What Family Resemblances Are Not: The Continuing Relevance of Wittgenstein to the Study of Concepts and Categories

Michael Ramscar (M.J.A.Ramscar@ed.ac.uk)
Department of Artificial Intelligence, University of Edinburgh
80 South Bridge, Edinburgh EH1 1HN, Scotland
Ulrike Hahn (U.Hahn@warwick.ac.uk)
Department of Psychology, University of Warwick
Coventry CV4 7AL, United Kingdom

Abstract

We argue that common interpretations of Wittgenstein's Philosophical Investigations within Cognitive Science misrepresent his account, underplaying its radical content. Appropriately interpreted, this account continues to challenge contemporary theories of concepts and categorisation. We illustrate the continued relevance of his position by directly applying its critique to current approaches to categorisation.

Introduction

> 66. Consider for example the proceedings we call "games". I mean board-games, card-games, ball-games, Olympic-games, and so on. What is common to them all? - Don't say: "There must be something common, or would they not be called 'games'" but look and see whether there is anything common to all. For, if you look at them you will not see something that is common to all, but similarities, relationships, and a whole series of them at that. To repeat, don't think, but look! (Wittgenstein 1953, §66, p31).

Cognitive science has a strong interest in categorisation: accounting for how the 'stuff of experience' is represented, manipulated and combined in the mind is a central concern of many researchers in the field. As can be quickly gleaned from a even casual perusal of the relevant literature, Wittgenstein's analysis of concepts and categories in the *Philosophical Investigations* (1953; PI) has had a great influence on the approaches taken in this area.

In Ramscar (1997) we examined in detail the veracity of the interpretation of Wittgenstein's view that is commonly held by researchers studying categorisation, comparing it with a detailed exposition of Wittgenstein's arguments. Although Wittgenstein is often presented as an opaque, difficult to interpret, and rather obscure philosopher sometimes leading to the *Philosophical Investigations* being seen as a philosophical pick 'n' mix, a series of gnomic quotables to be plundered in support of a thesis - PI sections §66 to §82 actually lay out a clear, if intricately connected, series of arguments detailing Wittgenstein's theoretical treatment of categories and categorisation in a fairly straightforward manner. The picture that emerges from a close reading of Wittgenstein's text is at considerable variance with the generally accepted account of Wittgenstein's position. At least one reason for this is a fundamental one: whilst Wittgenstein is often cited as a founding influence in cognitive approaches to concepts and categorisation, his concerns were markedly different than those of researchers in the modern cognitivist tradition. Whilst much categorisation research has been concerned with category representation - the encoding and structuring of objects together in some form of internal representation system (see Komatsu, 1992 for a review) - Wittgenstein was more concerned with word use, with the way that labels are used to pick out objects in the world as a part of the process of communication. In doing this, Wittgenstein was concerned with trying to specify the way in which the use of concepts and categories in communication imposes constraints on theoretical accounts regarding their nature - the 'looking' in §66 above strongly emphasising the need to fully understand the problem before tackling any solution to it.

Family resemblances?

The accepted interpretation of Wittgenstein's account within cognitive science is nicely summarised by Lakoff (1987; accounts which concur broadly with this can be found in Johnson-Laird, 1983; Medin & Ortony, 1989; Komatsu, 1992). Lakoff acknowledges Wittgenstein as the first theorist to notice what he terms a major crack in the classical theory of concepts and categories (e.g. Katz, 1972). Wittgenstein, says Lakoff, argues that categories such as *game* cannot be accounted for according to classical theories because there are no properties that are common to all games. Lakoff draws two key theses from this argument:
1: "Games, like family members are similar to one another in a variety of ways"; and
2: "That [family resemblances], and not a single well defined collection of common properties is what makes *game* a category" (Lakoff, 1987, pp 16-17)

Whilst 1 is an uncontentious statement of Wittgenstein's views, 2 is a rather more difficult interpretation to sustain. In PI §66 (p 31) Wittgenstein explicitly states that 'you will not see something that is common to all [games]'. Rather, he argues that what games have in common is the now notorious *family resemblances*: 'a complicated network of similarities overlapping and criss-crossing: sometimes overall similarities, sometimes similarities of detail' (PI, p 32). Lakoff, (and cognitive scientists in general) take this to be Wittgenstein's characterisation of what a category is. But what appears to escape these interpreters is the extreme negativity of this characterisation. In PI §67 (pp 31 -2)

Wittgenstein explicitly condemns this characterisation of naming categories as vacuous. Saying that the common theme that runs through a category is the continual overlap of family resemblances is directly analogous to saying that the common thing that runs through a thread is continuous overlapping of the fibres that make up the thread, and Wittgenstein dismisses both of these accounts as empty gestures: 'Now you are only playing with words' (PI p 32). There is, he says, no *thing* that runs through a thread in the form of overlapping fibres; a thread simply is a series of overlapping fibres. His view is a serious challenge to, rather than an endorsement of, Lakoff's formulation: if family resemblances are the common thing that run through *game*, just as overlapping fibres are the common thing that run through a thread, then *what* is this thing supposed to be? How is it supposed to do whatever it is it is supposed to do? *How long*, Wittgenstein asks, *is a piece of string?*

The length of a string - naming and boundaries

This question - 'how long is a piece of string?' - becomes important once the second part of Lakoff's exposition is introduced. Wittgenstein, as Lakoff notes, argues that the boundaries of categories are not fixed, commenting

> 68. "All right: the concept of number is defined for you as the logical sum of these individual interrelated concepts: cardinal numbers, rational numbers, real numbers, etc.; and in the same way the concept of a game is the logical sum of a corresponding set of sub-concepts." It need not be so. For I *can* give the concept 'number' rigid limits in this way, that is use the word "number" for a rigidly limited concept, but I can also use it so that the extension of the concept is *not* closed by a frontier. And this is how we do use the word "game". For how is the concept of a game bounded? What still counts as a game, and what no longer does? Can you give the boundary? No. You can *draw* one, for none has so far been drawn. (But that never troubled you when you used the word "game" before.). (Wittgenstein 1953, p32-3).

Lakoff interprets this discussion of *number* as follows: historically, says Lakoff, numbers were first taken to be integers, and then 'numbers' were successively extended to include rational numbers, real numbers, complex numbers, transfinite numbers, and all of the other numbers that mathematicians are wont to invent. But the concept of 'number' is not bounded in any natural way, and it can be limited or extended depending upon one's circumstances and purposes. Lakoff says that in mathematics, intuitive human concepts like *number* must receive precise definitions: Wittgenstein's point, he claims, is that different mathematicians give different definitions, depending upon their goal. Thus although the category *number* can be given precise boundaries in many ways, 'the intuitive concept is not limited in any of those ways; rather, it is open to both limitations and extensions' (Lakoff, 1987, pp 17).

The key question, on Lakoff's account, is how those limitations and extensions are governed - what factors determine the boundaries of categories in given circumstances. Lakoff answers this question in relation to *game* by saying that *game*'s boundaries are governed by resemblance to previous *games* in appropriate ways: a new thing can be a *game* if it is suitably similar to previous *games*. Lakoff cites the introduction of video games in the 1970s as a recent example of the boundaries of the *game* category being extended on a large scale.

Once again, subtle and not-so subtle discrepancies can be distinguished between Lakoff's characterisation of Wittgenstein's views and the content of Wittgenstein's stated arguments. In §68, Wittgenstein says that one '*can* give the concept 'number' rigid limits in this way, that is use the word "number" for a rigidly limited concept,' Lakoff's claim that in mathematics *number* must receive precise definitions appeals to this - 'but I can also use it so that the extension of the concept is *not* closed by a frontier.' Here, Wittgenstein is not talking about the extensibility of borders, but something far more radical: 'You can *draw* [a boundary], for none has so far been drawn. (But that never troubled you when you used the word "game" before)' (PI pp 32-3). Wittgenstein isn't talking here about the extensibility of boundaries; he is talking about their absence, a point developed in PI §69 to §73: categories do not have, or need, boundaries at all. In the context of Wittgenstein's overall discussion of categories, this is a vitally important point: it is one thing to seek to determine the length of a piece of string whose length isn't fixed (we might add a temporal dimension to our answer for instance); it is quite another thing to seek to find out how long a piece of string is when the string is of no particular length at all.

On this point, Wittgenstein is emphatic (PI §69). One *can* draw a boundary, for a special purpose, but it is just that, a drawn boundary. Important in the context of the special purpose, no doubt, but arbitrary to the concept or category in question. We do not *need* to draw boundaries, because we can happily use concepts where no boundary has been drawn; thus categories do not need boundaries to be usable. To further iterate this point, Wittgenstein considers the state of a user of a category (concept) who cannot specify that category's boundaries: is the user ignorant of those boundaries? - No, she does not 'know the boundaries because none have been drawn' (PI, p33). Not knowing the boundaries of *game* is not a state of ignorance - it is just reflective of the boundariless state of the category *game*.

The thesis that categories don't have boundaries is vital to Wittgenstein's position:

> 71. One might say that the concept 'game' is a concept with blurred edges. - "But is a blurred concept a concept at all?" - Is an indistinct photograph a picture of a person at all? Is it even always an advantage to replace an indistinct picture by a sharp one? Isn't the indistinct one often exactly what we need?
>
> Frege compares a concept to an area and says that an area without boundaries cannot be called an area at all. This presumably means that we cannot do anything with it. - But is it senseless to say: "Stand roughly there"? Suppose that I were standing with someone in a city square and said that. As I say it I do not draw any kind of boundary, but perhaps point with my hand - as if I were indicating a particular *spot*. And this is just how one might explain to someone what a game is. One gives examples and intends them to be taken in a particular way. - I do not, however, mean by this he is supposed to see in those examples that common thing that I - for some reason - was unable to express; but that he is now going to *employ* those examples in a particular way. Here, giving examples is not an *indirect* means of explaining - in default of a better. For any general definition can be misunderstood too. The point is that *this* is how we play the game. (I mean the language game with the word "game".) (Wittgenstein 1953, p34).

Again, Wittgenstein's rejection of boundaries - and not just the idea of fixing upon this boundary rather than that one seems to be both clear and unambiguous. We don't have to define boundaries in order to use concepts, nor is it clear that definite boundaries are always what we need; these points can be further drawn out if we contemplate §71 in conjunction with §76:

> 76. If someone were to draw a sharp boundary I could not acknowledge it as the one that I too always wanted to draw, or had drawn in my mind. For I did not want to draw one at all. His concept can be said to be not the same as mine, but akin to it. The kinship is that of two pictures, one of which consists of colour patches with vague contours, and the other of patches similarly shaped and distributed, but with clear contours. The kinship is just as undeniable as the difference. (Wittgenstein 1953, p36).

Categories do not have boundaries, and by defining boundaries we do not capture these categories, we create something new - call them bounded categories (in §68, Wittgenstein calls them 'rigidly limited' concepts, so we might call our bounded *game* a rigidly limited game) - which have some kind of kinship with our natural naming categories (e.g. *game*), but a rigidly limited *game* is markedly and importantly different to *game*. (This is similar to a point made earlier, our *use* of names is different form any theoretical view of them as constructs).

To return to family relations, these are the fibres that make up the threads that are categories: but Wittgenstein explicitly states that the length of these threads cannot be determined.

Categories and schemas: what's in a name?

In explaining what a game is, observes Wittgenstein, one gives examples of instances game, and one intends those examples to be taken in a particular way. What one does *not* do is expect the person to whom one is explaining 'game' to see the common thing - whether it be a core, schema or essence - which one cannot actually see oneself. It is true, says Wittgenstein, that when we give these examples our subject might see kinships between the examples, but these kinships are not in any way essential (hence the *differences* between the instances will be just as undeniable as these kinships). Giving these examples, says Wittgenstein, is not an *indirect* explanation; it *is* the explanation. We *don't* give a general definition, but this is not because we can't think of one, but because there is none to give.

> 72 *Seeing what is common.* Suppose I show someone various multi-coloured pictures, and say: "The colour you see in all these is called 'yellow ochre'" This is a definition, and the other will get to understand it by looking for and seeing what is common to the pictures. Then he can look *at*, and point to, the common thing
>
> Compare this with a case where I show him figures of different shapes all painted the same colour, and say: "What these have in common is called 'yellow ochre'".
>
> And compare this case: I show him samples of different shades of blue and say: "The colour that is common to all these is what I call 'blue'". (Wittgenstein 1953, p34).

It isn't just that there is no single 'thing,' common to all: Wittgenstein questions the way that 'commonalities' are supposed to be garnered in the first place. In the first example in §72 above, the commonality is easy to spot: provided the only common colour in the pictures was yellow ochre, and provided that the subject had grasped the meaning of colour, then she will be able to grasp what yellow ochre is - the colour that is common in all the pictures.

In example two, the subject could not proceed in the same way: although the figures all have colour (yellow ochre) in common, they also have other commonalities, such as being figures. Thus the subject could as easily learn to apply 'yellow ochre' to yellow ochre or to figures, or even to samples (all of the samples are 'samples' after all) from this example. Nothing in the definition picks out the particular commonality that 'yellow ochre' is supposed to pick out.

Finally, in example three, there is no *a priori* colour commonality to the pictures; rather, the commonality can only be perceived if one already has the concept 'blue' (Otherwise, one would see a riot of various 'colours'; since understanding this example is dependent upon an understanding of 'blue', the example could not serve as a explanation of, or a definition of 'blue'.

Wittgenstein poses a number of questions, albeit perhaps non-obvious ones, that are raised by the introduction of the idea of a generalised schema to serve as the basis for a category. Firstly, there is the question of the form that the generalisation should take: i.e. what shape should a generalised leaf be? Intricately linked to this is the question of the use of the schema. Even if we can answer the first question - how we say generate a generalised temperature for ice-cream - we are still left with the related question of how such a generalisation is to be used. Which particular aspects of the schema are general, and which are not (we might rephrase this question as asking which parts of the schema represent 'the generalised concept', and which are implementational details of the representation of this generalisation), and how in use are we supposed to know which is which. Is the generalised green shape a schema for green or a schema for generalised shape. Which raises the further question: provided one could generate answers to these very challenging questions, what is supposed to be intrinsic to such a schema that would cause it to be used differently to an *example* of that which it was supposed to be a generalisation of? Wittgenstein makes it quite clear in PI that satisfactory answers to these questions cannot be provided. Thus he doesn't advocate schemas as a theory of category representation (as argued by Johnson-Laird, 1983), but rather he seeks to demonstrate that schemas alone *cannot* provide an account of how concepts are represented

Wittgenstein's account

We can state the broad outline of Wittgenstein's arguments as follows:

1. That categories have no necessary or sufficient defining characteristics: rather that kinships "family resemblances" can be traced across categories (§65-7)
2. That these category spaces are unbounded i.e. there are no boundaries to the space across which "family resemblances" can be traced (§68, 69, 70, 71, 73)
3. That learning a category such as game does not involve extracting an essence or schema from instances. (§71-83)
4. In learning a "category" such as game, one learns examples (instances) and appropriate ways of using these examples (§69,71, 73, 81, 82)

Wittgenstein's arguments, as examined so far, do not advocate a particular view of concepts and categories - what has become known loosely as 'family resemblance theory' - but rather they represent a thorough attempt to elucidate the deep problems inherent in trying to account for concepts and categorisation. To Wittgenstein, the problems involved in explaining how categories are defined stem not from the phenomenon under examination, but the way this phenomenon has traditionally been defined (hence, perhaps, the famous 'don't think, but look!'). If we 'think' - i.e. if we assume that the existence of things called games entails the existence of, say a central schema (defined in some as yet to be determined way) in virtue of which the things can be considered games - we do not explore categorisation: we merely predetermine the explanations we can formulate.

Implications for cognitive theories of concepts and categorisation

Trying to spell out more clearly Wittgenstein's views on naming is, in our opinion, far more than a matter of setting straight the interpretive record. Wittgenstein's views and arguments continue to be relevant to current theories of categorisation. Here we trace out a Wittgensteinian perspective, so to speak, on current accounts.

1. Prototype theories. First and foremost, it seems from the above that renditions of category structure in terms of similarity to one or more central "prototypes" is incompatible with Wittgenstein's perspective and the very real theoretical problems embodied therein - despite the frequent appeals to him from proponents of prototype theories (e.g., Taylor, 1995). This holds both for versions of prototype theory which view the prototype as an abstracted central tendency or schema, and for those versions which take prototypes to be particular, privileged exemplars (for discussion of the different variants of "prototype" see Barsalou, 1987).

That the schema version is irreconcilable with Wittgenstein's position has already been argued at length. Such accounts, which seem particularly popular within Cognitive Linguistics (see, e.g., Taylor, 1995), but have also been proposed in psychology (see e.g., Smith & Medin, 1982 on the "probabilistic view"), are very explicitly at odds with the PI.

By contrast, the conflict between his position and the "prototype-as-privileged-exemplars" stems from the fact that such prototypical exemplars would, in fact, provide the glue to hold the category together in a way that Wittgenstein denies. Central exemplars *would* constitute a central thread or focal point around which the category is organised. Items would all obtain category membership by virtue of the single, simple fact that they are sufficiently similar to a central exemplar.

This is not the "criss-crossing" associated with Wittgenstein's idea of family resemblance (and indeed not the way real families, viewed over multiple generations, are structured). Thus the popular equation of "the family resemblance" view of category structure, which claims direct descendence from the PI, and "prototype theory" must be rejected (but see, Komatsu, 1992; Taylor, 1995).

2. Exemplar theories assume that our mental representations of categories consist simply of stored exemplars, i.e., known members. There is no abstraction of schemas or central patterns. Despite this very extensional "feel", exemplar accounts nevertheless allow classification of novel, previously unencountered objects by virtue of their similarity to known exemplars. In the most basic version, a novel item is simply given the classification of the known exemplar to which it is most similar, i.e., classification is based on the single "nearest neighbour" in similarity space. Despite its simplicity, this approach to categorisation proves remarkably successful in machine learning contexts (Cover & Hart, 1968). As a cognitive model, however, it conflicts with Wittgenstein's claim that natural language categories have no boundaries. The simple nearest neighbour approach produces well-defined category boundaries which run along the paths of equal distance between members of competing categories.

However, as a class of account, exemplar models need not posit such boundaries and, in fact, the most prominent exemplar model in the psychological literature does not. This model, Nosofosky's (1986) Generalized Context Model (GCM), is one of the leading psychological models of categorisation, and has provided remarkable data fits to human behavioural data in a variety of contexts (e.g., Nosofsky, 1986, 1988). This model, too, assumes that categories are represented in terms of stored exemplars. Specifically, exemplars are represented as points in a multi-dimensional "psychological space". Coordinates for points are determined by their value along the particular psychological dimensions in question; these dimensions can be things like "loudness" or "size" or more complex, composite dimensions. Similarity between exemplars is a function of distance in psychological space (specifically an exponential decay function, see Nosofsky, 1986). Classification decisions are governed by a *probabilistic* response rule. The probability of categorising an item as a member of a particular category corresponds to the weight of the evidence for this category. In contrast to the nearest neighbour algorithm, the evidence takes *all* exemplars into account. Specifically, the strength of the evidence for a category C, corresponds to the summed similarity between the novel item and all known exemplars of C, divided by the summed similarities to all stored exemplars, that is, not just members of C, but also the relevant competing categories.

Consequently, the model does not impose discrete category boundaries in psychological space, but rather *probability distributions* over the entire space.

The GCM seems to fit with all 4 points extracted from Wittgenstein's argument above: first, categories have neither necessary nor sufficient boundaries; second, category spaces are unbounded; third, learning does not involve extracting an essence or schema; fourth, in learning a category such as "game", one learns examples and appropriate ways of using these examples. Thus, at least at first blush, there is a contemporary cognitive account of categorisation which is compatible with Wittgenstein's description of categories and category structure.

3. *Connectionist models* have not only been widely used for general classification, but also specifically for cognitive

accounts of human categorisation (e.g., Small, 1997). Connectionism as a general framework is too loosely defined to allow broad generalisations with respect to Wittgenstein's points. Thus we limit ourselves to the basic approach to categorisation one might take with a standard feedforward multi-layer perceptron. We will assume that the inputs to the network are feature vectors representing different exemplars, and that the networks outputs are activation patterns that represent particular classification decisions. In training, the network is presented the example patterns and learning proceeds through incremental adjustment of weights in response to error signals derived from output errors. The network must find a set of weights which singly satisfies (if possible) all patterns, such that when presented with an input, the activation flow through the network produces the correct output.

Such a network would learn defining characteristics only in exceptional cases. Thus it sits happily with the rejection of a definitional story. Unbounded category spaces emerge if, as is most frequently the case, units have sigmoidal activation functions rather than simple thresholds. Sigmoid functions induce smooth distributions over the instance space in a way that is analogous to GCM.

More subtle is the relationship to Wittgenstein's other two claims. The net "learns examples", albeit in a loose sense. While learning is driven by examples, the actual examples themselves have no privileged status for the network as they would for a system with discrete exemplar representations. The net might respond equally strongly and accurately to patterns it has never seen before. The least straightforward issue is whether or not such networks extract schemas or essences in order to solve their task. Uncontroversially, such networks can and will exploit statistical regularities in training sets; the question is, does this amount to schema abstraction, as has frequently been suggested (e.g., Bechtel & Abrahamsen, 1991)? We think not.

While there are systematic connections between prototype theories which base classification similarity to the central tendency (e.g., the "average item") and linear discriminant functions (see Langley, 1996), these connections do not hold generally. Multi-layer networks re-represent the input in whatever fashion enables the solution of the problem and the resultant hidden layer representations need in no way be interpretable as "schemata". Furthermore, the extent to which anything resembling extraction of statistical regularities happens at all is determined by the network's resources. Single-layer or multi-layer networks, given sufficient resources relative to the problem, will effectively implement a "table lookup". Thus the degree to which any sort of "abstraction" takes place or not depends on the specifics of the category and the network resources. Furthermore, the network does not, in fact, form a *representation* of a schema, even where its behaviour depends on extracted regularities, in the sense that connection weights are not readily conceived of as representational, and are better viewed as causal mediators of appropriate activation flow (Hahn & Chater, 1997a).

In summary, standard networks easily meet two of Wittgenstein's claims and present a differentiated picture with respect to the third and, particularly, the fourth, the issue of schema abstraction. Clearly though, abstraction of central patterns and regularities is not a *necessary* feature of their functioning, and thus not a general property of network categories.

4. *Theory based views.* The other main contender in current debate about conceptual structure is the so-called theory-based view (Murphy & Medin, 1985; Medin & Ortony, 1989). The theory-based view is defined primarily in contrast to any account, prototype- or exemplar-based, which seeks to ground real world categories in terms of perceptual similarity. It emphasises the role of background knowledge or "theories" in our everyday classification, in order to explain, for instance, the fact that, despite strong perceptual similarities, we do not classify bats as birds.

Due to its lack of computational explicitness the theory-based view is not that easy to align with Wittgenstein's claims. Given the widespread rejection of the definitional account of conceptual structure in current cognitive theory, one must assume that "theories" are not complete, i.e. allow deduction of classification decisions, but rather only "partial", that is form one component of a complex, non-deductive overall process (Hahn & Chater, 1997b). This overall process, however, which could involve similarity-comparisons, is not generally spelled out by advocates of the theory-based view. The simple claim then that "partial theories" or background knowledge are relevant to categorisation need not conflict with Wittgenstein. There is no statement about boundedness, nor is there a claim of definitional features. Though the theory-based view does suggest that learning and understanding a category also involves acquiring appropriate background knowledge, this does not directly contradict the role of examples in acquisition and use, but merely suggests an additional factor.

The greatest potential for conflict lies in the issue of essences or schemas governing a category. Complete theories, which enable a deductive classification process, would clearly provide essences. But what about partial theories, i.e. how partial does a theory have to be to not be stating "essences"? This is clearly an issue, but given that the theory-based view has done little to provide full accounts of *any* categories no concrete answers are possible. To the extent though, that too much faith is invested in the power of theories, another look at Wittgenstein's arguments and examples might be sobering.

5. *Categorisation as a bi-directional process.* What all of the preceding views have in common is that they view categorisation as an essentially unidirectional process. Exemplars have certain features and/or certain similarities hold between exemplars. Very recent work on categorisation has suggested that this picture is oversimplified. Objects don't come as ready-made bundles of features. Rather, the features objects are perceived to have are influenced in part by the categorisation context. If this is so then similarity relations between them can't yet be fully determinate either, and themselves do not exist entirely independently of the category level.

The case for the flexibility of featural descriptions has been made both on the basis of computational experiences with real-world stimuli(see Hahn & Chater, in press) which

have uncovered limitations of fixed, unchanging representation schemes for certain AI (Branting, 1989) and machine learning systems (Aha, 1992) and from recent experimental work with novel, artificial stimuli (Schyns, Goldstone & Thibaut, in press; Schyns & Roder, 1997). In these experiments, the decomposition of an object into component parts is directly affected by the classifications required of participants in a category learning task.

These issues are to some extent independent of Wittgenstein's concerns, but the rejection of unchanging representation schemes conflicts with definitional accounts, and the influences of the categorisation task resonate with Wittgenstein's emphasis on "use".

Is there really a single account?

Contrasting the substantive content of Wittgenstein's arguments with the leading models and perspectives in current categorisation research demonstrates clearly the continuing relevance of the issues he raised. Where conflicts arise, however, it is not immediately clear who is right. There is no space here to review all relevant evidence, but we would like to claim that Wittgenstein's detailed arguments should at least give pause for thought.

Reviewing these various models and accounts, one can't help but think that most, possibly all, make some important point. Background knowledge often does have a role to play; nevertheless similarity continues to allure; and the arguments for flexibility and top-down influences seem compelling too. One feels oneself pushed towards the inevitable conclusion of undergraduate essays on conceptual structure: that there is a little bit of truth in all accounts. This *is* glib; but might it not also be accurate?

Like these fragmentary insights, Wittgenstein's arguments bear down on any all-encompassing view of category structure. Together, the two appear to effectively explode the idea of the category as a unitary theoretical instrument: how likely is it that, even if categories aren't defining features, shared essences or some other common thread running through, that there is a fundamental unity in all categories? That clear cut members all have higher within category similarity than between category similarity (as predicted by GCM) or that all are based on partial theories, and so on?

Natural language categories and naming is our prime categorisation behaviour - are the products of collectives, not individuals. They develop over time and are subject to diachronic accidents. All of which might reasonably be expected to put a bound on whatever systematicity, at whatever level, we might hope to find. If category structures are like this, then the naming behaviour of the individual must to some extent follow, denying cohesion even at the level of processing. If category structure is variable, then processing too might be expected to be the product of multiple, even competing influences.

We have undoubtedly made great steps forward by developing and testing constrained models. But when we pause to look more at the untidiness of our real world categories, Wittgenstein's scepticism about category *structure* does not seem to find any happy resolution in unitary accounts of cognitive *processing*. The deep questions he poses loom large still.

References

Aha, D. (1992) Tolerating noisy, irrelevant and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies*, **36**, 267-287.

Barsalou, L. (1987) The instability of graded structure: implications for the nature of concepts. In Neisser, U. (Ed.) *Concepts and Conceptual Development* Cambridge University Press.

Bechtel, W. & Abrahamsen, A. (1991) *Connectionism and the mind.* Blackwell, Oxford, UK

Branting, K. (1989) Integrating Generalizations with exemplar-based reasoning. *Proceedings of the Eleventh Annual Meeting of the Cognitive Science Society.*

Cover, T.M & Hart, P.E. (1967) Nearest Neighbour Pattern Classification. *IEEE Transactions on Information Theory*, **13**, 21-27.

Katz, JJ (1972) *Semantic Theory*, Harper & Row, New York.

Komatsu, L K (1992) Recent views of conceptual structure. *Psychological Bulletin*, **112**(3), 500-526

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

Langley, P. (1996) *Elements of Machine Learning*, San Francisco, CA: Morgan Kaufman.

Hahn, U. & Chater, N. (1997b) Similarity and Rules: Distinct? Exhaustive? Empirically Distinguishable? *Cognition*, **65**.

Hahn, U. & Chater, N. (1997b) Concepts and Similarity. chapter in: *Knowledge, Concepts and Categories*, pp. 43-92, K. Lamberts & D. Shanks (eds.), MIT Press.

Hahn, U. & Chater, N. (in press) Real-World Categories Don't Allow Uniform Feature Spaces - Not Just Across Categories But Within Categories Also. Open Peer Commentary on Schyns, Goldstone, & Thibaut, The development of features in object concepts. *Behavioural and Brain Sciences*

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

Medin, D & Ortony, A. (1989) What is psychological essentialism? In S. Vosniardou and A. Ortony (Eds) *Similarity and analogical reasoning.* Cambridge University Press.

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

Nosofsky, R.M (1986) Attention, Similarity and the Identification-Categorization Relationship, *Journal of Experimental Psychology: General*, 115, 39-57,

Nosofsky, R.M (1988) Exemplar-based accounts of the relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, 700-708.

Ramscar, M.J.A. (1997) Wittgenstein and the nature of psychological categories. *Proceedings of SimCat 97*, Department of Artificial Intelligence Conference Proceedings, University of Edinburgh, Scotland, 205-211.

Schyns, P.G. & Rodet, L. (1997) Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory and Cognition, Vol. 23, 681-696.

Schyns, P.G., Goldstone, R.L., & Thibaut, J. (in press) The development of features in object concepts. *Behavioural and Brain Sciences*

Small, S.L. (1997) Semantic category imprecision: A connectionist study of the boundaries of word meanings. *Brain and Language*, **57**:181-194.

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

Taylor, J.R. (1995) *Linguistic categorization: Prototypes in linguistic theory.* 2nd edition. Oxford University Press.

Wittgenstein, L trans. *Anscombe, E* (1953). *Philosophical Investigations* Blackwell, Oxford.