Conceptions of Concepts

James A. Hampton
City University London

Hampton, J.A. (1999). Conceptions of concepts. *In* B.Kokinov (Ed.) *Perspectives in Cognitive Science, Vol 4,* pp.27-38. Sofia: New Bulgarian University Press.

Abstract
Anyone coming fresh to the literature would quickly observe that philosophers and psychologists have very different notions underlying their use of the word "concept". It is generally agreed by both that concepts are the building blocks or atoms from which thoughts are created. However the way in which concepts (and hence thoughts) can be described and investigated has become a source of conflict between psychology and philosophy. In this paper, I characterise and contrast two approaches that I term the Internalist (or Psychological) and the Externalist approaches to defining the notion of concept. I will seek to show that many recent criticisms of internalist theories have been based on misconceptions, and I will argue that a theory of concepts must in fact be grounded in psychological data.

Concepts: Inside or outside the head?

1. The inside and the outside worlds

A naive common-sense view of how thoughts relate to the world is to imagine a boundary or interface between the mind and the world. To one side of the boundary is a thinker. To the other side of the line is the external world. The thinker has thoughts, and we can safely suppose that many of these thoughts are about something. Moreover, what many of them are about is the external world. So how do we specify exactly what in the external world these internal thoughts are about? How is the interface "connected"? What exactly is the relation between mental contents and the world they are about?

To take the common sense view further, let us suppose that there is a (more or less) direct mapping between the two sides of the divide. When a thinker entertains the thought "All cats have tails", then this thought corresponds to or maps onto a corresponding state of the external world in which there exists a set of objects corresponding to the category of cats, and there exists a definable property of having a tail. The thought itself corresponds to the proposition that all members of the cat category have the property of having a tail. The thought will be true or correct if (and only if) the external world is such that:

(1) For every individual in some broader domain of reference, if that individual is a cat then that individual has a tail.

Hence there is a simple correspondence between the correctness of a belief that someone may hold (an internal state of affairs) and the truth of the logical proposition that is the content of that belief (in the external world). Furthermore the proposition itself can be represented in terms of the language of first order set logic, whose axioms and foundations are well understood.

As one might expect, the Internalist and the Externalist views seek to anchor the link of thoughts to the world to opposite ends of this mapping. Hence when providing an account of the compositionality of thoughts in terms of component concepts, they arrive at very different notions of the term "concept". These notions depart from the common sense correspondence view either by anchoring concepts to the external world, or alternatively by emphasising the thinker's internal representations. I shall take each in turn.

1.1 Externalism

The Externalist view starts with the world, and individuates concepts by reference to a metaphysical reality. There is a real world of which certain propositions can be truly asserted. But these propositions in turn must be frameable in terms of some elementary terms, and these terms are concepts. Concepts are thus (crudely) those ways of identifying real world entities which are needed to support all the true (and false) propositions that can be stated. Together they provide a full vocabulary for describing facts or beliefs about the way the world is, and the way that it is not. The classification of the world into categories and types is seen therefore as prior to the beliefs that people may have. It is claimed (e.g. Sutcliffe, 1993) that it is only possible to speak sensibly about a particular individual's thoughts being about some concept class, given that the concept class is determined independently of the thinker. Having an independent determination of the class allows us to test in an empirical way whether the person's thought corresponds correctly to that class (that is, whether their thought is to be taken as referring to the same class as the actual real world class). Only if there is an externally real reference class to compare it to, is it possible for my own concept to be a correct or incorrect representation of the world.

The contents of people's belief states, and their ability to classify and name objects in the world can then be investigated as a study of the degree of understanding or knowledge displayed by that thinker. Similarly, the meaning of words in a particular language, such as the word “cat” in English, can be empirically investigated by linguists who can attempt to discover how the word’s meaning in either common or specialised usage maps onto the world.

The research goal for psychology according to the Externalist program should be to provide an account of how the thinker manages to entertain thoughts about these external concept classes. When the thinker holds some belief about cats, then this belief is about the real world concept of cats. So the thinker who thinks about cats must actually be thinking about the real class of cats. We can tell what her thought is about, because we have independent reasons for believing that there is such a concept class. Hence it is possible for her thoughts or beliefs about cats to be wrong, fanciful or unfounded, and yet *still to be about cats*. A similar research program can be outlined for lexical semanticists, who need to determine facts about how speakers may use words to convey their meanings. Most semantic theory is based on the premise that there exist real-world categories onto which the reference of a word, phrase or sentence may be mapped.

The Externalist view of concepts, (as expressed for example by Rey, 1983, 1985, Margolis, 1994 and Sutcliffe, 1993, and more recently by Millikan, 1997, and Fodor, 1998) is often associated with a further assumption -- Essentialism. Any individual cat is a cat because it has the essential properties of cats (whatever they may be) -- a view that was also proposed by Medin and Ortony (1989) as characterizing people’s meta-beliefs about concepts. Of course, given that classes are defined externally, and without reference to the psychology of the thinker, this assumption of an essence may turn out to be either false or circular. If by essence we mean that the class can be defined in terms of common necessary properties - what Smith and Medin (1981) termed the Classical View of concept definition - then we can turn to physical science to determine whether that is the case. In the case of animal and plant species it is almost certainly not (Atran, 1997), and so essentialism turns out to be false. If on the other hand an essence is just the property of belonging to the class, then the notion is clearly potentially circular. Some philosophers, most notably Fodor (1998), have embraced this circularity. In Fodor’s most recent formulation, most conceptual classes are not constituted or defined in terms of any beliefs about them, essentialist or otherwise. Conceptual classes are just those categories of existence to which our minds become linked through exposure to prototypical instances of them. More precisely, "the property that a concept expresses is the one that minds like ours would lock to given experiences the intentional content of which is the stereotype [for that concept]." For a discussion of Fodor’s theory, and a robust defense by Fodor, see the Multiple Review of his book in *Mind and Language (2000).*

Finally, an important corollary of Externalism is that these essential properties need not be part of a thinker's mental representation of the concept class. The inability of respondents to provide clear explicit accounts of the criteria for category membership for any concept is not therefore evidence against concepts having essences or these essences being defined according to the Classical View. Nor is the vagueness or fuzziness of people’s categorization, since it is clearly possible for a thinker’s mental representation to be incorrect or partial in important respects.

For an Externalist psychology, the criterion for someone grasping a concept must therefore include a willingness on the part of the thinker to acknowledge that any belief that he or she may entertain about the class could turn out to be wrong. That is, they should show a willingness to defer to the External world and the possibility of discovery about it.

The thinker who possesses a concept of CAT has signed up to a commitment of the following kind:

> *(2) When I speak of cats, I intend to refer to the real class in the world that has as its essence the essence of catness. I may not really know what catness consists of, but I believe that there is some such essence that could in principle be discovered.*

Alternatively, someone who did not subscribe to essences could still hold the following externalist view:

> *(3) When I speak of cats, I intend to refer to the real class in the world that includes cats and no other creatures. I may not really know what catness consists of, and I may be unable to correctly identify members of the class in all circumstances, but I believe that there is some such criterion for generating correct classifications which could in principle be discovered.*

From this commitment follows Putnam's famous dictum that "meanings ain't in the head" (Putnam, 1975), and the argument that psychologists who talk of concepts as if they were individuated by their psychological content (i.e. the beliefs that we hold about them) are suffering from confused thinking. As Rey (1985) put it:

> "a) The representation of a concept is a summary description of the extension of the concept that includes a distinction between essential, defining conditions of that extension, and non-essential information that may form a 'conception' of some, all, or 'typical' members of the extension [BINARY HYPOTHESIS]
>
> b) These defining conditions are provided by the optimal account of the concept, and so may not be known to competent users of it, whose representations of those defining conditions would be marked by empty slots. [EXTERNAL DEFINITIONS]"

Wrapped up with this thesis is a strong distaste for the idea of vague or fuzzy concepts with unclear boundaries. Where there is vagueness in concepts, this should be treated as an aspect of human thought, rather than as something that one would wish to see in a true conceptual account of the world. The argument proposes that the real world is a rule-governed affair, of which true facts can be stated. What we have been led to believe by science about this world is that it can be described with logical and mathematical precision. There should be no place in it for confused or vague propositions.

This distaste for vagueness is clearly a reaction to the danger posed by fuzziness to the logical foundations of semantic theories (for a discussion of the difficulties that are caused by vagueness, and in particular the sorites paradox, see Keefe and Smith, 1997, Williamson, 1994). For example Frege (quoted by Sutcliffe, 1993) ruled out fuzziness in concepts by stipulation:

> *"A definition of a concept (of a possible predicate) must be complete; it must unambiguously determine, as regards any object, whether or not it falls under the concept (whether or not the predicate can be truly asserted of it). Thus there must not be any object as regards which the definition leaves in doubt whether it falls under the concept; though for us men, with our defective knowledge, the question may not always be decidable. A concept that is not sharply defined is wrongly termed a concept."*

Frege clearly needed to restrict concepts in this way in order to develop the mathematics of two-valued logic. The stipulation that concepts that are not sharply defined are not true concepts leaves open of course the important question of whether much of interest can be said about everyday human thought given this severe restriction.

While Externalism is clearly in sympathy with this view, classical definitions are not central to Externalism. One could still have a division into essential and non-essential information (Binary Hypothesis), and deference to an external definition of the essence (External Definitions), without presuming that the essence is itself a clear-cut concept, or one that can be defined in simple logical terms (see for example, Osherson and Smith, 1997, p204). Indeed the view propounded by

Williamson (1994) is that the real world determination of the membership of a class may be so complex and context dependent that we are prevented from ever discovering its true nature. According to this view the vagueness of our language reflects our limited epistemological powers to discover the correct rule for determining category membership.

In summary the Externalist view has two central tenets, and two optional additional tenets. The central tenets are:

a) Externalism - concepts are individuated externally and independently of minds, by the "true state of affairs" - the metaphysically correct way of dividing up reality in order to express true and consistent propositions, testable hypotheses, and so forth.

b) Deference - people's mental representation of concepts commonly defer to the external "correct" definitions of the concepts. Any of our mental "concepts" may be defeasible.

For some versions of externalism either (c) alone, or both (c) and (d) below are also held to be true:

c) Essentialism - concepts have essential "cores" that determine their true underlying nature and determine categorization. These cores can be distinguished from non-essential attributes, which may also be true of many or even all category members.

d) Classical definitions - concepts have essences that define category membership in terms of a conjunction of singly necessary and jointly sufficient criteria.

1.2 Internalism: Mental representations as concepts

The Internalist approach adopted by many psychologists starts at the other end of the thought-world relation. Psychologists are concerned with the psychological contents of people's minds - the information which they have in their heads that provides them with knowledge and understanding of the way the world is, thus enabling them to meet their goals, act effectively, and communicate with others. These contents can be investigated (among other things) by asking people to tell you their beliefs about the concept class. We can examine how these beliefs are formed, what kinds of situation might lead a person to change them, what deductive and inductive inferences they will support, and how people actually classify the world with respect to the concept term.

The thinker's concepts are individuated using behavioral criteria, such as whether they will judge that it is appropriate to apply a concept term to describe an object. In the case of non-verbal or pre-verbal thinkers like animals or small children, many other possible ways in which the thinker's behavior picks out a certain type of object, event or situation have been used for the exploration of conceptual understanding. Examples are sequential touching of objects in the same class, release from habituation on moving from one class to another, and the ability to learn classifications through instrumental conditioning and discrimination learning.

To study a concept such as CAT, a psychologist might start with the empirical observation that adults sampled at random from a common linguistic community of speakers of English use their natural language term "cat" in certain common ways. They can be asked about its reference, and they can be asked to generate information in the form of generalisations that they believe to be more or less true of the class. Performance in speeded classification of words or pictures, inductive and deductive inference and a host of other experimental procedures can be used to test alternative models of what information is represented, how it is represented, and how it is processed in information retrieval and decision making tasks.

How does the psychologist provide an account of what these "concepts" are concepts of? How is the link to the external world established? Of course this question has rarely been considered by many researchers, (which is why philosophers have been at pains to point out the difficulty). The experimenter is already typically a speaker of the same language, and so assumes that the participant in the experiment understands the term in the same way.

If one removes this privilege of sharing a common language, then an answer is still however possible. By observing and testing behavior (in its broadest sense) in different situations, a psychologist, like an ethnographer trying to understand an unfamiliar society, can come to a reasonable hypothesis about what aspects of externality are reflected in the thinker's conception.

As a scientist standing outside the World-Thought relation it is possible to observe and record the way in which behavior (like naming) relates to the world, and so hypothesize about the link between the two - just as an ethologist studying the instinctive behavior of birds might look at what kind of stimulus evokes a particular behavior. The scientist still needs a vocabulary for describing the corresponding states of the world, and so there will always be some inherent circularity if this is done using natural language (rather than technical objective measurement), but this need not render the analysis trivial. For example in the case of the psychophysics of perception we can turn to the physicist to provide a physical description of the stimulus arriving at our sense organs, and then relate the subject's psychological impressions and ability to discriminate and classify to these physical parameters.

Human thinkers obviously provide a much more complex range of behaviors than an ethologist would have to study. Furthermore our own self-knowledge places us in a uniquely privileged position as regards understanding the behavior of others. We can use introspection to provide us with clues about the concepts of another human, in a way that is impossible across species (Chater & Heyes, 1994). And yet even in the case of animal cognition we are not totally in the dark, since we can analyse behavioral patterns in terms of their function for the creature in the context of an evolutionary account.

Another part of the psychological research program that links thoughts to the world is the investigation of how the real world influences the psychological world of the thinker – both in the cultural evolution of conceptions and in the course of cognitive development. Thinker's conceptions are moulded in many ways. First, our understanding of the world is dependent on our native endowment of perceptual and cognitive apparatus, itself moulded through evolution to be adapted to relevant aspects of the external world. We do not start with a *tabula rasa*, nor a “buzzing, blooming confusion”. Second, the real world that we all inhabit has a strong influence in driving learning and adaptation in both perception and action. We are all born into a world with adults and children, sky and ground, food and drink. Third, being both social and verbal creatures, our developing conceptions are strongly influenced through the learning of the correct use of language terms, and through cultural and educational transmission of information and ideas.

What psychologists have discovered, tackling the problem from the inside out, could be considered a rather sorry tale. Just as when the reasoning powers of adults have been tested experimentally, even those of highly intelligent students turn out to be sadly lacking in many ways, (Evans & Over, 1996), so the way that people categorize and conceptualise the world shows up inconsistency, vagueness and uncertainty. This should be no surprise. Anyone who has read Plato’s Socratic dialogues will sympathize with the muddled conceptions of everyday language and thinking. Any teacher knows the variety of misconceptions that are possible about a subject, and the degree of muddle and incoherence that inhabit the typical student’s mind. The history of science clearly demonstrates what a long hard road it has been to arrive at a clear and consistent conceptual model of the physical world.

What most people have as conceptions of the world can best be described (so the psychologist's story goes) as <u>incomplete</u> or <u>partial</u> representations of classes. Prototypes, if properly extended and elaborated to increase their formal power to represent all relevant information in a structured way (Barsalou & Hale, 1993), serve very well to capture the sometimes muddled concepts that play a role in a thinker's thoughts. As long as the thinker is on familiar territory, dealing with typical situations that fit their concepts reasonably well, then the system runs along just fine. Thoughts about cats in actual situations pick out just the right set of things in the world, and the token "cat" in our language of thought serves the right purpose of supporting the expression of beliefs about those things. However if faced with a counterfactual situation in which the cluster of common attributes is broken up, the thinker is unable to draw any firm conclusions. It is not necessarily just that our concepts are vague (although they may be), it is also that their application to the world is underdetermined.

As a brief illustration, consider the concept of Bird (a favorite amongst concept theorists). The class of birds that happens to have a classical definition, even for naïve users of the term. Birds are the only creatures to have feathers, two legs and a beak, and to hatch from eggs.

Moreover all birds have all of these features. Our representation of the category has these as prototypical features, along with other less necessary attributes such as flying, nest building and so forth. But imagine a counterfactual case in which a Lost World is discovered in which the common defining features are not always found together. Would a creature with feathers, two legs, hatching from eggs, flying and nest building, but with a bony jaw with teeth be considered a bird or not? Or consider another creature that had evolved its forelimbs from wings into a pair of arms with hands and fingers, but was otherwise bird-like? Should this be counted as a bird? The Internalist theory argues that we cannot answer such questions except by some arbitrary decision, since there is nothing in our conceptual representation that allows us to do so. The point at which an atypical instance falls outside of a class and must become a class on its own is not clearly defined. Of course, the arbitrariness of the decision does not mean that it is unimportant. However the reasons for preferring one answer to another will have to do with the practical value of the classification system for various purposes.

1.3 What is so wrong with the psychology of concepts?

One may suspect that the philosophical critique of the psychology of concepts is partly a territorial response - the feeling that psychology is trespassing on sacred ground. Rey (1983), for example, stated that the philosopher's is the only serious way to define "concept", and proposed that psychologists should restrict themselves to the study of “conceptions”. If a major goal of philosophy is to arrive at ever better ways of understanding the world through the refinement of concepts, then a concept in this sense (it could reasonably be claimed) will not be illuminated by studying the confused and naive attempts of a group of untutored students asked to spend five minutes reflecting about their beliefs or their use of words!

If the debate is simply fuelled by a territorial dispute over the use of the word "concept" then it might be possible for both sides to agree to use different terms (Rey's “concept” and “conception” for example), or simply to acknowledge that the use of the term is different in the two disciplines (as must occur often enough in the borders between other disciplines). Enlightened psychologists might agree to only speak of concepts behind closed doors, or to acknowledge in footnotes that "concept" is being used as a shorthand for "mental conception".

There is however a strong reluctance from some philosophers to allow even this, since the very thesis they are defending is that “concept” itself like other concepts has an external real meaning – it is in effect a natural kind. It follows that psychologists who use the term to mean conception are using it not just differently but incorrectly, just as if I insisted on referring to whales as fish, in the face of received opinion to the contrary. More importantly, by not acknowledging the distinction between concept and conception, the psychologist may be vulnerable to all kinds of confused thinking and meaningless talk. After all a biological theory that placed whales and fish together in the same category would be highly restricted in scope. It seems then that a defence of the psychological research program is called for.

In this defence I will first claim that some of the arguments arrayed against the psychology of concepts depend on a misunderstanding of the basic notion of prototype representation. The relation between data and theory in this area is much less direct that is often assumed. In particular, neither typicality effects nor borderline cases in themselves can determine whether a concept should be modelled with a prototype representation. Second I address the issue of concept identity and stability, and whether the fact of successful communication of thoughts between individuals is evidence of a common external determination of conceptual content. Third, I argue that missing data in conceptual representations is probably the norm rather than the exception, but that such evidence cannot be used to argue that our concepts are externally determined. The intuition of deference to experts is not as primitively compelling to naive thinkers as it is to some philosophers. Thinkers can and do sometimes exercise a choice in the matter of what terms really refer to. Finally I argue that adoption of the full Externalist concept of Concept leaves most of our common sense concepts out in the cold - unconnected to the real world - and so is hugely over-restrictive as the basis of a psychological theory of thought.

2. In defence of a psychology of concepts – correcting misapprehensions

It is easy to point to inadequacies in psychological theories and writings about concepts. The current state of understanding is far from perfect, and current theorising is plainly inadequate. For example some of the early writings on prototypes were vague and non-specific about just what prototypes are and how they might work. Much of this vagueness has since been tightened up as the field has progressed. There are also many criticisms of prototype theory that are based on misunderstanding or misinterpretation. In this section I will briefly attempt to set the record straight about some of these points, as they continue to appear in recent writings (e.g. Margolis, 1994, Rey, 1992, Osherson & Smith, 1997). Further discussion of prototype theory can be found in Hampton (1993, 1995a, 1995b, 1997, 1999).

2.1 Typicality ratings

Much has been made in the literature of a study by Armstrong, Gleitman and Gleitman (1983) in which they showed that typicality ratings could not be pure reflections of family resemblance structure, as Rosch and Mervis (1975) had apparently supposed them to be. Well defined concepts like even number showed consistent typicality effects, although all even numbers are presumably equally "good" examples of the concept category. Armstrong et al. used these data to argue that the existence of gradations of typicality within a class is inadmissible as evidence that the class is not well-defined. Since typicality gradations occur equally consistently in classes that are by definition well-defined, finding that a category shows typicality effects throws no light on the question of whether or not the category has a prototype representation that determines category membership.

In fact detailed studies of gradedness effects in natural categories (e.g. Barsalou, 1985; Hampton & Gardiner, 1983) have shown that typicality is strongly influenced by similarity to a prototype, but may also be influenced by the familiarity of the object categorized, and its commonness as a member of the category. These factors have been identified as independent sources of variance explaining typicality effects in other measures (Glass & Meany, 1978; Hampton 1997). When one source of variance is held constant, then naturally the others will tend to assume a greater role as determinants of the outcome of the typicality rating task.

Armstrong et al.'s point is just that the observation of typicality effects *per se* is not a reliable indicator that a particular concept has prototype structure, and this point is well taken. To this could be added, that even in the case where typicality effects can be shown to depend on family resemblance similarity, we are still none the wiser about whether categorization depends on a classical or a prototype definition. Following the Binary View of category structure outline by Rey above, there could be a classical core definition which determines categorization while additional "characteristic" features determine the typicality of category members.

A second often misunderstood point concerning typicality is the false belief that variations in typicality for items that are all clear category members imply a differentiation between the defining core properties that determine category membership and the characteristic features that determine typicality (the Binary View). It is commonly argued that although penguins are atypical, they are nonetheless bona fide birds - that is to say that all would agree that all penguins are in fact birds. Hence it is argued that typicality must depend on stereotypical or characteristic features such as flight, whereas categorization per se depends on the classical core of necessary bird essence features, (or alternatively on some form of locking relation to the external class of birds). While typicality varies between penguins and robins, degree of membership in the category does not. Hence degree of membership cannot be based on typicality. (This fallacy is still current – see for example Osherson & Smith, 1997).

It can be easily shown that no such conclusion follows. Consider a standard model of prototype structure in which categorization depends on family resemblance similarity. An object's similarity to the prototype is determined by some process of measuring the informational overlap between the object's representation and the category prototype. A threshold criterion is then applied to this similarity, so that if the similarity is above a certain threshold level the object is classified as a member of the category. The threshold can vary within certain limits, but does not vary across the full range of the similarity scale. Clearly, if similarity can continue to rise above the maximum level for the threshold criterion, then items can vary in their similarity to the prototype, in spite of their all being clear category members. This pattern is indeed born out in experiments that compare typicality and categorization (Hampton, 1998). Typicality and categorization may be tied to the same underlying similarity function, but whereas typicality rises monotonically and continuously with increasing similarity, categorization probability starts at floor, rises fairly rapidly to ceiling, and then remains at ceiling.

The fallacy is illustrated by a simple example. Suppose that we ask whether a man is tall or not, and also ask how typical he is of a tall man. It appears obvious that both judgments must rely simply on his height. Above, say 1.90 meters or 6' 3" he is undoubtedly tall (generally speaking). Yet his typicality in the class will increase the taller he gets. Differences in typicality that are not associated with changes in category membership simply do not imply that the two judgments must involve different semantic information. They simply reflect the asymptotic level of categorization.

2.2 Borderline cases

If typicality effects are only indirect indicators of concept structure, then it might be hoped that borderline cases (fuzzy boundaries) would provide stronger evidence. These are cases in which there is variability in whether an item is classified in the category or not. This variability is typically found both between subjects asked the same question and also across occasions when a person is asked the same question twice (McCloskey & Glucksberg, 1978).

Unfortunately, borderline cases (or their absence) are neither a necessary nor a sufficient symptom of a prototype representation (or its counterpart, a classical concept). If a concept representation had a prototype structure, then it is perhaps likely that borderline cases will occur. Items that are of intermediate similarity to the prototype should fall within the range of variability of the threshold criterion and so lead to inconsistent categorization. However some prototype concepts may have thresholds with very limited variability. In other cases the world may be underpopulated with examples in this region of semantic space (the over-used example of birds is a case in point). So lack of borderline cases need not indicate lack of a prototype definition. Similarly, classical definitions would seem to imply clear-cut boundaries. However borderline cases may still arise if an item's possession of one of the criterial attributes is itself in dispute or difficult to determine. In effect a borderline case may arise because one of the attributes composing the classical definition is itself not clear-cut.

2.3 Prototypes

Another set of misunderstandings comes from an overly restrictive view of what a prototype can be. Writers often assume that prototypes may only contain superficial or directly observable perceptual features of objects, and similarity may only be computed across these features (Rips, 1989). Indeed they often appear to assume that only visual features (and two-dimensional static ones at that) can be represented - as when Rey (1992) claims that according to prototype theory a decoy duck should be a better bird than a penguin. It would only take the inclusion of some features such as the feel, sound, typical movement and behavior of the object to correct this particular misclassification. These are all sensory features, but there is no good reason why other features should not also be represented. Rosch herself included typical actions performed with objects in her studies, and no prototype definition for artifact categories could get off the ground unless functional information were also included (e.g. see Hampton, 1979). Some abstract concepts such as Science and Art also appear to have prototype representations (Hampton, 1981). Not since Posner and Keele's (1968) seminal study of random dot prototype learning has the notion been restricted to purely visual stimuli. The only reason to restrict the type of information contained in a prototype is the metatheoretical one of providing some constraints on the generality and power of the model.

The power of the model is well demonstrated by the fact that many well defined concepts can readily be represented as prototypes. As I have argued elsewhere (Hampton, 1993, 1995a), if the "defining" features are given sufficiently high weights (each must have a weight greater than the sum of the characteristic features) then the threshold for categorization can be set high enough to require that each defining feature is necessary for categorization, and that together they are sufficient, so that none of the remaining non-defining features can influence the result. In effect, classically defined concepts are in the class of linearly separable categories, (Medin, Wattenmaker & Hampson, 1987) which can all be represented by a simple prototype model using a linear similarity metric.

2.4 Family resemblances

A final misunderstanding relates to family resemblance structure itself. In some of the earlier descriptions of the theory, (Rosch, 1975) examples were given of concept structures where items could be chained together through similarity one to the next, all then belonging to the same category through their "family resemblance". For example AAAA, AAAB, AABB, ABBB and BBBB are all similar in a pair-wise manner, but the two ends of the chain have nothing in common. This way of describing prototype theory was misleading, and some critics have understandably found the notion difficult to accept (see Rey, 1992, p322). After all, the chain need never be broken, and one would quickly find oneself with every object in every category. The sensible way to define family resemblance is of course by reference to the prototype or central tendency of the category. In the above example, all four stimuli could be in the same prototype category if that were defined as "not more than two features different from the prototype AABB". The above would be members, but equally the exemplars BBAA, BBAB, BBBA and so forth would be excluded, in spite of having close similarity to members of the category. Family resemblance is constrained by measuring similarity to the centre of the category, (or to all the category members in the case of the related Exemplar Model) rather than to any one individual.

3. Why meanings may be “in the head” after all

Having defended Internalist psychological theories of concepts from some of the misapprehensions common in the literature, in the following sections I will turn my attention to addressing two more central aspects of the Externalist critique - Stability and Deference. I then consider a further problem with the generality of the Externalist account as applied to everyday common or garden concepts.

3.1 Stability

The externalist’s claim about stability is that without there being an independent means of individuating concepts, we cannot know whether any two people discussing a topic are actually referring to the same concept class. Furthermore, if all of our encyclopaedic knowledge is a part of the mental content that individuates a concept, then it is hard to see how it is possible for any two people to grasp the same content, and why our conceptual system does not change radically with every new small fact that is added. If I believe that domesticated cats originated in Mesapotamia, while you believe that they originated in Egypt, and we have a dispute about it, what fixes the term “cat” so that we are not simply talking at cross purposes (as would be the case if “cat” had two quite distinct meanings, and we were each intending a different meaning).

Rey's answer is that concepts must have externally based defining cores, which are held in common amongst the language community, and which act as the criterion of whether an individual has "the correct" conceptual grasp of the term. For a knowledge representation system to work (it is argued) there must be a clear distinction between the information that fixes or constitutes the reference of a term, and the information that is then known about the term. Others (e.g. Fodor, 1998) would eschew any defining core, and would instead fix the concept in terms of an external class, to which our minds become attuned. It is just stipulated that the class of cats exists, and the psychological study of concepts then becomes the study of how we become attuned to this class, and the questions of interest are the epistemological questions about how we acquire beliefs about it.

Agreement on the meaning of terms can however be achieved in other ways. It is not necessary to have a clear-cut distinction between constitutive core and other knowledge, where a distinction of degree may serve just as well. If some attributes are more central to a concept representation than others, then one can account for why some differences in belief are just differences of opinion about contingent facts, while others may be differences in conceptualisation. It is not necessary to draw a sharp line, provided there is some way to make the distinction. How these variations in the centrality of information in conceptions arise is a matter for empirical research. Obvious possibilities are a statistical or informational account, and a conceptual interdependence account. By the first account, central attributes will be those with the strongest cue and category validities, while by the second they will be those that support the most general and interesting theories of the world. Sloman and Love (19xx) have evidence that centrality of an attribute is determined by the number and centrality of other attributes that depend on it, rather than on its statistical properties in relation to the category.

Given some way of determining attribute centrality, then there are two factors that keep us honest - that is to say in line with each other conceptually. First there is the fact that we are of the same species, with common native intelligence, common perceptual and cognitive apparatus, living in the same social, cultural and physical world. Second there is the fact that we are socialised into a linguistic community in which our use of words is monitored and the conventionally correct use of terms is explained to us. Education has a large role to play. These conventions are even collected in advanced societies into dictionaries, which act as a brake slowing down the rate of change in meaning. It should be recognised however that dictionaries do not define meanings and usage in any fixed way, (contrary to a popularly held belief). Instead they reflect the efforts of lexicographers to track the way words are used (traditionally by reference to the cultural productions of the intellectual elite, although more recently there has been an increasing interest in reflecting word use by a wider range of society). These two factors – exposure to the same external environment, and communicating with a common vocabulary – will keep different members of the same community more or less in line with each other conceptually. For most purposes we can talk to each other, and understand the other's meaning correctly.

Why should we wish to allow degrees of centrality for attributes, rather than stick to all-or-none defining features? For a start it permits an account of conceptual change. The meaning of concepts can change developmentally as well as historically, by a gradual change in the relative centrality of different attributes. Change can be tracked through time, and so identity preserved, provided it is gradual (Millikan, 1992). Conceptual representations within a cognizer or a culture may be analogous to objects in the world. It is not their overt descriptive content that individuates them but their continuous path through space-time. When a frog turns into a prince, or a pumpkin into a carriage, it is the disappearance of the one at the same point in space and time as the appearance of the other that gives rise to the assumption of a continuous identity. Similarly, conceptual representations in a person's mind may be given stability through their continuity in time preserved by memory.

Varying degrees of attribute centrality also allow for people to have subtly different concepts. If two individuals have much the same set of attributes, linked up in similar ways, but with different weights, then they may still agree that they are discussing the same topic, but hardly agree at all on what is true about it. If the differences in attributes are too great then they may decide that they are talking at cross-purposes, and the prototype view of concepts argues that this can also be a matter of degree. Depending on the purpose of the conversation, it may or may not matter just how much your concept differs from mine.

The process is well illustrated in the case of technical vocabulary. Scientists frequently hold international conferences at which the meaning of particular terms is fixed by conventional agreement. For example in 1999, there was an international Internet poll of astronomers to determine whether Pluto should retain its status as a planet. The problem was that after Pluto had been discovered and labelled as a planet, it subsequently transpired that in terms of similarity to exemplars of the categories of Planet and Asteroid, it might more properly fit in the latter category, being a relatively small rock with an eccentric orbit at a great distance from the sun. Clearly even in scientific discourse, the meaning of terms is not fixed by the external class to which they refer. (Pluto survived the poll, and remains as a planet).

So the answer to the stability problem is two-fold.

(a) Stability is not dependent on essences, or on a common commitment to an external definition of terms. Indeed the cultural evolution of both concepts and word meanings is good evidence that concepts are not rigidly fixed, but instead are constantly being renegotiated in the interplay between individuals using them for communication and understanding.

(b) Stability and commonality of concepts between individuals are a matter of degree. Our concepts may be sufficiently similar for most purposes, and we may not notice any important differences. Yet problems of conceptual disparities between individuals frequently do arise in real life, and cannot usually be resolved by reference to some external standard.

3.2 Division of linguistic labour and deference to experts

Much has been made by Externalists of the fact that people may be willing to defer to Putnam's “linguistic expert” for the true definition of their terms (Putnam, 1975). When asked to classify pure metals as gold or not gold, I may be unable to do so, but I know that there is someone that I can trust to know the answer. This deference to experts is a direct result of the partial nature of many of our conceptual representations in particular domains. It is therefore clear that psychological accounts need to be provided with a means to allow for missing data to be flagged. If a prototype were to be represented by a frame with a set of slots, then only some of them would be filled with known values. After all, most people's knowledge is far from encyclopaedic.

The question of just what aspects of our conceptual system we defer to experts for, and what we retain under local control is however far more complex and remains an empirical question. We will need to investigate the factors affecting the degree of trust people will place in their own conceptual beliefs. We will also need to bring in a social psychological or even sociological account of when people are willing to place faith in others as sources of reliable information. When do we defer to the dictionary and when do we decide that the compiler has got it wrong? As social creatures we are committed to maintaining a degree of intersubjective agreement on conceptual structure. Part of this commitment involves delegating responsibility to others in certain domains, particularly in the areas of law, science and technology. But the commitment does not give carte blanche to expert determination of conceptual contents, and nor are we willing to hand over to experts all rights in the matter. Biologically speaking, cucumbers and peas are both fruits, but the general public has not shown any interest in adapting their concept of fruit to match that of the expert (unlike the cases of birds and fish).

Studies in which people are asked to imagine counterfactual situations, or categorize objects with unusual combinations of attributes suggest that people's willingness to assume that there are unknown essences and to defer to expert opinion may be restricted - even for natural kinds. Braisby, Franks & Hampton (1996) posed the following version of the Putnam problem to a sample of undergraduate students:

> *You have a female pet cat named Tibby. For many years people have assumed cats to be mammals. However, scientists have recently discovered that they are all, in fact, robots controlled from Mars. Upon close examination, you discover that Tibby too is a robot, just as the scientists suggest.*

Among other sentences, subjects were asked to agree or disagree with the following two:

(1) Tibby is a cat, though we were wrong about her being a mammal.

(2) Tibby is not a cat, though she is a robot controlled from Mars.

Putnam's claim is that common sense intuition dictates the first to be correct, since a cat is simply anything that shares the property of catness - which now turns out to be identifiable with Martian robotics. In fact opinion was divided, with 20% saying that the second was true, and a significant number of people actually agreeing with both statements.

Why should opinion be divided? I claim that when asked to suppose that one has been mistaken about the central attributes of a concept, then one has to set up a new concept. (The same is not true for changes in peripheral attributes - see Braisby et al, 1996). The thinker needs to set up a concept corresponding to <Robot from Mars with cat properties>. At the same time there is an existing concept - namely the furry terrestrial mammal that we previously believed this type of object to be. The question is then which of these two conceptions you want to label with the name “cat”. There is surely an element of arbitrariness about this decision. You could transfer the name along with the extensional class, and say that Cats are still cats. That is to say that you decide to use the word "cat" to refer to the same class of objects as before. You then need a different name (e.g. "mammal cat") for the concept you had previously, and which turned out to correspond to an empty set. You could then agree that cats are robots while mammal cats do not after all exist. Or you could equally well leave the name attached to the earlier concept (after all a "real" mammalian cat might yet turn up), and create a different name for the robots such as “robocat”. In which case you would agree that there are after all no cats, but there are a lot of robocats around. It's a question of whether the extension or the intension has the first claim on the name, and the answer would seem to be quite arbitrary, depending only on pragmatic criteria about the future usefulness of the concept name.

3.3. The generality of External definitions

Consider the following scenario. Some time in the seventeenth century we have two scientists with views about heat phenomena. One holds the view that heat is a substance – phlogiston – and that the hotness of a body depends on the quantity of phlogiston within it. Another holds that heat is a form of energy. At this stage, an Externalist would have to deduce that the latter scientist grasped a true concept, while the former had none. If Externalism is correct, then a “conception” or mental representation that has no corresponding external reality cannot be a concept. Our minds have become locked to a non-existent concept class, and so presumably all thoughts involving the conception are invalid and meaningless. But to a psychologist studying the two individuals at the time there could no way to distinguish who had the concept and who did not. In other words, the mental representations themselves could be identical in terms of their internal consistency and psychological structure, and could even be equal in their ability to make correct predictions about currently known phenomena. It was only with historical progress and the benefit of hindsight that one could differentiate the correct from the incorrect conception of heat. But it cannot be the case that progress in understanding the nature of mental contents (that is to say the psychology of concepts) must wait for the development of scientific understanding of the nature of the subject matter to which those contents refer. The relevant data needed to provide a full psychological account should not need to encompass future developments in other sciences. To take another example it should not require a full understanding of the aetiology and mechanisms of schizophrenia in order for us to decide whether or not a clinician actually possesses a concept of schizophrenia or not.

A second problem with Externalism is that it would restrict the universe of discourse for concepts to a very narrow range of scientific and academic subjects. There is a willingness to ignore or to gloss over the problem that if this strict definition of "concept" is adopted then most of the words in English do not have any conceptual content at all. Take concepts like BUG, SOAP OPERA, or FRIEND. The position being adopted argues that a competent speaker of English who knows the concept Bug has either some clear cut understanding of the essence of bugness, or has an empty slot that says that there is some essence of bugness in the real world, which scientists or other relevant experts either have discovered or are waiting to discover.

“Bug” is just one of a great many words, with well understood meanings, which make no contact with a scientifically describable world at all. To say that all thoughts about such things are empty of conceptual content is not good enough for a psychological account, but the Externalist position has no way of providing them with any other semantics. No one is going to waste their time exploring the essential nature of ICE CREAM or SAUSAGE or PARK (unless of course they happen to be employed by the European Commission in Brussels, to whom such questions are of paramount importance). Of course, broadly speaking there is a “correct” and an “incorrect” usage of such terms, but the only arbiter of correctness is the language community as a whole, and correctness itself will admit of many borderline disputes with no ultimately satisfying resolution. As discussed earlier, the determination of correct usage is not simply a democratic matter of a majority “vote”, as there are vested interests and institutional factors at play. It serves society as a whole to standardise word meanings and the concepts to which the words refer, through dictionaries and institutions such as the European Commission or the Académie Française. However this societally sanctioned function is never completely successful, acting merely as a damper on linguistic change rather than fixing the correct reference of terms for all time.

(Note that the issue here is not simply one of word meaning, but rather of conceptual categories. When the British sausage was threatened by the EC it was on the grounds that the non-offal meat content was insufficiently high to categorize it as a “Sausage”. But the common standard was only made possible by a linking of the reference of terms such as “saucisson” and “wurst” to a common conceptual category. The defence that the English word “sausage” doesn’t mean the same as the terms in other languages was not considered valid.)

The problem doesn't stop with non-technical terms like “bug”. Analysis of the meaning of natural kind terms suggests that even words like Fish or Tree correspond to no useful scientific taxonomic category (Atran, 19xx). Perhaps this just means that we are using the words incorrectly. Sometimes the damage can be repaired - generations of children can be educated into calling a Whale a mammal. But in general the degree of misfit is far too great for a re-education program. And in any case, however tempting the prospect, psychologists are not in the business of instituting training programs to get people to respond in just the right way to prove their theories correct. We have to accept that scientifically definable concepts have limited impact on the everyday conceptual world (see also Dupré, 1981).

If to qualify as a true concept, a term’s conceptual content must by definition be externally determined, then most of our thoughts are probably concept-free. Externalism is not going to do the job of accounting for concepts as the vehicles of everyday thought.

5. Conclusion

In conclusion, I have argued that a psychological account of concepts is interested in just those things which philosophy may wish to disregard as signs of irrationality and confusion. But the theory should not be mistaken for its subject matter. Just because psychology is interested in exploring the vagueness and incompleteness of conceptions, that does not imply that its theoretical approach needs to be vague or incomplete. If sometimes it appears to be so, then that may reflect the early stages of the research program, but it does not represent the essence of our approach, and I have tried to clarify a number of misconceptions of this approach. If philosophy and psychology are to cooperate on the study of concepts then it is important that conceptual differences in the notion of concept are clarified.

References

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

Atran, S. 1997 BBS

Barsalou, L.W., & Hale, C.R. (1993). Components of conceptual representation: from feature lists to recursive frames. In I. van Mechelen, J.A.Hampton, R.S.Michalski, & P.Theuns, (Eds.), Categories and Concepts: Theoretical Views and Inductive Data Analysis (pp. 97-144). London: Academic Press.

Barsalou, L.W. (1985). Ideals, Central Tendency, and Frequency of Instantiation as Determinants of Graded Structure in Categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 629-654.

Braisby, N., Franks, B., & Hampton, J.A. (1996). Essentialism, Word Use, and Concepts. Cognition, 59, 247-274.

Chater, N., & Heyes, C. (1994). Animal concepts: Content and discontent. Mind and Language, 9, 209-246.

Evans, J.St.B, & Over, D. (1996).

Dupré J. (1981). Natural kinds and biological taxa. The Philsophical Review, 90, 66-90.

Fodor, J. (1998). Concepts: where cognitive science went wrong.

Glass, A.L., & Meany, P.J. (1978). Evidence for two kinds of low-typical instances in a categorization task. Memory and Cognition, 6, 622-628.

Hampton, J.A. (1979). Polymorphous Concepts in Semantic Memory. Journal of Verbal Learning and Verbal Behavior, 18, 441-461.

Hampton, J.A. (1981). An Investigation of the Nature of Abstract Concepts. Memory and Cognition, 9, 149-156.

Hampton, J.A. (1993). Prototype models of concept representation. In I.van Mechelen, J.A. Hampton, R.S. Michalski, & P. Theuns (Eds.), Categories and concepts: Theoretical views and inductive data analysis, (pp. 67-95). London: Academic Press.

Hampton, J.A. (1995a). Testing Prototype Theory of Concepts. Journal of Memory and Language, 34, 686-708.

Hampton, J.A. (1995b). Similarity-based categorization: the development of prototype theory. Psychological Belgica, 35, 103-125.

Hampton, J.A. (1997). Psychological representation of concepts. In M.A.Conway & S.E.Gathercole (Eds.) Cognitive Models of Memory. (pp81-110). London: UCL Press.

Hampton, J.A. (1998). Similarity-based categorization and fuzziness of natural categories. Cognition, 65, 137-165.

Hampton, J.A. (1999). Typicality, graded membership and vagueness. MS under review.

Hampton, J.A., & Gardiner, M.M. (1983). Measures of Internal Category Structure: a correlational analysis of normative data. British Journal of Psychology, 74, 491-516.

Keefe, R., & Smith, P. (1997). Theories of vagueness. In R.Keefe and P.Smith (Eds.) Vagueness: a reader, pp1-57. Cambridge: MIT Press.

Margolis, E. (1994). A reassessment of the shift from the classical theory of concepts to prototype theory. Cognition, 51, 73-89.

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

Medin, D.L., & Ortony, A. (1989). Psychological essentialism. In S.Vosniadou & A.Ortony (Eds.), Similarity and analogical Reasoning. Cambridge: Cambridge University Press.

Medin, D.L., Wattenmaker, W.D., & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness, and category construction. Cognitive Psychology, 19, 242-279.

Millikan R. (1992). Presidential address to Society for Philosophy and Psychology, Memphis, TE.

Millikan, R. (1997) BBS

Osherson, D., & Smith, E.E. (1997). On typicality and vagueness. Cognition, 64, 189-206.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.

Putnam, H. (1975). The meaning of 'meaning'. In Mind, language and reality, volume 2: Philosophical papers. Cambridge: Cambridge University Press.

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

Rey, G. (1985). Concepts and conceptions: A reply to Smith, Medin & Rips. Cognition, 19, 297-303.

Rey, G. (1992). Semantic externalism and conceptual competence. Proceedings of the Aristotelian Society, 92, 315-331.

Rips, L.J. (1989). Similarity, typicality and categorization. In S.Vosniadou & A.Ortony (Eds.), Similarity and analogical Reasoning (pp. 21-59). Cambridge: Cambridge University Press.

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

Rosch, E., & Mervis, C.B. (1975). Family resemblances: studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Sloman, S. & Love, B.

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

Sutcliffe, J.P. (1993). Concept, class, and category in the tradition of Aristotle. In I.van Mechelen, J.A. Hampton, R.S. Michalski, & P. Theuns (Eds), Categories and concepts: Theoretical views and inductive data analysis. London: Academic Press.

Williamson, T. (1994). Vagueness. Routledge: London.