here.

1.1 Concepts and Categories

Interest in concepts has its roots in the philosophical tradition. Within experimental psychology interest in concepts began in the early days of Behaviourism, with the notion of categories of stimuli that evoked similar responses (equivalence classes, or generalisation sets). With the advent of cognitive psychology in the 1950s, notions of concept and category were developed in relation to the processes of concept learning. In concept learning tasks a subject is presented with a predetermined classification rule within a limited domain and has to learn to categorise the stimuli correctly, either by inducing the rule, or by any other means available (Bruner, Goodnow & Austin, 1956, Shepherd, Hovland & Jenkins, 1961). Following Rosch's (1978) insights into the mismatch between this laboratory based approach and the psychological realities of natural categories, (particularly those associated with the meaning of common nouns in natural languages), a modern tradition of psychological research has developed in which concepts and categories are seen as central to theories of knowledge representation and long term memory. Our definitions of terms derive most closely from the latter tradition (Smith & Medin, 1981, Neisser, 1987), and deviate strongly from recent philosophical treatments (see for example Putnam, 1975, Rey, 1983, Sutcliffe, this volume).

We will take the word 'CONCEPT' to refer to the idea or notion by which an intelligence is able to understand some aspect of the world (see for example Murphy & Medin, 1985). The word 'CATEGORY' we will use to refer to a class or set of entities (they could for example be objects, actions, states, qualities ....) which are grouped together on the basis of some criterion or rule.

Let us illustrate these definitions with a concrete representative example. The concept of a CHAIR is defined here as that psychological state by means of which a person (or other intelligent agent) is able to understand that a particular object is, or may be considered as a type of chair. Understanding something as being of a particular type here means being able to make some connection with previous knowledge, from which plausible inferences can be made. If an object is taken as instantiating a particular concept (that is to say, someone believes some individual object to be a chair), then a range of plausible inferences can be drawn, subject of course to the contextual constraints in the situation. The object will be expected to fulfill the functions of a chair, to have the parts of a chair, assembled in the correct relation to each other so as to enable those functions, to have been constructed by some agent with the intention of fulfilling those functions, and so forth. The category of CHAIRS then refers to the set of entities in the world that may successfully be categorised as a chair in that the concept of CHAIR can be used to understand them. [1]

Most concepts will provide a way of categorising the world into those entities that instantiate the concept, and those that do not. Given such a categorisation rule, we can then speak of the *category* associated with any concept, as the class of entities that pass the categorisation rule. (Although, as Michalski points out in his chapter, having a concept need not mean having an explicit way of categorising the world. I may have the concept of a billion digit prime number, but have no practical means of differentiating such a number from others).

The question of whether categories or concepts are the more primitive notion -- in the sense that concepts can determine categorisation rules, or alternatively concepts can be inductively derived to fit the naturally occurring categories in the world -- is a central issue in theories of concepts, and underpins much of the discussion of the different views presented in this book.

The distinction we have drawn between concepts and categories is mirrored in the distinction between intension and extension; while concepts are concerned with the intensional aspects of a concept/category relation (the information that is used for classification and the possible inferences that classification allows), categories concern the extensional aspects, the application of a term to refer to entities. The members of a category (those entities that fit the classification rule) are its extension. Individually they are also referred to as exemplars or instances of the concept/category. Note that exemplars of a category can be either individuals or classes. Both "Fido" and poodles are exemplars of the category DOG, although there is a more restricted use of the term exemplar (as in "exemplar" models of concepts) by which the term applies only to individuals.

1.2 Typicality

Traditional treatments of intension and extension involve a strict logical association between the two aspects of a concept/category. The intension is composed of just those attributes that are true of all (extensional) category members, while the extension is composed of just those objects that possess all the concepts (intensional) attributes. The symmetry of this dual relation underlies many of the data analytic procedures described in this volume (see for example Guenoche and Van Mechelen, Chapter ?).

More recently, psychological theories of concepts have been developed in which this direct logical association is made more complex in order to account for the phenomena of concept and category gradedness or typicality. Where the exemplars of a category vary in how well they appear to fit the category, then they are said to vary in typicality. For example, Rosch (1975) pointed out that a ROBIN is a very typical bird, whereas a PENGUIN is an atypical bird. Other terms also used to refer to the dimension of exemplar typicality (which plays a large role in the prototype theory introduced by Rosch) are prototypicality, representativeness, and goodness-of-example. Typicality is primarily defined as an extensional phenomenon, since it relates to an ordering that can be placed on the members of a category. The notion of gradedness can also however be applied to the intensional attributes of concepts. Some attributes will be commonly considered to be more central to the definition of a concept than others (for instance having feathers is more typical of birds than is rapid flight). This gradedness of attributes within the intension has also been variously termed attribute typicality, centrality, definingness and importance. The challenge for psychological theories of concepts is then to provide a specification of the relation between intension and extension that can account for the variations in observed typicality of category members and concept attributes. (Logico-philosophical treatments of concepts may of course treat such phenomena as irrelevant to the proper logical description of concepts, see Sutcliffe, Chapter ?). A similar challenge is also posed by typicality phenomena for theorists developing methods of inductive data analysis.

1.3 Properties, Attributes, Values, Features and Frames

If we start from the assumption that people have concepts and that they use them to determine categories, the next question is how this process of mapping works. That is to say, what kinds of rule are used in categorisation [2]? There will be no single answer to the question of how categorisation rules are formulated. Different concepts may differ widely in the kinds of categorisation rule they provide. However it would appear that almost all categorisation rules will be based on some form of descriptive property information (exceptions would be categories defined as arbitrary finite lists like letters of the alphabet). For example discriminating the category of chairs involves a consideration of descriptions of particular objects and their relation to some criterial description that is used to classify the world into chairs and non-chairs. Some descriptions will involve simple perceptual characteristics, like FLAT or HARD. Others such as SUPPORTS WEIGHT, or EXPENSIVE will involve more complex characteristics that may be deeply embedded in some implicit "theory" held by the cognizer. For example the description EXPENSIVE requires for its understanding a notion of exchange value as defined in the relevant cultural context.

The descriptions that form a part of a concept, and some (or all) of which may be involved in the categorisation rule are often called properties. The generic term property refers to any predicate that can be asserted of some or all of the members of a category.

The terms attribute and value refer to a specific type of property. An attribute is a property that has a number of mutually exclusive alternative possibilities termed its values. Thus the property "is red" can be thought of as an attribute COLOUR which is given a value RED. Attributes and values allow for the fact that properties often form contrastive sets. The attributes form the dimensions or aspects on which entities in a domain may differ, while the values provide the alternative forms those aspects can take (see Barsalou & Hale's chapter for more about attributes). The term feature is also often encountered in the concept literature, particularly in earlier accounts (e.g. Smith, Shoben & Rips, 1974). The notion of a feature derives originally from structural linguistics, as for example in the phonological notion of 'distinctive feature' (Jacobson, 1963). As used in semantic theories a feature usually refers to an attribute that has just two values - present or absent, or marked and unmarked (Bierwisch, 1971). Features were typically used to decompose word meanings into their components in a way that explained semantic relations such as synonymy, hyponymy and antonymy. The semantic theory proposed by Katz and Fodor (1963) used the featural approach. However, rather than taking the structuralist definition of a feature (as providing a generally relevant semantic contrast) Katz and Fodor (and others) considered features to be semantic/cognitive primitives, contrasting with the behaviourist tradition of treating semantic features as "fractional mediated responses" (Osgood, 1966, Dubois, 1989).

Since features are generally taken to be binary whereas many semantic dimensions can have more than two values, the more general notion of attribute has replaced that of feature in more recent theories of componential semantics in the American tradition.

Further development of ways of representing intensional information comes from the development of Knowledge Representation techniques within artificial intelligence research. If a set of attributes is collected together and placed in some structural relation to each other, then the resulting data structure is called a frame (the notion of a frame owes a lot to the earlier notion of a schema). Frames go a stage beyond simple attribute lists. They can embody default values and connections between attributes. For example there may be a rule that given value A(1) for attribute A, then value G(3) will be expected for attribute G, unless information is given to the contrary. To use a concrete example, given the value UNRIPE for the attribute RIPENESS of the concept ORANGE, then the value GREEN will be expected for the attribute COLOUR, and the value SOUR will be expected for the attribute

TASTE. Frames can also embody constraints defined across the attributes; for example given that attribute P has value P(5), the range of allowable values for attribute Q may be restricted to some subset of its normal range. The chapter by Barsalou and Hale gives more examples of frame representations.

2. Differing views of concepts

In presenting the views collected in the following chapters we propose to contrast them in at least two major respects, before then discussing more general issues.

2.1 Realism and Psychologism

First there is the important question of the ontological status of concepts. Many psychologists (as exemplified in this volume by Barsalou and Hale, Hampton, Murphy) typically adopt a representationalist approach to defining concepts. Michalski's chapter which deals with concepts from an artificial intelligence perspective shares many of the same assumptions, although Michalski prefers a different terminology.

Most psychological or cognitivist approaches takes a position which assumes that concepts exist *a priori*, either in the "real" physical world, or in an ideal (platonistic) one, waiting, in each case, for our minds to discover them. The alternative (constructivist) view that concepts are conventional creations of human societies, and are therefore relative to particular cultures and historical contexts, is less common in cognitive science. Concepts are thought of as the building blocks of knowledge representation, constructed at the interface between raw perception and sensation (a bottom-up conception of learning) and *a priori* understanding (top-down interpretation). Few psychologists would argue explicitly for a purely empiricist or a purely rationalist origin of our concepts, yet there is still a range of positions between these extremes large enough to allow a wide variety of views amongst psychologists.

If it is assumed that our minds can "contain" knowledge of the world describable in some representational format, then the task of the theorist is to construct and empirically evaluate a model of how this knowledge is represented, and how this knowledge representation operates in associated tasks such as learning, language comprehension, reasoning and communication. Note that on this account, different individuals may possess different representations, and representations may change over time as knowledge is added or lost from the system. Thus when we speak of the concept of chair, we are referring to some abstract societal norm, something like "that which is more or less in common to most individuals' concepts of chair that enables them to agree on what the term chair means." It will sometimes happen that people will not agree on the meaning of terms, because they have fundamentally different concepts for understanding a domain. Political issues are a common source of such disagreement, and politicians constantly seek to define the debate in their own terms; indeed one particular branch of psychology founded by George Kelly, 1955, Personal Construct Theory, is built around the assumption that each individual has a personalised way of construing the environment.

The cognitivist view is in stark contrast to the realist/logicist view of concepts, which defines them exclusively in terms of logical conditions. The sole philosophical contribution in this section (Sutcliffe) presents the realist approach to concepts - a position which relates the "real" categories in the world to the "real" properties of objects that define them, while omitting any consideration of a cognitive psychological representation (and indeed arguing that the very notion of mental representation is incoherent and hence untenable). These are concepts whose existence is by definition independent of any intelligence to grasp them.

The realist position is most commonly taken in philosophy (see for example Rey, 1983, Putnam, 1975, or more recently Woodfield, 1991) and in linguistics (for a detailed discussion and critique see Lakoff, 1987). It is less common within cognitive psychology, although traces of the view can be found within Piaget's work on logical development, or even in Rosch's work (see for example the analysis of Rosch's "classicism" in Dubois; 1991, Pacherie; 1991; Rastier, 1991). The key difference is that the task facing the realist is to discover the nature of the true concepts that best describe the world, whereas the psychologist, it is argued, can only discover what people may *know* or *falsely believe* about a concept (the difference between ontology and epistemology). The concept itself is an idealisation that goes beyond any individual's understanding of it. Concepts, in the realist view, are not to be described at a psychological level but at some more abstract level and in terms of some ideal theory of the physical world (Putnam, 1975). It then makes sense to develop a theory of concepts, in which not everything that we may naively think of as a concept (such as the ideas associated with many words in our language) is actually a real concept. We may use terms with imprecise understanding of what they mean, and indeed it may be that they have no clear meaning in any realist sense. Such terms would be pseudo-concepts, and one tack for philosophy to take would be to attempt to find ways of differentiating real concepts from pseudo-concepts, as a way of clarifying both thought and language. Woodfield (1991) prefers the term "conception" for the psychologist's notion of concept, which he distinguishes from the proper notion of a concept, while Sutcliffe talks of P-concepts and L-concepts.

This latter view of concepts is unconcerned with psychological notions like intuitions of exemplar typicality, and the information associated with concepts in memory. Concepts are defined by clear sets of "objective" criteria, which allow a direct relation to be specified between the concept and its category (the logical link of intension and extension described above). The view also places greater emphasis on conceptual structure as opposed to semantic contents.

2.2 Concept definitions

The second way in which the approaches described in the following chapters differ is in the nature of the categorisation rule, or how one uses the concept intension to determine the extensional category of the concept. Here, Sutcliffe develops the so-called classical approach to the concept-category link, in which concepts are constituted as a set of necessary properties or conditions for category membership. Entities that satisfy every condition are included in the category, while all entities that fail to satisfy any condition are excluded. Concepts can then be placed into hierarchical taxonomies, where the properties that serve to distinguish different category members at one level are then used to form common element definitions of the subcategories at the next level down the hierarchy. Given the realist approach taken in his chapter, there are clear advantages to using this way of defining the relation between concepts and categorisation. Taxonomic structure permits a high number of deductive inferences to be drawn from the membership of any entity in a particular class. Concepts with classical definitions can be made use of by classical set logic, providing for deductive and syllogistic reasoning. Taxonomic structure also lends itself to algorithmic methods for analysis of conceptual structure, given object by attribute data matrices of the kind described in Part II of this book.

The advantages of the classical approach are challenged by those contributors to the volume who take their goal as the more psychological one of describing concepts as the components of thoughts - as having a mental as opposed to a real or ideal existence. The difficulty, as pointed out by Michalski, is that many concepts cannot be associated with clear classification rules. A concept like BEAUTY as applied to a particular work of art may be hard for us to make explicit even where we clearly feel that the object exemplifies the concept. Also there may be many borderline cases where an individual decision could depend on the context or on our mood. Many of the concepts that people use to understand the world appear to have graded application. Some examples are clearly good and typical category members, while others fit less well, or may even be dubious members. It is to deal with this phenomenon that the Prototype View was developed by Rosch (1975), following philosophical analysis by Ryle (1951) and Wittgenstein (1953).

The chapter by Hampton describes one version of the prototype view in detail. In doing so, he attempts to clarify some of the misunderstandings that have arisen about the approach, such as the common claim that prototype models do not provide necessary and sufficient conditions for category membership. He also describes data that appear to challenge the value of the classical view as a psychologically valid model of all concepts. When people make categorisation judgments, it appears that their judgments are on occasion inconsistent with the tenets of classical set logic. They may classify A as a type of B, and B as a type of C, and yet be unwilling to make the transitive inference that A is a type of C, where A, B and C are concepts labelled by common nouns in English. Similarly in judging the membership of conjunctive and disjunctive combinations of natural categories, people make category judgments that are inconsistent with the classical definition of conjunction (as set intersection) and disjunction (as set union). Hampton argues that a source of such effects is people's reliance on prototype-based concepts, and on forming conceptual combinations by combining intensional descriptions of concepts, rather than by combining their extensional sets.

One clear aim of the prototype approach is to account for the vagueness of many of our concepts, while retaining a precise account of how and where that vagueness arises. Michalski's chapter addresses the important question of vagueness and gradedness in category membership. Michalski's proposal is that conceptual knowledge should be divided into two components (hence the title of the model, the Two-Tiered approach). The Base Concept Representation (BCR) is a strongly idealised representation of the clearest examples and standard usage of a concept in the understanding of familiar paradigm cases. Vagueness and uncertainty can be avoided within the BCR, so that some stability can be given to our view of the world. The second component of a concept is a further set of Interpretation rules, which provide constraints on how the base representation can be distorted or changed in the light of contextual information. The model thus aims to provide the best of both worlds in giving the knowledge system both stability and flexibility. As long as we are in familiar territory, we are able to comprehend the world rapidly and automatically using the fixed elements of the BCR. When novel or unfamiliar situations arise then we have the additional meta-knowledge in the Interpretation rules, which guide the process of adapting our knowledge to the new situation.

This two-tiered model presents a way of bridging the divide between classical and prototype representations. The former have been criticised for being two inflexible, and for not capturing the context-dependence and potential vagueness of concepts. The latter are often criticised for not providing a firm foundation for important aspects of rationality such as logical reasoning (Osherson & Smith, 1981). By having a dual component system, Michalski aims to provide both aspects. At present the system has been tested within a machine learning environment, using the criterion of efficacy of learning to argue for its validity. Of course, if the model is also to be a model of human concepts, psychological data will have to be collected to test the model's predictions. Systems that provide the best engineering solution to a problem are not necessarily those that will be the best models for how the mind tackles the problem. In particular, people's ability to do abstract rule-based reasoning is notoriously poor (see Johnson-Laird, 1983, for a detailed treatment).

The chapters by Murphy and by Barsalou and Hale take the prototype view as a starting point from which to argue for greater complexity and representational power in the psychological description of concepts. Murphy reviews a series of arguments that point up the inadequacy of simple prototype representations as models of concepts. Our understanding of different domains of the world involves much more than the kind of clustering by similarity on which prototype theory relies. Naive theories of objects, their history, the way their structure relates to their function or behaviour, and their relation to other objects, all require a more complex representational system. A clear example of the need for a rich source of background knowledge within the conceptual system comes from Murphy's studies of conceptual combination in noun-noun and adjective-noun compounds in English. When nouns are qualified by adjectives, it frequently happens that attributes other than the one specified by the adjective are modified. For example BOILED CELERY is not only cooked, but is also no longer CRISP (Murphy, 1990). Knowledge of the domain is involved in understanding these kinds of complex concepts.

Barsalou and Hale make a similar point, but attempt a more detailed and explicit exposition of the failings of simplistic representational systems. Starting with simple feature lists (properties), they show the need to employ increasingly powerful representations in order to capture essential conceptual structure. First they argue for feature lists to be replaced by attribute-value structures, and then they propose that these attributes be incorporated in frames. Within frames the complexity of the representation becomes much greater: instead of a set of unrelated properties, frame representations introduce an extended variety of explicitly labelled relations between properties. The representation further allows for constraints between the values of different attributes to be defined (for example for a bird to fly, a constraint will apply to its size, weight and wing span). Finally they propose that a full representation of concepts as frames will need the recursive embedding of frames within other frames, as well as links to other types of knowledge representations such as scripts and scenarios (Barsalou & Sewell, 1985).

3. Views of concepts and Data Analytic Methods

In this third section, we consider relation between the data analytic methods discussed in the second half of this book and the different views of concepts we have outlined.

The views on concepts adopted by Murphy and by Barsalou and Hale are clearly heavily influenced by artificial intelligence techniques for modelling knowledge representation. These powerful representational systems pose a serious problem for the types of data analytic approach that form the second section of this book. Data analytic techniques have been developed independently from contemporary psychological theories of concepts, in order to account for a variety of types of data. Many data analytic systems require as their input an object by attribute matrix where the attributes will typically have just two possible values (True/False, or Present/Absent), and all attributes are applicable to all objects. As such, the methods must limit themselves to domains in which the conceptual structures can be constrained so that they can be reduced to such a matrix. This is closer to the assumptions made by the Classical and the Prototype views discussed by Sutcliffe and Hampton respectively.

In the case of applying data analytic techniques to the analysis of taxonomic or prototype structures, the techniques can be used to reveal conceptual structure[3] . One recent attempt to use data analytic techniques to investigate the featural structure of non-verbal concepts from their extensional representation within a similarity matrix is Barthelemy's analysis of Dubois' categorisation data for photographs of roads and landscapes. As the relevant features that are involved in the categorisation task are unknown in advance by the experimenter, the underlying featural model (based on Tversky's contrast model, Tversky, 1977) is a source of new hypotheses regarding the featural structure of the average graded categories derived from the data (see Barthelemy & Guenoche, 1987; Dubois, 1991; Dubois, Barthelemy & Tenin, 1992). As this example shows, data analysis methods can provide useful inductive insights by deriving intensional structure from extensional structure in such a way as to account for the graded "similarity" structure of the extensional categories.

However if we take data analytic methods as providing models of cognition, rather than as a source of potential insights into data structures, then further difficulties arise. First the binary coding of features is not only insufficient for representing the complexity of semantic information, but is also unconstrained in its meaning. A plus versus a minus can be interpreted as TRUE/FALSE, as PRESENT/ABSENT or even as the poles of a bipolar dimension such as LARGE/SMALL. Since many techniques do not treat a plus and a minus in a symmetrical way (that is to say that exchanging plusses and minuses in the matrix would not yield an equivalent structure), the implicit semantics of the coding is important. (See also Guenoche and Van Mechelen, Chapter ?.)

A second problem is that data analytic methods have been developed with the aim of providing "bottom up" or "data driven" methods for revealing the structure within a set of objects. The methods reveal the different ways in which sets of objects can be formed on the basis of their attribute profiles. However this inductive procedure for forming categories cannot be taken seriously as a psychological model of concept formation. Much of our cognitive processing involves 'top down' theory-driven processes. Our perception, identification and interpretation of an object or event depend crucially on general and particular theoretical beliefs that we possess. Such processes would be closer to deductive processes than inductive ones (see Murphy, this volume).

There may be situations in which we lack any a priori beliefs, and have to rely on the data to show us inductive generalisations that can be made. The data analytic methods will provide us with a valuable tool for directing our attention to fruitful ways of thinking about the data. What kinds of naive data analytic method we possess ourselves is another issue, dealt with in the literature on concept formation and induction. Rosch (1975) for example believed that people can detect statistical correlation amongst the properties of objects, and hence derive prototype representations as a form of summary representation of the pattern of co-occurrence of properties across objects (see chapter by Hampton). Others (Keil, 1989) have pointed out that there is very little chance of this kind of inductive method yielding the right conceptual structure without very heavy constraints being applied to the generation of hypotheses.

Part II of this volume will argue that inductive data methods still remain powerful heuristics for both psychological and mathematical research. As we will argue in the following section, the methods for the collection of psychological data on concepts are themselves fraught with problems that must be taken into consideration for the interface with data analytic methods to be successfully exploited.

4. Discussion of psychological views of concepts.

Each of the chapters in the section aims to provide a fair and positive exposition of a particular view. There is consequently little space within each chapter for a discussion of wider issues spanning all the views. In this final section of our introduction some of these wider issues, particularly in regard to the psychological models of concepts, will be discussed. Two major issues will be raised: first whether current methodology can be relied on to provide us with an unbiased description of the cognitive phenomena, and second, from a theoretical point of view, what ontological status current models give to concepts as psychological entities.

4.1 Empirical and methodological issues

The psychological approach to concepts as represented by all the chapters except those by Sutcliffe and Michalski, is based on an empirical methodology in which the data provided by individual subjects are taken as the explicanda, frequently after being averaged. The application of this type of empirical method to the field of concepts is not however without its problems.

4.1.1 The value of introspective evidence

One problem can be formulated as a version of the traditional criticism of introspectionist psychology -- that what people can tell you about their internal state is highly limited, subject to strong situational biases, and may be wildly inaccurate. As such, so the argument goes, one is trying to discover the answer to a scientific question by taking a majority vote from a random sample of informants. If a person is asked a vague, meaningless or ambiguous question then one can expect a vague or ambiguous answer. Since the goal of psychological science is to provide a characterisation of the workings of the mind, there is an imperative need to obtain different converging lines of evidence for any theory. Unfortunately, in the case of concepts the kind of evidence available is very limited. Psychologists can investigate intensions by interviewing people about their beliefs and thoughts concerning a concept, and by eliciting naive descriptions or lay definitions, and they can investigate extensions by asking for judgments about classification and category membership, and collecting incidental information about how well exemplars fit their categories, or how rapidly and accurately people can make the membership decision. However the amount of variability in both types of data means that linking intension to extension at the individual level is very difficult, and is unlikely to yield a reliable picture. The result is a reliance on averaged data, which leads to the second problem.

4.1.2 From group measures to individual concepts

Using group means can clarify the picture of a concept by reducing random noise, but then the status of the results is questionable in two ways. First, are we still describing the psychological phenomenon of an individual's concept, or might we be discovering a cultural and socially shared representation of it? Gradedness in a group mean may just reflect the variability with which individuals have appropriated the social norm, and structural properties such as extensional gradedness within a category may reflect distributions of different representations within the group, rather than individual structure. Some studies have addressed this issue (e.g. McCloskey & Glucksberg, 1978, showed that disagreement about category membership was paralleled by within-individual inconsistency in the membership decision, and McNamara & Sternberg, 1983, analysed concepts at the individual level), but the general worry remains.

4.1.3 Generalisability of results

Third, there is the question of how justified we are in taking studies of highly homogenous and restricted subject populations, such as college students, and generalising to universal aspects of human minds. In answer to this question, there are some studies that address the issue of universality, notably Rosch's own work with the Dani, (Heider, 1972). Indeed there is a whole area of psychology devoted to cross-cultural studies of cognition (see for example, Cole et al., 1971). To fully answer the question it would be necessary to address more empirical research to the concepts of different cultural groups both within a particular culture (for examples experts and novices) and across different cultures.

4.1.4 Concepts or word meanings

A further major cause for concern which goes well beyond the scope of this introduction is the fact that most techniques for studying concepts rely very heavily on language (see for a discussion Murphy, 1991). Words are used to stand for concepts, and it can be argued that the results reflect facts about word meanings rather than about concepts. Particularly when there are demonstrations of flexibility and context-dependence in categorisation, it should be remembered that flexibility can be equally well attributed to contextual effects on meaning. Understanding the relation between word, concept and object is fraught with difficulty and controversy. Words (with the exception of homonyms) can be individuated in terms of their phonetic and orthographic form. They then have meanings that allow them to stand for, or symbolise concepts in written or spoken communication[3]. The concepts are in turn related to mental representations either of entities in the world or of more abstract relations. By showing that the relation between word and object is subject to flexibility and vagueness, one does not thereby show that it is the concept itself that is changeable. A word may refer to its intended referent very precisely in the right discourse context, even though the word is only very loosely related in meaning (as normatively defined in, for example, a dictionary) to the object in question. For example, on hearing a student singing loudly in the corridor outside a classroom, the lecturer may ask a student at the back to go out and ask Pavarotti to practise somewhere else. The meaning of the term "Pavarotti" is sufficient in the context to identify the referent, although it does not have that specific meaning (cf Nunberg, 1979).

As applied to concept research, these issues of communicative language function are particularly important when subjects are asked to provide sets of ratings or categorisations in the absence of any explicit context or communicative goal. There is a need for researchers to take account of the flexibility of language use in their interpretation of categorisation results. To put it simply, the question is whether subjects in categorisation tasks are being asked to judge the objective truth of some proposition about a state of the world ("Is the set of objects called CHAIR entirely included in the set of objects called FURNITURE?") or whether they are being asked to judge the acceptability of certain language utterances in some supposed context ("Is it normally accepted within the conventions of English to assert that a chair is a type of furniture?"). Often it seems that subjects may be responding to the latter question, but theorists are taking them to be providing their beliefs about the first question (Hampton, 1982; Dubois, 1991). Cross-linguistic studies may be of critical importance here, since it may be expected that individual languages differ in the limits of acceptability allowed for different assertions, or even that lexical items in each language segment a particular semantic domain differently (a frequent problem for translators). Such comparative research, as well as the use of other non-linguistic modalities such as pictures (Dubois & Denis, 1988), may therefore be able to separate issues of conceptual content from issues of linguistic convention. In addition, such research could provide a way of identifying more precisely the relation of conceptual structure to lexical structure. To what extent is conceptual organisation language-free? As an illustrative example, the translation of Murphy's noun-noun compound examples EXPERT REPAIR and MOTORWAY REPAIR into French, would result in the phrases REPARATION EXPERTE, and REPARATION SUR AUTOROUTE. In French, unlike English, the semantic relation within the complex noun phrase must be marked in the surface structure of the phrase. For French speakers therefore, the problems posed by ambiguous compounds are greatly reduced. This simple example illustrates the need for caution in attributing the properties of language comprehension within a particular language to properties of the conceptual system per se.

4.1.5 The neglect of processing issues

A final critique of the methodology used in psychological investigations of concepts is that insufficient attention is paid to the actual judgment processes involved when subjects provide typicality ratings, lists of attribute values, or categorisation decisions. It is often assumed that the subject can give the experimenter a direct "read-out" from stored knowledge in memory in the manner of a memory retrieval model. The emphasis in these studies is on the structure of the knowledge representation: what attributes are present with what range of values? While there has been interest in the decision processes involved in making rapid instance-category verification decisions (i.e. "Is a robin a bird?"), there has been relatively little study of the processes involved in making typicality judgments, or in generating attributes for concepts. One thing that we do know about such processes, is that they are highly unstable in the results that they produce, leading to interesting questions about whether the instability lies in the process or in the database on which the process works. Barsalou (1987) has shown that typicality ratings or rankings within a category are liable to change quite radically from occasion to occasion, even within the same subject. Bellezza (1988) has shown similar instability in other tasks such as generating exemplars of categories or producing attributes for categories. The working hypothesis adopted by many psychologists in this area appears to be that such instability reflects randomness in the process of accessing semantic information. Hence if averages are taken across a number of occasions, or subjects, then the degree of randomness in the data can be reduced. It is likely that where subjects have to generate and recall information themselves (as in exemplar production or attribute listing tasks), the availability of retrieval cues and the general difficulty of memory retrieval will lead to instability. With rating tasks, such as categorisation itself and typicality judgments however, the instability in responding presents a more radical challenge. Barsalou (1987) interpreted his results as providing evidence that not only the procedures for accessing information, but the information itself was inherently unstable, so that conceptual categories and decisions about typicality and membership are often constructed 'on the fly', and are subject to uncontrolled contextual influences.

Given the critical importance attached to the question of vagueness and indeterminacy in categorisation (see for example Hampton's chapter on prototype theory), it is very important to discover the source of this instability. Is it in the structural level of the representations themselves, in the processes that interrogate that level to retrieve information, or in decision processes operating on the retrieved information? At present, this question is unresolved.

4.2 Final theoretical issues

Each of the psychological models makes the common assumption that the semantic content of a concept can be specified in terms of some set of attributes or features. The main differences amongst the approaches concern just how those attributes define the concept, and how powerful the representational tools have to be to capture the semantic content of our conceptual system. One can however question the general aim of this process of conceptual analysis. In particular, it sometimes appears that there is a danger of circularity in the tendency to explain one concept in terms of another (Fodor, Garrett, Walker & Parkes, 1980). For example if I define CHAIR in terms of its SEAT, BACK and LEGS, then I have replaced one problem (what is a chair?) with three new ones (what are seats, backs and legs?) As psychologists have long been aware (Rosch, 1978, Smith & Medin, 1981), the process of decomposing concepts into attributes is only sensible if those attributes can be grounded in some more primitive set of representational elements. In terms of the psychology of concepts, it is commonly supposed that these elements would presumably refer to 'low-level' perceptual and behavioural components, mainly determined at a sub symbolic perceptual level, and constrained by physiological structure.

But the danger of circularity also relates to the relation between concepts and beliefs. One of the important functions for concepts is to provide the building blocks for the construction of thoughts of different kinds. Thus if I have the thought 'the cat is sitting on the mat' I can have this thought by virtue of having concepts for the terms CAT SIT and MAT. However, we then find psychologists explaining what it is to have the concept of a CAT in terms of the set of beliefs that can be held about cats (cats have four legs, cats purr, cats are mammals, cats sit on mats and so on). One way to escape this higher level circularity is to look for external grounding of concepts in other psychological entities. Possible candidates for grounding of concepts would be platonic concepts -- ideals provided a priori by the structure of the nervous system -- or in culturally transmitted ideas, in particular those that are embodied within a particular language, which may provide top-down input of which concepts should be taken as basic.[4]

Psychological models of concepts thus try to ground concepts in some cognitive representational system. Realist approaches on the other hand ground concepts in the real world. The world happens to display a taxonomic structure of classes with associated intensions. Our concepts can then be grounded extensionally, perhaps through ostensive kinds of definition -- "Chairs are the class of all objects that share the same essence as THAT" (said pointing to a chair). In fairness to the cognitive view, we can also question the status of the philosophically realist position on concepts. What is the status of the analysis of concepts offered by the theoretician? If the analysis is not grounded in empirical data taken from language users, then there is little constraint for choosing one analysis of a concept from another. Any conjunction of intensional attributes can be taken to define a class, and the world can be divided into taxonomic structures in indefinitely many ways. There is nothing in the realist position to prevent the construction of arbitrarily absurd conceptual structures, provided that they are logically consistent. (This problem derives from the reliance on logic, which explicitly eschews semantic content.)

To justify a realist conceptual analysis one could choose to defend the conceptual structure in terms of the part that it plays in true (or as yet undisproved) scientific theories of the world, thus indirectly appealing to empirical data. Concepts would then be grounded in scientific theory. For example the modern scientific concept of heat as random kinetic energy at the molecular level provides a more useful conception than did the earlier theory of phlogiston. Within the realm of scientific concepts, the scientist has to refine and develop her terms in order to provide better understanding and more accurate predictions of experimental and observational outcomes. If this type of concept is taken as the central concern of the psychology of concepts (as some psychologists would argue (Armstrong, Gleitman & Gleitman, 1983)) then the classical view of concepts appears to be better suited to model them. Polythetically defined concepts are not generally found in the physical sciences.

However if we widen the conception of contemporary paradigmatic science even so far as the social sciences, like psychology or economics, we find that theoretical understanding relies heavily on concepts that simply do not have clear classical definitions. Psychiatric diagnosis for example (as embodied in the Diagnostic and Statistical Manual III) is now officially acknowledged to rely in part on the partial matching of patients to sets of symptoms, very like a prototype theory view (although diagnostic classification rules are often more complex than simple sums of symptoms). A realist might answer that the reasons for this state of affairs lie in our partial understanding of such matters, and that given a few more centuries of progress in the field, the common root causes of mental illness will be uncovered, and the classical approach to defining disease categories will be once more appropriate. However, in the meantime, if one takes a less optimistic view of scientific progress, one could also argue that there is very little evidence to suppose that expert concepts used in psychiatric diagnosis are radically different from other classificatory concepts that we use in everyday life such as CHAIR or RABBIT (see for example, Dubois et al, 1992, for an experimental analysis of expert knowledge). As Chater and Oaksford (1990) have argued, the status of such concepts may in principle be considered as no different from that of concepts like PHLOGISTON which have been discredited as scientific terms (and who knows what other current scientific concepts will follow?). Since there is no clear definition for the terms, and almost any true statement that can be made using the terms can be shown to be defeasible (that is exceptional circumstances can be imagined in which the statement is no longer true) Chater and Oaksford argue that our common-sense ontology will never provide a proper basis for a correct understanding of the world. The philosophical realist endeavour to discover the nature of concepts would do well therefore to distance itself from the common sense world and everyday language terms. Take the concept CHAIR for example. Following Wittgenstein's argument, any universally quantified statement about chairs that one cares to make can be shown to be false under exceptional circumstances. It follows that no universally true facts can be known about chairs. There is no real concept of CHAIR, since the conception of CHAIR can play no part in any true theory of the world, any more than the notion of PHLOGISTON has a part to play in the physics of heat.

Having said this, there must also be the real possibility that there are structures in the world that simply do not show the taxonomic pattern of classes assumed in the classical model, and yet do have a theoretical role to play in explaining and understanding the world. For example the biological notion of species no longer has a classical definition, (Mayr, 1984) and yet it nonetheless provides an extremely useful level of classification for biology. If we choose to drop all use of terms that have apparently polythetic definitions as being unscientific and likely to mislead our theorising, then we may often be left with no theory of the domain at all.

References

Armstrong S.L., Gleitman, L.R., & Gleitman, H. (1983). What some concepts might not be. Cognition, 13, 263-308.

Barsalou, L.W. (1987) The instability of graded structure: implications for the nature of concepts. Chapter in U.Neisser (Ed.), Concepts and conceptual development: Ecological and intellectual bases of categories. Cambridge: Cambridge University Press.

Barsalou, L.W., & Sewell, D. (1985). Contrasting the representation of scripts and categories. Journal of Memory and Language, 24, 646-665.

Barthelemy, J.P., & Guenoche, A. (1988) Les arbres et les representations de proximites, Paris: Masson.

Barthelemy, J.P. (1991) Similitude, arbres et typicalite, in D.Dubois (Ed.), Semantique et cognition: Categories, concepts et typicalite. Paris: Editions du CNRS.

Bellezza (1988) Bulletin of the Psychonomic Society.

Bierwisch, M. (1971). On classifying semantic features. In L. Jakobowitz, & D.Steinberg (Eds.) Semantics. Cambridge: Cambridge University Press.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). A study of thinking. New York: Wiley.

Chater, N. & Oaksford, M. (1990) Logicist Cognitive Science and the Falsity of Common-Sense Theories. Unpublished MS, University of Edinburgh.

Cole, M., Gay, J., Glick, J.A., & Sharp, D.W. (1971) The cultural context of learning and thinking. New York: Basic Books.

Dubois, D. (1989) Contribution de la psychologie aux sciences du language. In Histoire, Epistemologie, Langage: Sciences du langage et recherches cognitives, 11, 85-104.

Dubois, D. (1991) Catégorisation et cognition "10 ans après: une évaluation des concepts de Rosch. In D.Dubois (Ed.): Sémantique et cognition: Catégories, prototypes, typicalité, Paris: Editions du CNRS.

Dubois, D., Barthelemy, J.P., & Tenin, A. (1992) Tree represenation of similarity and categories of situations from photographs and pictures. Manuscript under review.

Dubois, D., Bourgine, R. & Resche-Rigon, P. (1992) Connaissances et expertises finalisées de divers acteurs économiques dans la categorisation d'un objet perceptif, Intellectica. (in press)

Dubois, D., & Denis, M. (1988). Knowledge organization and instantiation of general terms in sentence comprehension. Journal of Experimental Psychology: Learning Memory and Cognition, 14, 604-611.

Fodor, J.A., Garrett, M.F., Walker, E.C.T., & Parkes, C.H. (1980) Against definitions. Cognition, 8, 263-367.

Hampton, J.A. (1982) A demonstration of intransitivity in natural concepts. Cognition, 12, 151-164.

Heider, E.R. (1972) Universals in color naming and memory. Journal of Experimental Psychology, 93, 10-20.

Jacobson, R. (1963) Essais de linguistique génerale. Paris: Editions de Minuit.

Johnson-Laird, P.N. (1983). Mental Models. Cambridge: Cambridge University Press.

Katz, J.J. & Fodor, J.A. (1963). The structure of a semantic theory. Language, 39, 170-210.

Keil, F.C. (1989) Concepts, Kinds, and Cognitive Development. Cambridge, MA: MIT Press.

Kelly, G. (1955) The Psychology of Personal Constructs. New York: Norton.

Lakoff, G. (1987). Women, Fire and Dangerous Things. Chicago: University of Chicago Press.

Mayr, E. (1984) Species Concepts and Their Applications. In E.Sober (Ed.): Conceptual Issues in Evolutionary Biology. pp. 646-662. Cambridge, Mass: MIT Press.

McCloskey, M., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? Memory and Cognition, 6, 462-472.

McNamara, T.P., & Sternberg, R.J. (1983). Mental models of word meaning. Journal of Verbal Learning and Verbal Behavior, 22, 449-474.

Murphy, G.L. (1990) Noun phrase interpretation and conceptual combination. Journal of Memory and Language, 29, 259-288.

Murphy, G.L. (1991) Meanings and Concepts. In P.J. Schwanenflugel (Ed.): The Psychology of Word Meanings. pp 11-36. Hillsdale NJ: Erlbaum.

Murphy, G.L., & Medin, D.L. (1985). The role of theories in conceptual coherence. Psychological Review, 92, 289-316.

Neisser, U. (1987) Concepts and conceptual development. Cambridge: Cambridge University Press.

Nunberg, G. (1979) The non-uniqueness of semantic solutions: Polysemy. Linguistics and Philosophy, 3, 143-184.

Osgood, C.E. (1966) Meaning cannot be Rm? Journal of Verbal Learning and Verbal Behavior, 5, 402-407.

Osherson, D.N., & Smith, E.E. (1981) On the adequacy of prototype theory as a theory of concepts. Cognition, 11, 35-58.

Pacherie, E. (1991) Aristote et Rosch: un air de famille? In D.Dubois (Ed.): Semantique et cognition: Categories concepts et typicalite, Paris: Editions du CNRS.

Putnam, H. (1975) The meaning of 'meaning'. In K. Gunderson (ed.) Language, Mind and Knowledge. Minnesota Studies in the Philosophy of Science, Vol.7. Minneapolis: University of Minnesota Press.

Rastier, F. (1987) Sémantique interpretative, Paris: P.U.F.

Rastier, F. (1991) Categorisation, typicalite et lexicologie, In D.Dubois (Ed.): Sémantique et cognition: Catégories, prototypes, typicalité, Paris: Editions du CNRS.

Rey, G. (1983). Concepts and Stereotypes. Cognition, 15, 237-262.

Rosch, E. (1975). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-232.

Rosch, E. (1978) Principles of categorization. In E. Rosch & B. Lloyd (Eds.) Cognition and Categorization. Hillsdale NJ: Erlbaum.

Ryle, G. (1951) Polymorphous Concepts. Proceedings of the Aristotelian Society (supplementary series), 25, 63-65.

Scholnick, E. (1983) New trends in conceptual representation: challenges to Piaget's theory? Hillsdale N.J.: Erlbaum.

Shepherd, R.N., Hovland, C.I., & Jenkins, H.M. (1961). Learning and memorization of classifications. Psychological Monographs, 75, no. 517.

Smith, E.E., & Medin, D.L. (1981). Categories and Concepts. Cambridge MA: Harvard University Press.

Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and process in semantic memory: A feature model for semantic decisions. Psychological Review, 81, 214-241.

Wittgenstein, L. (1953). Philosophical Investigations. New York: MacMillan.

Woodfield, A. (1991) Conceptions. Mind, 100, 547-572.

Footnotes

[1] There remains an ambiguity as to whether Category as we define it is to be understood as a set of real world objects, or a set of mentally represented objects - either recallable from memory, or potentially perceivable. Unfortunately a discussion of this question would take us well beyond the scope of this introductory chapter. We will be using the term primarily as referring to a class of real world objects.

[2] The process of how an individual actually decides on a categorization in any particular situation may well differ in many respects from the 'idealization' of the rule that best maps intensions to extensions. This issue is raised later.

[3] The methods are perhaps not purely inductive, since the data analyst must specify the objects and attributes to be used before the technique can be applied.

[3] At this point classical semantic theories try to relate the meaning of words directly to sets of referents in the world - missing out any cognitive element of a mental representation. More recently however the discipline of Cognitive Linguistics has attempted to incorporate psychological perspectives in theoretical semantics (see for example Jackendoff, 1983, Lakoff, 1987, Langacker, 1986)

[4] The same criticism also applies to the question of the ontological status of the conceptual structures identified in psychological models. Should they be considered as depicting the actual contents of the mind, or are they more abstract theoretical entities, employed by the psychologist as a way of accounting for the data?