THE ROLE OF SITUATIONAL INFORMATION IN CONCEPTUAL KNOWLEDGE

Susan Huia Anthony

A thesis submitted in partial fulfilment of the requirements of the University of Hertfordshire for the degree of Doctor of Philosophy

The programme of research was carried out in the Department of Psychology, University of Hertfordshire

October 2004

ABSTRACT

This thesis investigated the influence of situational knowledge on the performance of two common tasks; category member generation under a free-emission procedure and the judgement of similarity between two items using rating scales. In both tasks, self-report protocols were used to identify the strategies that people seemed to be using to complete the tasks. The main goal was to identify the role of situational knowledge in the organisation of semantic memory. Traditional models would not predict a role for situational knowledge in either of the target tasks. In the category member generation studies (Chapter 2) participants frequently instantiated situations or perspectives to cue retrieval of category members for both taxonomic and ad hoc categories. Chapter 3 investigated the factors that determine subjective similarity: category type, typicality, context and presence or absence of self-report. The quantitative data analysis showed the need for careful qualifications to previous claims concerning the effect of context on similarity (Barsalou, 1982). Specifically, ad hoc category members were rated more similar with context only when judgements were made without self-report and when items were relatively typical. Self-report protocols showed that co-occurrence of items in a situation frequently entered into judgements of similarity. Chapter 4 investigated the role of events in determining the strength of this 'thematic' similarity. Individual indices of association strength between the items and an event were shown to predict similarity ratings – thus confirming that thematic similarity is driven, at least partially, by the association of items to common settings. The findings lend empirical weight to theoretical positions that present memory for situational information as an integral part of conceptual knowledge. This approach may underpin a new direction for research into concepts in both normal and clinical adult populations.

Acknowledgements

I would like to express my gratitude to Neville Austin and Fred Vallée-Tourangeau for their support and assistance in the early stages of this work and to Diana Kornbrot for her invaluable advice and commitment in the later stages; to Stefanie Schmeer for proof-reading and to Brian, my family and friends who have lost out on my time so often.

CONTENTS

APPENDICES

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

The Theoretical and Empirical Scope of the Thesis

This thesis is concerned with conceptual knowledge. By this term is intended the knowledge that humans have of entities in the world. It will be taken as self-evident that this knowledge is garnered from the individual's experience of those entities. Arguments from philosophy that we are able to have knowledge beyond that which is experienced will be disregarded here since they seem to apply principally to mathematical and scientific concepts and not to the common concepts which tend to be the of interest to cognitive psychologists. It will be assumed therefore that knowledge may be gained from first hand experience resulting from direct interaction with an entity or second hand experience through receipt of information from a third person or media. It seems that memories of these experiences become associated with each other such that separate encounters with, say, a tree, do not remain distinct and unique in our minds but rather form part of a more inclusive memory of all things that we believe to be classified as trees. It will be taken that such close associations between memories of experiences relating to trees, which will entail accumulating knowledge of these items, constitute a concept of tree. Concepts "capture the notion that many objects are alike in some important respects, and hence can be thought about and responded to in ways we have already mastered" (Smith and Medin, 1981, p.1).

The aim of the programme of work was to address a particular question concerning conceptual knowledge. Namely, what is the role of 'situational knowledge' in elementary cognitive operations such as the generation of category members and assessments of similarity between familiar items? 'Situational knowledge' refers to that information which relates to the events, environments and situations in which items are commonly encountered. Theoretical arguments concerning the organisation of conceptual knowledge will be considered in the light of the empirical data. This is an epistemic investigation in that it focuses on demonstrations of what people <u>know</u> about existing, real world entities and how that information is accessed and utilised to support performance in specific cognitive tasks. In addition to the stored knowledge that one may hold in relation to a given item, knowledge is implicated in the study of concepts from the point of view that to recognise an item as x is to experience a 'moment of knowing' (Smith and Samuelson, 1997, p.161). That is to say to recognise a rabbit as a rabbit is to know that you have seen a rabbit. Not everyone is content to assume that in studying concepts we are asking questions about conceptual knowledge. Indeed, Fodor (1998; 2004) is highly critical of assumptions that 'questions about what a concept may be' depend upon answers to questions about 'what it is to possess a certain concept'. In Fodor's view, equating concepts with epistemic capacities has represented a 'wrong turn' on the part of psychologists and cognitive scientists. Fodor would agree that concepts carry information but not in the form of a related list of propositions. The issue of how researchers may best understand what constitutes a concept will be returned to in the final discussion of this thesis. It will be argued that psychologists may legitimately pursue questions of what sort of knowledge people access and utilise when they engage in tasks that require them to think about categories of items without necessarily taking a particular stand on which parts of that knowledge (if any) may reasonably be called a concept.

In an attempt to maintain clarity and consistency in a minefield of terminology, this introduction will next deal within interpretation of the key terms *concept* and *category* followed by related issues. A summary of traditional theories concerning the nature of conceptual organisation and also more recent, relevant theoretical developments will be provided, followed by a brief background to the two main tasks that have been chosen to investigate the role of situational knowledge; category member generation and similarity judgements. The conclusion of this chapter will highlight the aims and preview the empirical content of the thesis.

Clarification of terms

Concepts and Categories

It has been claimed that *categories* are classes (kinds) of things that exist in the world, and *concepts* are corresponding mental representations of them. Medin (1989), however, counselled against this intuitively appealing notion for two reasons. The first is the existence of empty concepts, i.e. concepts that have no real world referent, such as *unicorn* and *Father Christmas*. The second is that the claim implies realism about categories in the world that may not be appropriate. This latter objection alludes to a philosophical debate concerning whether there are in the world objective kinds of entities based on essential properties that humans notice and learn, or whether maximally informative mental arrangements of information are projected, as categories, out into the world (see Malt, 1995). Medin prefers the following distinction; " a *concept* is an idea that includes all that is characteristically associated with it. A *category* is a partitioning or class to which some assertion or set of assertions might apply." (Medin, 1989, p.1469). By contrast, Barsalou (1993) rejects the use of the term concept for all the information associated with an idea. He maintains that engagement in cognitive tasks requires only partial activation of knowledge of any given entity, and wishes to reserve the term concept for the temporary representation of that subset of knowledge which is constructed in working memory to meet a specified goal. However, many authors often fail entirely to make explicit the intended referents of the terms; *concept* and *category* are frequently used interchangeably in the psychology literature.

The term concept is also on occasions conflated with meaning, so that one's concept of *cat* is often taken to be what one 'understands' by the term *cat*. Laurence and Margolis (1999) move between theories from philosophers of language concerning the meaning of words (Kripke, 1972; Putnam, 1975), linguists (Jackendoff, 1987) and theories of concepts. The authors freely acknowledge this; " Throughout we'll ignore certain differences between language and thought, allowing claims about words to stand in for claims about concepts" (Footnote 12, p.11). However, the authors make no argument for the legitimacy of this position. Fodor (1998) also states his intention to "move back and forth pretty freely between concepts and word meanings; ...it may turn out in the long run, for the purposes of the present investigation word meanings just are concepts" (p.2).

It is perhaps unsurprising that meanings of words and concepts are often seen as equivalent given that influential early work in empirical psychology concerning concepts (Collins and Quillian, 1972) sprang from attempts to explain text comprehension (Quillian, 1968). Furthermore, theories of word meaning frequently appear to be readily applicable to concepts. To equate meaning with a mental particular (i.e. a concept) is to lose its essential role in reference. Putnam (1975) maintained that a stereotype, i.e. a linguistic description (as in a feature list) of an item forms part of its meaning but only a part. The actual entity in the world that is picked out by its linguistic label also forms part of the meaning. Barsalou,Yeh, Luka, Olseth, Mix and Wu (1993) indeed argue strongly that meanings are not equivalent to concepts. Under their view of a concept (a frame that encompasses specialised models of individuals), concepts may assist in establishing reference but since they do not have physical counterparts in the environment (individuals do have counterparts but concepts do not) it is not possible that they should refer. If reference is accepted as an essential part of meaning and if concepts, as envisaged by Barsalou et al. do not refer then concepts cannot be equivalent to meaning. Naturally one may take issue with the need for meaning to involve reference and/or with Barsalou and his colleagues' notion of concepts but it is clear that a considerable amount of debate has taken place around this is issue and a consensus has yet to be reached on its resolution. It is not intended that the issues investigated in this thesis should be discussed in relation to meaning and so to err on the side of caution, it will be assumed that whilst it is not necessarily the case that theories of word meaning cannot explicate concepts, the equivalence of word meanings and concepts must be carefully questioned and argued for rather than assumed to be the case

Throughout this thesis the Medin (1989) use of *concept* will be adopted, that is " a concept is an idea that includes all that is characteristically associated with it" (p.1469). To think of a concept, as Medin does, as the total knowledge associated with an idea, represented in some format in memory, is the most traditional way of using the word. However, it will be seen from the literature review that the scope of the 'total knowledge' is often limited to perceptual and functional features. The term *category* will be used throughout to refer only to the accepted culturally driven partitioning of items in the world. There are clearly a number of topics that are closely related to concepts, such as semantic memory, representation and categorisation. The nature of their relationship to concepts will now be briefly addressed.

Semantic Memory

In many texts on cognitive psychology, semantic memory is described as that part of memory that deals with 'general knowledge' about the world. This may include general facts ('the present Queen is Elizabeth II') and also facts about items such as 'trees have branches'. This, then, would be the natural home of conceptual knowledge. What characterises semantic memory is that the information therein can be accessed without recall of the specific episode in which it was first learned. Tulving (1972) first contrasted semantic memory with episodic memory. The latter was used to refer to "memory for personal experiences and their temporal relations, while semantic memory is a system for receiving, retaining, and transmitting information about meaning of words, concepts, and classification of concepts" (p.401-402). Tulving considered that there were sufficient "fundamental differences" (p.384) between the two to merit separate consideration. Tulving originally suggested this classification as a heuristic which could aid memory researchers by facilitating laws and principles which may explain one set of related phenomena independently of those which govern the other domain. He did not, however, rule out the possibility of reciprocal influence between the two systems. Despite his disclaimer that the naming of these systems did not necessarily express "any profound belief about the structural or functional separation of the two", the distinction became increasingly reified in the literature over the next 10 -15 years. Tulving subsequently updated his conceptual model (the 1984 version proposes that episodic memory is embedded in semantic memory as opposed to being a distinct system), but the distinction was maintained despite heavy criticism (see McKoon, Ratcliff and Dell, 1986) and the proposal of single storage models which would account for data previously held to support the distinction (Anderson and Ross, 1980).

Glenberg (1997) in a radical review article, claimed "I am explicitly equating episodic and semantic memory systems in the sense that there are no separate episodic and semantic memory, hierarchically arranged....or otherwise." (p.8) and further "thus the distinction between episodic and semantic memory probably reflects a difference in the frequency with which memories are used, the methods of assessment, and the context of the information, rather than any intrinsic differences in memory systems" (p.9). The need to specifically refute the distinction bears witness to the influence that it continues to exert in people's thinking about the organisation of memory, and of conceptual knowledge. This issue is raised because throughout this thesis, a case is made for greater recognition of the role played by knowledge of situations and events in the way in which we think about objects. Whilst the knowledge of relevant situations may well be itself generalised and not limited to specific episodes, this approach almost certainly requires that semantic memory should not be conceived of as a separate form of memory.

Concepts and Mental Representation

The relationship between concepts and representation is dependent on a particular perspective of what characterises cognition. Van Gelder (1995, p.375) claims that " Cognition is distinguished from other kinds of complex natural processes (such as thunderstorms, subatomic processes, etc.) by at least two deep features: on one hand, *a dependence on knowledge*: and distinctive kinds of complexity..." (emphasis added). The dependence on knowledge is subserved by mental structures that can represent that knowledge. These structures, as symbols, can be systematically manipulated. This is the basis of the computational view of mind that pervades cognitive psychology and cognitive science. The notion of representation is thus central to this view and the mental representation of x is one's total knowledge of x, held in long-term memory and is thus equated with one's concept of x.

Concepts and Categorisation

One of the most highly investigated functions of concepts is the way in which they permit categorisation. By virtue of having a concept, one is able to make decisions about which things fall under that concept. Performance on categorisation tasks has been therefore used to infer the nature of conceptual structure. Note that identifying an item as falling under a particular concept, e.g., that is a table, that is a chair etc., automatically involves assessing the willingness of participants to apply a particular word (concept label) to a particular item and therefore overlaps again into areas of word meaning and use. One should accordingly be wary of whether these tasks really offer an insight into the organisation of conceptual structure (by indicating concept boundaries etc.) or whether they indicate only something about the way in which people are prepared to apply naming words.

An alternative means of investigating conceptual structure is to examine the way in which participants learn classifications of artificially created stimuli. As Smith and Medin (1981, p.6) argue, " if natural concepts are hypothesised to have a particular structure, one can build this structure into an artificial class and see if people use this class in the same way they use a natural one; if they do, we have added support for the hypothesised structure of natural concepts". It is possible that such methodologies have had the effect of placing an extreme emphasis on the role of intrinsic features of objects in categorisation tasks, which in turn has led to a similar over-reliance on these features in constituting the concept of the object. Consequently, the methodology in this thesis does not rely on artificial classifications created by the experimenter.

The following section will provide a review of the most influential theories regarding the organisation of conceptual knowledge.

Theories of the Organisation and Structure of Conceptual Knowledge

Psychological theories concerning the organisation of human conceptual knowledge have evolved through a process of empirical testing over the past 50 years. The earliest psychological research adopted, from classical philosophy, a rule-based notion of category membership that required objects to possess defining features. The inability of such theoretical positions to account for empirical data from human conceptual systems that operate to meet the demands of a complex, real world has obliged the development of more sophisticated theories. In common with Komatsu (1992) and Smith and Medin (1981), this review will initially consider psychological theories of conceptual organisation to fall under one of five descriptions; the classical, the family resemblance, the exemplar, the schema, and the explanation or theory based view. With perhaps the exception of explanation-based theory of concepts, all of these theories place the notion of shared or matched features at the heart of the theory. Items are organised in such a way as to maximise the intra-category similarity and minimise the inter-category distance. The similarity is based on features. It is the nature of those features that is of particular relevance to this thesis. Although these theories are well known, they are described again here with a particular view to highlighting the type of features that have been assumed to form an appropriate basis for conceptual organisation. It will be seen that these tend to be perceptual and functional properties. There is little, if any, mention of the settings and situations in which the objects appear. A consideration of more recent and possibly controversial views of concepts will follow these traditional positions.

The Classical View

The classical view argues that the features of an object are individually necessary and jointly sufficient to guarantee membership of the category defined by those co-occurring features. For example, the concept of square may be defined by possession of the following four properties, (1) it is a closed figure, (2) it has four sides, (3) its sides are equal in length, and (4) its internal angles are equal. These properties jointly describe all items falling under this concept and any item under this concept will exhibit all these properties. While it has been pointed out that few psychologists have supported this view in its baldest form (Komatsu, 1992), early experimental work was at least partially based on the assumption that category membership can be defined by rules and that these rules can be cashed out in terms of features. The most famous example of this approach is Bruner, Goodnow and Austin (1956). In this study, participants were shown a range of geometrical shapes that could be grouped in various ways, i.e. by the number of shapes on the stimulus (1, 2 or 3) or by the nature of the shapes (cross, circle, square etc.). Participants were presented with one instance of a category which the experimenter had in mind and then had to identify the category (rule) by asking if other stimuli were also in that category and receiving feedback. Participants adopted strategies for learning the category that were in line with a defining feature approach. The 'features' in this paradigm were purely perceptual.

By the 1960s, models of conceptual knowledge were still heavily based on feature-lists but admitted of members that did not share all features with all other members. For example, Collins and Quillian (1969) proposed a model of the organisation of the type of memory required to verify sentences such as "A canary can fly". In their model, categories are arranged in a hierarchy as shown in Figure 1.1.

Properties which are true of all category members are stored only once at the highest level of category to which it applies, e.g., 'has wings' and 'can fly' are associated not with *canary* but with its superset, *bird*. In order to decide whether 'A canary can fly' is true, it would be necessary to establish first that canary is a bird and then to infer that it can fly from the fact that birds can fly. This proposed organisation could claim advantages in terms of economy of storage (it is not necessary to store all properties with all members of any given category) and led to testable hypotheses about the length of time it would take participants to verify the sentence 'a canary can fly' compared to the sentence 'a canary is yellow', where yellow would be a property stored with *canary* and not at the higher level of *bird* since it is not true of all birds. Of course, it is also not true of all birds that they can fly but notable exceptions, such as *ostrich* and *penguin* would have this fact stored at the level of *ostrich* and *penguin* to block the inference from the higher category.

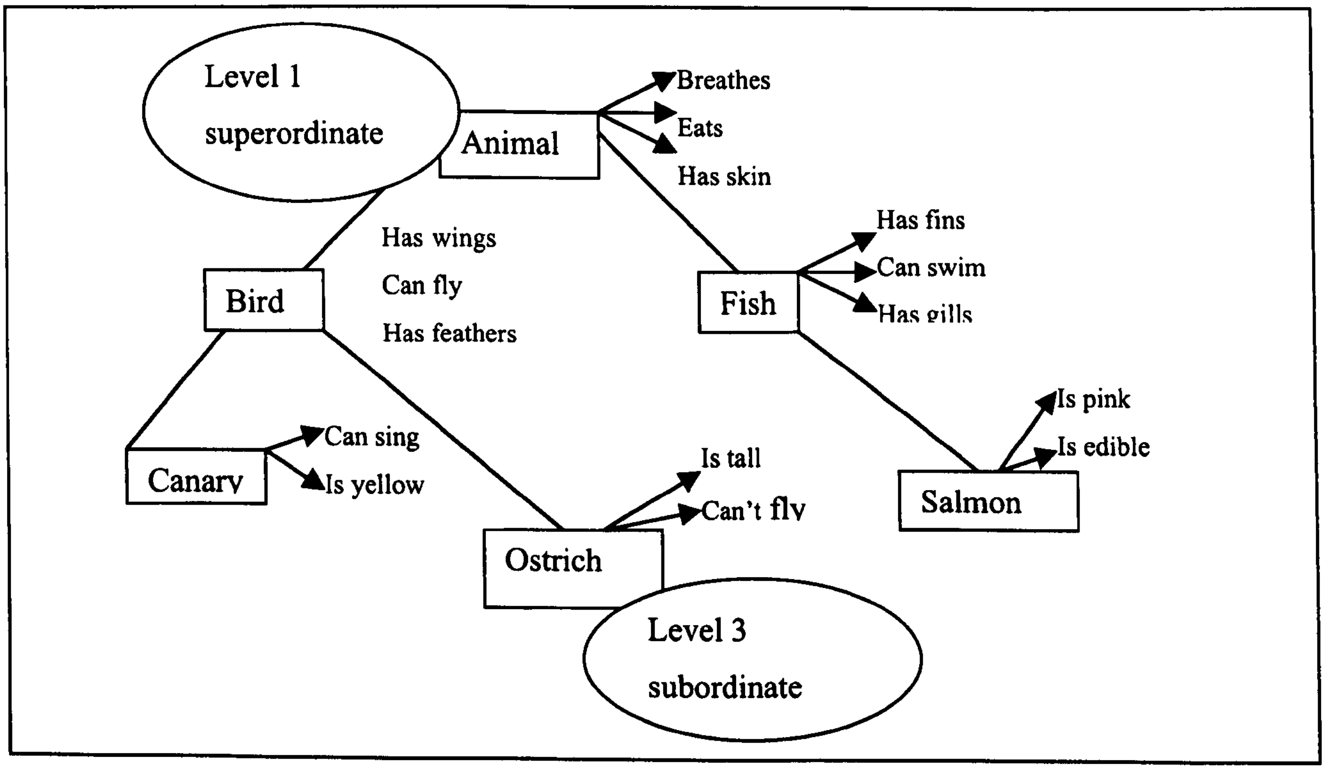


Figure 1.1 Hierarchical model of concept knowledge (Collins and Quillian, 1969).

In summary, the important features of the model were:

1) It is hierarchical
2) Features (or attributes) are associated with the concept at different levels of the hierarchy
3) Subordinate nodes inherit features from superordinate nodes
4) A departure from the classical view is marked by the recognition of members which fail to share a feature with all other members e.g. ostriches can't fly
5) The distance shown diagrammatically in the model is also intended to represent psychological distance
6) The features mentioned in the model are a mixture of perceptual ('is yellow") and what have become to be known as 'functional' (e.g. "can fly").

Collins and Quillian used two types of statements in sentence verification tasks to test different aspects of the model. They referred to these as 'superset relations' sentences, e.g. 'A canary is a bird' and property relations sentences, e.g. 'A canary can fly'. Data from sentence verification tasks using statements like these lent support to claim (5) above and in general to the organisation depicted by the model. That is to say participants were able to assert the veracity of the sentence 'Is a canary a bird?' more quickly than they could the sentence 'Is a canary an animal?' With regard to properties, participants took longer to verify the sentence 'a canary can fly' than they did 'a canary is yellow'. In a further test of the Collins and Quillian model, Conrad (1972) found that when participants were asked to describe named categories, certain properties were mentioned more frequently than others. Moreover, the properties mentioned most frequently tended to be those associated with level 1 superordinates and those mentioned least frequently were those associated with level 3 subordinates. This suggested that the trend of increased reaction times on the property relation sentences (from level 1 to level 3) could have been due not to the need for inferences to be drawn from the higher level but to the comparative difficulty of retrieval of these properties. In other words, the fact that it takes longer to verify that a canary can fly than it does to verify that a canary is yellow is because the property *fly* is less frequently associated with canary than the property yellow. Conrad (1972) ran two experiments that showed reaction time to be a function of property frequency that was not controlled for in the Collins and Quillian study. Conrad claimed support was found for the hierarchical nature of the model but not for the storage economy hypothesis. Conrad (1972) concluded by suggesting that "properties are stored in memory with every word which they define and can be retrieved directly rather than through a process of inference." (p.154). In as much as the Conrad work suggested that not all properties are equally salient to people in describing categories, this also challenged the notion inherent in classical theories that all features weigh equally in the concept.

Two further problems for the classical view arose during the 1970s. Firstly, if being a member of a category is just a matter of possessing the appropriate set of features, then all members of any given category should be equally representative of that category. Rosch (1973) found that participants reliably rated some category members as "better" (more typical) than other members. Rosch and Mervis (1975) asked subjects both to generate features to common category members and to rate the typicality of the category member. They found that the typicality rating of any given category member was a function of the number of overlapping features between that item and other members.

Secondly, McCloskey and Glucksberg (1978) found evidence of inconsistency in participants' judgements of category membership for given items. Participants were presented with pairs of members and categories, e.g., fly *-insect*, strawberry-*fruit*, fork-*kitchen utensil* etc. Their task was simply to indicate (yes or no) whether they thought the item (named first) was a member of the category (named second). Participants performed the same task with the same materials again after one month. Both between subjects at session one and within subjects across the two sessions, a high degree of consensus in response was found for items which had been independently rated as highly typical or highly atypical. However, there was far less consistency for items which received intermediate ratings of typicality, with the highest level of disagreement occurring for items rated at typicality level 4 on a 10 point scale where 10 was highly typical. These results suggested that boundaries between common categories were not clear and distinct, as implied by classical theory, but were "fuzzy".

Perhaps most importantly, it transpired that it simply was not possible to identify a set of defining features for most everyday categories.

The Probabilistic View and Prototypes

The theories that largely superseded classical views retained an important role for shared features in explaining the constitution and coherence of concepts but discarded the idea that a certain set of features was necessary and sufficient to guarantee membership. Instead a notion of *family resemblance* was invoked. Under this view, members of categories shared some overlapping, characteristic features but there were no features that were necessarily common to all members. Most notably these ideas were expounded by Eleanor Rosch, following the famous observations from Wittgenstein (1953) concerning the absence of defining features for the category 'games'.

To explain the mental representation of a category based on family resemblance, *prototype* theories were invoked. The notion underlying prototype theories was that as one encounters category exemplars in the world, certain repeating features are recognised and a representation of the group, which captures the 'average' or modal features of its instances, is abstracted from these experiences. The representation may include non-necessary features and the features are weighted to reflect their salience and the conditional probability of their association with the concept. This representation is called a prototype and category decisions about new instances of entities are a function of the degree of similarity between the instance and the prototype. Under this theory, category membership is not "all or none" but is more or less probable - for this reason the theory falls within the group of probabilistic theories of categorisation. The family resemblance approach predicted the "fuzziness" of category boundaries as found by McCloskey and Glucksberg (1978). A further prediction of the theory was that some members of a category would be considered to be "better" members by virtue of being closer to the prototype than other members. As mentioned above, that some category members are perceived by subjects to be better examples (more typical) of the category than others was found to be the case by a number of researchers (Labov, 1973; Rosch, 1973, 1975; Rosch and Mervis, 1975). The phenomenon whereby members of a category are seen as varying in "typicality" relative to that category gives rise to 'typicality effects', i.e., typical members have a privileged status across a range of tasks. Typical members are generated to the category label more easily and are identified as category instances more quickly in comparison to less typical members. The fact that a category has members that vary in terms of their perceived typicality is referred to as the category having "graded structure". It should be noted that the discovery by Armstrong, Gleitman and Gleitman (1983) that well-defined categories such as *even numbers* also exhibit graded structure cast some doubt on the assumption that typicality effects demand explanation by prototype.

Hampton (1998) asserted that the similarity involved in assessing candidates for category membership against a category prototype is not restricted to a comparison of perceptual features. He stresses that even in Rosch's (1975) seminal work on prototypes, categorisation was envisaged to involve an evaluation of a number of different types of feature including functional features, origins and common ways of interacting with the item. Hampton's paper was inspired by the studies conducted by Rips (1989) which purported to demonstrate a dissociation between similarity and categorisation. That is to say that items could be judged to be more similar to one category and yet considered be more likely to be a member of a different category. Using previously published data of typicality and likelihood of category membership (McCloskey and Glucksberg, 1978), Hampton showed that a high proportion of variance in category membership judgements could be accounted for by typicality (similarity to prototype). There were deviations from the monotonic relationship between typicality and categorisation and Hampton undertook a further study to test post hoc explanations for these deviations. He found that for biological kinds, there was sometimes a conflict between typicality and what participants felt may be technical reasons for including or excluding an item in a category. For artifacts, a highly typical item within one category might nonetheless receive a low categorisation score as participants were aware of other categories where it may be classified more appropriately (e.g., a hammer may be fairly typical as a weapon, holding many appropriate features, but be felt to be better classified as a tool.

Hence, prototype theories do not always predict category membership as a simple function of similarity to the prototype but Hampton would argue that they need not lose credibility as models for categorisation for all that.

Prototype theories can deal with some of the empirical data which classical theories struggle to explain (typicality effects, fuzzy category boundaries), but these theories are also subject to criticism. The main thrust of the attack on prototypes as a comprehensive theory of concepts is on their inadequacy in conceptual combination. Fodor and Lepore (1996) maintain that not only do many complex concepts fail to give rise to a prototype altogether (e.g., new species) but even in cases where they do, the prototypes of the constituent concepts do not combine to produce the prototype of the complex concept. For example, the prototype of *pet fish* would probably resemble a goldfish, but a goldfish is likely to be far removed from the prototype of either *pet* or *fish*. One option for prototype advocates is to adopt a Dual Theory. A dual theory proposes that concepts comprise both a conceptual core, which is classical in nature, and a prototype component that supports identification procedures. Whilst this allows the retention of prototypes, alternative views to prototype theories have been proffered.

Exemplar-Based Theories

Prototype theory proposes that the representation of a category abstracts and captures the most frequently appearing features and characteristics of its members. In contrast, some theorists have seen no need to posit a stored, central representation of a category.

Exemplar theories, in their most extreme form such as that of Reed (1972), propose that information concerning <u>all</u> encountered instances of a category is retained. The mental representation of the category coheres through the similarity of the accumulated instances. This approach does not rule out the possibility of the construction of an abstracted central representation but suggests that this would occur at point of use (i.e., when identifying a potentially new member, judging typicality etc.) and therefore need not form part of one's stored concept. However, the effect of typicality on speed of classification need not be explained by construction of a central representation at all. When confronted with a new instance for classification, the exemplar(s) that come to mind will be the ones encountered most frequently (these are likely to be the most typical, e.g., apple for fruit), so if the instance to be classified is also typical (e.g. orange), the verification response will be fast but if the instance is atypical (e.g. kumquat) then the response will be slowed by its dissimilarity to the apple exemplar. Brooks (1987) argued for the influence of the processing of prior episodes on classification and identification tasks. Brooks wished to distinguish between analytic and non-analytic knowledge and processing. By nonanalytic, he refers to a reliance on analogy "...everyday analogizing is ...nonanalytic in that a great deal of information is inherited in a flood and little effort is made in the current processing to selectively process categorically relevant information. Relative emphasis on attributes is dictated more by what had been important for accomplishing the prior purposes rather than what is important for the category as a whole." (p.142) Brooks (1987) argues that models that prioritise the role of specific experiences in handling exemplars can offer a more natural account of the demonstrable flexibility of typicality ratings than can theories which propose that typicality is tied to more stable, abstracted representations.

A further advantage of exemplar theories is that they preserve information about the correlation of features within categories. Prototypes, in abstracting "average" features from a range of exemplars, lose this type of information. Consider the following example from Reisberg (1997, p.320). Two softball teams, Red and Blue, have members who vary in their height and weight (seeFig.1.2.).

	Blue Team			Red team	
	Height	Weight (lbs.)		Height	Weight (lbs.)
George	6'3"	120	Fred	6'4"	250
Tina	4'1"	240	Sam	4'1"	110
Lee	6'4"	250	Susan	6'3"	240
Alyson	4'2"	110	Jane	4'3"	90
Tom	4'3"	160	Chris	6'2"	240
Mary	6'2"	170	Jeff	4'2"	120
Average	5'2.5 "	175	Average	5'2.5 "	175

Figure1.2 Using correlated features in categorisation (from Reisberg, 1997).

The average weight and height of red and blue team members are equal. On the Red team however, there is a positive correlation between weight and height (tall members are heavy and short members are light). This correlation does not hold in the Blue Team. A new player arrives on the field who is tall and thin - to which team is this player most likely to belong? Consulting a representation of each team which captures the average value of the salient characteristics, weight and height, would make membership of the Blue and Red teams equally probable. Comparison of the mystery player to individual exemplars of each team would incline one to opt for the Blue team since this team has at least one individual who is both tall and light-whereas in the Red team all the tall players are heavy. Medin, Altom, Edelson and Freko (1982) provided an empirical demonstration of this sensitivity to feature correlation in the context of diagnosing fictional diseases from symptoms.

Exemplar models over prototype theories also capture information about the variability of members in a way that prototype models fail to do. One of the explanations invoked for the earliest demonstration of a dissociation between similarity and categorisation (Rips, 1989) was subjects' knowledge of intra category variability. In the Rips study subjects were asked (for example) whether a 3-inch round object was more similar to a pizza or a quarter, and whether it should be categorised as a pizza or a quarter. Subjects judged the object to be more similar to a quarter but categorised it as a pizza. Subjects' knowledge that pizzas are more variable in their size than are quarters was cited as the reason for the preferred categorisation.

Exemplar models can also account for a phenomenon that prototype models cannot, i.e., the research findings that the perceived typicality of category members may vary with context (Barsalou, 1985; Roth and Shoben, 1983) and with points of view (Barsalou, 1987). If the prototype of a category is a relatively static representation of the category, then how can comparison of the same item with the prototype yield different decisions on different occasions? The exemplar view, however, can account for this by proposing that different exemplars can be retrieved from memory against which to compare the new instance, depending on the circumstances in which you are making the comparison.

It should be noted that exemplar models vary in the degree of abstraction that they incorporate. As mentioned, the most extreme views propose the storage of every experienced instance. Such a position is best encapsulated by Reed's Proximity model (1972). However, in other models exemplars are subsets of larger categories (e.g. a robin is an exemplar of *bird*). In these models, a degree of abstraction must have already taken place across all instances of robin to reach the exemplar for this subset. It is not the case then that exemplar models necessarily offer no role for abstraction but as Medin and Smith (1981) express it " our safest conclusion is that exemplar-based representations show a substantially greater lack of abstraction than representations based on classical or the probabilistic view" (p.146).

It appears then that the exemplar view can explain the same data (typicality effects etc.) as can the prototype view but it can additionally account for phenomena (such as knowledge of variability, awareness of feature correlation, flexibility of typicality judgements) which pose a problem for prototype theory. However, it is clear that people do have knowledge of prototypes and that their performance on certain tasks is influenced by representation of central tendencies rather than by individual exemplars (Posner and Keele, 1968; Franks and Bransford, 1971). Though as previously stated, since exemplar theory protagonists do not deny subjects' ability to form prototypes "on-line", these data do not necessarily speak against their position. People's knowledge of both central tendency information for categories and detail of individual instances has led some to claim that conceptual knowledge is comprised of both these types of information which can be used as the task demands (Anderson, Kline and Beasley, 1979). Yet again, a combination of theoretical positions is invoked to offer a more comprehensive account of the data. It should be noted that these theories offer no particular view as to what type of features would enter into comparisons between newly encountered items and stored exemplars. It is also important to note, for the purpose of the argument to be made later, that although exemplar models do not explicitly claim that the exemplars are stored without contextual information, this is certainly implied.

Schema Views

Rumelhart (1980) claimed that " a schema...is data structure for representing the generic concepts stored in memory. There are schemata representing our knowledge about all concepts" (p.34). It is easy to see why this theoretical form of conceptual structure would be a candidate for describing conceptual representation. By aligning the variables posited in schema theory (slots) with concept features or attributes, it can be seen how the schema view of concepts can capture both information abstracted from repeated experiences (default values) and information pertaining to specific experiences (exemplars). Schema models can therefore be considered exemplar hybrid models that allow for information of central tendency and individual instances to be stored simultaneously (Komatsu, 1992). The key advantage of schema over both prototype and exemplar models is that they permit the representation of knowledge about relationships between concepts and of relationships between the constituent attributes of concepts. The need to include such information was at least partially demanded by the realisation that a simple feature matching form of similarity could not fully account for category membership decisions. Reisberg (1997) illustrates this with the following examples. A lemon which has been painted with red and white stripes, injected with sugar to make it sweet and squashed flat is still categorised as a lemon despite not sharing any superficial features with either a prototype of lemon or any experienced exemplar. Contrarily, a counterfeit paper note ($10 bill) shares a full set of superficial features with a real currency but would not be classified as such.

It seems in such cases that category membership is determined by knowledge not only of superficial features but also of " deep" properties (genetic constitution in the case of the lemon and the relevant printing history of the counterfeit bill) which somehow reveal the "essence" of the item. Keil (1989) offered evidence of these types of judgements in children. Pre-school children could not be persuaded to reclassify a skunk as a racoon simply because it had been treated in such a way as to superficially resemble a racoon. These data do not rule out the role of similarity in categorisation but if it is to be claimed that categorisation is still a function of similarity to a prototype or an exemplar, then it must be similarity based on shared "deep and essential " properties rather than surface features. The question remains as to how these important characteristics come to be known as the ones most salient to category decisions. This is explained in terms of "a widespread fabric of beliefs" (Reisberg, 1997, p.335). It is knowledge of functional and causal relationships between concepts - a relevant theory of interconnections - which supports the weighting of some features as being more important to categorisation than others. Schema theory can encompass the representation of this form of knowledge as well as of features and attributes. In this respect, schema theory can be seen as a form of "explanation based theory".

Explanation Based Theories or Theory-Theories

In 1985, a highly influential article was published by Greg Murphy and Doug Medin in Psychological Review. The authors argued that theories relying heavily on similarity to explain the coherence of mental categories, such as the prototype and exemplar views, were too simplistic in that they failed to account for the knowledge we have of the relationship <u>between</u> features of entities (e.g. things fly <u>because</u> they have wings) and of the relationship between those entities in the world. Their proposal was that concepts are organised by people's theories (explanations rather than scientific theories) about the world and are not just collections of features or attributes. Murphy and Medin use the example of seeing a man at a party jump into a swimming pool fully clothed. They argue that "jumps into a swimming pool fully clothed" is not a feature of the category "drunk" which can be compared to a prototype or previous exemplars (you may categorise this man as drunk without ever before having seen a drunk behave in that particular manner). The category decision is guided by a theory about inebriation and behaviour that explains the man's actions. Empirical evidence for the theory view of categorisation arises from work with young children. For example, as previously mentioned, Keil (1989) reported an experiment in which children were told a story of an animal that looks exactly like a racoon but it has the internal organs of a skunk, it has skunk parents and gives birth to baby skunks. The older the children, the more likely they were to say that the animal is a skunk and not a racoon. Younger children seemed to be more influenced by perceptual similarity. These results were interpreted in terms of the older children having developed more sophisticated theories about the world which guide them to pay attention to the biological and genetic features of the animal when making their category decision and to disregard the shared superficial features.

Studies with adults concerning the effect of background knowledge have also lent empirical support to the idea that theories relating features to one another may account for category coherence better than notions of similarity between members based on shared features. For example, Kaplan and Murphy (2000) examined the effect of learning novel categories that consisted of differing sets of 5 real characteristic features with each set having one idiosyncratic feature that was related to a theme or 'consistent knowledge structure'. For example, one exemplar for a hypothetical vehicle category could be described as having vinyl seat covers, an airbag, a front licence plate, two doors and manual transmission (characteristic features). It would additionally be described as having one of a set of idiosyncratic features that linked it either to tropical use or arctic use (e.g. is made in Africa or is made in Norway). Participants saw six exemplars from each category to be learned. The exemplars from any given category either included idiosyncratic features from the same theme or from one of two different themes. It was found that where the idiosyncratic features all related to the same theme the categories were learned faster than where the category members had idiosyncratic features from two different themes. Post experimental questionnaires showed that those who had noticed the themes learned faster than those participants who did not report having noticed the themes. Furthermore, the participants who saw consistent themes were better at learning the idiosyncratic features than the characteristic features. Kaplan and Murphy suggest that normal models of categorisation based on numbers of shared features would not predict these effects as all exemplars shared 5/6 of the same values. The authors argue that it was the background knowledge, linked to the thematic features that facilitated learning in the single theme conditions. Importantly the idiosyncratic features themselves varied and had to be noticed in relation to each other during learning in order for a theme to be constructed and used. These findings, that defy explanations of category learning based on similarity of features between exemplars and illustrate the role of theories in such tasks, lend weight to theory-theories of categorisation.

It should be noted that this theory does not claim that features/properties are unimportant to the categorisation judgement but that the way features are weighted in category decisions depends upon one's background knowledge and theories about how the world works.

Alternative views

It was claimed on page 3 that Medin's (1989) notion of concept included "all that is characteristically associated with it". However, it can be seen from the review so far that the favoured candidates for "associations" are features and possibly relations between features. Other "associated" knowledge is either not mentioned at all or is accommodated by 'joining it on' to the concept in some manner. Barsalou (1991) proposed that world models are constructed around entities (standard object categories) and that the models are people's knowledge of locations in the environment, together with the knowledge of entities and activities that exist currently in those locations" (p.53). So although this body of extra conceptual knowledge was acknowledged it was not treated as part of the concept itself. Features and featural relations have been accorded a special status in this respect. Theories of concepts have been based on the "theoretical assumption of descriptionism" (Millikan 1998, p.55). There are, however, two theories that have a radically different view of the role of features altogether – essentialism and conceptual atomism.

Essentialism

The term 'psychological essentialism' was coined by Medin and Ortony (1989) to capture the observation that people seem to reason about category membership, not on the basis of observable, external properties but on the assumption of category members sharing some unobservable, underlying essence. Moreover, they seem to believe that it is the nature of that essence that gives rise to those observable, external features that are commonly used for identification. Under this view, being a member of a category is a matter of possessing the right kind of essence of that category, rather than a matter of sharing superficial features. The theory-theory of concepts discussed earlier can be seen as linking to this view by suggesting that theories can account for the relationship between features and thus explain the coherence of the category. A body of evidence from conceptual developmentalists has suggested that even young children demonstrate beliefs in such essences. Gelman and Wellman (1991) presented 4-year olds with stories such as that of a newborn cow that was taken to a farm to be raised with pigs and was never allowed to see another cow. The children at both the lower and upper limit of the age group expressed beliefs that the newborn cow would exhibit cow-like properties when it was grown, rather than features of pigs. Gelman and Wellman describe the children as having beliefs about the 'innate potential' (p.364) of different kinds of species. Gelman and Wellman (1991) showed that children as young as 3 years old could identify two members of a superordinate category as having the same 'insides' (e.g. a pig and a cow). This can be seen as laying the foundations of beliefs about internal essences as they relate to category membership. It has been argued that belief in essences is not necessary to account for the empirical findings in the developmental literature. Strevens (2000) suggested that beliefs in what he terms K-laws will equally explain the observed effects. K-laws are "causal laws that connect kind membership with observable properties" (Ahn, Kalish, Gelman,Medin,Luhman,Atran, Coley and Shafto, 2001, p.59). Under this 'minimalist' position, Strevens claims that it is sufficient for people to believe that it is something about being a cat that causes it to purr but it is not necessary to hold any particular view of what that 'something'; might be. Ahn et al. (2001) refute this view and claim that the Strevens position fails to fully explain the theoretical and empirical evidence.

Much of the empirical evidence related to arguments around essentialism seems to arise from studies with children. This may be because adults' understanding of biological domains is assumed and the question has therefore been – how early are these beliefs acquired? As mentioned earlier (Reisberg, 1997), adults are not expected to be fooled by storied of item transformation into believing that a member of a biological category has changed category by virtue of undergoing superficial changes (e.g. the red, striped lemon.). However, Hampton (1995) reported that when adult participants were told of an offspring from two zebras that was provided with a special diet grew to look, and behave, just like a horse (having no stripes and being of a uniform brown colour), two thirds of the participants did not agree that this animal was really a zebra.

It should be noted that these views deal with the way people seem to reason about concepts rather than addressing what concepts might actually be (in contrast to the project of Fodor's conceptual atomism – to be discussed shortly). However, essentialism clearly sees features or properties as effects of *whatever it is that consitutes the concept* (author's italics) rather than the playing the key role in determining category membership in the way that classical or prototype theories posit. This means that arguments, such as that made in this thesis, concerning the role of situational features in theories of categories and concepts would not impact on a theoretical positions of essentialism at all.

Conceptual Atomism

Classical, prototype and exemplar theories differ in their view of which features are constitutive of a given concept and which are 'collateral' (Laurence and Margolis, 1999) – but they all allow that features play some role in our concepts. Conceptual atomism is distinguished by its assertion that the content of a concept is "constituted, exhaustively by symbol-world relations" (Fodor, 1998). Fodor maintains when the proponents of most of the well-established theories in psychology posit features as the concept, they are working under an assumption of Inferential Semantics. This is to say that concepts are what they are (at least partly) by virtue of their inferential relations to other concepts. For example, holding the concept *apple* infers the holding of other concepts such as *fruit, round, skin, pips* etc. One of Fodor's principle reasons for rejecting this as a plausible theory of concepts is that it has not been possible to produce a principled way of identifying precisely which inferences should be included in order to individuate the concept. Fodor (1998) dismisses classical defining feature theory, prototypes and theory-theories as possible candidates for what concepts might be as follows;

1) Concepts can not be definitions as "there are practically no defensible examples of definitions" (p.45.)

   Furthermore, the concepts in the definition – even if it existed, would be held in an inferential conceptual network of the type that he rejects.
2) Prototypes can not be concepts because concepts must be productive and systematic. That is to say that we must be able to compose new concepts from the concepts that we possess and that believing a proposition composed of concepts such as Jane loves John automatically entails being able to entertain the proposition John loves Jane. It is the property of compositionality that allows both these facilities. As previously mentioned in this chapter (p.25)

Fodor maintains that prototypes do not compose and therefore they can not be concepts. If prototypes composed then the new prototype of a complex concept such as PET FISH would include all the most typical features of the prototypes of each of its constituents, PET and FISH. Since a goldfish is a typical pet fish but is neither a typical pet or fish then prototypes fails to meet the demands of compositionality.

3) Fodor asserts that theory –theories fail to offer any coherent response to the question "what's a concept" at all (p.117). He concedes that supporters of theory-theories seem unconcerned with addressing this question (unlike Fodor, himself) but are rather attempting an explanation of what causes the features to cohere within a concept.

All three of these theories are, in any case, fatally flawed in Fodor's view as they all presume that the majority of concepts have a structure – this is an assumption that that Fodor does not accept. Fodor's alternative proposal is that concepts are atomic– that is to say that they have no component features and therefore no structure. It is not possible to do justice to his detailed arguments for this view here but his controversial position is that what it is to be the concept of x is to express "the property that our kinds of minds lock on to from experience with good examples of instantiated" x (p.137). Ultimately, Fodor feels that there should be no difference in explaining how we come to have the concept *doorknob* from experiences of doorknobs from the explanation of how we come to have the concept *red* from our experience of things that prompt redness.

There have been a number of critical responses to Fodor's arguments (e.g., Laurence and Margolis, 1999) and these are not necessarily germane to this summary. However, if Fodor is right, then discussions of whether situational features need to be accorded a larger role in concepts will become moot.

Multiple categorisation procedures ?

With the exception of conceptual atomism, the different theories of categorisation suggest that deciding whether or not x is a member of category y could be a matter of

a) seeing whether a rule that determines membership of Y is met by x

b) Evaluating similarity between x and numerous exemplars of Y
c) Evaluating similarity between x and a protoype of category y
d) seeing whether the features of x can be best explained by an underlying theory of what it is to be a Y

Smith, Patalano and Jonides (1998) reviewed evidence from various types of study that, when taken together, suggest that more than one of these procedures may be applied when making decisions about the same category; specifically they examined the rule –based and the exemplar-similarity based procedures. Given that the application of rules to decisions about category membership may reasonably be expected to relate to defining feature models of categorisation – and given that the plausibility of defining feature models had been placed under considerable doubt, one might be surprised that rule based procedures were given such attention and consideration. However, many studies of the acquisition and use of artificial categories continued to suppose the application of rules by creating novel categories that varied on a limited number of dimensions and supplying participants with specific criteria for identifying members from this limited pool of features. For example, Allen and Brooks (1991) gave participants two new categories to learn. The potential members were presented pictorially as bird-like creatures but only three visual features were presented as being salient to their category membership as either builders or diggers - body shape, leg-length and body markings. One group of participants was given the rule that if a item had 2 out of the three following attributes then it was a builder- otherwise it was a digger – long legs, angular body, spotted markings. The second group was simply exposed to a number of exemplars from both categories. Both groups were trained with a number of exemplars from both categories until they could correctly classify ten animals. In the test phase they had to classify a subset of exemplars that they had already seen plus some new exemplars. The two groups showed a different pattern of response on the new test items (in particular the negative matches) – demonstrating two different categorisation procedures with the same materials. However, the results suggested that the rule-based group had also based their decisions on exemplar similarity. Furthermore, the response time date indicated that both procedures had been utilised on the same trial. The finding that the group instructed with a rule spontaneously made use of exemplar learning suggested to Smith et al. that assessing category membership through similarity to learned exemplars is a relatively automatic process that may take some time to 'kick in' but that once activated it is quicker than applying a rule. Smith et al. also point out that requiring participants to perform a secondary task concurrently with the categorisation task results in greater interference with rule based categorisation than with similarity based categorisation (Smith and Kemler, 1984; Smith and Shapiro, 1989). The review then considered comparable evidence for the two procedures from the 'natural' category literature. The type of categories involved in the studies that Smith et al. considered were those where partial textual descriptions of putative members were provided to participants. As Smith at al. point out " the relevant attributes are denoted by distinct words in the description, they are readily separable and salient, just as in the case with artificial materials" p. 179. It should be noted that this situation differs considerably from the category decisions that one may make under everyday conditions where the decisions about <u>which</u> features will be relevant to the classification or identification of the item will be key to the subsequent categorisation. Nevertheless, Smith et al. found that studies with a restricted notion of natural categories also suggest the parallel adoption of similarity and rule based procedures. In a set of influential studies mentioned earlier in this chapter, Rips (1989) demonstrated dissociation between judgements of similarity to a category and judgements of category membership. Rips' participants were presented with limited (sparse) descriptions of 36 items such as *this object has a 3-inch diameter* and then asked to decide whether the item is more <u>typical</u> of a pizza or a quarter, whether it is more <u>similar</u> to pizzas or quarters and whether it <u>is</u> a pizza or a quarter. The value of the key variable (in this case, size) was predetermined asking the same participant in advance to think of the largest quarter that they could remember and the size of the smallest pizza. The hypothetical object was then described as being of a size half-way between these two values. The decisions varied depending upon the question asked. That is to say, responses to the category membership question tended to result in a decision in favour of the category with most variability around the key variable (e.g. pizza) whereas the similarity question tended to result in a decision that the item was most similar to the category where the key variable was fixed (e.g. the quarter). Typicality judgments seemed to fall roughly equally between the two options. This result was interpreted by Smith et al. as a further demonstration of participants using a rule based procedure to make their category membership decisions instead of similarity; something along the lines of 'quarters can never be 3 inches in diameter so it can not be a quarter'. Smith and Sloman (1994) extended the Rips' method to investigate the effect of providing richer descriptions of the item under consideration, and in particular to include in its description a feature of the non-variable category e.g. an object that is three inches in diameter and silver in colour. This description has a feature that virtually rules out its membership of the fixed category (quarter) but includes a feature that is characteristic of the same category (quarters are usually silver, pizzas are not). Participants answered similar questions to those posed by Rips for either sparse or rich descriptions. The results of this first experiment failed to replicate Rips' reported dissociation and instead found that the category decisions reflected the similarity judgements. This was the case for both sparse and rich descriptions. However, in a second study, a condition in which the participants had to think aloud whilst making their decisions was included. This was the method used in the original Rips study. This time category/similarity dissociation was found for the sparse items only – replicating Rips findings under the same conditions. This suggests that providing think aloud protocols encourage participants to engage in a more rule-based categorisation procedure than they would have done otherwise. This can be seen as supporting Smith at al's contention that similarity assessments are relatively automatic. The implications of this study for some of the work reported in this thesis will be returned to in Chapter 5.

Recent Developments

"The only people who regularly consider categories like clothing, fruit and furniture in their abstract, decontextualised senses may be categorisation researchers!"

Barsalou, 1991, p.45

More recently, researchers are being exhorted to take a fresh look at the area and to allow for the possibility that a radically new perspective is required if the field of enquiry is to expand in useful directions. Although the ideas come from several different areas of research within cognition (Funnell, 2001; Barsalou, 1999; Smith and Samuelson, 1997; Whittlesea, 1997), the focus of attack is the same, that is the notion that concepts can be adequately studied and understood as isolated entities. The recent work of two proponents of these ideas (Funnell, 2001; Barsalou, 1999, 2003) will be presented. Barsalou and Funnell are not the only authors who have challenged a 'decontextualised' view of conceptual knowledge (e.g.Whittlesea, 1997; Smith and

Samuelson, 1997; Glenberg, 1997) but they are presented in some detail since they offer specific new theoretical frameworks within which the empirical findings can later be discussed. It should be noted that neither of these theories drove the formulation of this programme of work since most of their development took place relatively recently- nonetheless it may be considered that they provide theoretical support for the investigation of 'contextualised' concepts.

Perceptual Symbols and Situated Concepts

Barsalou has been developing a theory of concepts as situated simulations through various publications over a number of years. Barsalou (1999) proposes that many of the problems associated with traditional models of concepts can be overcome by thinking of conceptual knowledge as being stored not amodally and propositionally but as being constructed from a store of perceptual symbols, extracted from experience (see also Barsalou, Weh, Luka, Olseth, Mix and Wu, 1993). Given that the symbols are drawn directly from experience and are memorised in the same form in which they are initially experienced, i.e., as perceptual memories, Barsalou maintains there is no longer a problem in explaining the relationship between the representation and what it is purported to represent in the world (frequently referred to in the philosophical literature as the problem of intentionality and one which has been dealt with by psychologists by ignoring it or denying its importance for psychology). The perceptual symbols are alleged to capture not only sensory information but also a record of internal states that accompanied the original experience, e.g., emotional reactions. Under this view, the feature lists associated with objects (which participants can provide on demand) reflect not a list of propositions stored in memory but a "sequential description of an experiential image". They are descriptions people give, *not of objects*, but of the manner in which they have constructed a schematic image. The degree of consensus in these descriptions between a range of people can presumably be explained by the similarity of their perceptual apparatus and their shared experience within a culture. Barsalou uses his earlier metaphor of a world model (Barsalou, 1991) to describe the new theory, " A world model is a person's beliefs about the current state of the world. It is not beliefs about the types of things in the world, such as taxonomic knowledge about birds and tools; instead it is beliefs about particular individuals in the world, along with their current states and locations" (pp. 30-31). Within the world model, individuals, as ontological entities, are represented as frames that capture both the enduring aspects of the individual entity over time as well as the variability. All information (in the form of perceptual symbols) appertaining to an individual is integrated into a single frame. Over time, the frame for any given concept comes to contain " extensive multimodal information of what it is like to experience this type of thing" (p.12). The frame may be subsequently partially activated in order to simulate a particular experience. In this theory, the simulator carries much of the explanatory force. Since the simulator is equivalent to a concept, the simulator must be able to fulfil all the putative functions of concepts, for example, permitting categorisation, re-identification procedures and supporting categorical inferences. This new model of conceptual knowledge can be seen as emerging from a move to re-unite perception and cognition (see Goldstone and Barsalou, 1998).

An interesting aspect of this work for the interpretation of these studies, which will be presented in the following chapters, is the emphasis on the multi-modal nature of the perceptual symbols. The symbols capture, through attentional processes, any number of aspects (not all aspects since they are inevitably partial records of an experience) of experiences with entities or events. So *all* sensory modalities are deployed (not just vision on which so much rests in traditional accounts of object recognition) including proprioception and introspection. This opens up the possibility of conceptual knowledge that extends far beyond the awareness of features and functions required for re-identification playing an equal role in concept-based cognition.

More importantly for this thesis, Barsalou's recent development of these ideas (Barsalou, 2003) stresses the situated nature of the simulations that he posits. In fact, the theory is now referred to as 'situated simulation theory'. It is proposed that the simulation of an object, which supports thought about that object and is provoked by a goal, is always situated. That is to say, the object or entity is not simulated as an object, free of any background setting but is always integrated with information about its external environment. The nature of that background will vary depending upon the nature of the stimulus for the simulation. Barsalou claims that the primary purpose of the conceptual system is to support 'situated action' and that it is therefore necessary that it should be organised around the action/environment interface (see also Franks and Braisby [1997] for a view of concepts that prioritises the goal of guiding reliable action in the environment).

Since developing his theoretical position, Barsalou has reviewed earlier experimental evidence from this perspective and has additionally reported experimental evidence of his own that lends support to the theory. For example, he has investigated whether perceptual simulations are used by participants when they are asked to generate properties of objects. Wu and Barsalou (in press) predicted that participants given no instructions on how to achieve the task should perform similarly (in terms of the type of properties generated) to participants who were explicitly told to use mental imagery. This was found to be the case. Wu and Barsalou also demonstrated that the participants given no instructions were affected by the manipulation of perceptual variables that one would not expect to alter performance if amodal symbolic representations (such as feature lists) were being accessed during the task. Perhaps the most convincing demonstration was the increased degree of access to certain properties when the noun was preceded by a modifier. For example, more internal properties were generated when the noun was preceded by a modifier that would make internal properties more "visible" in a perceptual simulation e.g. *rolled-up lawn* compared to when the noun was presented alone e.g. *lawn.*

The theory makes a number of claims about the conceptual system that mark it out from previous models. Firstly, perception and conception are not autonomous, modular systems. Secondly, a concept should be thought of more as a *facility* for producing representations (simulations) to meet particular immediate goals, rather than as the retrieval of a stored, decontextualised representation. Thirdly, the simulations include information about environments, likely goals and introspective states. The Wu and Barsalou paper also provided support for the claim that when concepts are simulated they are simulated within relevant settings or situations. In all four experiments that involved property generation under differing manipulations, participants produced physical settings and introspective states in their protocols as well as properties. The amount of situational information generated during the tasks ranged from 19% to 35%, The authors take this as being informative of conceptual processing rather than dismissing it as 'error'. This offers a key role for the

relationship between items and their settings within cognition that contrasts with older models of semantic memory that focus heavily on intrinsic features of objects.
The studies reported in the following chapters can therefore be viewed as support for concepts being situated. Indeed, the published paper resulting from chapter 2 (see Appendix A) has been cited by Barsalou as further evidence for his theory (Barsalou, 2003).

Levels of Meaning

Memory for events and situations has been the subject of a considerable body of research, much of it under the auspices of autobiographical memory (see Conway, 1990b). It has been claimed that memory for events is organised by activities, that it has multiple organisations (Lancaster and Barsalou, 1997) and has a basic level, similar to that of object taxonomies (Morris and Murphy, 1990). When discussing and researching event knowledge, one notion that remains very influential is that of scripts (Shank and Abelson, 1977). Scripts have been proposed as organising structures for information concerning events. These have been thought to form hierarchies of increasingly abstract levels of information and incorporate knowledge of the relevant agents, objects, activities, locations and goals most commonly associated with the event. For example, the script for shopping at the supermarket might include trolleys, store assistants, fruit, vegetables, frozen foods, tins, credit cards, paying for goods, carrier bags etc. One might reasonably expect that that retrieval of an event frame or script would activate the relevant objects, but it would not necessarily be predicted that when one formed an intention to activate knowledge of certain types of objects (e.g. *fruit*) that one would do so through an event frame or script, since information regarding types of objects would be assumed to be more readily available through a deconxtualised repository of facts about the world such as 'what types of things are called fruits?'. Physical scripts have been described by Shank (1982) as providing a 'snapshot' of a visual field that would include objects. A review of the evidence for the involvement of scripts in studies of cognitive breakdown in both Alzheimer's patients and semantic dementia patients led Funnell (2001) to formulate the levels of meaning model of semantic memory, which is intended to be descriptive of memory in the absence of pathology but that can explain the patterns of performance and behaviour noted in dementia patients.

Patients with semantic dementia typically show poor semantic knowledge in clinical tests of word meaning and use, but retain comparatively good abilities at using everyday objects. It is not unusual for such patients to be unable to specify the function of an item when shown it under test conditions but to use the same type of object appropriately in a variety of familiar, everyday activities. As the disease progresses, the patients' ability to generalise knowledge about objects beyond the most immediate context deteriorates. Funnell (1996) documents a severe case of a semantic dementia patient who retrieved a ballpoint pen and calendar from her kitchen in order to arrange the next appointment having only just failed to recognise an identical pen when tested. Snowden, Griffiths and Neary (1994) also described the way in which only items experienced in repeated, personal contexts were recognisable to semantic dementia patient, KE. The function of precisely the same objects could not be specified by KE if it was moved from its usual location to a different place in her own home. It seems then that the condition involves a progressive dissociation between the ability to show the researcher how an object can be used under formal testing and the appropriate use of it in everyday, goal-based activities. New learning has been shown to take place even in advanced cases of semantic dementia but Funnell notes that it is only successful in a very narrow context and does not generalise to new situations. Funnell concludes that ' the effect of new learning on conceptual representations might be explained more successfully within a model that includes both abstract, generalisable knowledge and knowledge tied to a specific context rather than the traditional approach that isolates semantic memory entirely from contextually based information." (p.333).

She consequently suggests that meaning, in the absence of pathology, is represented by a 'continuum of processing levels'. The model proposed by Funnell consists of three levels of meaning for concepts - the specific event level at which knowledge of an object is tied to specific encounters with it; the general event level, at which knowledge of an object is linked to repeated and varied experiences of the situations in which the object plays a role and the conceptual level at which knowledge of the object is abstracted from the other levels and forms a 'repository of facts' about the item that more closely resembles the usual notion of semantic memory (see figure 1.3.). The extreme ends of the continuum are information that has, across time, been completely abstracted from the contexts in which it was originally experienced, and at the other end, information that is embedded in a specific personal experience. In between are differing levels of abstraction.

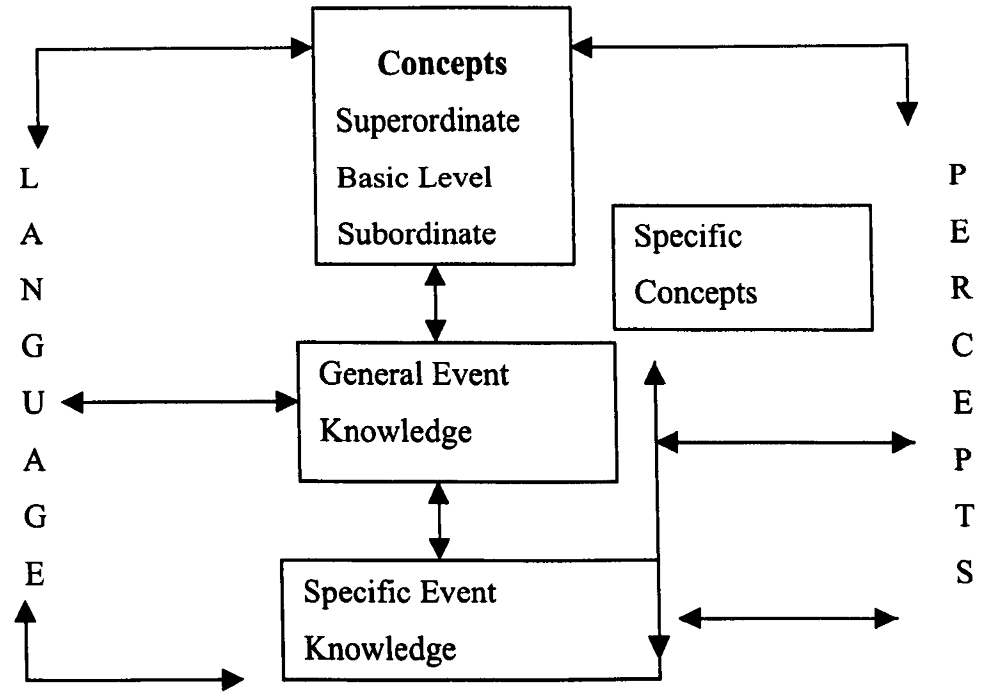


Figure 1.3 Levels of Meaning Model (Funnell, 2001).

Funnell maintains that such a model of conceptual knowledge can account for the apparent anomalies in dementia cases regarding their semantic knowledge and their interactions with objects. Semantic dementia can be seen as the initial breakdown of the highest level of abstraction (the 'encyclopaedic' knowledge of items), so that meaning devoid of context becomes increasingly difficult to access. It is this level of meaning that would be exercised by standard clinical tests of semantic knowledge. The more general event knowledge in which objects are embedded in memories of repeated experiences of similar settings may be preserved for longer, allowing patients to make use of objects (for which they can no longer describe the function) providing that the contexts in which they appear are familiar. Eventually, any new learning may take place only at the level of specific episodes and may not generalise even to other similar events.

The model also has developmental support in that the initial interface between infants and objects will be in the context of specific episodes but repeated episodes allow the formation of structures that retain knowledge of the principal components but lose the phenomenology of specific episodes. This would be represented in the model by the general event knowledge. The level of abstracted features and function-based categorical knowledge would be the last to develop. There is evidence that initial groupings of objects by children tend to be around shared scripts for commonly occurring events (Lucariello and Nelson, 1985) and that groupings made by young infants often cohere through shared events and situations (Markman, 1989).

Funnell proposes that loss of access to abstract knowledge of the item is compensated by knowledge of how to interact with the object in well–rehearsed situations, thus allowing them a facility for object use that would not be predicted by the patients' performance on tests of their 'factual' knowledge about that same object. The same inference has been drawn by Snowden, Griffiths and Neary (1994) based on similar observations of semantic dementia patients.

Given the slippery notion of 'meaning' and its complicated relationship with concepts (see p. 16), it may be preferable to think of this model in terms of levels of abstraction as opposed to meaning. Initial encounters would be experienced as specific episodes from which can be abstracted the nature of regularly repeated episodes which in turn provide the 'database' from which repeating features of frequently occurring objects can be drawn.

It can be seen that this model, arising from work in cognitive neurospychology, also moves away from the notion of conceptual knowledge as solely de-contextualised feature lists that characterised earlier theories of concepts. However, it concerns itself principally with the possible relationships between different types of memories and unlike the theoretical framework developed by Barsalou, it is silent about the way in which the components of the memories are represented in memory, i.e., as amodal propositions or in units that are analogues of experience.

Summary of Theories of Organisation

The literature appertaining to categories and concepts is immense. The preceding review does not claim to be exhaustive, but seeks to convey something of the development of theories of conceptual structure. It is notable that, with the possible exception of classical defining-feature views, the other theories described continue to co-exist and compete. It can be seen that the earlier theories envisaged concepts as repositories of facts about entities and that those facts were, almost exclusively, descriptive of their physical features and their function. Context, in the sense of differing environments and differing actions, was seen as something to be 'added' to core representations. In fairness to theories that attempt to present mental concepts as relatively stable, abstracted structures- this perspective is not an arbitrary one but has been driven by the non-trivial observation that people have the ability to repeatedly recognise and perform appropriately towards an object despite "varying local circumstances" (Smith and Samuelson, 1997. p 161). That an object can be recognised as a member of a particular category across a wide range of different environmental contexts implies that what is known about that object must be abstracted from the environments in which it has been encountered and that regularities, rather than contingent variations, constitute the stable base of the representation of that object. The focus on isolating the aspects of an entity that would permit this reliable identification (i.e. its features) has perhaps been responsible for the implication that this body of knowledge is itself necessarily isolated from other information.

Both essentialism and conceptual atomism deny features or properties of the type of role afforded by other theories. For an essentialist the features usually discussed as being included in definitions, exemplars or prototypes would be causally linked to the essence of the concept rather than serving to individuate it and proponents of conceptual atomism would claim that features, whilst clearly related to the atomic concept, in no way form part of it.

The later theories (Funnell, 2001; Barsalou, 1999, 2003) both propose that information about situations and events play a more important role in thinking about entities than has been implied by previous theories. One of the primary differences between the two models is that Barsalou's does not rely on stipulating types of organisation in memory. Under his theory, knowledge of objects is stored as perceptual symbols that are accessed to simulate a wide range of perceptual information regarding an object for the purposes of thought. Funnel proposes a more standard 'storage structure'.

Features

The focus of this review has been the different ways in which 'features' of entities have been supposed, or not supposed, to enter into the organisation of conceptual knowledge. Before detailing the aims of this programme of work, It is worth spending a little time to consider what this important term may be considered to include.

What may count as a feature?

One may be forgiven for assuming that identifying features of entities would be non-problematic. However, using abstract figures, Medin, Goldstone and Gentner (1993) have demonstrated that what counts as a feature when, for example, assessing similarity between items, may change depending upon the items under comparison. As shown by the review provided in this chapter, the term 'feature' is key to virtually any discussion of concepts in the psychological literature. But what is covered by the term 'feature' is by no means a non-controversial issue. As Murphy (2002) "Any object can be conceived of in many different ways, and it would be impossible to encode each one and store it as part of a concept" (p.174). Implicit in this quote is the notion that not all attributes or characteristics of an item can be considered to be a feature of the item as concept. The large groupings of features that are most frequently referred to are perceptual (sensory), functional and encyclopaedic. Perceptual properties or features are those that can be perceived directly through the senses (such as size, colour, shape, texture, sounds) whilst functional features are usually thought of as those that pertain to how the item acts or is used. The documentation of cases of individuals who show selective deficits in their conceptual knowledge has led to speculation that the relative importance of these different types of feature is differentially weighted for living and non-living things and that category specific impairment may result from damage to either type of feature. This has become known as the sensory/functional hypothesis (Farah and McClelland, 1991;Warrington and Shallice, 1984). Much of the impetus for thinking of semantic knowledge as a distributed network of overlapping discrete featural units arose from the investigation of category specific impairments and the collection of normative data concerning features has therefore often been undertaken by researcher in this area.

McRae, De Sa and Seidenberg (1997) obtained normative data from a feature generation task and classified the resulting features as 'aspects' (to include perceptual features); functional (relating to use); classification (indication membership of some other category); information related to a situation in which it takes part (e.g. grows on trees) and people's related cognitions. The situational features represented only 9.9% of the features coded compared to 56 % aspects and 22.5 % functional. Hence situational features are indeed considered here as features but appear to be far less salient than other types of feature in a feature-listing task.

Garrard, Lambon Ralph, Hodges and Patterson (2001) collected normative 'feature' information for 64 items from 8 superordinate categories. Their method differed from a simple 'list the attributes' instruction. Participants had to complete sentences referring to the items such as 'X is......, X has...... and X can........' . The resultant attributes were coded as sensory, functional, encyclopaedic or categorising information. Functional examples were an *owl can fly, a suitcase can be carried, a cat can catch a mouse.* Encyclopaedic examples included *a tiger is found in India* and *a toaster is kept in a kitchen.* Overall , 50.5 % of the features generated were classified as sensory, 27.6% as functional, 14.7% as encyclopaedic and 7.2% as categorising. The authors were surprised to find that when examining the distinctive features (as distinct from the shared features) it was only the encyclopaedic features that showed a higher proportion of distinct features for the living compared to non-living domain. They note that patients exhibiting a deficit for non-living things should , according to their results, also show a relative deficit in their knowledge of encyclopaedic knowledge.

The type of knowledge about items that is highlighted in this thesis is referred to t as situational knowledge and it is may cover some of the features that have been previously identified in cognitive neuropsychology as functional (*can be caught* as a functional feature of a mouse would link it to a mousetrap as something that *can catch mice*) or encyclopaedic (*is found in cold climate* may link *penguin* to *igloo*). Locations and situations are often difficult to disentangle. It could therefore be argued that this type of information has long been recognised and included amongst 'features'. However, situational features have not attracted the same degree of interest in cognition literature (although see Wu and Barsalou, in press).

It will be seen in Chapter 3 that these situational features (e.g., is taken to weddings for *camera*) can be used as a basis of similarity. If one is content to extend one's usual notion of features to include these situational features then situational knowledge can be represented within a distributed model of semantic memory (e.g., Moss, Hare, Day and Tyler, 1994). One may argue then that when people make use of thematic relations to judge similarity between items or to form the basis of categorisations (Lin and Murphy, 2001) then they are simply making judgements based on these features rather than other types of feature (e.g., perceptual). However, it will be argued in the final chapter that this approach cannot account for the type of thematic similarity that seems to arise from items 'featuring' in the same event or situation (e.g., between jelly and balloons). A feature of Jelly could be thought of as 'eaten at children's parties' and of balloons as 'decorating children's parties' but this is not the same feature- yet this co-occurrence in an event seems to account for some measure of similarity between items and may lead to their shared categorisation.

It may be therefore that rather than thinking of situational knowledge as a different type of feature, situational knowledge may be more usefully thought of as arising from an additional, script-like organisation of object information revolving around events and situations.

AIMS

The purpose of this research programme was to investigate the influence of situational information in two specific cognitive tasks. The first task comprised the generation of members of a given category in a restricted time. People were also asked to explain the rationale for their choice of items. The second task comprised the rating of similarity between category members. The categories included both those with members that share perceptual and/or functional features and categories (taxonomic) with members that share few, if any, perceptual and/or functional features (ad hoc). These tasks were chosen precisely because they have traditionally been explained in terms of traditional notions of "semantic memory" unaffected by situational knowledge.

The Empirical Tasks

It is clear that we have, for any given entity, a wide range of information that may assist us in predicting its **usefulness** in a particular context, that is the manner in which we can interact with it to fulfil specified goals and where in our environment it is to be found if needed. However, this introduction has suggested that the range of non-featural knowledge has, in older theories of conceptual structure, been implicitly accorded a role of secondary importance as background knowledge or 'associated ideas'. In this thesis, the two chosen tasks required extensive preliminary investigation to generate appropriate stimuli. Furthermore, both tasks led to different kinds of analyses on qualitative and quantitative aspects of the data. This approach led to inferences concerning several related theoretical issues. A brief background to the two tasks will be given in the following sections, with a fuller introduction preceding the relevant studies.

Category Member Generation

This task was chosen as an elementary cognitive operation that demands, one assumes, access to conceptual knowledge, and has been traditionally thought of as tapping semantic memory. Participants are provided with a series of category headings and asked to list as many members of that category as they can. This is known as a free emission procedure. The output of such tasks has been examined for evidence or organisational dimensions (Henley, 1969; Rips, Shoben and Smith, 1973). However, there have been relatively few studies that deployed the free emission procedure with a view to examining the nature of participants' retrieval strategies, namely, Gruenewald and Lockhead, 1980; Williams and Hollan, 1981; Walker and Kintsch, 1985. Of these, only the last two collected protocols from their participants as to how they completed the task. Williams and Hollan's participants were asked to recall names of people they attended school with. Only Walker and Kintsch (1985) used a natural and an object category as the original stimuli.

There is an extensive literature that investigates the clustered output from learned word lists. Clusters have been found to take advantage of categorical connections between items in the original to-be-learned list, and to represent the imposition of an organisation on lists of unrelated words. Since this has been taken to be informative as to the organisation of semantic memory, one might reasonably argue that output from free emission tasks should also reflect organisational themes. For this reason, the older, list-learning literature is also reviewed in the introduction to the free emission study. The thesis study departs from the previous literature by requiring protocols of retrieval strategies for an extended range of both taxonomic and ad hoc categories.

Similarity Judgements

Similarity is believed to underlie a number of vital cognitive activities, such as categorisation, reasoning by analogy and inductive reasoning and as such it has been the subject of much theoretical debate (Goldstone, 1994; Goodman, 1970) and empirical investigation.

In everyday situations, it is rare that one would make a judgement of similarity between two items for no express purpose. The comparison might be made in the course of problem solving or reasoning; or there may be a practical goal that one is trying to meet. Certain laboratory tasks infer similarity from the achievement of other goals, such as sorting tasks or perceptual confusability tasks. A number of studies, however, have made use of direct assessments of similarity between items using scaled ratings (e.g. Barsalou, 1982). Following on from the category generation task, it was decided to examine these direct judgements of similarity for the influence of situational knowledge for the following reasons. Firstly, the degree of similarity between two items has long been believed to be a function of the number of properties that they share and the number of properties on which they differ (Tversky, 1977). There have been many debates about the nature of the constraints that may operate upon the properties involved in the comparison. The focus of investigation into similarity has been principally one of identifying the factors that affect ratings of similarity, presumably by altering the range of properties under consideration such as context (Tversky, 1977; Medin & Shoben, 1988) and expertise (Sjöberg, 1972; Suzuki, Ohnishi & Shigemasu 1992). Until quite recently, there was never any doubt that the process was one of feature comparison and that the features were perceptual/functional. If situational information were found to be influential within such a process, it could therefore be argued that theories of similarity should be amended to include a role for this other type of shared 'attribute'.

Methodological considerations

Taxonomic and Ad Hoc Categories

The "standard" taxonomic categories such as *birds, furniture, vehicles, vegetables* etc. exercised a stranglehold on research for a number of years. It is useful to remind ourselves of their origins as appropriate concepts for research. Rosch's seminal experiments reported in 1975 commenced with an experiment in which reliability of subjects' ratings of internal structure was verified, and norms were collected for ratings of the goodness of example of 50–60 members of 10 "systematically chosen categories" (p.196). The ten categories were the only ones remaining at the end of the following procedure. Firstly, all concrete nouns with a word frequency of greater than 10 were identified from Kucera and Francis (1967). From these, categories were eliminated if, a) they bore a part-whole relationship to the superordinate, b) if there was linguistic ambiguity in the superordinate, and c) if the superordinate cut across a number of other taxonomic structures. The final 10 categories were the familiar, *fruit, bird, vehicle, vegetable, sport, tool, toy, furniture, weapon* and *clothing*. In 1983, Lawrence Barsalou drew the attention of concepts researchers to the existence of a very different kind of category that is prevalent in everyday use. These categories are used to describe collections of items which are not normally thought of as being in the same category (indeed, the members are drawn from other existing standard categories) but often are grouped together for a specific purpose. Barsalou's earliest examples of these categories included *things to take from one's home during a fire, things to eat on a diet, things to take on a camping trip.* Barsalou refers to these categories as ad hoc or goal-derived categories.

Barsalou (1985) found that these categories, like standard categories, also had typical and atypical members (i.e., exhibited graded structure). He proceeded to investigate the basis of that typicality. Since these members are perceptually very different from each other it seemed clear that typicality could not be assessed by an enumeration of perceptual features shared between them and some abstracted representation of the most frequently occurring features of the category. What would the prototype of *Things dogs chase* be like? These category members do not share many properties-sometimes they appear to have in common only the goal for which they have been grouped together to serve (usually made explicit in the category name). Barsalou (1985) found that the typicality of these ad hoc categories was determined by two factors. Firstly, 'frequency of instantiation' which is a measure of how often the item has been encountered 'as a member of that particular category'. Note that this is different from a measure of pure frequency; the typicality of torch as a member of *Things you take on a camping trip* depends not on frequency of experience of a torch per se but upon frequency of experience of torch in the context of *things you take on a camping trip.* A second factor that determines the structure of these categories is the extent to which the member exhibits the characteristics of an "ideal" member of that category. For example *things to eat on diet* could contain both lettuce and yoghurt, but since an ideal food to eat on diet has few calories, the member that comes nearest to that ideal will be seen as a more typical member, i.e., lettuce.

The significance of ad hoc categories has yet to be fully explored. They are frequently cited as illustrating the flexibility of the conceptual processes but are rarely included in theoretical arguments about knowledge organisation. They can be seen as a distinct type of category altogether from standard categories in that being developed " on line" to meet a temporary goal, they do not demand accounts of conceptual structure. However, Barsalou has suggested that these categories, given sufficient reason for use, can evolve into more standard categories (*things to give to someone on their birthday* becomes *birthday presents*). This implies a continuum between ad hoc and standard categories. Lucariello and Nelson (1985) found that ad hoc categories tend to be developmentally prior to standard taxonomic categories. Ross and Murphy (1999), whilst not referring to them as Ad hoc categories, found that participants were willing to sort foods into categories such as *snack foods* (foods which can be eaten between meals and eaten on the move) and to use these categories to support particular inferences. It may well be that in everyday contexts, categories such as *things to take to work with me* feature in thinking and planning more frequently than do categories of *vegetable* or *toys*. This is one reason for including ad hoc categories in the materials for the studies to be presented here, in addition to the more standard representatives of 'category'. There has been a considerable focus on the differences between these types of categories. In his seminal paper on ad hoc categories, Barsalou (1983) showed that taxonomic categories have stronger concept-to-instance associations than do ad hoc category members leading the taxonomic categories to be more firmly entrenched in memory. Ad hoc categories have been said to have a heavy reliance on goals in a way that is rarely associated with taxonomic categories. Ad hoc category members tend to share fewer perceptual features than do taxonomic category members. By using both category types within the same framework in this thesis, it was possible to investigate whether the role of situational information also differs for categories that have been believed to differ in a number of other respects.

The term 'taxonomic' will be used in this thesis to describe both natural kind and artifact categories that admit of a hierarchical organisation, exhibit a graded structure based on shared properties and are presumably well established in memory. This may extend beyond the somewhat limited range of 'Roschian' categories. Barsalou (1983) uses the term 'common' category rather than taxonomic and elsewhere the term 'common taxonomic categories' has been used (e.g. Medin, Lynch and Solomon, 2000). The term 'taxonomic' will be used here throughout for the sake of consistency. The phrase 'ad hoc' will be used in preference to 'goal derived' to describe those categories whose members cut across taxonomic categories, share few perceptual features and are often depicted as being created 'on-line' to meet a specific aim.

The Use of Protocols

For both the category generation tasks and the first of the similarity studies, it was decided to make use of written protocols to provide information concerning the strategies used by participants to complete the experimental tasks.

There has been a long history of debate concerning the extent to which protocols can be relied upon to represent a window into cognitive processes. Nisbett and Wilson (1977) and Ericsson and Simon (1980) most famously debated the value of verbal reports as data. This argument focused around verbal protocols recorded concurrently or retrospectively, as opposed to written protocols as obtained in the present studies. However, substantially the same objections (and defence) can be made in both cases. Nisbett and Wilson, in reviewing a wide range of studies in which verbal reports were treated as data, concluded that "when people attempt to report ......on the processes mediating the effects of a stimulus on a response, they do not do so on the basis of any true introspection" (p.231). They propose that in the absence of availability of genuine introspection, people give reports of their cognitive processes that are based on their existing knowledge of plausible relationships between stimuli and responses. Ericsson and Simon countered that verbal reports, collected carefully under particular conditions, <u>could</u> be treated as valid data. They drew attention to the fact that the studies reviewed by Nisbett and Wilson were "neither designed for nor primarily concerned with determining subjects' memories of their cognitive processes" (p.246). Nisbett and Wilson, from the outset of the article dismiss the possibility of participants being able to report on perceptual or memorial processes. The data considered by Nisbett and Wilson were collected under a wide range of experimental conditions for differing purposes but for the main part involve asking participants to give an account of their behaviour under particular conditions. For example, in the "insufficient justification studies" examined by Nisbett and Wilson, participants required to undergo a second series of electric shocks (while performing a learning task) with no real justification showed lower galvanic skin responses and better learning than subjects who were provided with a justification for the second series of shocks. However, the former group did not report the shock to be any less painful than did the latter group. In reviewing other similar studies, Nisbett and Wilson report that in the majority of studies of this type, behavioural differences between groups of participants were not matched by differences in their verbal reports. Nisbett and Wilson also make reference to studies in which attitudinal shifts are identified in participants as result of discussion but participants' protocols showed little or no awareness of these changes (e.g. Goethals and Reckman, 1973). Nisbett and Schachter (1966) subjected their participants to a series of increasingly painful electric shocks. Experimental participants were administered a placebo and were told that the pill would produce physical symptoms which were actually those which would resulting naturally from the shocks. These participants, attributing their physical experience to the pill rather than to the shocks took approximately four times as much shock as the control subjects. The participants were subsequently asked about their ability to withstand shock. For example, " I noticed you took more shock than average. Why do you suppose you did?". Only 3 of the 12 subjects in this condition explained this in terms of their attribution of the unpleasant effects to the pill rather than to the shock. Typically they looked to an explanation from somewhere in their past which could account for their stoicism. Ericsson and Simon, however, pointed out that the form of the probe question made it not only *possible* for subjects to account for their responses *without consulting their memories* on how they did so, but highly likely that this would be the case. Subjects were not asked, for example, what they were thinking about when they experienced the symptoms or what they had believed to be their source. Furthermore, in some of the work critiqued by Nisbett and Wilson, the interval between completing the task and providing the protocol was, in the view of Ericsson and Simon, so long as to cast doubt on whether the subjects' memory for how they undertook the task was still available to them with any degree of accuracy. Ericsson and Simon are also wary of studies in which the protocol is recorded at the end of a number of trials, obliging the participant to abstract a general strategy across several occasions or to generalise the last strategy used to all trials. Clearly the informativeness of such a protocol is possibly limited and unrepresentative. In the current studies, participants were probed after each trial. This approach reduces the time lapse between task and protocol and leaves any identification of general, repeated strategies to the experimenter during protocol analysis. This should increase the accuracy and the specificity of the protocols. Additionally, the instructions given to participants in these were entirely open and offered no example of what might constitute "the correct answer", thus it should not have been possible for participants to divine any requirement on the part of the experimenter

Thesis Structure

Generating Category Members

Chapter 2 reports on category member generation and protocols of participants' explanations of their item generation strategies. If semantic memory acts as a repository for category information that has been extracted across numerous experiences and stored in a manner that will serve the demands of categorisation and identification regardless of context, then one would predict non-strategic generation of category members. That is, one would predict that accessing one category member would activate other close members of the category in a relatively automatic manner. Furthermore, one would predict that participants would be unable to report the nature of the retrieval, or at most would report semantic groupings within the category. If, on the other hand, knowledge of category members were contextual, then one would predict that familiar contexts would act as useful retrieval strategies.

Judgements of Similarity

Chapter 3 reports the quantitative and qualitative results of direct judgements of similarity between category members. Pairs of items, varying in typicality, from both common and ad hoc categories were presented with or without context. Participants gave judgements of similarity by means of scaled ratings. A subset of participants also provided written protocols explaining their ratings. A primarily feature-based account of concepts would predict that judgements would be based on a computation of shared and unshared properties. However, if situational knowledge is utilised in such judgements, one would predict that perceptually dissimilar items could still be found to be similar if they both feature in the same type of situation or event. Evidence of this would be predicted in both ratings and protocols.

Predictors of Thematic Similarity

Chapter 4 details empirical work that set out to further investigate the type of similarity that arises from participants linking items together, thematically, in a situation. When perceptually related items are viewed as being more or less similar to each other, it can be explained as a function of their number of shared properties. Hence a sparrow and a robin are likely to be rated as more similar to each other than an eagle and a robin. What could account for variations in ratings of similarity for items that are thematically linked? If these items are seen as sharing a situation or script then one might predict that one factor affecting the ratings could be the likelihood of each item appearing in that situation. Chapter 4 describes an investigation based on this prediction. The relationship between ratings of thematic similarity and the relationship of the items, not directly to each other, but to a common event was examined. The manipulation was based only on data of frequency of generation from the event to the item and the item to the event.

Theoretical Conclusions

In the concluding chapter, a summary of the main empirical findings is presented for each study. Each set of results is discussed with reference to the two recent theories of conceptual knowledge, Barsalou (1999, 2003) and Funnell (2001), described earlier in this chapter. The direction of future studies arising from this work follows and the chapter concludes with an assessment of the contribution made by the thesis.

CHAPTER TWO

Accessing conceptual knowledge: Retrieval Strategies in Free Emission Tasks

"Memory, as measured by free recall, has a certain organization, and that organization is related to previously acquired conceptualisations"

Kendler (1966, p. 198)

Overview

This chapter reports two studies. In study 1, participants generated category members for 2 taxonomic categories and 8 ad hoc categories under a time constraint (90 seconds). After each category, the participants provided written protocols in answer to the question "how did you go about thinking of items for this category?". Data concerning fluency and item dominance are reported. However, the data of principal interest is the classification of retrieval strategies arising from the protocols. Study 2 remedied two possible artefacts in the first study by rephrasing the titles of the ad hoc categories and by increasing the number of taxonomic categories within the materials. The studies aim to identify the strategies that participants report using whilst listing members of both taxonomic and ad hoc categories and to consider these in the light of predictions from standard theories of semantic memory organisation (also published as Vallée-Tourangeau, F., Anthony, S. H., & Austin N. G., 1998).

Background

During the 1940s, 50s and 60s, there was substantial interest in verbal learning experiments in which participants typically had to study, and subsequently recall, word lists. It was consistently found that the recall output consisted of temporal clusters or "bursts" of words (Bousfield, 1953; Pollio, Richards and Lucas, 1969;

Puff, 1970). Furthermore, examination of the cluster contents suggested that participants made use of semantic organisation afforded by the content of the lists to organise their recall. Even in the absence of any organisation of the stimuli in the mind of the experimenter, participants imposed their own (Tulving, 1962). The methodology deployed was predominantly the learning/recall of predetermined words but some of this work included a paradigm in which the subject was simply asked to list as many members of a given category as possible. This procedure is referred to as a "free emission task" and resulted in an output which was similarly characterised by clusters of apparently related words produced with relatively short inter-item pauses, separated by longer pauses (Bousfield, 1944). For both these tasks it has been posited that the nature of the organisation demonstrated in the output is likely to reflect organisational principles in memory (Tulving, 1962, 1964,1966; Tulving and Pearlstone, 1966; Kendler, 1966). This is a key assumption in the conclusions that will be drawn from the two free emission studies reported in this chapter. The focus of the present studies will therefore be traced back through part of the list learning literature, which has historically been closely intertwined with studies on free recall from semantic memory. Studies making specific use of free emission tasks, which resemble the focus of the current work more closely, did not appear until the 1980s. These will be dealt with towards the end of the introduction.

Previous Work

Amongst the earliest experimenters to make use of a free emission procedure were Bousfield and Sedgewick (1944). In this early study participants were simply asked to name as many quadruped animals or US cities as they could in 18 minutes. The characteristics of the responses most relevant here were a) the output did not emerge as a string of items separated by equal temporal intervals but in bursts, or clusters, of items separated by relatively longer pauses, and b) participants appeared to make use of relationships between words to organise the material in recall. That is to say, the constitution of the response clusters was not arbitrary but demonstrated "various types of contiguity" (p. 153). This feature of the data was noted by the authors but was not pursued since the stated purpose of the study was to examine the temporal characteristics of sequences of responses. Bousfield (1953) went on to extend this line of work to recall of learned word-lists which consisted of 15 instances from each of 4 categories, animals, names, professions and vegetables. It was found that category members tended to be recalled together even though they had been presented in a random order.

The Relationship between Clustered Items

The clustering or grouping of items in the recall output seemed to reflect relationships between the to-be-remembered items. An obvious question to be addressed was the nature of those relationships. Lists organised by the experimenter to fall into a number of superordinate categories seemed to result in clusters of categorically related items as found by Bousfield (1953), but Jenkins and Russell (1952) reported 'associative clustering' in the responses of participants who were asked to recall pairs of highly associated items (chosen from word association norms) which did not necessarily share category membership, e.g., Black-White, High-Low. Although the word pairs were separated and randomised for the learning phase of the experiment, members of the same associated pair tended to be produced together in the recall outputs. Overall mean recall was 24 words from a total of 48 and approximately half of these appeared in their associated pairs. Marshall (1967) and Cofer (1965) attempted to tease these factors apart to try and offer a definitive answer to which type of association between items formed the basis of organisation in recall. Marshall (1967) used pair lists, similar to the Jenkins and Russell study; these were chosen such that the components of each pair were associated with each other but half the pairs were categorically related (e.g. Table-Chair) whilst the remainder did not share category membership (e.g. Sit-Chair). Six different levels of 'associatedness' between the paired items were presented to six different groups of subjects. Marshall found that the higher the degree of association between paired items, the more likely they were to be recalled together. However, there was an even stronger tendency for this to be the case when the pairs were drawn from the same category. There was additionally an interaction such that the category effect was stronger with low association strength pairs than for highly related pairs. Both category associations and non-category based links seemed to be used as the basis of organisation. Cofer suggested that this implied an active search for ways of organising the material in as much as use would be made of any perceived relationship between list items with non- categorical association being utilised in the absence of a categorical link; the latter being favoured. The importance of this line of research was that it was viewed as a means of garnering evidence for competing models of semantic organisation in long-term memory.

Subjective Organisation

Tulving (1962) looked at what occurred in recall if subjects were asked to learn lists of unrelated words and tested over repeated trials of the same stimuli. He reasoned that if two items from the list are recalled adjacently on a number of different trials then they had probably been organised by the subjects in some manner. His means of measuring this was referred to as a measure of subjective-organisation (SO). Sixteen words were presented on different orders for 16 trials. There was a steady increase in the number of words recalled on each trial plus an increase in SO scores. This lent strength to Tulving's growing conviction that the improvement which was witnessed in recall over trials with the same materials was not due to repetition per se but was due the opportunity that repetition afforded for the subject to organise the stimuli in a meaningful manner which aided recall. The difficulty with studying subjective organisation is that it is not easy to identify the particular basis for organisation used by any one subject and for this reason Tulving and others depended more heavily on methods using pre-determined organisations decided by the experimenter. Pollio, Richards and Lucas (1969) found that in recall of a list of 25 words, comprised of five members from five categories, participants produced clusters of category members irrespective of whether or not their attention was drawn to the categorical nature of the word-list.

Gronlund and Shiffrin (1986) examined the effect of various enforced retrieval strategies on recall from natural categories. The authors concluded that free recall probably involved " use of idiosyncratic strategies that provide a roughly ordered sequence of cues somewhat superior to alphabetic or size cues" (pp. 555). Three strategies were compared, size (recall to begin with the smallest physical category member, increasing in size) alphabetic (recall to begin with members starting with the letter 'a' and cycling through the alphabet) or free (no strategy enforced). Superior recall in terms of total items generated was found for the free recall condition. In one experiment (Gronlund and Shiffrin, 1986, exp.3) participants had to learn lists of 25 words from each category. This time there was no effect of strategy. The authors conjectured that this may reflect a change in the nature of the free recall strategy used when a list is to be recalled as compared to free recall strategies used when the task involves drawing on long-term semantic memory. This distinction is worthy of note, as it will be returned to later (see Discussion). The method of testing recall of word lists presented in experimental settings was believed to be tapping episodic memory.

The existence of episodic memory as a distinct memory system was famously espoused by Tulving (1972). In its weakest form, the distinction represents "a hunch that there is something different about the way people remember their own past and how they store and retrieve general information about the world" (Tulving, 1986, p.308). However, in its strongest form, episodic information "stores information about temporally dated episodes or events and temporal-spatial relations among these events" (Tulving, 1972) and constitutes a separate memory system to semantic knowledge which "involves general, abstract, timeless knowledge that a person shares with others" (Tulving, 1986, p. 307). The learning of word lists, despite being comprised of words the meaning of which would be presumably the bailiwick of the semantic memory system, was considered an episodic task because it is the experimenter's particular choice of the subset of words presented in the task which has to be learned and recalled. This methodology was the primary manner in which the characteristics of episodic memory could be examined. Free emission tasks, on the other hand, were seen as drawing on semantic memory, since the knowledge base supporting the task is the participants' knowledge of the world that predates the laboratory experiment. The following group of studies specifically made use of free emission procedures.

Free Emission Studies

Gruenewald and Lockhead (1980) suggested that one of the reasons for any similarity of the output from both list learning and free emission tasks is that in both cases the participants must continue over time to produce new items from a pool of possible responses which decreases in size with each response. Gruenewald and Lockhead sought to establish whether the parametric characteristics of recall from list-learning data (e.g., exponential decay) would apply equally to data from a free emission task. The second aim of their paper was to determine a way of establishing the presence of clusters of related items in an output which extrudes from a wide knowledge base and not from a carefully chosen list of words where potential clusters can be determined a priori. The 4 categories used were *animals, birds, foods* and *cold foods*. A period of 15 minutes was allowed for each category. Clusters of contiguous items were noted in the output. The authors were, once again, concerned with producing an algorithm that characterised the temporal clustering of the data. In line with previous researchers (e.g., Graesser & Mandler, 1978; Pollio, Kasschau and DeNise, 1968), they assumed that the free recall process was comprised of two stages and appealed to the readers to introspect on how the task might be performed. "Consider recalling food words. Perhaps you would consider that supermarkets have foods and attempt to search the store aisles in your imagination; then you might recall foods on the dining room table; perhaps next to vegetables, then to fruits, to foods which are disliked, and so on...." (p. 229). Gruenewald and Lockhead suggested that participants search a "semantic field" in order to retrieve clusters of semantically related items. One or two clusters from their participants are itemised anecdotally but yet again, no inquiry was made into whether participants actually made use of the strategies suggested above in completing the task.

Unlike the studies discussed so far, Williams and Hollan (1980) were interested in the think-aloud protocols provided by the participants and in the strategies deployed in fulfilling the demands of a free emission task. Williams and Hollan conducted a detailed observational study in which four subjects over periods of between 4 and 10 hours had to recall names of high school peers. For one participant this meant recalling names from 19 years ago. Williams and Hollan conceived the recall task as a "complex problem-solving activity" (p.87). Their data was strongly suggestive of what the authors called a "retrieval cycle" which consisted of three recursive stages. First participants have to establish a retrieval context that is "designed to focus on a relevant subset of the initial information provided". Having found this context, it is used to guide the search of an indefinitely large database. Having retrieved items using this context as a cue, the item is verified as meeting the requirements of the original instruction. The process is strategic, effortful and reconstructive. The notion of a mediating cue in facilitating the memory search is consistent with some of the other suggestions considered here such as Gruenewald and Lockhead's 'semantic fields' and Graesser and Mandler's (1978) 'conceptual dimensions'. Williams and Hollan noted that two of the most extensively used contexts were locations and activities, such as where classmates lived or the sorts of activities the recaller took part in with the target, "people I played tennis with".

Walker and Kintsch (1985) also collected protocols (both concurrent and retrospective) from their participants who verbally listed members of two categories, *animals* and *furniture* over a time period of 12 minutes per category. Walker and Kintsch looked to model retrieval of 'real world knowledge' using a model by Raaijmakers and Shiffrin (1981) which had been developed to account for data from list-learning experiments. The protocols showed that category name was not the only cue utilised by participants; rather participants generated their own retrieval cues and produced items to those particular cues. For example, for the category *Soup*, one subject reported thinking of 'Soups I eat most often' and " I am trying to think of different vegetables that might be in soups". This led Walker and Kintsch to propose a two-stage model of retrieval under these conditions. Participants must first produce appropriate cues and then use these to search for retrieval of appropriate category members. The authors' comments on the nature of their findings presage the results of the present study. The category that is generated in response to the task is not " in any sense a memory unit that can simply be retrieved. Furthermore, these intermediary memory nodes that are involved in retrieval are typically not abstract, context-free and general, but can be very personal and idiosyncratic" (p. 280). However, frustratingly they did not investigate, or even document in any detail, the actual strategies reported by their 12 subjects.

What is the Value of the Free Emission procedure?

The importance of studying the output response and reported strategies in free-emission tasks is the possibility of drawing inferences from these data to the organisation of the knowledge base in memory. There is a good precedent for the assumption that the clustering seen in the output obtained through free recall tasks provided an insight onto the underlying memory organisation (Barsalou & Sewell, 1985; Chase & Ericsson, 1981; Graesser and Mandler, 1975; Mandler, 1967; Reitman, 1976; Tulving, 1962,1964, 1966; Tulving & Pearlstone, 1966). Lancaster and Barsalou (1997) in examining the organisation of people's memories of events, suggest that when participants are asked to learn lists in laboratory conditions "the most parsimonious and simple approach is to adopt existing organisational systems" (p.597). Under this view one would expect similar patterns of organisation to appear in both list learning experiments and free emission studies. This has been found to be the case in some studies (e.g. Rubin and Olson, 1980) but note Gronlund and Shiffrin (1986, Experiment 3), in which there was no advantage for a free recall compared to an alphabetic strategy in recalling word lists whereas such an enhancement of recall had been found in the previous two experiments using a free emission procedure. This suggests an interaction of strategy and task in as much as the strategy most effective in generating items from one's conceptual knowledge might not be useful in assisting recall of a predetermined word list. Gronlund and Shiffrin did not investigate the 'idiosyncratic' strategies used by participants in the free recall condition and were therefore not able to say with any certainty that the nature of these had changed between the two different task types. It seems that the nature of these strategies, referred to briefly in the studies discussed above and examined more closely by Walker and Kintsch (1985) and Williams and Hollan (1980) for a very limited range and type of category, bear closer examination for a wider range of categories.

The new studies described in this chapter investigate the extent to which subjects report the use of search strategies in retrieving members of a diverse range of categories. Furthermore, the intention is to document and attempt to classify the strategies reported and to consider whether they can shed any light on the organisation of conceptual knowledge. Bear in mind that many of the studies discussed above point to a hierarchical, taxonomic organisation of semantic memory. But what of the organisation within categories? Henley (1969) reported that dimension of size and ferocity characterised the output of free recall from the category "animals". It should be noted, however, that these dimensions were the author's interpretation of the patterns produced through multi-dimensional scaling. A dimension showing *tiger, bear* and *lion* at one end and *sheep, deer* and *donkey* at the other could well be indicative of a perception of similarity in terms of ferocity but *cat* and *dog* (both household pets) were nearer to the top of this dimension than the bottom. An equally valid interpretation might be one of perceived intelligence in the animals. Rips, Shoben and Smith (1973) found similar dimensions relating to *birds*, as well as sub-domains of *predators, domestic birds, domestic edible* or *farm birds*. One might therefore reasonably expect that participants will report strategies involving searching for different 'types' of animals and that the same may be true of other well established, well structured taxonomic categories such as *fruit, furniture* and *vehicle.*

Taxonomic categories are not the only sort of categories of which we have knowledge. Barsalou (1983) pointed out that we are able to construct a wide range of ad hoc categories such as *things to save from your house in a fire, things to pack for a holiday* etc. These are categories that can be contrasted with well-rehearsed taxonomic categories and the two category-types have been differentiated in various ways. Although both exhibit graded structure (certain members are viewed as more typical of the category than others), typicality is determined largely by similarity to a central tendency in the case of taxonomic categories but the typicality of a member of an ad hoc category will depend upon how often that member has been encountered as a member of that category (frequency of instantiation) and whether it exhibits characteristics which make it ideal in meeting the goal expressed by the category title . Taxonomic category members have been said to share context independent properties whilst ad hoc category members share properties rendered salient only by context. Taxonomic categories are generally named by a single lexeme whereas ad hoc categories are labelled with a "complex phrasal construction" (Barsalou, 1991). Barsalou (1991) has also hypothesised that the means of acquisition for these two categories may vary; "The acquisition of common taxonomic categories, such as *apple, bird, shirt* and *chair*, relies heavily on experiences with exemplars" (p.2) but he goes on to claim that acquisition of a category like *things to pack in a suitcase* does not rely on such experiences. Furthermore, these different types of category are alleged to play different roles in the cognitive system. Barsalou (1991) talked of taxonomic categories as providing "building blocks for world models" while goal derived categories "provide interfaces between world models and event frames for achieving goals" (p.53). Given these structural and role differences between taxonomic and ad hoc categories, it would be reasonable to expect that participants would draw on different strategies if asked to generate instances of each of these types of category.

If 'strategies' of retrieval are required at all for generating members from taxonomic categories, then access through semantically related subgroups, such as those witnessed by Henley (1969) and Rips et al. (1973), would be anticipated. In contrast, performing the same task with ad hoc categories may be a matter of 'constructing' members by considering the ideal characteristics defined by the category title. This would particularly be expected to be the case for categories of which participants would have little experience. Barsalou (1997) has proposed that "ad hoc categories construe entities as playing roles in events" (p. 30) .For example, *things to stand on to change a lightbulb* may access *chair, stool,* and *table*, through recreating experiences of the event, changing a lightbulb. Therefore, one might expect access to more familiar ad hoc categories to be mediated by consideration of those events. Alternatively, it may also be possible to access members of this category by deciding upon the characteristics needed to meet the goal of the category, i.e. objects of a particular height, which will bear your weight, which may be nearby in locations where you are likely to be changing a bulb etc. Having generated suitable items, these may be then tested for suitability as a response. There are reasons then to anticipate differences in strategies that will be used by participants to complete a free emission task for taxonomic as compared to ad hoc categories.

Study 1

The participants in this study generated instances for each category for a 90s period. After each category the participants wrote down the strategies that they had used in the process. One of the aims was to assess the extent of the prevalence of experiential strategies documented in Walker and Kintsch (1985) across a more representative range of categories, and in particular to assess their usage for taxonomic categories. Previous research has established a close connection between autobiographical memories and ad hoc categories (Conway, 1990a) and it was therefore expected that instance retrieval from 'familiar' ad hoc categories would be mediated to a greater extent by experiential strategies. Instance retrieval for categories that are better established in long-term memory, for example, the taxonomic category *fruit*, might not be mediated by experiential cues to the same degree as familiar ad hoc categories: for example, as suggested in Barsalou (1983), the strong category-instance associations might render such mediations superfluous. In addition, if experiential cues are used spontaneously for familiar ad hoc categories, it is of interest to investigate what cues subjects employ when they have had little or no experience with a category (e.g., animals found on the Galapagos).

Method

Participants

Fifty undergraduates from the University of Hertfordshire volunteered for this study. The modal age of participants was 19yrs.

Categories and Design

Ten categories were selected, two taxonomic categories taken from Rosch and Mervis (1975), namely *vehicle* and *fruit,* and eight ad hoc categories. These eight ad hoc categories were made up of three subgroups of categories. The first group was one for which the category instances were concrete objects whose grouping was likely to be familiar to the participants, namely *things people keep in their pockets, things dogs chase, things people take to a wedding, things made mostly of plastic.* For ease of exposition we will refer to these four categories as *pockets, dogs chase, wedding, and plastic,* respectively. The second group of ad hoc categories was one for which the category instances would cover a different ontological range including, for example, moods, feelings, life events, causal explanations, although participants were unlikely to think the groupings odd. These were *reasons for going on a holiday*, and *excuses for arriving somewhere late.* For convenience these will be referred to as *holiday* and *late.* The final group of ad hoc categories was made up of concrete entities whose groupings were unlikely to be familiar to the participants. These were *animals found on the Galapagos*, and *things sold on the black market in Russia* (these two categories will be referred to as *Galapagos* and *black market*).

Procedure

All participants received all ten category names in one of ten randomised orders. The participants were then asked to answer in writing the question "how did you go about thinking of items for this category?". Answers to this question provided the data on the nature of the retrieval strategies. Participants were also requested to rate how difficult they found it to produce items for that category on a scale of 1 – 10 where 1 indicated not difficult at all and 10 indicated very difficult. Test materials were assembled in a booklet. Single category names were printed on the top of separate lined sheets. After each category sheet, a sheet with the three questions was inserted.

The front page of the booklet informed the participants about the nature of the task and that for each category they would be given 90s to list the names of as many items they could think of (see Appendix B for full instructions and sample page of materials).

Participants were run in groups ranging in size from five to ten. Once the experimenter had answered any queries, participants were instructed to turn over the cover page and start listing items for the first category. After 90s had elapsed, participants were instructed to turn over the page; they were given another 90s to answer the question concerning retrieval strategy. This procedure was repeated for the remaining nine categories. We chose to segment the experimental procedure into those time intervals in order that test booklets would take no longer than 30 minutes to complete.

Measures

Three measures are reported. These are: 1) the number of items produced for each category (*output fluency*) 2) the difficulty ratings provided for each category 3) the proportion of participants who generated each of the ten most generated items for each category, a measure of consensus termed *item dominance* and 4) The segmentation and classification of participants' answers to the question "how did you go about thinking of items for this category?". The answers to this question were first segmented into parts, where each segment appeared to correspond to a distinct idea or approach to item generation. (In many cases the answers were naturally segmented by the participants with the use of terms such as "and" and "then".) Two investigators independently segmented all protocols and met to compare their resulting segmentations and resolve disagreements. The number of protocol segments ranged from 93 (*holiday*) to 117 (*pockets*), with a mean of 102 segments per category (s.d. = 9.5). All protocol segments were then transcribed onto index cards labelled with a subject number and the category name. Triplicates of the cards were produced and a set given to the experimenter and two colleagues. The cards were independently sorted into groups. In a first classificatory stage, segments were sorted largely on the basis of their surface form, i.e., on a minimal interpretation of their meaning. For example, for *fruit*, protocol segments like "fruit I eat" and "fruit I like" were sorted initially in different groups. In a second phase, groupings of protocol segments were identified that appeared to reflect similar approaches to item generation. The final classification of the segments was agreed by all three judges. Three broad classes of strategies were identified: 1) experiential mediation, 2) semantic mediation, and 3) unmediated retrieval. Eighty-eight percent of the protocol segments could be classified under these headings (8% resisted such groupings and 4% were uninterpretable).

Results

Output Fluency

The mean number of items generated for each category and each category type are reported in Table 2.1 and illustrated by category in Figure 2.1. Participants generated most items for the taxonomic categories (mean of 12.98) and fewest for concrete unfamiliar ad hoc categories (mean of 5.87). To illustrate the differences and similarities both between and within category types, a number of pairwise comparisons were carried out. Thus, within the taxonomic categories, fluency for *fruit* was significantly greater than for *vehicle,* $t(49) = 4.86$, $p <. 0001$ (all tests at 95% significance level two-tailed). Means for *vehicle* and *pockets* were not significantly different, but fluency for *plastic* was significantly smaller than for *pockets*, $t(49) = -10.3$, $p < .0001$. Across ad hoc category type the mean for *plastic* was smaller than for *holiday*, $t(49) = -3.44$, $p < .001$. Finally, the fluency for *Galapagos* was significantly smaller than the fluency for *black market*, $t(49) = 2.23$, $p < .03$

Table 2.1

Mean output fluency (standard error in parentheses) for the 10 categories in Study 1

		Taxonomic Categories		
Fruit	*Vehicle*			Overall mean
14.80	11.14			12.98
(0.63)	(0.61)			(0.48)
		Concrete Familiar Ad hoc categories		
Pockets	*Dogs Chase*	*Wedding*	*Plastic*	
11.14	9.10	8.04	6.78	8.77
(0.40)	(0.30)	(0.41)	(0.35)	(0.21)
		Abstract Familiar Ad hoc categories		
Holiday	*Late*			
8.20	8.18			8.19
(0.36)	(0.36)			(0.26)
		Unfamiliar Ad hoc categories		
Galapagos	*Black Market*			
5.00	6.74			5.87
(0.37)	(0.53)			(0.34)

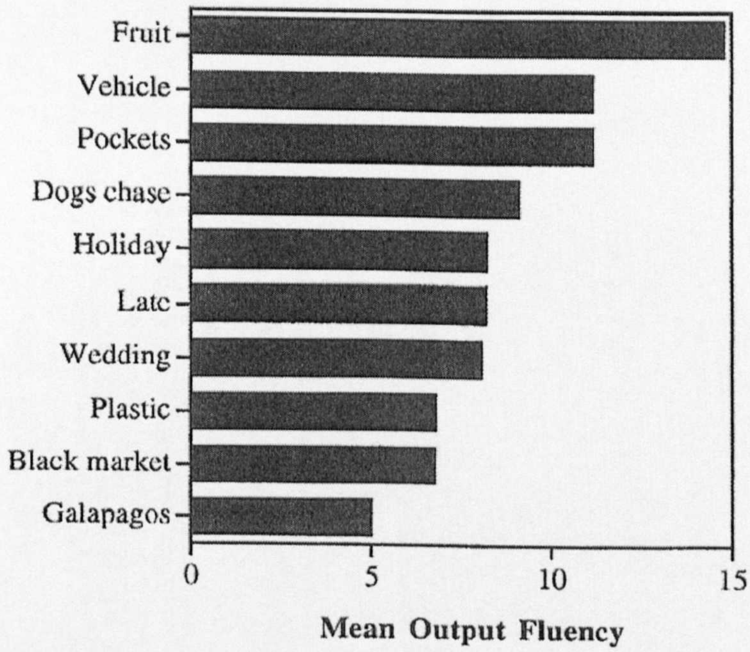


Figure 2.1 Rank order of mean output fluency for each of the 10 categories.

Difficulty Ratings

After generating members for each category, participants were asked to indicate how difficult they had found the task on a scale of 1 – 10 that was anchored at each end by not difficult at all (1) and very difficult (10). Mean ratings were calculated for each individual category and for each category type (Table 2.2).

Table 2.2

Mean Difficulty Ratings (standard deviation in parentheses) for the 10 categories in Study 1. The higher the rating, the more difficulty is indicated.

		Taxonomic categories		
Fruit	Vehicle			Overall mean
1.75	2.49			2.13
(1.36)	(1.66)			(1.25)
		Concrete Familiar Ad hoc categories		
Pockets	Dogs Chase	Wedding	Plastic	
2.93	3.86	5.29	6.25	4.58
(1.72)	(1.88)	(2.48)	(2.4)	(1.41)
		Abstract familiar Ad hoc categories		
Holiday	Late			3.92
4.27	3.59			(1.81)
(2.4)	(1.94)			
		Unfamiliar Ad hoc categories		
Galapagos	Black Market			
7.94	6.73			7.34
(3.31)	(2.71)			(2.06)

A one–way repeated measures ANOVA showed a significant difference in the ratings for the four category types, Wilks Lambda $F(3,46) = 80.606$, $p<0.001$, eta squared = 0.84. Participants rated the unfamiliar abstract categories as most difficult ($M = 7.34$) and pairwise contrasts showed this to be significantly more difficult than the next most difficult category type which was concrete familiar ($M = 4.58$) Wilks Lambda F

(1,48) = 83.99, p<0.001, eta squared = 0.636. (The largest range within a category type occurred in this group - between *pockets* and *plastic*, 3.32) This mean in turn was significantly higher than that of the abstract familiar categories (M = 3,92) Wilks Lambda F(1,48) = 6.96, p= 0.01, eta squared = 0.127. The lowest difficulty ratings were found for the taxonomic categories (2.13) and this was significantly lower than the mean for the abstract familiar categories Wilks Lambda F(1,48) = 39.622, p< 0.001, eta squared = 0.693.

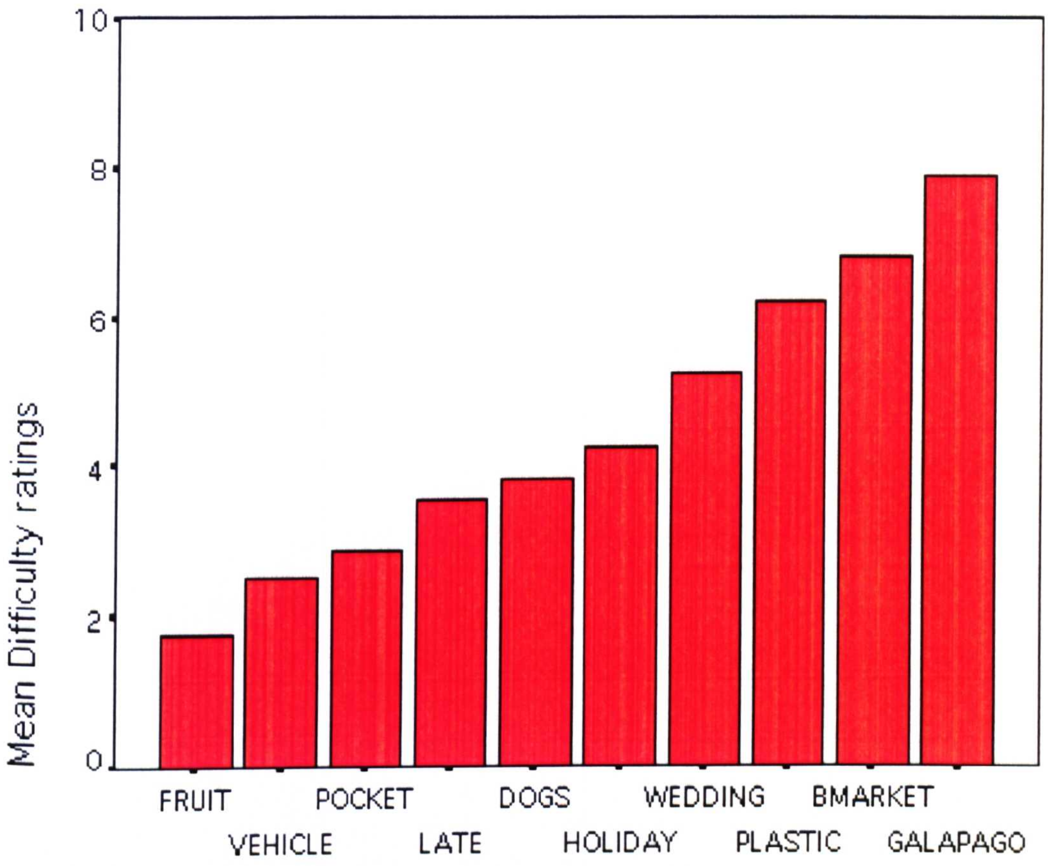


Figure 2.2 Mean difficulty ratings for the 10 categories in Study 1 arranged in ascending order.

It can seen that the individual category means are anchored by the taxonomic at one end (easiest) and the unfamiliar categories at the other (most difficult) but the two familiar ad hoc category types are somewhat interleaved with *Wedding* rated as more difficult than *holiday* $F(1,47) = 4.973$, $p = 0.031$, eta squared = 0.096. Pairwise contrasts showed no significant difference between *vehicle* vs. *pockets* ($p=0.194$), *Late* vs. *dogs* ($p = 0.387$), *dogs* vs. *holiday* ($p=0.255$) and *plastic* vs. *black market* ($p=0.199$). The difficulty ratings correlated significantly and negatively with fluency for every individual category except *dogs* and *holiday* and overall mean difficulty rating for each category correlated significantly and negatively with overall mean fluency $r = -0.909$, $p< 0.001$.

The participants' perceptions of how difficult they found it to generate members for the different categories seemed to be reflected in the number of items they actually generated with the most difficult categories resulting in the lowest fluency. This supports the notion that taxonomic categories are better established in memory than ad hoc categories and are therefore easier to access, however these findings suggest that there are also significant differences in the ease with which people may generate members of differing types of ad hoc categories. It should be noted that each question on difficult followed each generation task so the difficulty ratings can be seen as a post hoc justification of performance and would therefore expect to be correlated with the number of items generated. The difficulty ratings predicted a significant amount of variance in all categories except for *dogs* and *holidays*. Figure 2.3 shows regression lines for all four category types.

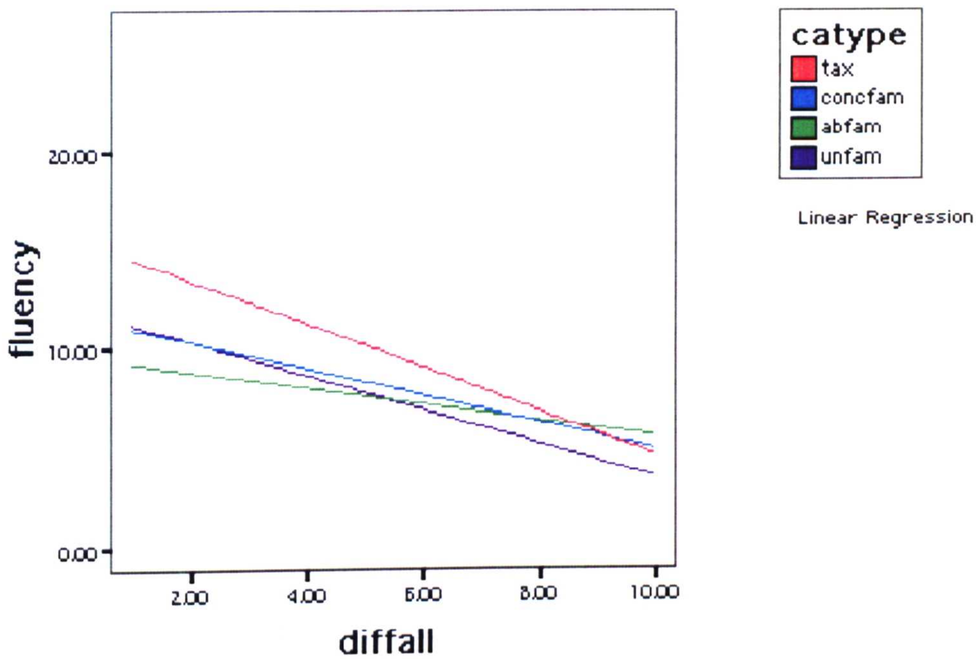


Figure 2.3 Regression slopes for difficulty ratings as a predictor of fluency for taxonomic, concrete familiar, abstract familiar and unfamiliar category types.

Table 2.3

Values of adjusted r2 for all categories and category types. All values are significant unless so indicated by n.s.

		Taxonomic categories		
Fruit	Vehicle			Overall
0.073	0.163			0.177
(7.3%)	(16.3%)			17.7%
		Concrete Familiar Ad hoc categories		
Pockets	Dogs Chase	Wedding	Plastic	
0.105	0.049 n.s.	0.081	0.328	0.028
(10.5%)		(8.1%)	(32.8)	2.8%
		Abstract familiar Ad hoc categories		
Holiday	Late			
0.046 n.s.	0.200			0.109
	(20%)			(10.9%)
		Unfamiliar Ad hoc categories		
Galapagos	Black Market			
0.293	0.427			0.406
(29.3%)	(42.7%)			40.6%

The relationship between difficulty ratings and fluency is strongest for the unfamiliar ad hoc categories where participants tended to produce few items and give high difficulty ratings. In the taxonomic categories participants were most fluent and tended to give low difficulty ratings.

An ANCOVA using difficulty as a covariate showed that difficulty accounted for a significant amount of variance in fluency rating $F(1,480) = 129.9$, $p<0.001$, eta squared = 0.213. There was a significant main effect of category type $F(3,480) = 23,758$, $p<0.001$, eta squared = 0.129. There was also an interaction between difficulty and category types $F(3,480) = 4.01$ $p = 0.008$, eta squared = 0.024, showing that the slopes shown for each category type in Figure 2.3 were not the same.

Item Dominance

Figure 2.4 shows the proportion of participants who generated the *i*th most generated items (*i* ranged from 1 to 10) for each category (right panel) as well as the average by category type (left panel). The picture conveyed by the left panel is clearer. Item dominance distributions for taxonomic categories showed the most consensus while the least consensus was observed for the unfamiliar ad hoc categories. The right panel portrays a fuzzier picture where the item dominance distributions for each category overlap a great deal. On the basis of Kolmogorov-Smirnov analyses (Wilkinson, 1990), the item dominance distribution for *fruit,* which reflected the highest consensus, was significantly different from the distributions for all other categories ($p<0.0001$) categories. . In turn, the item dominance distribution for *plastic,* which reflected the least consensus, was significantly different from all categories except *Galapagos* ($p=0.749$) and *black market ($p=0.271$).* For the categories between these two extremes, the item dominance distributions did not differ significantly.

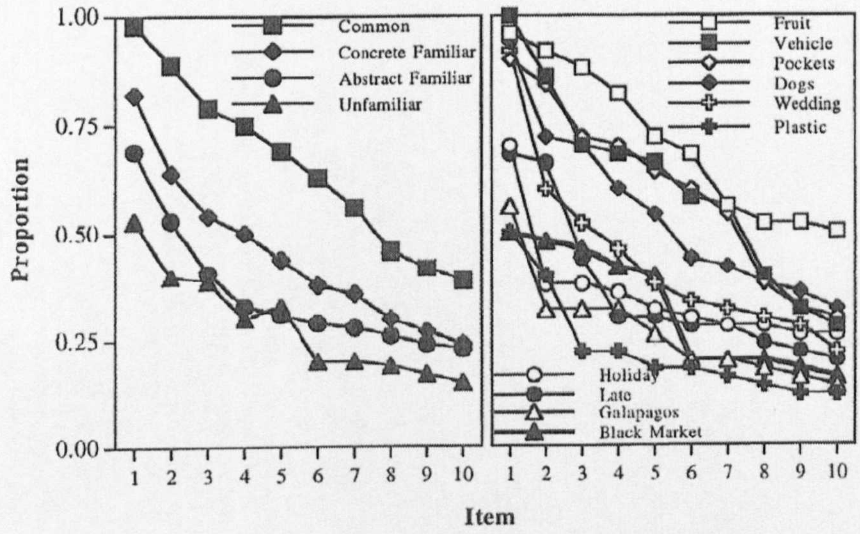


Figure 2.4 Proportion of participants generating each of the 10 most generated items for the 10 categories (right panel); average proportions for each category type shown in left panel.

Retrieval Strategies

Table 2.4 reports all the subgroups that were identified in each of the three categories for *fruit* from the protocol segments of Study 1. Protocol segments classified within the experiential mediation category indicated that the generation of category instances was based on autobiographical facts or mediated by the recreation of specific or generic personal experiences during which these instances are encountered. This latter distinction is based on the one made in Conway (1990c) between specifically dateable experienced events and "experienced events (...) which subjects could only date to a lifetime period and which were abstracted from specific experiences." (p. 134). Barsalou, Yeh, Luka, Osleth, Mix and Wu (1993) similarly argue that "(...) an episodic situation represents a single event that occurred at a specific time, [and that] a generic situation generalised over related episodic situations." (p. 38). Thus, the participants reported using specific personal experiences to aid retrieval, such as, for *wedding*, "the last wedding I attended" or more generic experiences, such as for *dogs chase*, "dogs I've seen in the park". Finally, strategies that described second hand personal information, that is, being told directly of the experience (e.g., by friends, via the television) were also included in the experiential mediation category. Segments classified within the semantic mediation category suggested that instance retrieval was mediated by access to subtaxonomic groupings, or by the consideration of the abstract characteristics of the category, or by analogy to related indices ("I just thought of the type of things which are black market in other places"); or by using a distinctive reasoning strategy, like a recipe for a fruit salad (or thinking of the contents of yoghurt pots). The unmediated retrieval category grouped together protocols that implied that instances were not retrieved via an intermediate cue of which subjects were aware at the time of report or which implied an appreciation of the graded structure of the category (i.e., that some instances were "common"). Segments classified in that category specified that instances “popped into my mind”, or “came unbidden” or were “common knowledge”.

Table 2.4

Subgroups of Protocol Segments for *fruit* within each of the three classes of retrieval strategy.

Experiential	Semantic	Unmediated
1. Fruit I eat	1. Consideration of subcategories: citrus fruits, exotic fruits, garden fruits	1. Popped into head
2. Fruit I like or don't like	2. Reasoning about features (sweet vs. Bitter).	2. General knowledge ("I just knew them")
3. Fruit I buy	3. Recipe	3. Notion of typicality
4. Fruit I've seen	4. Cued by other products, e.g., types of yoghurt, fruit juices.	4. Inter-item cueing
5. Fruit seen in supermarkets	5. Linguistic cues like "berry".	
6. Fruit bowl at home		
7. Specific memories of eating a fruit		
8. Fruits I've heard of		
9. Media		
10. What I know other people eat		

Table 2.5 reports representative examples of actual protocol segments for *fruit*. The protocol segments classified in the experiential mediation category ranged from autobiographical facts (e.g., "what I hate"), reflections on specific life episodes where fruits were encountered (e.g., "the fruit salad my friend made the other day"), to generic episodes (e.g., "fruits I see in the supermarket"), to experiences of others (e.g., "fruits my children like"). The protocol segments classified in the semantic mediation category indicated that instances were generated on the basis of strategies that made no reference to personal experience, such as a reflection on taxonomic organisation (e.g., "summer fruits"); a consideration of where they grow (e.g., "garden") or how they are utilised (e.g., "yoghurt pots"). The unmediated retrieval category was more

liberally constructed, including any protocol segments that did not refer to a deliberate strategy and/or that indicated that the category instances were retrieved automatically. As it will become apparent below, because the proportion of protocol segments fitting this category was quite small, adopting such a liberal classification criterion did not artificially inflate the importance of these protocol segments.

Table 2.5

Protocol Segments that described Instance Retrieval Strategies for *fruit* in Study 1.

Experiential Mediation

"What I hate"

"What I eat."

"First thinking of all my favourite fruits."

"Thought about the fruits I bought."

"I though of the fruit salad my friend made the other day"

"My auntie's fruit bowl"

"What I have growing in my garden."

"I took the easy way out and wrote what I see in the supermarket."

"Fruits that my children like or dislike."

"Visualised media adverts."

Semantic Mediation

"I though of categories, e.g. summer fruits."

"Tropics, orchard, garden."

"I thought of fruit which you can grow in the garden."

"Fruits to make fruit salad."

"Yoghurt pots."

Unmediated Retrieval

"Fruits -all different kinds are common knowledge."

"General knowledge."

"By the word association, i.e., one fruit brought back recollection of the next."

"Whatever came to mind."

"I listed the fruit I knew of."

The percentages of experiential, semantic and unmediated protocol segments for each category as well as averages for each category type are reported in Table 2.6. Since most subjects reported more than one strategy, protocol segments were not independently produced. Consequently, no inferential statistics are reported. However, clear patterns in the segment distributions can be observed. For seven of the eight categories, experiential strategies greatly outnumbered both semantic strategies and reports of unmediated retrieval. In terms of individual categories, the percentage of experiential protocols outnumbered semantic protocols for *fruit, pockets, dogs chase, wedding, plastic, holiday*, and *late*; for *vehicle* the percentages were nearly even. Striking reversals of this pattern were observed for *black market* and *Galapagos,* as illustrated in Figure 2.5.

Table 2.6.

Percentage of Protocol Segments classified as indicating Experiential Mediation, Semantic Mediation and unmediated retrieval for each category and each category type.

	Experiential	Semantic	Unmediated
Taxonomic Categories			
Fruit	67	16	17
Vehicle	36	40	25
Mean	52	28	21
Concrete Familiar Ad Hoc Categories			
	Experiential	Semantic	Unmediated
Pockets	83	10	7
Dogs chase	54	31	15
Wedding	62	25	14
Plastic	74	12	14
Mean	68	19	13
Abstract Familiar Ad Hoc Categories			
Holiday	60	27	13
Late	69	10	22
Mean	65	19	18
Unfamiliar Ad Hoc Categories			
Galapagos	24	67	9
Black market	32	65	4
Mean	28	66	7

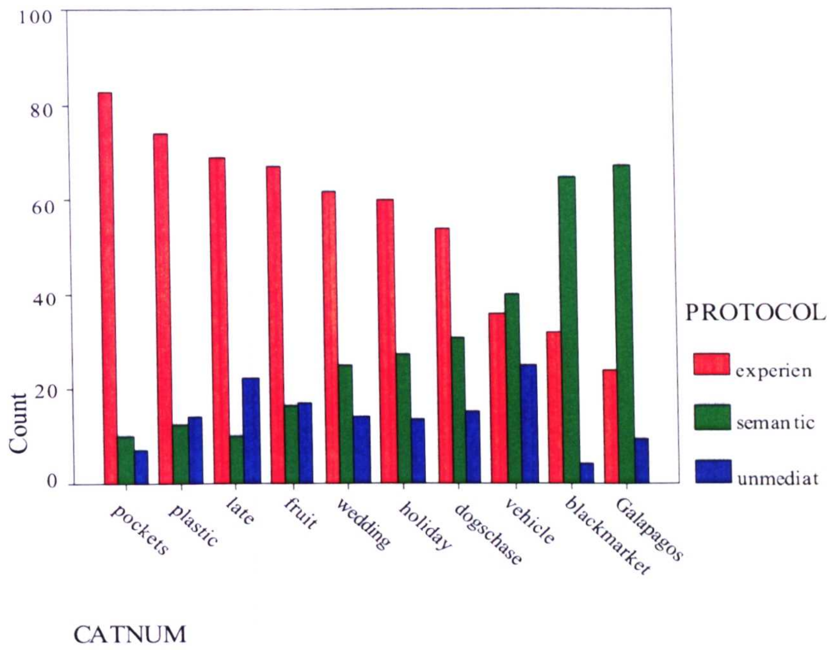


Figure 2.5 Percentage of Protocol Segments classified as indicating Experiential Mediation, Semantic Mediation and unmediated retrieval for each category.

Discussion

The data on output fluency, difficulty and retrieval consensus will be discussed first, followed by a discussion of the data on retrieval strategies. Both the difficulty ratings and the output fluency indicated that participants found it easier to generate items for the taxonomic than for the ad hoc categories, and within ad hoc categories, they found it easier to generate items for the familiar types. Since the difficulty ratings were made on completion of the output for each category and participants would be aware of how few or how many members they had managed to recall, it is unsurprising that the ratings accounted for such a large amount of variance overall in the fluency. This does suggest that participants were generally reporting the difficulty with a high degree of honesty and also gives one some confidence that they took the task seriously when responding to the questions asked. The results concerning output fluency are also unsurprising and are in general agreement with those reported by Barsalou (1983) where the average output fluency for his nine taxonomic categories and the average for nine ad hoc categories differed reliably. These statistical differences may suggest a clear dichotomy between taxonomic and ad hoc categories. However, if the mean output fluency for each category is plotted in rank order, as in Figure 2.1, a more subtle picture emerges: While *fruit* and *Galapagos* anchor the two extremes of the graph, the smooth transition within these two poles suggests no important discontinuities corresponding to a processing boundary between taxonomic and ad hoc categories (nor between the different types of ad hoc category). In this respect, it is interesting to note that our item dominance distributions overlapped considerably (see right panel of Figure 2.4).

The data on the reported strategies (see Table 2.6) suggests that the mediators that participants often used were more closely related to personal experiences that establish the meanings of everyday concepts than to the decontextualised semantics in which theories of everyday concepts are generally framed. The percentage of protocols indicating unmediated retrieval was lower than 20% in eight of the ten categories. For seven out of the ten categories in this study, the proportion of protocol segments indicating an experiential strategy was much greater than those indicating a semantic strategy. This pattern was noticeably reversed for *Galapagos* and *black market.* In these two categories, less than a third of the protocol segments indicated an experiential strategy; where these did occur they tended to describe memories of watching television programmes or other media sources. Participants were more obliged to reason explicitly about the membership criteria for these two unfamiliar categories. Furthermore, the distribution of experiential/semantic strategies for these two categories indicate that the question "how did you go about thinking of items for this category" did not bias the retrospective description of the strategies in a way that would artificially favour experiential over semantic strategies.

Semantic strategies were about as frequent as experiential strategies for *vehicle.* It appears that despite the taxonomic usage of this category label, in both everyday discourse and in concept research, a number of participants experienced some uncertainty in defining the scope of category membership. For example, one participant wrote: "Thinking that vehicle includes a moving means of transport on four wheels. Is it true? I'm not sure." and another wrote: " I think I may have got a bit carried away and put things which would come under transport rather than vehicles.....". This uncertainty over category membership seemed to result in some participants seeking to establish a definition of the category in terms of properties, for example "Transportation over distance for one or more human beings which requires a source of power". Protocols such as these were classified as semantic. This was the only category which appeared to cause widespread 'boundary disputes'.

Certain considerations might have led us to expect a greater proportion of semantic strategies for both taxonomic and ad hoc categories. In some traditional accounts, the long-term memory representation of categories is characterised as being structured along conceptual dimensions (e.g., Henley, 1969; Rips, Shoben, and Smith, 1973; Rubin and Olson, 1980; Lund and Burgess, 1996). These dimensions are uncovered through analyses of item proximity in subjects' free recall output; or on the basis of pairwise similarity ratings; or on the basis of lexical co-occurrence in text. Matrices of item proximity are cluster analysed (e.g., Friendly, 1977) or subjected to a multidimensional scaling analysis that yields 'interpretable' dimensions along which items vary. For example, Rubin and Olson (1980) report that size and ferocity are dimensions that underlie *mammals*. Given that participants' output for a taxonomic category can be characterised by a small set of semantic dimensions, it is perhaps surprising that participants referred to them so rarely when generating instances. It was decided that a proximity analysis of the items generated by participants in this study would not bear on the issue of interest here (i.e. the strategies that they claim to be using to complete the task) for the following reasons. As mentioned above, these analyses have been used to infer underlying conceptual dimensions in the output. A similar procedure here may have resulted in some clusters of items that could, post hoc, be interpreted as representing some dimension such as citrus fruits or exotic fruits. If so, it could have been argued that these are the dimensions that characterise the representation of the category in memory and that the strategies claimed in the protocols are a) either not a valid account of how the task was actually achieved or b) are indicative only of retrieval strategy and do not imply anything in particular about the nature of the category representation. Both of these possibilities will be discussed later in this chapter and the reader will be given reasons to discount them. It would also not be possible to use the results of a proximity analysis to validate the reported strategies since the same strategy may lead to different items being reported - e.g. fruits I buy, fruits I like etc. It has been generally assumed that the nature of the cues which provide the most successful access to category exemplars shares important similarities with the way these exemplars are represented in memory (e.g., Barsalou and Sewell, 1985, p. 651; Morris, Bransford, and Franks, 1977; Tulving and Thompson, 1973). A proximity analysis would neither support nor rule out the possibility that the situational, self-generated cues identified in this study are successful because experiential aspects of the context in which objects as instances of a category are encountered are an intrinsic part of the category representation.

As Barsalou (1991) has demonstrated, an explicit consideration of the ideals defining an ad hoc category often guides the generation of instances during planning. For example, in planning a holiday, people need to consider the category *vacation location*. Barsalou reported data that indicated that people tailor the candidate instances of *vacation locations* along certain ideals, reflecting their current goals and constraints. Furthermore, the graded structure of an ad hoc category is determined in part by the extent to which its instances satisfy the ideal or goal associated with the category (Barsalou, 1985). In this respect, one might surmise that ad hoc categories are to some extent perhaps even more amenable to a conscious consideration of the semantic dimensions that define them than are taxonomic categories. Thus, in the case of *pockets*, it may not have been unreasonable to expect participants' protocols to display an appreciation of the practical nature of the objects carried in pockets in terms of quotidian needs and goals. Yet such considerations very rarely surfaced in the participants' protocols, in stark contrast with the panoply of personal experience descriptions.

Study 2

The results of Study 1 suggested that the retrieval of category members from semantic memory is mediated to a large extent by personal experience, for both ad hoc and taxonomic categories. However, two possible artefacts were identified. The first concerns the way in which ad hoc categories were defined to the participants. For example, the formulation of *things people keep in their pockets* might be interpreted to mean "things people I know -myself included- keep in their pockets". Instead of focusing on the nature of the instances that belong to the category, participants might have been encouraged to retrieve personally relevant memories of things kept in pockets. It is possible to rephrase ad hoc categories in a way that focuses more clearly on the nature of the instances; e.g., *things that may conveniently be kept in pockets* instead of *things people keep in their pockets.* This formulation not only eliminates reference to "people" but also emphasises the nature (and perhaps even the function) of the category instances. It was therefore decided in this study to include a condition in which the titles of the ad hoc categories were rephrased in this way. One might consequently expect to see reports of participants thinking of items that fulfil the requirements of the category, i.e. small, light, needed on a daily basis.

The second potential artifact concerns the relative proportion of ad hoc and taxonomic categories, namely 4 to 1, and their order of presentation. On average, across all the randomised orders of presentation, participants were much more likely to have generated instances for ad hoc categories before generating items for the taxonomic categories. If the wording of the ad hoc categories encouraged an experiential retrieval strategy or otherwise did not sufficiently stress the conceptual dimensions that characterise the category members, the participants might have simply continued using experiential strategies for taxonomic categories. To address that concern, this study made use of four taxonomic and five ad hoc categories and additionally fixed the order of presentation such that some participants generated items for taxonomic categories before they were presented with the ad hoc categories.

If genuine differences in strategies exist for this task depending on category type, these alterations to the design should afford the opportunity for this to become apparent. Other than these changes, the methodology replicated study 1.

Method

Participants

Ninety-two undergraduates from the University of Hertfordshire volunteered for this study. There were 31 participants in Condition A, 31 in Condition B, and 30 in Condition C. The modal age of the participants was 18 yrs.

Materials

Nine categories were used, four taxonomic and five ad hoc. The taxonomic categories, taken from Rosch and Mervis (1975), were *fruit, vehicle, bird*, and *furniture.* Three of the five ad hoc categories were of the concrete familiar type, namely *things people keep in their pockets (or the alternative wording, things that may conveniently be kept in pockets), things dogs chase (*or *things that might be chased by dogs)*, and a new category *things people put on walls* (or *things that can be put on walls).* The fourth one was a new abstract familiar ad hoc category, namely *things people hate when they are ill* (or *things people would hate when they are ill*), and the fifth one was the unfamiliar *things sold on the black market in Russia* (or *things that could be sold on the black market in Russia).* For convenience the five ad hoc categories will be referred to as *pockets, dogs chase, walls, ill,* and *black market.*

Design

The order of presentation of the categories and the wordings of the ad hoc categories were manipulated in three conditions. In all three conditions, the categories were blocked such that the participants were presented with the four taxonomic or the five ad hoc categories in consecutive order. Four different randomised orders of presentation were created within each block of categories.

Condition A. The four taxonomic categories were presented first followed by the five ad hoc categories defined with the original wording of Study 1.

Condition B. The five ad hoc categories, defined with the original wording, were presented first, followed by the four taxonomic categories.

Condition C. The same presentation order as in B was used. However, the ad hoc categories were now defined with an alternative wording that might encourage the participants to consider the nature and/or the function of the category members, for example *things people keep in their pockets* was presented as *things that may conveniently be kept in pockets.*

Full instructions and sample pages of the materials can be seen at Appendix C.

Procedure

The same experimental procedure as in Study 1 was employed.

Results

Output Fluency and Item Dominance

The mean output fluency for each category and the average output fluency by category type in the three conditions of Study 2 are reported in Table 2.7. Highest fluency was observed for some of the taxonomic categories (e.g., *fruit*). For the taxonomic categories, fluency seemed relatively stable across conditions. In contrast, for the three types of ad hoc categories, output fluency seemed to differ across the conditions. A 4 (category type)* 3 (condition) ANOVA showed a main effect of category type $F(3,86) = 62.83$, $p < 0.001$ eta squared = 0.687. There was a main effect of condition $F(1,88) = 3.65$ $p = 0.03$ eta squared = 0.77 and a significant interaction $F(6,172) = 4.9$, $p<0.01$ eta squared = 0.146.

Overall fluency was highest for taxonomic categories ($M = 11.33$) and lowest for abstract unfamiliar ($M = 8.01$). Overall there was no difference in fluency for the two abstract ad hoc categories ($p = 0.366$). Across category types, the highest fluency was seen in condition A and lowest in condition C (Sheffe post hoc comparison $p = 0.031$). The only significant simple effects of condition occurred for the abstract familiar, $F(2, 90) = 8.52$, $p < .0004$, and concrete familiar ad hoc category types, $F(2, 176) = 6.86$, $p < .002$. This gave rise to the interaction seen in Figure 2.6.

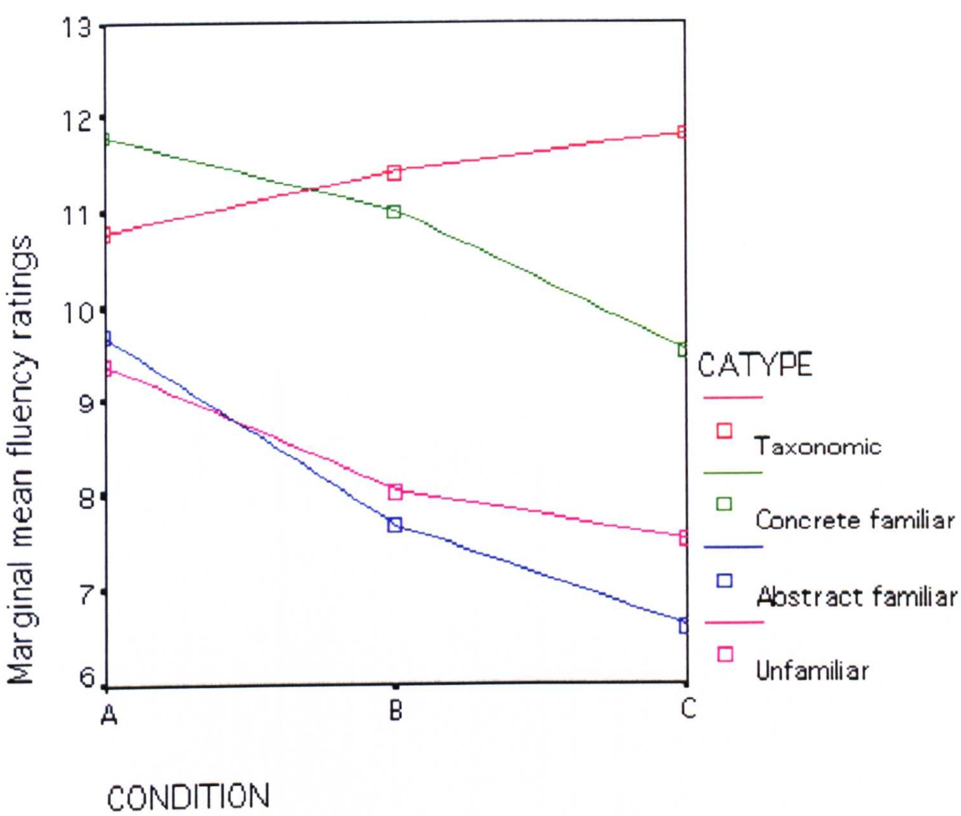


Figure 2.6 Output Fluency for taxonomic, concrete familiar, abstract familiar and unfamiliar category types across all three conditions of presentation.

The experimental manipulations had little effect on the item dominance distributions: within each category, item dominance distributions did not differ significantly (on the basis of Kolmogorov-Smirnov analyses); production frequencies for the categories used in Study 2 are reported in Appendix A of Vallée-Tourangeau, Anthony and Austin (1998*)*. As in Study 1, there were few salient differences between categories. Thus *fruit* showed the highest consensus and *ill* the lowest. Among the taxonomic categories, consensus for *bird* and *vehicle* were the lowest and the item dominance distributions for *bird* differed significantly from the ones for *fruit* in all three conditions (Condition A $p < 0.0001$; Condition B $p = 0.006$; Condition C $p = 0.006$) The three familiar concrete ad hoc categories showed very similar distributions across conditions which, generally, did not differ significantly from the distributions for the taxonomic categories. Finally, the item dominance distributions for *ill* differed significantly from the item dominance distributions from all other categories with the exception of *bird* ($p = 0.112$*)* *black market* ($p = 0.309$*)* and *walls* ($p = 0.12$) in Condition A, and *black market* ($p=0.112$) in Condition C.

Table 2.7

Mean output fluency (standard error in parenthesis) for the 9 categories in each of the three conditions of presentation.

	A	B	C
	Taxonomic categories		
Fruit	13.65 (0.52)	13.97 (0.43)	14.39 (0.64)
Vehicle	8.23 (0.82)	8.40 (0.93)	10.04 (0.57)
Bird	9.19 (1.05)	10.83 (0.97)	12.52 (1.00)
Furniture	11.74 (0.54)	12.57 (0.57)	10.75 (0.40)
Overall	10.70 (0.43)	11.44 (0.42)	11.9 (0.37)

Table 2.7 contd.

Concrete Familiar Ad hoc categories			
Pockets	11.97 (0.59)	11.43 (0.71)	9.03 (0.64)
Dogs Chase	11.43 (0.58)	10.50 (0.61)	9.77 (0.48)
Walls	11.94 (0.53)	11.07 (0.57)	9.40 (0.31)
Overall	11.79 (0.32)	11.00 (0.36)	9.40 (0.31)
Abstract Familiar Ad hoc categories			
Ill	9.71 (0.66)	7.67 (0.52)	6.60 (0.41)
Unfamiliar Ad hoc categories			
Black Market	9.39 (0.68)	8.03 (0.78)	7.50 (0.45)

Difficulty Ratings

As in Study 1, participants recorded the degree of difficulty attached to the generation of members for each category by means of a scaled rating. The mean ratings for all categories in all conditions can be seen in Table 2.8. High ratings indicate a greater degree of difficulty.

Table 2.8

Mean Difficulty Ratings (standard deviation in parentheses) for the 9 categories in Study 2

	Taxonomic categories				
	Fruit	Vehicle	Bird	Furniture	Overall
Condition					
A	2.03 (1.96)	3.67 (2.41)	2.9 (1.86)	3.53 (2.01)	
B	1.71 (1.18)	2.69 (1.83)	2.9 (2.11)	3.0 (2.16)	
C	2.45 (1.67)	3.86 (2.34)	3.13 (2.32)	3.49 (1.73)	
Overall mean	2.08 (1.65)	3.41 (2.25)	2.98 (2.08)	3.34 (1.97)	2.94 (1.46)
	Concrete Familiar Ad hoc categories				
	Pockets	Dogs Chase	Walls		
A	3.57 (1.99)	4.10 (2.02)	4.37 (1.81)		
B	3.03 (1.94)	3.90 (2.00)	3.77 (2.16)		
C	3.90 (2.41)	3.32 (2.39)	3.52 (2.29)		
Overall mean	3.50 (2.13)	3.77 (2.16)	3.88 (2.11)		3.72 (1.71)
	Abstract familiar Ad hoc categories				
	Ill				
A	4.73 (1.98)				
B	5.83 (2.08)				
C	4.13 (2.60)				
Overall mean	4.89 (2.33)				4.89 (2.33)
	Unfamiliar Ad hoc categories				
	Black Market				
A	5.97 (2.59)				
B	6.43 (2.79)				
C	5.74 (2.84)				
Overall	6.04 (2.73)				6.04 (2.73)

A 3*4 mixed ANOVA showed a significant main effect of category type Wilks Lambda $F(3,86) = 54.48$, $p < 0.001$ eta squared = 0.65. The unfamiliar Ad hoc category received the highest difficulty ratings ($M = 6.04$) which was significantly higher than the ratings for the abstract familiar ad hoc category ($M = 4.89$) $F(1,88) = 11.46$, $p = 0.001$ eta squared = 0.115. This in turn was significantly higher than for the concrete familiar categories ($M = 3.72$) $F(1,88) = 31.43$, $p < 0.001$ eta squared = 0.263 and the taxonomic categories received the lowest difficulty ratings ($M = 2.94$) which were significantly lower than the concrete familiar categories $F(1,88) = 27.5$, $p < 0.001$ eta squared = 0.238. There was no main effect of condition ($p=0.506$) but a significant interaction of category-type and condition was found Wilks Lambda $F(6,172) = 3.4$, $p = 0.003$, eta squared = 0.103. This seemed to be due to the ratings for the abstract familiar and the unfamiliar categories rising more steeply in condition B than they did in the other 2 conditions as can be seen in Figure 2.7.

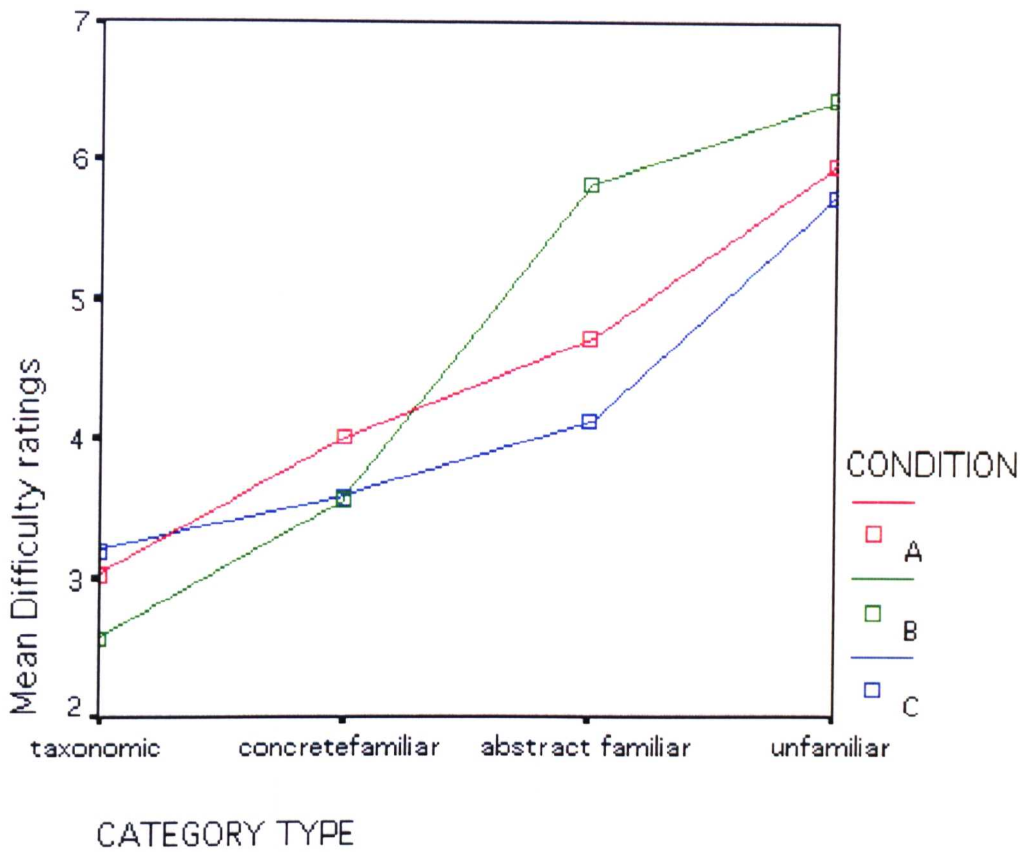


Figure 2.7 Difficulty ratings for all four category types in all conditions of presentation.

The difficulty ratings for the individual categories collapsed across condition can be seen in Figure 2.8 displayed in ascending order.

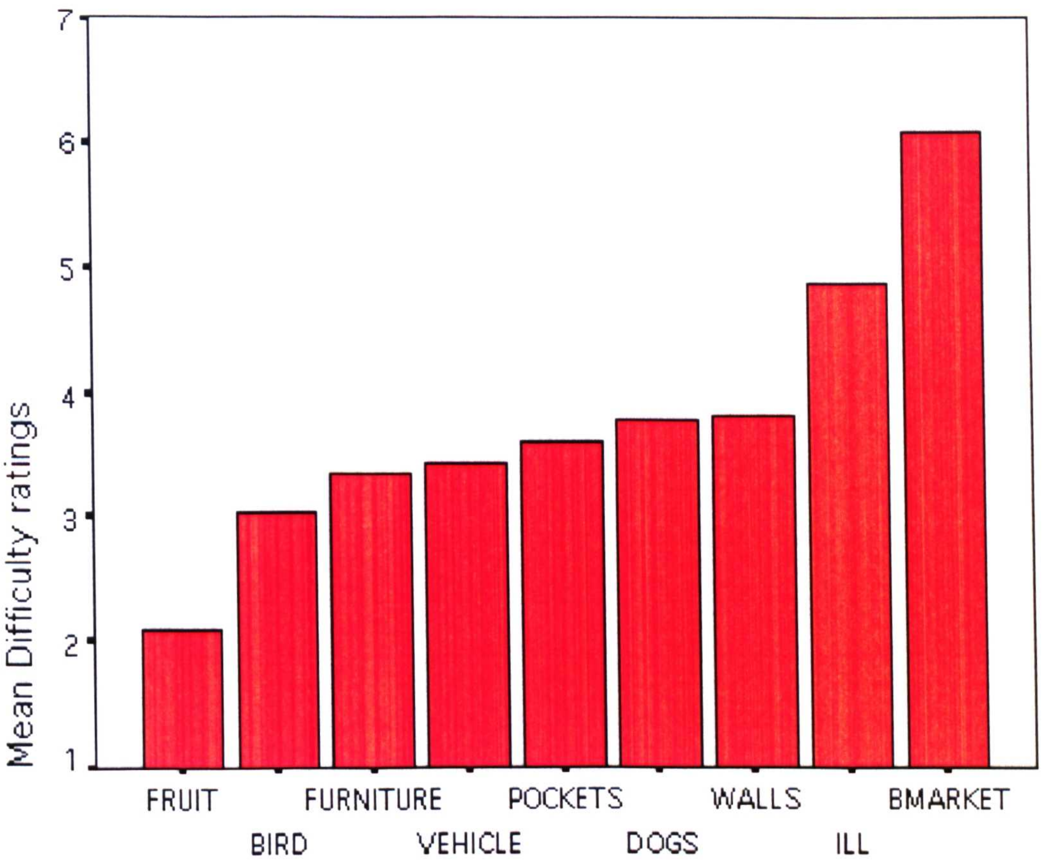


Figure 2.8 Mean difficulty ratings for all 9 categories in study 2 shown in ascending order.

A 3*9 mixed ANOVA showed a main effect of category Wilks Lambda F (8,76) = 25.02 p<0.001 eta squared = 0.725. Participants found *black market* the most difficult in terms of generating category members and *fruit* the least difficult. However, the only significant contrasts between categories were *fruit* vs. *bird* F(1,83) = 28.71 p< 0.001 eta squared = 0.257, *walls* vs. *ill* F(1,83) = 18.27, p< 0.001 eta squared = 0.0.18 and *ill* vs. *black market* F(1,83) = 12.03 p =0.001 eta squared = 0.127.

Three was no effect of condition on the ratings (p = 0.859) but there was a small interaction between category and condition F(16,152) = 1.99, p = 0.016 eta squared = 0.174. This appeared to be due to the rating for *ill* increasing in condition B to a greater extent than in condition A or C. No particular reason can be found for this finding.

Difficulty ratings correlated negatively and significantly with fluency for all categories except Fruit and Vehicle (p = 0.463 and p=0.127 respectively) . Figure 2.9 shows the regression lines for each category type

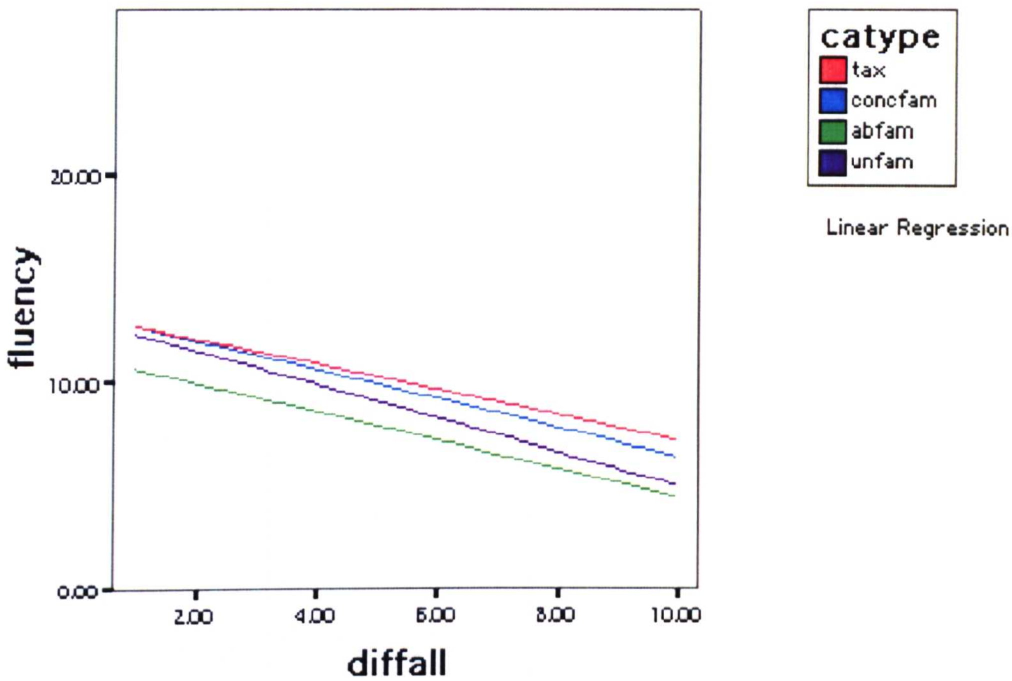


Figure 2.9 Regression slopes for difficulty ratings as a predictor of fluency for taxonomic, concrete familiar, abstract familiar and unfamiliar category types.

Table2.9

Values of adjusted r2 for individual categories and category types. Percentage of variance in fluency accounted for by difficulty shown in parentheses.

		Taxonomic categories		
Fruit	Vehicle	Bird	Furniture	Overall mean
0.005 n.s.	0.017 n.s.	0.123 (12.3%)	0.106 (10.6%)	0.096 (9.6%)

	Concrete Familiar Ad hoc categories			
Pockets	Dogs Chase	Walls		
0.088 (8.8%)	0.302 (30.2%)	0.266 (26.6%)		0.207 (20.7%)

	Abstract familiar Ad hoc categories			
Ill				
0.234 (23.4%)				0.234 (23.4%)

	Unfamiliar Ad hoc categories			
Black Market				
0.357 (35.7%)				0.357 (35.7%)

As in the first study, the relationship between difficulty ratings and fluency is strongest for the unfamiliar ad hoc category where participants tended to produce few items and give high difficulty ratings.

An ANCOVA, using difficulty as a covariate, showed a significant main effect of difficulty $F(1,787) = 142.6$, $p < 0.001$, eta squared = 0.153 and a very small main effect of category type $F(3, 794) = 11.286$, $p < 0.001$ eta squared = 0.041. There was no significant interaction between difficulty and category type ($p = 0.648$).

A broadly similar pattern of difficulty ratings is seen here as in study 1. The difficulty ratings are clearly related to fluency as one would expect. The difficulty also varies with category type with participants finding taxonomic easiest in this task and unfamiliar categories the hardest.

Retrieval Strategies

Across all three conditions, 83% of the protocol segments were classified within one of the three retrieval strategy categories (i.e. experiential, semantic, or unmediated), 9% resisted the classification scheme and 8% were uninterpretable. The average number of classified protocol segments, across categories, varied little within each condition: means of 41.2, 41.0, and 40.7 were observed for Conditions A, B, and C respectively.

The percentage of protocol segments coded as reflecting an experiential strategy, a semantic strategy, or no strategy (unmediated retrieval) is shown in Figure 2.10 for each category in all three conditions. To facilitate the comparison with the percentages observed for the categories that were also used in Study 1 (i.*e., fruit, vehicle, dogs chase, pockets, and black market*) they are plotted alongside the percentages observed in Study 2. The experimental manipulations appeared to have had no systematic effect on the relative distribution of the protocol segments: As in Study 1, for all but the unfamiliar ad hoc category *black market*, experiential strategies outnumbered semantic strategies in all conditions, including the two new ad hoc categories, *ill* and *walls*, and, importantly, the two new taxonomic categories, *bird* and *furniture*. The sole exception to this pattern was for *vehicle* in Condition A. As in Study 1, retrieving instances for *black market* was done mostly on the basis of semantic strategies.

The influence of the prior presentation of the ad hoc categories (using the wording of Study 1) on the percentage of experiential strategies in taxonomic categories can be assessed by examining the data for the four taxonomic categories between Condition A, in which they were presented first and Condition B in which they followed the ad hoc categories. *Vehicle* registered an important difference in that experiential strategies were used much more after the presentation of the ad hoc categories (Condition B and also in Condition C) than when the taxonomic categories preceded the ad hoc ones (Condition A). For the remaining three taxonomic categories, *fruit, bird,* and *furniture,* the presentation order seemed to have little influence on the extent to which experiential strategies were employed.

A comparison of Conditions B and C for the ad hoc categories assesses the effect of rewording the ad hoc concepts in a way that emphasised object-centred properties. As can be seen in both Figure 2.10 and Table 2.10, the preponderance of experiential strategies did not vary across Conditions B and C for walls and *ill*; it increased slightly for *pockets*, but decreased slightly for *dogs chase*.

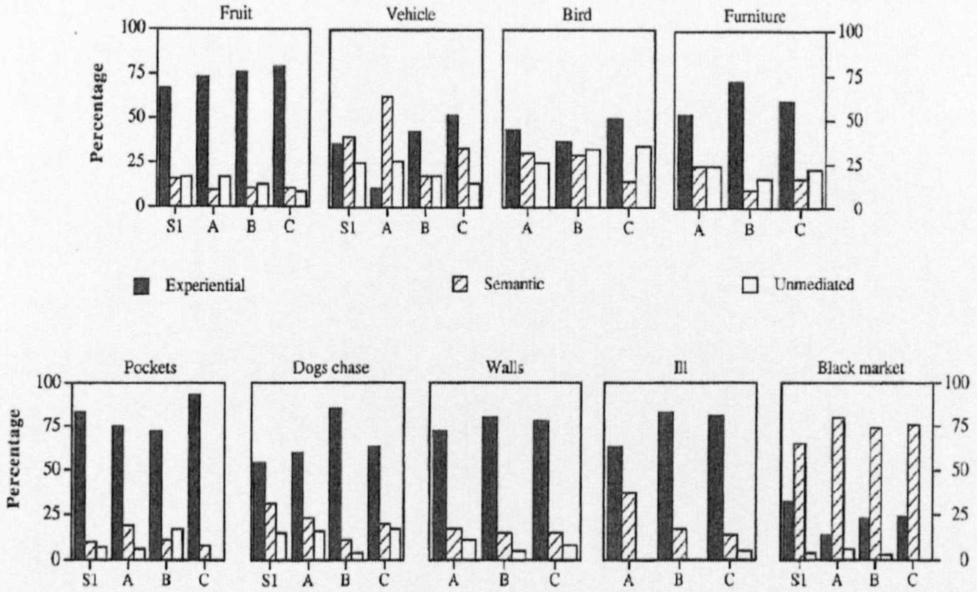


Figure 2.10 Percentage of experiential, semantic and unmediated protocol segments in conditions A, B and C in Study 2. For the categories, *fruit, vehicle, pockets, dogs chase* and *black market*, the percentages of each segment type from study 1 are also shown.

Table 2.10

Percentage of each protocol type by category and condition for Study 2 and from Study 1 where available. These are the frequencies depicted graphically in figure 2.10

	A	B	C	Study 1
BIRD				
Experiential	44	37	50	
Semantic	31	30	15	
Unmediated	25	33	35	
FURNITURE				
Experiential	53	72	61	
Semantic	24	11	17	
Unmediated	24	17	22	
FRUIT				
Experiential	73	76	79	67
Semantic	10	11	11	16
Unmediated	17	13	9	17
VEHICLE				
Experiential	11	43	52	36
Semantic	63	39	34	40
Unmediated	26	18	14	21
DOGS CHASE				
Experiential	60	85	63	54
Semantic	23	11	20	31
Unmediated	16	4	17	15

Table 2.10 contd.

POCKETS					
	Experiential	75	72	93	83
	Semantic	19	11	8	10
	Unmediated	6	17	0	7
WALLS					
	Experiential	72	80	78	
	Semantic	17	15	15	
	Unmediated	11	5	8	
ILL					
	Experiential	63	83	81	
	Semantic	37	17	14	
	Unmediated	0	0	5	
BLACK MARKET					
	Experiential	14	23	24	32
	Semantic	80	74	76	65
	Unmediated	6	3	0	9

Since these classifications are post hoc and violate assumptions of independence, analysis by Chi-Square would not be appropriate. It is possible that the data are unduly influenced by individual participants who gave many more examples of certain types of strategy than other participants. An analysis that included participants as a factor would be appropriate but this would require data collected for each participant indicating the number of each type of strategy deployed for each category. The data were not, unfortunately, originally collected in this form and are therefore not available. In the absence of appropriate inferential statistics, the results must be viewed cautiously. It should be noted, however, that there is only one reversal of the ratio of experiential strategies to semantic or unmediated strategies shown in the categories that were also presented in study 1. For the category vehicle, the semantic strategies seem to outweigh the experiential strategies in condition A but this reverses in conditions B and C. It should be recalled that the purpose of study 2 was to see whether the high proportion of experiential strategies for all but the unfamiliar categories were an artifact of the high ratio of ad hoc categories to taxonomic categories in study 1. Even in the absence of inferential analysis, these strategies do not seem to have been eradicated for the taxonomic categories in this study, most importantly they still feature in the reports for the taxonomic categories in Condition A where these categories were seen in a block before the ad hoc categories and the participants' reported strategies could therefore not have been influenced having performed the task for ad hoc categories first. It also seems that the experiential strategies for the ad hoc categories did not dramatically disappear when the wording was altered to defuse any implication that the participants should refer to their own personal experience (Condition C).

Clusters in Output

Since the output for each category was written and not recorded verbally, it was not possible to identify the items that may have been produced in short bursts together or where any intervals between groups of items occurred. It was therefore not possible to assess the direct correspondence between clusters of items and reported strategies. Having said that, when the reported strategies were based on norms or knowledge shared by both participants and experimenters (e.g., arrangement of bedroom furniture) analysis of the protocols revealed some striking correspondences. Table 2.11 shows items generated by three participants (from Study 2), shown in the left portion of each column, along with their retrospective protocols, shown in the right portion of each column (the back slashes indicate how these were segmented).

Table 2.11

Protocols and output from 3 participants in Study 2.

65c bird		82a vehicle		62c furniture	
eagle	*I thought of*	car	*The modes of*	table	*My own*
owl	*big birds /*	lorry	*transport*	chair	*house - going*
------	*then of birds I*	bus	*seen on the*	dresser	*room to room*
bluetit	*would see in*	train	*road. /*	piano	
robin	*the garden /*	coach	*Followed by*	sofa	
chaffinch	*then in a zoo.*	------	*two wheeled*	easy chair	
sparrow	*/ When I*	motorbike	*variety / and*	coffee table	
blackbird	*thought of*	bicycle	*then the*	wardrobe	
thrush	*pigeon the*	------	*building*	bed	
swift	*next two came*	tractor	*trade. My*	dressing table	
house martin	*automatically*	Land Rover	*husbands*	hi-fi unit	
------	*to mind.*	JCB	*works in this*	stool	
ostrich		crane	*field.*	ottoman	
------		van		desk	
pigeon		------			
dove		furniture van			
seagull		------			
		scooter			

The protocol from 65c accounts for all the items generated, although the series from *bluetit* to *housemartin* could have included smaller clusters produced by local strategies or processes. The protocol from 82a matches the clusters only to some extent: presumably *furniture van* was triggered by *van,* and *scooter* is a member of a previous category. Finally, the protocol from 62c, while perfectly interpretable, is not specific enough to suggest clusters, and without knowing the arrangement of this participant's house, it is hard to be confident that he was being quite as systematic as he claimed.

General Discussion – Studies 1 and 2

The alterations to the wording of the ad hoc category titles and the increase in the proportion of taxonomic categories amongst the stimuli in Study 2 did not result in any notable change of reported strategy.

To my knowledge, this was the first study to investigate the way in which people produce lists of items belonging to a wide range of both taxonomic and ad hoc categories. This task clearly requires the tapping of one's conceptual knowledge. Post hoc, the reliance on the situations and contexts in which the participants customarily encountered the exemplars of the categories as a retrieval strategy may seem perfectly reasonable and unsurprising. However, given the complete absence of this type of information in most theories of the way in which conceptual knowledge is structured and represented, it can be argued that these theories would not have predicted these findings. Indeed, traditional theories would have predicted greater use of semantic strategies, in particular for taxonomic categories.

The issue of the validity of protocols in offering a genuine window on cognitive processes was discussed in Chapter 1. It was pointed out that the short time delay between generation and protocol on each trial in these studies should at least alleviate concerns about any difficulty that participants may have experienced in accessing their strategy over an extended period of time. Nisbett and Wilson (1977) suggested that "when reporting on the effects of stimuli, people may not interrogate a memory of the cognitive processes that operated on stimuli; instead, they may base their reports on implicit á priori theories about the causal connection between stimulus and response" (p.233). However, the specificity of some of the protocols collected here, as discussed shortly, speaks against the application of an á priori theory.

Despite the measures that were taken to increase the informativeness of the protocols, there remains the possibility that the protocols are the result of a constructive process at the time of writing, rather than a reconstructive memory process of what happened when the category members were generated. There are however, several points that argue against this view of the data. Firstly, the specificity of some of the protocol segments. Secondly, the presence of identifiable clusters in the outputs which appear to correspond to the reported strategies. Thirdly, the documented tendency of subjects to group items into semantic, not experiential, categories when engaged in a sorting task (Walker and Kintsch, 1985) and finally, the similarities found by Walker and Kintsch between strategies reported in concurrent and retrospective protocols.

Protocol Specificity

If one assumes that the participants truly had no access to the means by which they had generated items for each category, then they would be obliged to produce an explanation that would seem "acceptable" to both themselves and the experimenter. Consider the following protocols for *fruit* and *bird* respectively:

" If you go to Tesco's in Royston you meet the fruit and vegetable section first of all. I imagined I was shopping and can remember the layout and most of the fruit"

" I do a lot of sailing and many boats in the type of sailing boats I sail in are named after birds; so I went through my memory of the opposition's boat names"

It seems unlikely that such specific and personal accounts would present themselves to participants as suitable self-explanations to lay claim to if they were not indeed what the participant had actually done.

Clusters in Output

Unlike the early studies of output performance discussed in the introduction to Chapter 2, these studies were not designed to allow a systematic cluster analysis. Even if clusters had been identified temporally, it would have been clearly much harder to establish that their intra- cluster similarity is based on an idiosyncratic strategy such as ' what I see at the supermarket' than it was in earlier studies to ascertain that the clusters were based on shared category membership or word-association. That said, some of the protocols did imply a series of different strategies that seemed to be reflected in the output (see Table 2.11).

Sorting versus Generating

It may be thought that participants were not able to explain their own output from memory and were looking at their own items, seeking to account for meaningful groupings in their own output. However, if this had been the case, it could be argued from Walker and Kintsch (1985) that they would be most likely choose to report semantic strategies and not experiential. Walker and Kintsch (1985) presented participants with slips of paper showing the names of items that they had themselves generated in an earlier free emission task. They were asked to sort the slips into "natural categories" as they saw fit. The groupings formed in this task were compared to the clusters identified in the same participants' verbal generation output; the groupings in the sorting task overlapped only 19% of the clusters from the generation task. Furthermore, the basis for the groupings was different for each task. While the majority of the clusters (86%) in the sorting task were described as semantic (e.g., small cars, foreign cars, luxury cars for *car*), most of the clusters produced in the generation task (77%) were related to 'episodic cues' such as friends' cars, my cars, cars I have wrecked etc. Indeed, the similarities between what Walker and Kintsch refer to as episodic cues and what I refer to as experiential strategies are striking. This suggests that the experiential nature of the protocols collected here were not imposed post hoc by participants but do genuinely reflect the strategies deployed.

Concurrent versus Retrospective Protocols

Although the protocols were collected retrospectively, it is likely that concurrent protocols would have shown similar reports. Walker and Kintsch (1985) collected both concurrent and retrospective protocols in a between subjects design and found that the relative distribution of semantic and 'episodic' retrieval cues was similar for both conditions. This suggests that the occurrence of the situational strategies found here was not an artifact of the retrospective (and possibly reconstructive) nature of the protocols. Interestingly, Walker and Kintsch did note that participants producing concurrent protocols produced far more non-category retrieval cues than did the retrospective protocol participants. They suggest that this was due to the retrospective participants having less access to their strategies as a result of the time lag in reporting them. The retrospective protocols then are likely to be incomplete, but not necessarily inaccurate.

Experiential Contexts as Effective Access to Abstract Knowledge Structures

The studies reported here confirm the hints from the studies of Gruenewald and Lockhead (1980) and Walker and Kintsch (1985) that subjects make use of personal experiences to set contexts in which the names of semantic category members may be retrieved. The careful protocol analysis across a much wider range of categories than studied by Walker and Kintsch (1985) has established that, in completing free emission tasks, people tend to instantiate personal experiences to recreate the contexts (situations) in which they may come across the category in question. The names of items are retrieved through this context. The studies here are unique in revealing the extent of this strategy. If one accepts that retrieval from semantic categories informs us of the mental organisation of this knowledge, then the import of this work lies in the challenge that it could cast to the notion of conceptual knowledge as de-contextualised, as discussed in Chapter 1. To launch such a challenge on the strength of these findings alone would seem premature. Even if one accepts the arguments, outlined above and accepts the protocols as an accurate account of how the task was undertaken, they need not necessarily be indicative of underlying conceptual structure - they need be no more than effective retrieval mechanisms. If this were to be the case, one might wonder as to why widespread, successful retrieval cues should be unconnected to the organisation of the information being retrieved. Using an organisation to retrieve information that differs from the organisation of the to-be-recalled- information itself is unlikely to be successful. As many undergraduates are encouraged to demonstrate for themselves, reciting the months of the year in alphabetical order is more difficult than reciting them in the order in which they occur throughout the year. Successful retrieval cues are likely to be those that have been encoded with the target information (Tulving and Thompson, 1973). However, the simultaneous encoding of common situational information alongside a category member (e.g., apples are bought in supermarkets) need not necessitate that supermarket form part of the conceptual structure of apple. Hewitt (1973) first made an interesting distinction between extrinsic and intrinsic context. Under this distinction the intrinsic context was seen as a subset of features encoded about an item that form part of the item itself and the extrinsic context included features that may be present when the item is encountered but do not form an integral part of the item. This would allow supermarket to be an extrinsic contextual feature of apple that would act as an efficient cue to retrieve apple in the same way that recreating an odour at recall has been shown to enhance recall of material encoded in the presence of the same odour (Herz, 1997; Pointer & Bond, 1998). In other words, the participants in this study are engaging in self-generated, state-dependent recall.

If this is the case, then the situations used as strategies by the participants are extrinsic contexts and not intrinsic features of the categories. This is indeed a plausible explanation for the strategies found here but it begs the question as to why the situational features (or extrinsic) should be distinguished from the intrinsic features and considered to be extrinsic in the first place. As previously discussed in Chapter 1, there may be good reasons for not wishing to include features that do not individuate concepts in the concept itself (*supermarket* will equally form a part of *newspaper* and *shampoo*) but this position has to be argued for and one's view of this debate will naturally lead to a greater or lesser inclination to accept the finding of this study as suggesting anything at all about conceptual organisation. Of course, one need not make a strong claim about the status of this information as a conceptual 'feature' of an object. These results could equally point to an alternative, additional, organisation of knowledge about objects that revolves around events and situations and that provides an easier access to the names of category members than the organisation based on intrinsic features. The remaining chapters of this thesis will build upon an argument for taking both of these possibilities seriously when studying how people draw on their conceptual knowledge to complete simple tasks and the author finds herself in good company in accepting that successful retrieval paths for category members are successful because of their overlap with the nature of the category representation (e.g., Barsalou and Sewell, 1985).

The remainder of this discussion will continue this line of argument by presenting different sources of empirical data that, taken together, can be seen as offering support for this position.

Barsalou and his colleagues have offered empirical evidence supporting the idea that aspects of the situations in which objects are encountered become integrated in the knowledge of objects and categories. In a concept learning experiment, Yeh and Barsalou (1996) presented their participants with instances from the same category but in different situations. Each situation was correlated with a different set of relevant properties. If the category representation that develops as participants experience different instances across situations is principally decontextualised, reflecting an abstraction of properties across contexts, then the recognition of situation-relevant properties should not be faster when it takes place in that situation. Yet, situations primed the recognition of situation-relevant properties. Such results suggest that aspects of the situations in which category instances are encountered form an integral part of their representation.

The fact that participants are able to produce apparently context-free knowledge of categories has possibly misled us into believing that this reflects the nature of their mental representations of these categories. In a second experiment, Walker and Kintsch (1985) examined retrieval from script-like memory structures. That is to say participants were asked to say what typically happens in certain taxonomic situations such as *going to a restaurant for a meal, going to a grocery store to buy groceries* and *going to a doctor's office for a check-up*. In Shank and Abelson's (1977) notion of scripts, this knowledge of various culturally specific situations is developed across a number of experiences and, not unlike the notion of natural categories, is abstracted until it serves well to represent any similar situation but does not equate to any one specific event or episode. However, Walker and Kintsch's analysis of retrospective verbal protocols, in which participants explained how they had met the requirements of the task, revealed that a) participants reported using more than one strategy and b) "Only 10% of the retrieval strategies reported were normative, abstract, context-free retrieval cues i.e., the classical notion of scheme or script - whereas the great majority of strategies that were reported were clearly of an episodic nature" (pp.277). Participants reported either recalling a particular visit (especially to a restaurant or doctor) or a particular place in the case of the grocery store script. Walker and Kintsch, however, noted a discrepancy between the strategies reported and the actual descriptions given. The descriptions did not describe particular locales or give evidence of the personal nature of the alleged strategy that gave rise to them. Walker and Kintsch suggest that this does not undermine the validity of the protocols but that participants are drawing on one or more specific memories which are then edited to produce a suitable generalisation to meet the demands of the instructions. One could extrapolate from this to suggest that subjects engage in a similar process when performing 'semantic' tasks. It is possible that conceptual representations are entwined with memories for events and situations and the abstraction takes place at retrieval rather than storage level, if this is necessary to meet the demands of a particular task. Several recent lines of theoretical and empirical work speak to this possibility.

As mentioned in Chapter 1, Barsalou, Yeh, Luka, Osleth, Mix, and Wu (1993) sketch the outlines of a broad ranging theory of conceptualisation and meaning in which "the fundamental conceptual representations in the human cognitive system are schematic perceptual images extracted from all modes of experience" (p. 26) (see also Barsalou, 1997). More relevant to understanding why experiential context frequently mediates retrieval in our experiments is their equally radical proposal that "concepts are neither context independent nor universal but are situated and local" (p. 38). Mental concepts, of both particular individuals and generic types of individual, are always established within and include the perspective of the experiential scenarios in which they are encountered or imagined. Rather than having an essential meaning, or even a univocal core, they are "collections of all specialised models for a particular type of individual together with their associated generic situations" (p. 47). Context and context effects are not an optional extra to context free theories of concepts and categorisation but are intrinsic to the mental representations constituting concepts. Since the publication of the data reported here, Wu and Barsalou (in press) have demonstrated that participants performing property generation tasks unexpectedly generate situational properties, albeit to a lesser degree than entity properties.

Lancaster and Barsalou (1997) investigated the organisation of memory for events. The background to their studies was the view that event memories are organised by knowledge of activity (e.g., memories of eating at a restaurant are stored together with abstract knowledge about eating at restaurants etc.), that is to say, the cue used to access event knowledge is activity. Barsalou (1988) proposed an alternative conception which assumed that event memories are "cross classified simultaneously in a variety of global organisations, including organisations for locations, participant, time and **objects** - as well as for activity" (emphasis added). This type of organisation would permit retrieval to an event memory through a number of different cues aside from activity. Thus an event memory could be accessed through an object cue (e.g., beach ball for summer vacation memory). The corollary of this is that an object may be accessed through an event memory (or a location, or activity etc.). In Lancaster and Barsalou's studies, the participants had to learn and then recall a list of sentences that described fictional events incorporating a well-known person and an activity. The clustering of the participants' recall outputs suggested that participants formed clusters around locations, times, agents and activities, with the latter two being the dominating tendency. Lancaster and Barsalou concluded that people's memories for events are cross-classified in a variety of ways giving rise to multiple organisations that can be accessed flexibly during retrieval. There is no reason to believe that object representation is not equally dynamic and flexible.

Further support for the interleaved nature of 'different' memory systems has come from the autobiographical memory literature. Conway's research has also pointed to a close interdependence between autobiographical memory and semantics. Conway (1987) asked participants to verify true or false semantic facts and autobiographical facts. Semantic verifications were of the form "is an apple a fruit?", while autobiographical verifications took the form "are apples your favourite fruit?". These questions were preceded either by a neutral word or a by a category-name prime, for example, "fruit". Semantic verifications were, as expected, faster when preceded by the presentations of the relevant category-name primes. Importantly, autobiographical verifications were also primed by the presentation of a relevant category name, e.g., "fruit" speeded responses to "are apples your favourite fruit?". Conway (1990b) suggested that "frequent and current autobiographical knowledge might be stored in memory with semantic knowledge and so help 'customize' the semantic system." (p. 176).

Conway (1990c) provides evidence that memories of specific and generic personal experiences are closely connected to representation of taxonomic and ad hoc categories. The general procedure employed in that study involved giving participants names of categories and asking them to "bring to mind an image of whatever [you take] the word or phrase to refer" (p. 134). Conway supplied participants with three predetermined categories in which to classify their images: (1) specifically dateable autobiographical memory images; (2) loosely dateable generic images derived from specific experiences; and (3) images not based on specific or generic experiences, which Conway labelled "semantic" images. The majority of images elicited by ad hoc categories (e.g., *things to take on holiday*) were dateable autobiographical memories (53%; Experiment 1) and 23.5% were generic images. In the classification of the protocol segments in our studies, the specific and generic image categories were fused into the experiential mediation category. On average, 72.6% of the strategies for the (familiar) ad hoc categories were experiential, a figure that closely corresponds to the sum of the two percentages above (76.5%). This convergence is important since, unlike in Conway (1990c), we did not provide our participants with predetermined categories that would help them identify their retrieval strategies. Furthermore, the sum of the percentage of the specific (33.8%) and generic (36.8%) images generated by taxonomic category labels summed to approximately the same value, namely 70.6%

(note the correspondence with the figure reported in Walker & Kintsch [1985] of 77% episodic retrieval strategies). Thus, Conway's studies provide further evidence of the close connections between memory of personal life events and concepts.

Similar inferences can be drawn from work with neuropsychological patients with semantic dementia reported in Snowden, Griffiths, and Neary (1994; see also 1995). These researchers have shown that even if these patients' performance on general tests of word comprehension is extremely poor, their understanding of objects, places, and people that are encountered on a regular basis is relatively intact. Snowden et al. argued that it is "the experiential input into the meaning system" (p. 287) that sustains their meaning. The meaning of objects, places and people that were part of the patients' lives some time in the past is lost. This difference "highlights the role of autobiographical experience in the maintenance of meaning, and underlies the interrelationship between episodic and semantic memory" (p. 265). Once again, the emphasis is on memory systems that are, at the very least, closely related in a way that has tended to be underestimated in the investigation of any one particular 'type' of memory.

A Developmental Perspective

Further support for the role of events/situations in supporting conceptual organisation can be found in the developmental literature. Nelson and Gruendel (1981), in investigating young children's knowledge structures posited a central role for events as being the initial object of mental representation. In early studies, Nelson (1978) examined the responses of pre-school children to questions like " what happens when you go to McDonalds?" or "what happens when you have dinner?" in addition to questions about events such as having a birthday party or getting dressed. Children as young as 3 years old were able to reproduce component actions of events in the order in which they genuinely occur. One of the other striking characteristics of the scripts they tended to produce as responses was the generality of the scripts. The children tended not to personalise the report by use of the 1st person but rather to use the pronoun "you" in its impersonal sense of "one". They additionally utilised the tenseless forms of verbs ("you eat etc.."). When asked about specific personal events, children were able to recall less material. Both 5 yr. and 3 yr. old children gave good generalised accounts but they had more difficulty in producing accounts of specific events, showing a tendency to slide into the more generalised form of the script. The authors suggested that these scripts develop from the children's "experiences in events and from observation" (p. 145). Nelson and Gruendel note that the existing division of episodic (specific episodes) and semantic memory (general knowledge) does not sit well with what they come to refer to as Generalised Event Representations (GERs); by this term they mean "memory for events that is not specific to a particular experience" (p. 147). Nelson and Gruendel propose that GERs are derived from episodic memory although the process by which this takes place is unspecified.

Nelson and Gruendel make the strong claim that GERs are the building blocks of cognition. From these structures, which necessarily hold "slots" for objects as props in the event scripts (e.g., food at a birthday party), object categories are abstracted in due course. This process would involve a pattern-matching search that compares elements within and across GERs. Under this view, the abstraction and construction of object categories is a process that operates not on the world but on acquired representations of repeated experienced events. Given this framework, it would not be surprising that even as adults, one of the easiest ways to generate category members is to reconstruct the generic episodes/events in which these items have been regularly encountered. The situations instantiated by the participants in the present studies did indeed tend be of a generalised nature (" being in Tesco's" - "bird I would see in the garden") rather than of specific single episodes.

Also adopting a developmental perspective, Lucariello and Nelson (1985) tested the theory that event schemas provided the basis of categorical organisation for young children. Using the same recall and clustering paradigm discussed in the introduction to Chapter 2, it was shown that memory (list recall) and organisation (clustering in recall) in 3- 4 yr. olds was better for slot-filler categories (e.g. foods eaten at lunch) than for taxonomic categories (food) or complementary categories (items associated spatially and or temporally such as milk and cup).

Summary

The use of a simple procedure combined with a detailed protocol analysis in the present studies has demonstrated the following:

- When participants are asked to generate members of categories, they report conscious use of strategies as opposed to 'downloading' the contents of their semantic memory in an effortless and automatic manner.

- The strategies documented by participants were, in the main experiential, based on participants mentally consulting a range of situations and contexts in which they come across the members of the category. This was found to be the case for <u>both taxonomic and ad hoc categories</u>. This finding is interpreted as being inconsistent with theories that posit conceptual knowledge as a decontextualised repository of abstracted features.

- Participants are, on average, able to produce more members of taxonomic categories than from ad hoc categories. This would be predicted by the view that ad hoc categories are less well established in memory and have weaker concept-to-instance links. However, the large degree of overlap in the output fluency between the two category types speaks against a clear distinction on this measure.

The results of these studies would not, in themselves necessarily give cause to question traditional models of conceptual knowledge. However, it has been argued that in conjunction with both theoretical and empirical evidence from the autobiographical memory, developmental and abnormal literature, they add weight to the claim that the map of conceptual organisation may require to be redrawn. The notion of separate knowledge structures (e.g., semantic, episodic) needs to be greatly reduced and replaced by a picture of multiple overlapping representations of events, objects and activities which are largely generic in nature but are so arranged as to enable abstraction if required by the perceived goal of the task. Knowledge need not be abstract in storage and contextualised in use but could be richly contextual in storage and only detached from its usual contexts at retrieval if the current goal so demands it. However, further evidence is required of situational knowledge playing a role in the performance of other tasks that one might have expected to be free from such influences. Chapter 3 presents a study that was designed to see if any such evidence could be found.

CHAPTER THREE

The Role of Situational Knowledge in Judgements of Similarity

Overview

This chapter reports a study in which participants were asked to provide ratings of similarity for pairs of items. The items were either presented in the context of their category label or without any context. The category type (taxonomic or ad hoc) was manipulated, as was the relative typicality of the item pairs. The ratings were examined for effects of these factors. Associated protocols offering explanations of the ratings were collected from a subset of participants. The protocols were analysed with a view to offering an explanation of the ratings and to investigating the extent to which the same type of situational knowledge illustrated in the category-member free-recall task may influence the way in which people judge similarity between items under experimental conditions.

Background

The results of the studies reported in Chapter 2 strongly suggest that when faced with a particular cognitive task, namely the free generation of category members, people draw on a range of personal, generic experiences that serve to situate their conceptual knowledge. These types of memories are not episodic, in that they rarely refer to a specific dateable experience; neither are they semantic in the sense in which that term is usually applied to mean one's repository of facts about entities in the world. It appears that in order to meet the goal of the task, the situations in which the participants customarily encounter the categories of objects acted as the best self-generated cue for producing the names of the items themselves. The data sit well within recent perspectives on conceptual knowledge (see Chapter 1), suggesting a greater role for situational knowledge of concepts.

Conceptual knowledge is believed to underlie a number of cognitive operations such as the assessment of similarity between two items, inductive or inferential reasoning

(Osherson. Smith, Wilkie, Lopez & Shafir, 1990) and reasoning by analogy (Gentner, 1988). All of these processes rely upon the support of a knowledge base that links entities to their properties and to each other. The category member free-recall task used in study 1 seemed to elicit use of situational knowledge; The focus of the next study is the extent to which that knowledge-base may play a role in the performance of other cognitive operations, specifically judgements of similarity

Similarity has traditionally been thought to depend upon a comparison process between two entities involving an assessment of matched and mismatched features (Tversky, 1977). Questions have been raised as to what actually constitutes a feature and it has been shown that the answer to this may be altered by the nature of the comparison itself (Medin, Goldstone and Gentner, 1993). The process of comparison has been further defined by Markman and Gentner (1993) as being initially one of comparison along shared dimensions, which brings the entities into alignment. A further comparison of values along these dimensions then takes place in addition to a comparison of the extent to which the items share dimensions at all. This gives rise to the notion of alignable and non-alignable differences which are weighted differentially in the evaluation of similarity. For example, if comparing a lion and a rabbit, the shape of their tails would be an alignable difference whereas the mane of the lion would be a non-alignable difference since the rabbit does not have one at all. Markman and Gentner (1993) found that the non-alignable features were more detrimental to the perception of similarity between items than the alignable ones. This formulation of the comparison process also gives rise to the somewhat counter-intuitive result that the more similar two items are the more differences that can be listed between them. This is because the more dimensions they share, the greater the number of potentially alignable differences between them.

Throughout these interesting incursions into similarity, few researchers seem to have doubted that when one makes a similarity judgement the features under consideration are a) perceptual and/or b) functional. They are intrinsic properties of the object.

In the study reported here, a subset of participants were presented with pairs of perceptually dissimilar items, such as *clock* and *photo*, with no context offered for the comparison. The main question of interest was whether these participants would report drawing on their everyday experience of the way in which these items are normally encountered and use this as a reason for finding them similar. Where the items were both highly typical of the ad hoc category from which they had been generated, (e.g., Money and Wallet from *things that may be conveniently kept in pockets*), it was expected that the participants may refer to this very context as a basis for similarity even when it was not made explicit. These expectations were heightened by two papers that drew attention to the limitations of feature matching models of similarity in explaining judgements of similarity made under experimental conditions (Bassok and Medin, 1997; Wisniewski and Bassok, 1999).

Bassok and Medin (1997) set out to test whether the alignment model of similarity (Gentner, 1983) could be extended from the comparison of objects to stimuli consisting of semantically constrained combinations of interrelated objects i.e. simple noun-verb-noun statements such as "The carpenter fixed the chair" and "the carpenter sat on the chair". They found that the logic of structural alignment could be used to predict the similarity ratings between sentences but to their surprise found that their participants were making also use of processes other than comparison. "For certain stimuli they systematically integrate the paired stimuli into a common thematic scenario" (p. 312). These authors were the first published to observe and comment upon this and the first to draw a distinction between comparison and what they call thematic integration. As an example of thematic integration, the authors cite the protocol of the participant who said of the sentences "The carpenter fixed the chair" and "The carpenter sat on the chair" that they were "similar because he sat on the chair to see whether he had fixed it well".

Wisniewski and Bassok (1999) started from the premise that objects are alignable if they can be compared on many dimensions and non-alignable if they do not share dimensions along which comparisons can be made. They claimed that alignable entities evince comparisons by their very nature whilst non-alignable items will evoke thematic integration. That this should be the case seems non-contentious to the authors, but their empirical work set out to show that these processes are adopted even if they are not task appropriate or consistent with instructions. Pairs of items were systematically manipulated on the basis of whether they shared only taxonomic relations, i.e., attributes (A+T-, e.g., milk- lemonade), taxonomic plus thematic relations (A+T+, e.g., milk- coffee), only thematic relations (A-T+, e.g., milk-cow) and neither taxonomic nor thematic relations (A-T-, e.g., milk-horse). In experiment 1 it was found that the similarity ratings of the pairs in descending order were A+T+; A+T-; A-T+; A-T-. Where items could be integrated into a common scenario, participants in their studies reported taking this into account when making their ratings, and the ratings reflected this influence. The study also included a protocol (self-report) and no protocol condition and it was found that the ratings were very similar in both of these conditions. Wisniewski and Bassok (1999) was published after the inception of this study and involved a manipulation of the materials based on prior assumptions concerning thematic relations between the stimulus pairs. Since their notion of thematic similarity hinged on the linking of two items through a shared scenario or situation, it is clearly related to the search in this study for the spontaneous use of situational similarity in judging similarity.

A further reason to expect that similarity judgements may draw on thematic links between the items, in addition to evaluating shared/unshared properties, is a series of experiments reported by Lin and Murphy (2001). In these studies, adult participants demonstrated a preference for thematic grouping over taxonomic groupings in a variety of tasks. In experiment 1, participants were presented with a series of word triads consisting of a target word and two other words, one of which was taxonomically related to the target and the other thematically related. They were asked on successive trials with different word groups to pick one of the options that goes best with the target to form category. Across participants, the thematically related word was chosen in 62% of the triads. In the second study the same materials were used but no one word was identified as a target. Instead, participants were told to decide which two out of the three items best formed a category. The average percentage of thematic categorisation dropped to 49% but the authors noted that this was still a high percentage when previous research had suggested that adult participants show a strong preference for taxonomic groupings in similar tasks (Smiley and Brown, 1979). Experiments 2 – 5 involved re-testing a subset of the Smiley and Brown stimuli under differing modes of stimulus presentation. Thematic categorisations still occurred at a much higher rate than that reported by Smiley and Brown. Further manipulations designed to highlight the properties shared between taxonomically related item pairs (listing commonalities) and thus to encourage taxonomic pairings did increase the percentage of taxonomic groupings (by 33%). Importantly, Lin and Murphy also demonstrated that the thematic relations between items could support some kinds of inductive inference and could also prime verification of a taxonomic category member. Once again, this work was published after present study had been completed but Lin and Murphy's findings clearly speak to the expectation that thematic or situational relationships between items may influence other tasks that involve the manipulation of conceptual representations, such as evaluating similarity.

A number of other factors have been shown experimentally to affect subjects' assessments of similarity, including the co-presentation of different information (Kelly and Keil, 1987; Medin and Shoben, 1988; Medin, Goldstone and Gentner, 1993; Tversky, 1977); expertise of the person making the judgement (Sjöberg, 1972; Suzuki, Ohnishi and Shigemasu, 1992) and the time course of the judgement (Ward, 1983). It appears that such manipulations alter the focus of the participants' comparisons and lead to systematic variation in similarity judgements.

Barsalou (1982) investigated the effect of presenting the items, with or without their category label, on direct measures of similarity (scaled ratings). Barsalou reported that ratings of similarity were higher for pairs of items from ad hoc categories when the category label was present than when it was absent. No such effect was found for taxonomic pairs. Barsalou explained this finding as the effect of context on the properties of the items rendered salient in the comparison process. Barsalou claimed that certain properties of entities were context-independent (CI) in as much as they would be automatically activated in the mind of the cogniser simply by presentation of the concept name. Other properties, however, were proposed to be context-dependent (CD) in that they would only be made salient in specific contexts. For example, the property of flammable would only become activated for newspaper in the context of using it to light a fire but not when considered as reading material for a train journey. In a previous unpublished manuscript, Barsalou (1981) had argued that properties associated with taxonomic category members were generally context-independent whilst the properties shared by ad hoc category members tended to be context-dependent. This was used then to explain the differential effect of context upon similarity ratings for pairs of items from taxonomic and ad hoc categories.

Barsalou claimed that similarity ratings of ad hoc category items were enhanced by the presence of a category name because common properties have been activated and rendered salient to the comparison. How does this particular operalisation of 'context' relate to the type of situation instantiation that was seen in chapter 1? The ad hoc category labels in this study 'things you take to a wedding' 'things you find in an attic' can be seen as offering a common situation in which largely perceptually linked items may appear to be linked in a way that may not be easily apparent in the absence of that label. If the items are highly typical of that category then they may well suggest this situation to the participants even if it is not made explicit. This may particularly be the case in the protocol conditions where participants have longer to seek out similarity. If participants make use of situational knowledge in this task then these category members should afford them the opportunity to do so. However, the same can not be said for taxonomic category items. The taxonomic category labels are not suggestive of situations in the same way as the ad hoc category labels and under a common understanding of 'context' may not be considered as contextual cues at all. These labels relate more to the use of context that implies ' think about orange and banana in relation to all fruit' as opposed to having regard to a situation. A context for fruit that would be more analogous to the ad hoc category names would be something like 'eating fruit after a meal'. Even if certain pairs of fruits were linked together in this way, whilst others were not, one would not expect this to be made salient by a category heading of simply "Fruit". One might therefore predict that 'situational similarity' will be evident in the protocols for ad hoc category pairs but less so for taxonomic item pairs.

It can be seen that much of the research on similarity has focused on identifying the factors that affect similarity judgements. It should also be noted that outside of experimental settings, people rarely make an assessment of similarity between two items for its own sake. Assessment of similarity tends to be in the service of further cognitive processes such as categorisation, the projection of properties onto new category members, or problem solving. In everyday situations, there may be a need to make practical judgements of similarity, such as when finding that one does not have the correct tool for a particular job and looking around for something similar. In such cases, it would be the goal of the assessment that drives the comparison process. In contrast, when participants in experimental situations are presented with the names or pictures of two entities and asked to make a rating of similarity, some sense has to be made of the task. This may require the instantiation of a plausible reason for making the comparison (based on past experience) or an attempt to anticipate the aim of the experimenter. Judgements of similarity may therefore be influenced by pragmatic factors as well as purely cognitive ones.

The process of making similarity judgements in an experimental setting should be seen then as an active, constructive one. Participants are not passively observing and evaluating similarities but are engaged in an active search for similarity. Hassin (1997), for example, reported that asking subjects for similarity judgements or difference judgements of pairs of drawn figures led to differences in interpretation of ambiguous features, and the subsequent weighting of those features in the comparison.

There are various ways of attempting to measure similarity. It may be inferred from the outcome of sorting tasks; participants can be asked to make same/different judgements or may be presented with a range of items and asked which is most similar to a target (Tversky, 1977). Alternatively, scaled ratings of similarity between pairs of items may be requested. In the study reported here, a simple procedure of requiring participants to provide judgements of similarity between two items as ratings on a scale of 1 –9 was utilised (where 1 = not similar at all and 9 = highly similar). It was decided to require participants to provide a written protocol explaining their considerations in making the rating. The protocols would be subsequently examined for evidence of the situational concept knowledge under interest.

The materials for the study followed from those used in studies reported in the previous chapter. By using items from taxonomic and ad hoc categories, it would be possible to make comparisons between the two category types: both in terms of the explanations given for the ratings and the ratings themselves. It was decided to design the study so as to allow a replication of the Barsalou (1982) results. It was noted, however, that the Barsalou study had not commented on, nor reported, the typicality of the items used as stimuli. Typicality has been characterised as a function of how similar the exemplar is to a prototype of the category (Rosch and Mervis, 1975). Barsalou (1985) showed that similarity to the central tendency of the category correlated highly with the 'goodness of exemplar' judgements given by participants for taxonomic categories. He reported that ad hoc categories also exhibited graded structure. However, this was not determined by similarity to a central tendency but partly by a measure of how familiar the subject was with the item as a member of that category ('frequency of instantiation') and partly by the degree to which the member is felt to match the 'goal-ideal'. "Ideals.... are characteristics that exemplars should have if they are to best serve a goal associated with their category" (p. 630). Since members of both types of categories can vary in typicality, it was reasoned that their perceived similarity could be influenced both by the context of the comparison and the typicality of the items. It was therefore decided to include typicality as an additional factor to see whether this in any way modified Barsalou's earlier findings regarding the effect of context.

Requiring the production of protocols would necessarily affect the time course of the judgements. Time allowed for similarity judgements has also been found to affect the process. Goldstone and Medin (1994) demonstrated that abstract structural features are weighted more heavily in similarity judgements than superficial features when participants are given more time to respond. It has elsewhere been suggested (Smith and Kemler Nelson, 1984; Ward, 1983) that a primitive, holistic assessment of similarity is made by participants when cognitive resources are constrained by speed; this is contrasted with comparisons along particular dimensions which appear to be made when time is not a factor. These studies have involved indirect measures of similarity such as sorting tasks or same/different judgements, rather than the scaled ratings involved in this study and also utilised comparatively small time differences between conditions. Nevertheless, it was decided to compare ratings from the protocol conditions to those from a condition in which no protocol was requested, which would result in a much smaller amount of time being spent on the task.

These considerations gave rise to a four-factor design that had two aims. The principal aim was to assess the role (if any) of situational concept knowledge in direct ratings of similarity – both the protocols and the ratings in relation to the protocols would provide the necessary data. The secondary aim was to investigate whether previously reported context effects would be modified by the consideration of category member typicality – only the quantitative date would speak to this question.

The factors investigated here were; (i) the presence or absence of an explicit context, (ii) the type of category from which the entities under comparison were drawn (taxonomic vs. ad hoc) and (iii) the relative typicality of the items being rated. Self-reports were collected from a subset of participants, with type of self-report constituting a fourth factor. It was predicted (following Barsalou, 1982) that the ratings of similarity for pairs of ad hoc category items would be higher in the presence of the category label (with context) than when presented without context. It was further predicted that this effect would be mediated by typicality. Most importantly, it was predicted that the protocols provided by those participants who were asked to explain their rating would show that their considerations had included knowledge of the situations in which the items are often encountered and not solely feature matches and/or mismatches. Since the protocol analysis was a substantial activity in its own right, it is intended to report the method for the study overall and then the results and a brief discussion of the quantitative data followed by the protocol analysis.

Method

Participants

Seventy-two undergraduates from the University of Hertfordshire and 58 sixth form school students volunteered for this study. Additionally, 40 sixth form students from local colleges volunteered to provide typicality ratings to assist with the compilation of the materials. The modal age of the sixth form students was 17 years.

Design

Four factors were manipulated in this study. The first was category (taxonomic or ad hoc). This was a within-subject factor. The second within-subject factor was pair type. There were three levels of this factor, TT, where both items in the pair were typical of the category; AA, where both items in the pair were of low typicality (atypical) within the category and TA where one item in the pair was typical of the category and the other was of low typicality. The third factor was context, a between-subject factor which referred to whether the pair of items were presented in the presence of the category label (with context) or not (no context). The fourth factor was self report, this was a between-subject factor with three levels; participants provided a protocol at the time of making the rating (concurrent SR), provided a protocol after making the rating (retrospective SR) or did not provide a protocol at all (NSR).

Materials

Ten categories were used; 4 taxonomic categories drawn from Rosch and Mervis (1975) namely, *bird, fruit, vehicle* and *furniture*, and 6 ad hoc categories namely, *things you find in an attic, things you find on a shelf, things people take to a wedding, things that may be conveniently kept in pockets, things dogs chase* and *things sold on the black market in Russia.* A range of members of these categories had been generated by undergraduates in the two previous studies. Forty sixth-form college students rated a set of these items from each category for typicality ('goodness of example'). See Appendices D, E and F for details of instructions and stimuli. Appendix G shows the mean typicality ratings and standard deviations for all stimuli in all ten categories. On the basis of these ratings, three of the items rated as most typical and three of the items receiving the lowest typicality ratings were selected from each category. Superordinates (such as *food*) were avoided. The classification of the items as typical and atypical was not based on a median split. Typicality is often judged relatively within categories and so there was no attempt to match typicality across categories. The classification could be viewed as fairly crude. The items were combined in pairs to give rise to a pool of 3 typical/typical (TT) pairings, 3 atypical/atypical (AA) pairings, and 9 typical/atypical (TA) pairings for each category. (See Appendix H) The ordering of the items within any given pair was kept constant.

Test booklets were constructed. These consisted of 30 pairs of items, one of each type of pairing for each of the 10 categories. Individual booklets were compiled by drawing pairs of items from the pool in such a manner that maximum variation in materials between participants was ensured. The order of the pairs within each booklet was separately randomised. Participants were asked to rate each pair of items for similarity "to each other" on a scale of 1 to 9 where 1 represented a judgement of "not similar at all" and 9 indicated "highly similar". At the top of each page appeared the names of the two items under comparison. The names of the items were on the same line but separated by a short space. For half the booklets, the items appeared just below the name of the category from which they had been drawn (the Context condition) while for the remaining booklets, the names of the two items stood alone (the No Context condition). These materials were then used for three different groups of participants. For the participants in the No Self Report (NSR) condition, participants were required only to rate the similarity of the items to each other in response to the written question "How similar are an X and a Y?" which appeared on the page just below the names of the items and was followed by the scale described. Participants worked through the materials at their own pace, having been told to give each judgement some thought but not to dwell over it for too long. In the Concurrent SR condition, participants saw the names of the items, X and Y, which were followed by the same question, " How similar are an X and a Y?" and the instruction "write down the things you are considering in making your judgement of similarity" then appeared. A space was provided for the self-report and below that was the instruction to make their rating on the scale at the bottom of the page. This format was repeated for every pairing. In the Retrospective SR Condition, the instruction to make the rating and the scale appeared before the written instruction to "Write down the things you considered in making your judgement of similarity". This format was the same for each pair of items (see Appendix I for sample materials).

Protocol analysis

The classification was initially made independently by two judges on a sample of 100 protocols. For protocols with no context the classification was:

Instantiated - protocol made <u>direct reference</u> to the category label from which the items had been originally generated and which was presented with the item pair in the context condition.

Close - protocol made an indirect or <u>close reference</u> to the category label from which the items had been originally generated and which was presented with the item pair in the context condition.

Not instantiated - protocol made <u>no reference</u> at all to the category label from which the items had been originally generated and which was presented with the item pair in the context condition.

For the protocols with context the classification was:

Mentioned - protocol made <u>direct reference</u> to the category label that was presented with the item pair
Close - protocol made an indirect or <u>close reference</u> to the category label that was presented with the item pair
Not mentioned- protocol made <u>no reference</u> at all to the category label that was presented with the item pair

The level of agreement between the two judges was 98% - the disputed protocols were resolved by agreement. This classification was not a highly controversial one and so one judge classified the remainder of the protocols.

The above classifications were subsequently further divided on the basis of whether the protocols made reference only to similarities between the items in the pair or to only differences between the items or to a mixture of both similarities and differences. Protocols that did not fall into one of these three categories were classified as 'other'.

Once again, a trial classification was initially made independently by two judges on a sample of 100 protocols (these were the same judges and the same protocols as mentioned above). The level of agreement between the two judges was again 98%; consequently one judge classified the remainder of the protocols on the same basis.

Procedure

Most participants were run in groups ranging in size from four to ten, a small number were run individually. Participants were asked to read the instructions on the front of the test booklet – these were also read aloud to them. Participants were allowed the opportunity to ask questions before starting the task. In the two self-report conditions, a period of 90 seconds was allocated for the completion of each page of the test materials; subjects were told when the 90 seconds had elapsed and were instructed to turn over the page. The whole procedure for the self-report conditions thereby taking 45 minutes per participant to complete compared with about 10 minutes for the self-paced, NSR condition in which protocols were not collected.

Results

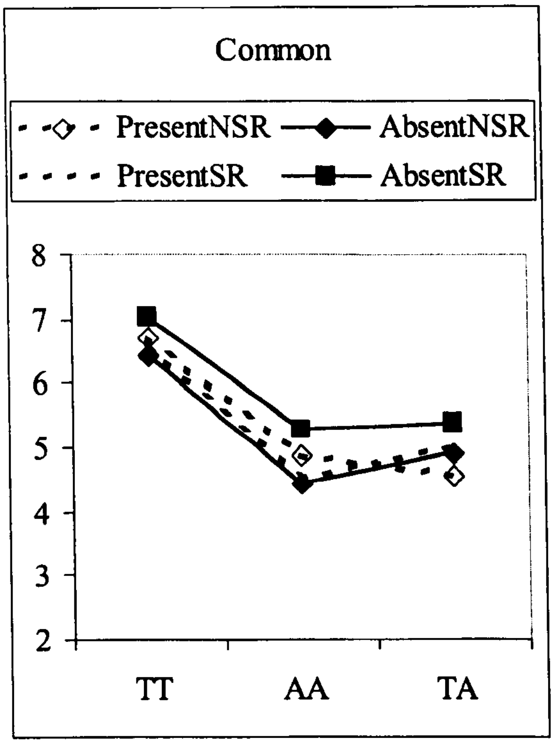
The results of the quantitative data analyses will be reported first to address the hypotheses concerning the effects of context and typicality upon similarity ratings. The analysis of the protocols collected in the self-report conditions will then be reported. As explained earlier (see materials), the typicality ratings given by a previous set of participants were used to classify the items as typical or atypical. This classification was then used to create the different pair types. Although typicality was originally collected as a continuous variable, it was very far from normally distributed and is therefore not suitable for use as a covariate in the quantitative analyses. The category classification of pairtype was used instead.

Firstly, overall mean similarity ratings in the two different kinds of self-report conditions, concurrent (4.34, sd = 0.96) and retrospective (4.45 sd = 0.77), were compared. They were very similar and did not differ reliably (F 1.24, $p = 0.3$). Neither of these conditions entered differentially into any significant interactions with the other three factors, and they were therefore combined into one Self-Report condition, and contrasted with the ratings in the No Self-Report condition. Table 3.1 and Figure 3.1 show the mean similarity ratings for the different pairs of items varying in typicality for both categories in the presence and absence of context. It should be noted that where there was no self-report, the variability in the ratings, as indicated by the standard deviations, seems to be both objectively small and relatively far less in the absence of context compared to in the presence of context and compared to all conditions with self-report. This was the case for both taxonomic and ad hoc categories. This suggests that for all category types, participants have the greatest consensus around the similarity when a) their judgements are made fairly quickly and with no need to justify them ratings and b) the items are presented without any other information (i.e. the category label). When more time is given and a protocol required, or when the additional information of the category label is provided, participants may then vary in the extent to which they attempt to make some use of that information and may use it in different ways, leading to greater variability in the ratings. This point will be returned to in the discussion when the pragmatics of the task facing the participants are considered.

Table 3.1.

Mean similarity ratings for taxonomic and ad hoc pairs of items, for all three kinds of pair types in the No Self-Report and With Self-Report conditions, (standard deviations in brackets).

	No Self Report					
Context	Taxonomic			Ad hoc		
	TT	AA	TA	TT	AA	TA
Present	6.69	4.88	4.55	4.49	3.33	3.14
	(1.37)	(1.47)	(1.48)	(2.02)	(1.21)	(1.47)
Absent	6.43	4.44	4.91	2.74	2.21	2.08
	(0.34)	(0.33)	(0.35)	(0.17)	(0.16)	(0.17)
	With Self Report					
Present	6.59	4.53	5.01	3.99	2.92	2.89
	(1.25)	(1.11)	(1.22)	(1.58)	(1.06)	(1.06)
Absent	7.02	5.29	5.37	3.57	2.82	2.77
	(1.18)	(1.16)	(1.35)	(1.08)	(094)	(1.04)



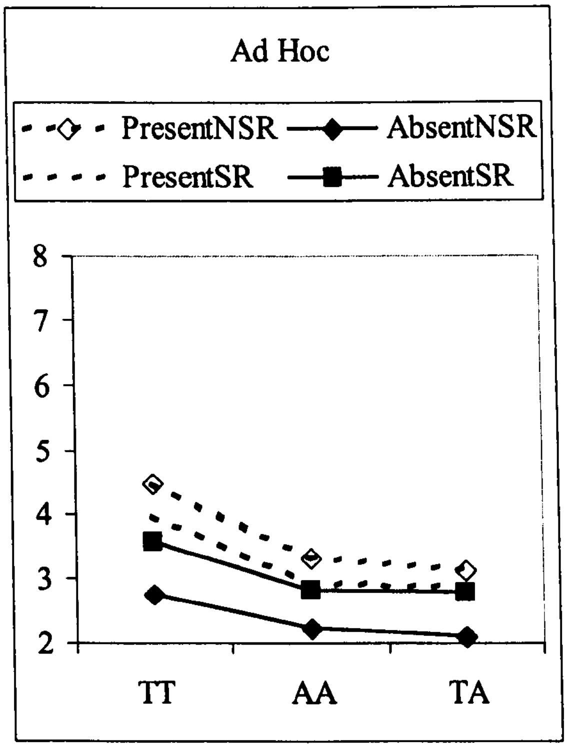


Figure 3.1 Mean similarity ratings for all typicality combinations of Taxonomic and Ad hoc item pairs; with and without context and self-report

The data were subjected to a Category Type (2) by Typicality (3) by Context (2) by Self-Report (2) mixed ANOVA. The Category Type main effect was reliable, $F(1, 126) = 465.185$, $p < .0005$, eta squared = 0.787. Overall, taxonomic items were rated more similar (5.52) than ad hoc items (3.11).

The main effect of Context was not reliable, $F(1, 126) = 3.34$, $p = 0.070$. There was a reliable main effect of Typicality, $F(2, 252) = 231.57$, $p < .0005$, eta squared = 0.648. Overall, pairs of items made up of two highly typical items produced higher similarity ratings than pairs of items made up of two atypical items, or one typical and one atypical item. The means for the latter two pair types were remarkably close, and not reliably different. For taxonomic categories they were 4.82 and 5.02 for AA and TA pairs, respectively; for ad hoc they were 2.83 and 2.75 for AA and TA pairs respectively. This pattern was true for both kinds of categories, with or without context. The main effect of SR was not reliable, $F(1, 126) = 2.521$, $p = 0.115$.

There were a number of interesting interactions. First, there was a reliable interaction of Context by Category Type, $F(1, 126) = 18.51$, $p = 0.01$, eta squared = 0.128. The rated similarity of taxonomic items was slightly *lower* in the presence of context ($M = 5.4$) than in the absence of context ($M = 5.7$), whereas the rated similarity of ad hoc items was slightly *higher* in the presence of context ($M = 3.4$) than in the absence of context ($M = 2.8$). Individual post-hoc contrasts of the taxonomic and ad hoc data showed the effect of context to be reliable in the case of the ad hoc category members, $F(1,126) = 9.7$, $p = 0.002$, but not in the case of items from taxonomic categories, $p = 0.134$.

Secondly, the nature of the Context by Category Type interaction in the absence and presence of self-report was rather different. In the NSR condition, where subjects made similarity ratings at a faster pace than in the SR conditions, the effect of context is absent for taxonomic items $F = 0.224$, but the perceived similarity of ad hoc items is reliably higher in the presence of context compared to the ratings made without context, $F(1, 41) = 14.5$ $p < 0.01$ (see Figure 3.1). In the conditions which primed subjects to think more deeply about the comparisons, namely the SR conditions, the perceived similarity of ad hoc items was now uninfluenced by the context, $F = 1.2$, but the perceived similarity of taxonomic items was reliably lower in the presence than in the absence of the category label, $F(1, 87) = 5.5$ $p = 0.021$ (see Figure 3.1).

When Context exerted an influence on ratings of ad hoc items, namely in the NSR condition, the interaction with Typicality approached significance, $F(2,80)=2.9$, $p=0.061$, eta squared = 0.068. The most pronounced effect was on the pairs of highly typical items. When Context had an effect upon ratings of taxonomic category members, namely in the SR condition, there is no such interaction with Typicality $F(2,172) = 1.73$. All types of pairing were equally rated more similar in the absence of context. To summarise, the enhancing effect of context on the ratings for ad hoc pairs found by Barsalou appeared only when there was no self-report and most strongly when both of the items in the pair were typical of the category.

An attempt was made to use a range of items pairs from within each category in order to increase confidence in the potential for generalising any conclusions to other categories and category members. Participants could have seen any one of three pairings of TT items, 9 pairings of TA and 3 pairings of AA items. However, a further analysis based on items (as a between factor) as opposed to participants showed small item effects in both conditions of category type (taxonomic and ad hoc) and under every condition of typicality as follows:

Ad hoc TT: $F(17,703) = 9.89$, $p < 0.001$, eta squared = 0.193

Ad hoc TA: $F(53,556) = 2.98$, $p < 0.001$, eta squared = 0.221

Ad hoc AA: $F(17,703) = 12.05$, $p < 0.001$, eta squared = 0.237

Taxonomic TT: $F(11,466) = 6.88$, $p<0.001$, eta squared = 0.140

Taxonomic TA: $F(36,365) = 3.07$, $p < 0.001$, eta squared = 0.232

Taxonomic AA: $F(11,466) = 15.44$, $p < 0.001$, eta squared = 0.270

There were no significant main effects of self-report (as was the case in the analysis by participant) except for the taxonomic TA pairs where there was a small effect of protocol $F(1,365) = 6.38$ $p=0.012$ eta squared = 0.017.

Most importantly, item did nor enter into any interactions with either protocol or context with the exception of two small interactions between item and protocol in the taxonomic AA condition $F(11,459) = 1.96$, $p = 0.031$, eta squared = 0.045 and TA conditions $F(36,365) = 1.77$, $p = 0.005$, eta squared = 0.149.

There was still an effect of context for all pairings of ad hoc items, TT , $F(1,703) = 42.612$, $p<0.001$ eta squared = 0.057; TA, $F(1,556) = 14.97$, $p < 0.001$ eta squared = 0.026: AA, $F(1,700) = 19.04$, $p < 0.001$ eta squared = 0.026 and significant interactions with self-report due to the ratings with context being higher for pairs in the no self -report condition but the pairs in the self-report condition not reflecting the same pattern. This is consistent with the story of the effect of context and protocol on ad hoc item pairs shown in the participant by analysis. (For example 100% of AA item pairs showed higher ratings with context in NSR but only 47% showed this pattern in the SR condition with the other item pairs showing the opposite pattern)

There was no effect of context for the TT and AA pairs but there was a small effect of context for TA pairs across both conditions of self-report $F(1,365) = 9.00$ $p<0.003$ eta squared 0.024. There was no interaction of context and protocol for that condition but there was such an interaction for the TT, $F(1,466) = 5.08$ $p= 0.025$ eta squared = 0.011 and AA pairs $(1,459) = 10.634$ $p=0.001$ eta squared = 0.023. Where these interactions occurred it was due to the ratings being slightly higher without context than with context in the SR condition but the pattern not being repeated in the No SR condition.

Despite the item effects found, the overall pattern of findings concerning context and self-report are also found at an item level. The findings from the participant analysis concerning the differential effects of these factors on TT ad hoc pairs compared to TA and AA pairs should perhaps be treated with some caution due to the small interaction effects of protocol and item at some levels of typicality,

Discussion of the Similarity Ratings

Participants in the self-report conditions were informed at the outset that they would be asked to explain their rating. Irrespective of whether the explanation was given before or after the similarity rating, participants in both conditions approached each judgement expecting to have to think more deeply about the comparison involved. While the primary aim of collecting the self-reports was to afford an insight into participants' reported considerations, it transpired that self-report was itself a causal factor, entering into interesting interactions with the other variables.

Typicality had the expected effect. However, context had an effect only for categories that were ad hoc and where there was no self-report. To put it another way, instructions to self-report wiped out the effects of context found in previous studies.

Barsalou (1982) reported a reliable increase in similarity ratings for items from ad hoc categories in the presence of a category label but not for taxonomic category items. As Figure 3.1 illustrates, this interaction was replicated in the NSR condition, which was procedurally closest to Barsalou (1982). Subjects in the NSR condition were not encouraged to dwell upon their ratings, neither were they required to describe their reasoning. In contrast, when subjects were required to describe their thoughts as they formulated their ratings in the SR conditions, the presence or absence of category label had no effect on the ratings of similarity of the ad hoc category items (see Figure 3.1). The self-report condition may have encouraged participants to generate their own contexts for the judgements. This possibility will be pursued when the protocol contents are examined.

The results of this study suggest that the effect of providing context on ratings of ad hoc pairs found by Barsalou (1982) may be confined to particular conditions under which the ratings are made, namely, relatively speeded and unreflective. Additionally, this effect of context on the ratings of ad hoc items found in the NSR conditions was most marked for pairings of typical/typical items. This suggests that the ad hoc items rated in Barsalou's study were good exemplars of the category. It is unclear from the present study whether it is the extended time course of the judgements alone or the requirement to report on the thoughts underlying the rating that explains the loss of the effect when participants were asked to give protocols. However, the results support the notion that different "kinds" of similarity judgements are made about the same entities depending upon the time and effort afforded to the task (Goldstone, 1995).

It will be suggested throughout this discussion that the co-presentation of the category name with the items under comparison can be viewed as providing subjects with a uniform framework for the comparisons- this will be referred to as *explicit* context. Moreover, on a pragmatic basis, this information will be perceived as having been supplied by the experimenter for a reason (cf. Hilton, 1995).

The mean ratings for items with an explicit context remain fairly stable across the SR and NSR conditions. The phenomena of greater interest are the ratings in the *absence* of explicit context. In the case of taxonomic category members, the mean ratings without context are higher in the SR condition than they were in the NSR condition. For ad hoc items, the Barsalou (1982) effect is lost at least partly because the mean ratings made in the absence of explicit context are higher than they were in the NSR condition. It appears then that an explanation of the results will hinge largely on what is happening when participants are asked to rate items in the absence of explicit context. It is proposed that, under these circumstances, the active construction of similarity leads participants to instantiate a context for themselves, suggested by the co-presentation of the items. Items that are more typical of a category are more likely to evoke a context than items that are less typical. This implies that all judgements are, in effect, made "in context", whether it is the explicit context provided by the experimenter or the context instantiated by subjects in their search for similarity.

Patterns of Effects of Typicality

A distinctive pattern of ratings for the varying degrees of item typicality was shown overall and across both category types, with and without explicit context. Typical pairs were rated more highly than the other two pair types, which were rated at a very similar level.

The provision of an explicit context in the form of the category name directs comparisons; pairs of typical items will then be more similar than pairs of less typical items by virtue of sharing more features. When no category name was provided, the high salience of the category membership of taxonomic items makes it likely that this will nevertheless be the context most easily instantiated by subjects, yielding a similar pattern of results, in this respect, to the explicit context condition. Indeed this was found to be the case. While these results do not refute Barsalou's notion of Context Independent properties, they are equally supportive of the notion that these highly salient properties are *dependent* upon the activation of the equally salient context of their category membership.

When ad hoc items were presented with a category name, pairs of typical items were again rated as more similar than the other pairings; under a feature-matching model of similarity, this would be due to sharing of a higher number of 'ideal' attributes. In the absence of explicit context, the co-presentation of typical ad hoc items may have evoked a plausible context for comparison. For example, camera and confetti (*things people take to weddings*) would be likely to suggest the context Wedding even if not specified and direct subjects' attention to a potential framework for the judgement; these good exemplars will therefore be rated as more similar compared to the situation in which less typical items, such as cake and hat are presented. Such items may be less likely to evoke a frame of comparison to direct the judgement and therefore be rated as less similar since they have few commonalities outside of this context.

The Effect of Context on Ad Hoc Category Members

When judgements were unreflective (NSR), participants who saw pairings of typical, ad hoc category members with explicit context rated the items as reliably more similar than did participants who saw the items in the absence of context. It is possible that the relatively speeded nature of the judgements meant that some participants failed to find an association to direct their comparisons if one was not explicitly stated; even if subjects did instantiate a context for themselves, it could have been a heterogeneous set of contexts having an unpredictable effect upon ratings.

Barsalou explained this effect in terms of the activation of context-dependent properties. This is a plausible explanation for the example Barsalou cites of a basketball for which "floats" is a context-dependent property activated only in a particular context. However, for items such as Confetti and Camera or Wallet and Money, (*things people take to a wedding* and *things you keep in your pocket*, respectively), it is difficult to identify a context dependent property that they share and which would thus render them more similar once a context was established. One of two explanations can be offered as to why these items are rated as more similar in context. Firstly, in the search for similarity and to oblige the experimenter, participants find ANY commonality a reason to rate perceptually dissimilar items as more similar than at first thought. Alternatively, the fact that items both play roles in a common scenario or situation is considered in the same way as a shared property of any other kind. That is to say 'is used at weddings' is considered by many participants to be a 'feature' of confetti and one that can be potentially shared by other objects. The 'respects' in which the items could be considered similar has simply been extended to include external factors such as situations.

The effect of context on ratings of ad hoc items was not found to the same extent for pairings of atypical items or typical/atypical items. Even if a context is found, or made explicit for these items, at least one item is a poor exemplar of the category and as such may rarely have been experienced as a member of that category and may not exhibit any characteristics associated with an ideal member; ratings of similarity may therefore remain low. For example, the provision of the context *things chased by dogs* could be seen to do little to enhance the similarity of cow and bicycle.

As observed, no effect of explicit context on ad hoc pairs was found when participants made more careful considerations of their judgements in the SR conditions. This may be explained then by subjects in this condition having more time in which to instantiate their own context that could direct the comparison; this would then defuse the impact of explicit context. Additionally, when judgements are made relatively quickly, the provision of explicit context (by virtue of which two seemingly dissimilar items may be found similar in some unexpected way) may result in a rating which, if given more time to consider, would be reduced; for example, although *cats* and *sticks* are indeed both items which may be chased by dogs, they are, as objects, quite different.

The Effect of Context on Taxonomic Category Members

For the taxonomic category items, the effects of context were quite different. When judgements were relatively speeded, provision of explicit context had no reliable effect upon ratings of these items. When participants see *apple* and *orange*, these items have a high default similarity by virtue of their shared properties as *fruit*; the membership of that category being (presumably) activated automatically. Co-presentation of the category name did not, therefore, result in higher similarity ratings.

The reduced similarity of taxonomic category members in the presence of context when ratings were made with self-report may have been a genuine effect of being prompted to reflect upon the rating task more deeply. As suggested, under NSR conditions the search for similarity between *apple* and *orange* produces 'fruit' (and consequently fruit-like qualities) quite automatically, the judgement is made against this background and the participant moves on. Possibly when the participant provided with no context has time to ruminate upon the comparison, 'fruit' may seem too obvious as the field of comparison intended by the experimenter and the items may be judged against a wider category membership of, for example, *food.* Since this category has a larger and more diverse membership, this may have the effect of rendering two fruits within this instantiated context more similar than if the relevant basis for the comparison is considered to be *fruit* (cf. Sjöberg and Thorslund, 1979).

Comments

The explanations of the complex interactions found in this study rely upon understanding similarity judgements obtained under these conditions as an active, goal-related construction; this is consistent with recent dynamic, process-based views of similarity. Participants sought meaning for the task and made use of any information provided in the explicit context of the comparison; in the absence of such information, a context of their own was likely to have been instantiated. It could be argued then that similarity judgements are never 'context-free'. A context free judgement would be one without constraints and would be subject to the famous Goodman criticism that everything can be found similar to everything else (Goodman, 1970). The process of constructing a framework to identify relevant respects in which items could be similar might be mediated by semantic, conceptual knowledge and also may reflect pragmatic considerations whereby the participants seek to anticipate the intentions of the experimenter. As mentioned earlier, recent studies of similarity have shown that judgements are not always based solely on feature matching. Bassok and Medin (1997) found that when participants were asked to assess the similarity of paired statements which had matching nouns, for example "The carpenter fixed the chair" and " The carpenter sat on the chair", subjects offered responses such as "similar because he sat on the chair to see whether he fixed it well". Such a response clearly refutes the conception of similarity as linear computations of shared and different properties.

Medin, Goldstone and Gentner (1993) suggested that a distinction can, and should be made between indirect measures of similarity (same-different judgements of stimuli), direct measures of similarity (as in scaled ratings) and similarity as a theoretical construct. However, the literature is replete with claims concerning the flexibility or stability of similarity that ignore the distinction advised by Medin et al. Researchers must be mindful of the likely differences between the way in which direct measures of similarity can be influenced and the constraints which may operate upon similarity as it is deployed in the service of other cognitive processes. At the very least 'context' effects upon measures of direct similarity are likely to include the experimental context; that is the meaning that subjects attribute to the task and the goals they formulate.

The data collected here, through an expanded set of experimental conditions in comparison to Barsalou's much cited study, have shown that claims about the effect of context on similarity should be carefully qualified.

In the next section, the analysis of the protocols will be used to investigate whether the participants' considerations varied under different conditions and whether they drew on other sources of similarity altogether, such as event or experience-based similarity

Protocol Analysis

The primary aim of the protocol analysis was to look for reports of other sources of similarity apart from feature–matching and in particular the use of situational knowledge. However, it was also deemed interesting to see if protocols could shed any light on the pattern of ratings seen with and without context for the taxonomic and ad hoc categories in the SR conditions, namely increased ratings in the no context conditions for both ad hoc and taxonomic pairs compared to the NSR conditions. This will be addressed first.

Each of the eighty-seven participants was asked to provide protocols for 30 similarity judgements, giving rise to a total of 2610 ratings and accompanying protocols. In seven instances participants provided neither a rating nor a protocol for an item. This left a total of 2603 protocols (1563 for ad hoc item pairs and 1040 for taxonomic item pairs). In a few cases (1.3%) the participant provided a rating and no protocol – these were subsequently coded as such and remained part of the analysis. Occasionally a participant provided a protocol but no rating. The initial analysis of the protocols was conducted to see whether they could offer any substantiation for the speculation with regard to the cause of the context effect. The first question to be addressed concerned the ad hoc category members and the loss of the context effect when protocols were provided. Although, the basis for the ratings cannot be compared to the NSR condition (the considerations of the participants in the NSR condition are, by definition, unknown to us) it was felt that looking at the protocols for the ad hoc item pairs with and without context may lend a partial understanding of the ratings. There were two possibilities; firstly, that participants in the no context condition may have had more chance to instantiate contexts for themselves and to find similarities that would not be apparent when making a faster judgement. Secondly, participants in the 'with context' condition may have had more opportunity to dwell on the superficiality of the category membership provided, and to consequently focus on perceptual and functional differences between the items. The pattern of the ratings seemed to suggest that the loss of effect was due to the 'without context' ratings and so the first explanation was expected to be more consistent with the ratings. The initial classification of the protocols therefore was in terms of whether or not they instantiated the context when it was not explicitly provided and whether or not they made reference to it as a basis for similarity when it was provided. Ad hoc and taxonomic item pairs were separated for the purposes of this breakdown. For protocols with no context the classification was Instantiated, Close or Not Instantiated. For the 'with context' protocols the classification was Mentioned, Close or Not Mentioned (see Method section for further definition of these terms). For clarification purposes, the following examples of protocols from each of the four main categories are presented:

Table 3.2

Examples of each classification type. All examples relate to items from the category *things people take to a wedding*.

No Context	items	
Instantiated	Cake/Hat	Associate with occasions – weddings, summer days, special occasions. Even a similar shape (could be!) S79NCRP
Close	Present/ Camera	Camera is a machine. A present is a gift given to someone, they are both used at special occasions. S7NCCP
Not instantiated	Cake/Hat	Different uses. You eat a cake. You wear a hat. Made of different materials and come in different sizes. S76 NCCP
With Context		
Mentioned	Rice/ Cake	Not so similar. So what if they are both at a wedding. S3WCCP
Close	Rice/Hat	They are both thrown! Specially, if you're a man. Being hit by either can be painful. S41 WCRP
Not mentioned	Camera/Confetti	A camera can take a picture of a moment in time, whereas once confetti is thrown it cannot be used again and is thrown away. S55WCCP

Ad hoc items pairs

Table 3.3.shows classification of protocols for ad hoc item pairs

Table 3.3

Number and percentage of protocols for ad hoc item pairs that did and did not instantiate the context even when it was not provided (no context condition) and did or did not mention the context when it was provided (with context condition).

NO CONTEXT			WITH CONTEXT		
Instantiated	Close	Not instantiated	Mentioned	Close	Not mentioned
79	29	665	244	33	513
10%	4%	86%	31%	4%	65%

Since it was the classification of the protocols that were of interest, as opposed to the classification of participants, and since the participants had an equal number of opportunities to produce protocols of any particular type, the protocols were treated as the unit of analysis and a multinomial logistic regression was performed on the data using the classifications as the dependent measure. This produced a highly significant result from the 1563 protocols - $\chi^2(2) = 108.14$, $p < 0.005$. Unsurprisingly, the proportion of references to the context was higher when context was provided and the proportion of protocols failing to mention the context was lower when context was provided. More importantly, it can be seen that when the context was provided, 65% of the protocols made no reference to it and when it was not provided, 14% of protocols instantiated it (or a category very close to it) anyway. However, since the assumption of independence of cells was violated, it was possible that a small number of participants may have contributed disproportionately to the pattern seen in Table 3.3. The pattern of protocol classification for each participant was therefore examined. It was found that in the 'no context' condition, 100% of participants showed the pattern reflected in table 3.3 where a significantly higher number of protocols failed to instantiate the context ($M = 15.47$, s.d 1.5) than to instantiate it ($M = 2.07$, s.d. $= 2.04$) $t(42) = -27.507$ $p < 0.001$. In the 'with context' condition 70%

of participants showed the protocol classification pattern of Table 3.3 with participants producing significantly more protocols that failed to mention the context ($M = 11.29$, $sd=5.51$) than mentioning it ($M =5.9$, s.d. $= 5.38$) $t(43) = -3.298$, $p = 0.002$. It is consequently concluded that the protocol pattern reported in table 3.3 is representative of the performance of participants and has not been unduly influenced by the data from a small subset of the sample.

However, instantiating a context, or mentioning one provided, need not necessarily result in finding items similar – note the example for Rice/Cake in Table 3.2.
The above protocol classifications were therefore subsequently further divided on the basis of whether the protocols made reference only to similarities between the items in the pair or to only differences between the items or to a mixture of both similarities and differences. Protocols that did not fall into one of these three categories were classified as 'other' (see Table 3.4).

Table 3.4
Number and percentage of protocols with and without context, mentioning only similarities between items, only differences or both similarities and differences.

	N	Similarities only	Differences only	Both	Other
No Context					
Instantiate it	79	29	1	32	17
	%	37%	0%	41%	22%
Do not instantiate it	665	140	122	212	191
	%	21%	18%	32%	29%
With Context					
Mention it	244	67	53	72	52
	%	28%	22%	29%	21%
Do not mention it	513	120	165	114	114
	%	23%	32%	22%	22%

It can be seen that where no context was provided to participants and yet they still instantiated that context, there were virtually no protocols referring only to differences between the items whereas 18% of the protocols in the no context condition mentioned only differences when the context was not instantiated by the participants themselves. Approximately a sixth of the protocols showed evidence of participants instantiating their own context which matched or was close to the one the experimenter had in mind and this did seem to have the effect of causing them to focus somewhat less on the differences between the items. Perhaps more interestingly, of the protocols showing no instantiation of the original context, 21% of these still listed only similarities for these perceptually different items, and a further 32% mentioned both differences and similarities. The question of what might constitute similarity between these dissimilar items, other than the context label originally envisaged, is addressed later.

When context was provided to participants, only 31%of the protocols contained a specific reference to it (see Table 3.3) and of those protocols, 22% still mentioned only differences between the items. The greatest number of 'difference only' protocols occurred where context was provided but was not referred to at all.

As before, a multinomial logistic regression was performed on the data using the classifications as the dependent measure, χ^2 N(6, 1501) = 67.49, p<0.005. The classifications were associated with both context, χ^2 N(3, 757) = 50.635, p<0.005 and reference to context χ^2 N(3, 744) = 27.299, p<0.005. The effect arose mainly from

i) the higher than expected similarities only protocols, where a context was instantiated even though it was not provided (37%);

ii) the lower than expected difference only protocols where a context was instantiated even though it was not provided (0%)

iii) the higher than expected difference only protocols where a context was both provided and mentioned (22%)

The same caveats apply to this analysis as to the previous analysis and so the pattern of protocol classification for each participant was therefore examined with particular reference to the proportion of 'similarity only' protocols versus 'difference only' protocols produced. In the 'no context' condition where the participants instantiated the context in any case, 53% of the participants did not produce any 'similarity only' or 'difference only' protocols: 44% produced more 'similarity only' than 'difference only' protocols and only 3% produced the opposite pattern. Participants produced significantly more 'similarity only' ($M = 0.69$) than 'difference only' protocols ($M = 0.07$) $t(42) = 4.33$, $p < 0.001$.

In the 'no context' condition where the participants did not instantiate the context, 7% of participants did not produce any 'similarity only' or 'difference only' protocols: 9% produced the same number of each: 37% of participants produced more 'similarity only' than 'difference only' whilst 46% of participants produced the opposite pattern but there was no significant difference between the mean number of similarity only ($M=3.26$) and the mean number of difference only ($M=2.81$) protocols produced by participants in this condition $t(42) = 0.552$, $p=0.584$.

In the 'with context' condition, where participants did mention the context provided 36% did not produce any 'similarity only' or 'difference only': 12% produced equal numbers of each protocol type: 27% of participants produced more 'similarity only' than 'difference only' whilst roughly the same percentage of participants (25%) produced the reverse pattern and there was no significant difference between the mean number of 'similarity only' ($M=1.64$) and the mean number of 'difference only' ($M=1.34$) protocols produced by participants in this condition $t(43)$ $t=0.634$ $p=0.530$.

In the 'with context' condition, where participants did mention the context provided, 9% did not produce any 'similarity only' or 'difference only' protocols: 7% produced an equal number of each protocol type: 32% produced more 'similarity only' than 'difference only' and 52% showed the opposite pattern but there was no significant difference between the mean number of similarity only ($M=2.79$) and the mean number of difference only protocols produced by Ps in this condition ($M=3.45$) $t(43)$ $t=0.818$ $p=0.418$.

With the exception of the final condition, this suggests that the pattern of 'similarity only' and 'difference only' protocols shown in table 3.4 is representative of the performance of the participants and does not reflect an undue influence on the part of few participants. In the case of the 'with context but no mention of the context' situation, more participants did produce the pattern shown but the difference in mean numbers of 'similarity only' and mean number of 'difference only' protocols was not significant . This suggests that there were a relatively high number of similarities produced by those participants who did favour mentioning similarities only.

With regard to the explanations for the pattern of ratings discussed previously, it does seem as though ratings in the 'no context' condition for ad hoc item pairs may have been raised as a result of a combination of factors. Some protocols showed that the participant had instantiated the experimenter's context for themselves and seemed consequently to have focused somewhat less on differences between the items whilst others did not instantiate that context but still showed consideration of only similarities between these perceptually dissimilar items (e.g. clock and photograph) Clearly, precisely the same considerations could have been made by participants who did not provide protocols but since time has been shown to affect similarity ratings (Ward, 1983) it is perhaps reasonable to speculate that the reflections of the self report participants may differ from those in the relatively speeded condition.

Once again, it must be acknowledged that the protocols probably offer at best only a small window on the processes involved in making similarity judgements; there are likely to be more considerations taking place than participants report. However, in summary for the ad hoc item pairs, there is some evidence to suggest that participants found more differences than one might have expected when context was provided and more similarities than one might have predicted when no context was provided. The breakdown shown in Table 3.4 also suggests that participants are finding many similarities between these items that are not directly related to the context that the experimenter had in mind. The nature of these will be investigated later in this section.

Before moving on to look at protocols for the taxonomic item pairs, it is interesting to note that amongst the ad hoc item pairs, it was found that some contexts were more easily suggested to the participants by the item pairs than others. An examination of the 108 protocols in which the original context (or something very close to it, e.g., played with by dogs as opposed to chased by dogs) was instantiated in the 'no context' condition showed the following breakdown by category

Table 3.5

Number of protocols in which each ad hoc category was instantiated even when no context was provided.

things people take to a wedding	things that may be conveniently kept in pockets	things you find on a shelf	things you find in an attic	things dogs chase	things sold on the black market in Russia.
40	27	21	12	8	0
N=129	N=129	N=129	N=29	N=129	N=129

The ad hoc categories of *things people take to a wedding, things that may be conveniently kept in pockets* and *things you find on a shelf* were the most easily suggested by the items themselves.

Taxonomic item pairs

The same classification as reported above was also made for the taxonomic category item pairs as follows:

Table 3.6

Number and percentage of protocols for taxonomic category item pairs that instantiated the context even when it was not provided (no context condition) and mentioned the context when it was provided (with context condition).

NO CONTEXT			WITH CONTEXT		
Instantiated	Close	Not instantiated	Mentioned	Close	Not mentioned
314	42	158	252	38	236
61%	8%	31%	48%	7%	45%

As was the case for the ad hoc protocols, a multinomial logistic regression was performed on the data using the classifications as the dependent measure. This produced a significant result - $\chi^2(2)$ (N=1040) = 22.411, p <0.005. The pattern of classification differed with and without context, with the proportion of protocols instantiating the context when none was given being higher than those mentioning it when it was provided and the proportion of protocols failing to instantiate a context when one was not given being lower than the proportion failing to mention it when it was present. The pattern of protocol classification for each participant was examined as for the Ad hoc category results. It was found that in the 'no context' condition, 82% of participants showed the pattern reflected in table 3.6 where a significantly higher number of protocols instantiated the context (M = 7.3, s.d 2.45) than to failed to instantiate it (M = 3.67, s.d. = 2.63) t (42) = 4.78 p < 0.001. In the 'with context' condition, the mean number protocols that mentioned the context (M = 5.73, s.d = 2.91) was not significantly different to the mean number of protocols that did not mentioning it (M =5.36, s.d. =3.14) t (43) = -0.405 p = 0.687. 43% of participants in this condition produced more protocols that instantiated the context than protocols that did not whilst 57% produced protocols showing the reverse pattern. It is consequently concluded that the protocol pattern reported in table 3.6 is representative of the performance of participants and has not been unduly influenced by the data from a small subset of the sample

It can be seen that when the context was provided for taxonomic category items, it was mentioned and not mentioned in roughly an equal number of protocols. When no context was provided, the context was instantiated in 61% of cases. However, when it was provided, 45% of the protocols made no reference to it at all when explaining their judgements.

A further breakdown of the protocols was made as for the ad hoc category item pairs to see what proportion within each of the categories shown in Table 3.6 were mentioning similarities or differences between the items.

Table 3.7

Number and percentage of protocols, with and without context, mentioning only similarities between items, only differences or both similarities and differences.

	N	Similarities only	Differences only	Both	Other
No Context					
Instantiate it	314	112	1	179	22
	%	36%	1%	57%	7%
Do not instantiate it	158	46	13	57	42
	%	29%	8%	36%	26%
With Context					
Mention it	252	95	10	132	15
	%	38%	4%	52%	6%
Do not mention it	236	67	40	98	31
	%	28%	17%	42%	13%

A multinomial logistic regression was significant χ^2 N(6, 960) =112.401, p<0.005

The classifications were associated with both context χ^2 N(3) = 20.123 p<0.005 and reference to context χ^2 N(3) = 88.1 p<0.005. The effect arose mainly from

i) the lower than expected difference only protocols , where a context was instantiated even though it was not provided (0%)

ii) the higher than expected protocols listing both similarities and differences, where a context was instantiated even though it was not provided (57%).

The pattern of response was different with and without context.

Once again, the pattern of protocols produced by each participant was examined to see if the apparent differences between the 'similarities only' and 'difference only' protocols were the result of an extreme performance by only a small subset of the sample. In the 'no context' condition where the participants mentioned the context in

any case, 23% of the participants did not produce any 'similarity only' or 'difference only' protocols: 3% produced the same number of each protocol type:
72% produced more 'similarity only' than 'difference only' protocols: Only 2% showed the reverse pattern and participants produced significantly more 'similarity only' (2.60) than 'difference only' protocols. There was a significant difference overall between the mean number of similarity only and the mean number of difference only protocols (0.07) $t(42) = 5.77$, $p < 0.001$.

In the 'no context' condition where the participants did not instantiate the context, 33% of participants did not produce any 'similarity only' or 'difference only' protocols: 5% produced the same number of each: 49% of participants produced more 'similarity only' than 'difference only' whilst only 14% of participants showed the reverse pattern and there was a significant difference overall between the mean number of similarity only (M=1.07) and the mean number of difference only (M=0.30) protocols produced by participants in this condition $t(42) = 2.78$, $p = 0.008$

In the 'with context' condition, where participants did mention the context provided, 25% did not produce any 'similarity only' or 'difference only' protocols: 2% produced equal numbers of each: 66% of participants produced more 'similarity only' than 'difference only' whilst 7% participants produced the reverse pattern and there was a significant difference overall between the mean number of similarity only (M=2.14) and the mean number of difference only (M=0.45) protocols produced by participants, $t(43)$ 4.75 $p<0.001$

In the 'with context' condition, where participants did not mention the context provided, 25% did not produce any 'similarity only' or 'difference only' protocols: 6% produced an equal number of each: 41% produced more 'similarity only' than 'difference only' and 27% produced the reverse pattern but there was no significant difference overall between the mean number of similarity only (M=1.5) and the mean number of difference only protocols produced by Ps in this condition (M=0.93) $t(43) = 1.475$ $p = 0.148$.

With the exception of the final condition, the pattern of similarity only and difference only protocols seen on a participant basis reflects the pattern shown in table 3.4 on a protocol basis. In the 'with context but did not mention it' condition, more participants did produce protocols referring to similarities only than differences only but the failure of the difference in the means to achieve significance suggests that those participants showing the reverse pattern produced a relatively higher number of differences.

Recall that in this self report (protocol) condition, pairs of taxonomic items were rated as more similar without context than with context. It was speculated that this could have resulted from participants finding more similarities, or fewer differences between the items without context compared to with context. There seems to be some evidence from the analysis that there were fewer differences only listed by participants who saw the items without context but instantiated it anyway and that the same participants listed more similarities and differences than would have been expected. It is very difficult to know how the weighting of the similarities and differences in such protocols took place when deciding upon a rating.

It was suggested in the discussion of the similarity rating data that participants seeing the items without the category name, when allowed time to consider the judgement more carefully, as in the SR condition, instantiated a category context that was broader than the one envisaged (e.g., food rather than fruit) and consequently found the items more similar in this wider classification. An examination of the protocols for *fruit* was made to investigate this - there were 14 references to the property "edible" as a basis for similarity in the 'with context' condition compared to 37 references to the same property in the 'no context' condition. Although it is possible that participants made their comparisons against the background class of 'edible things' – this difference again does not seem remarkable enough to account for the effect found. Participants seemed to generate few differences in the 'no-context' conditions -whether or not they instantiated a context of their own.

Situational similarity

A further examination of the protocols provided in the 'no context' condition for ad hoc pairs was made to establish the nature of the similarities that were reported when the context intended by the experimenter was not instantiated. It will be recalled that there were 66% of these protocols that mentioned some point of similarity between the items. It was found that some protocols listed featural/functional similarities despite the largely perceptually dissimilar nature of the items (e.g., lipsticks and pens are both long and can be used to write with). There were, however, a number of examples of participants citing a connection or link between the items that was not based on standard feature matching at all. This included references to the items being linked together in terms of physical contiguity (e.g. where you find spiders you find dust) or a functional relationship (e.g. you might use screws to fix the telephone to the wall). There were also a number of protocols that referred to the items appearing in a common situation. For example, for phone and newspaper (things you would find on a shelf), one participant in the 'no context' condition wrote:

"A reporter for the newspaper might use a telephone to obtain some information for their story/report, Or over the telephone, someone may tell someone else their news" S4NCRP.

An examination of the total set of protocols for ad hoc item pairs was conducted and 142 protocols for ad hoc item pairs were found to contain these types of links or associations in the proportions for the 'context' and 'no context' conditions shown in Table 3.8

Table 3.8

Frequency of protocols for taxonomic and ad hoc item pairs in both conditions of context showing thematic links (non-featural considerations)

		Thematic link	No thematic link	Total
With context				
Context mentioned	Taxonomic	11	151	252
Context not mentioned	Taxonomic	18	218	236
Context mentioned	Ad hoc	5	239	244
Context not mentioned	Ad hoc	34	479	513
No Context				
Context instantiated	Taxonomic	7	307	314
Context not instantiated	Taxonomic	21	137	158
Context instantiated	Ad hoc	3	76	79
Context not instantiated	Ad hoc	100	565	665

These thematic link protocols broke down by category as shown in Table 3.9

Table 3.9
Breakdown of protocols for ad hoc item pairs in the 'no context' condition, which showed evidence of situational links between items other than the category itself.

Ad hoc category	Number of situational protocols (N=142)
things dogs chase	29
things sold on the black market in Russia.	28
things that may be conveniently kept in pocket	30
things you find in an attic	28
things you find on a shelf	18
things people take to a wedding	9
Total	142

It is interesting to note that the two categories giving rise to the highest number of protocols that linked the items through a common scenario were those in which the context envisaged by the experiment was rarely instantiated (see Table 3.5), i.e., *things dogs chase* and *things sold on the black market in Russia*. It appears that the discovery of common situations was most likely to occur when no context was provided for perceptually different ad hoc items and the experimenter's context was not instantiated. It should be noted that the ad hoc category names themselves constitute situations and that similarity based on membership of this category (they are both kept on shelves) can be seen as thematic. Taking the "context instantiated" protocols from the 'no context' condition (n = 108) together with the "context mentioned" protocols from the 'with context' condition (n = 277) and the set of protocols mentioned in Table 3.9 that refer to situational links other than these expected contexts (n = 142), it can be argued that drawing on knowledge of shared situations in assessing similarity is not a rare occurrence (total n = 527 from a total of 1563 protocols).

Perhaps surprisingly, there was a similar proportion of thematic links between items mentioned in the protocols for taxonomic category pairs.

As in the case of ad hoc item pairs, the highest number of protocols showing links was in the 'no context' condition when the context was not instantiated in the protocols. However, whereas these links occurred in more than one of the ad hoc categories, it should also be noted that all the cases of thematic similarity occurred for the category *furniture*, for example:

Chair/Piano

" They are similar because you need a chair/something to sit on if you want to play a piano. Both made of wood" SINCCP rating (6)

Cushion/Footstool

" My first thought was that they weren't related at all but then I realised that you might put a cushion on a footstool so I circled a number a little higher than 1" SIWCRP rating (3)

Thematic differences

It was also noted that some protocols contained references to thematic difference between items, for example for the pair cat /stick (things that dogs chase)

"Cat is wily and unpredictable, stick has quite a predictable trajectory of flight"

......and for cat/ball (same category)

"Cat's may retaliate, ball won't"

It will be argued later that such protocols run counter to the notion that participants are merely giving expression to rather non-specific feelings of association when they evaluate thematic similarity.

Personal references

The protocols were examined for evidence of specific personal references that would suggest that participants were not just consulting their general knowledge of objects and situations but were thinking about their own objects. There were found to be no such personal references for the taxonomic category pairs and a total of 25 such references in the protocols for ad hoc item pairs which broke down as follows: 14/236 in the 'with context' condition (context mentioned): 0/453 in with context condition (context not mentioned): 4/78 in 'no context' condition (context instantiated) and 7/558 in 'no context' condition (context not instantiated).

Individual Differences

The analysis of the protocols reported in the previous sections suggests that when participants are asked to judge similarity between items they engage in a wider process than feature matching- often noting thematic connections between the items. However, whilst classifying the protocols, it was noted that there seemed to be considerable individual difference in participants' willingness to take thematic similarity into consideration when making their rating, even for the same items and the same situation. Consider the following protocols which were both provided with ratings for Money and Wallet (things that can conveniently be kept in pockets) with no context:

"Money and a wallet are quite similar because they are connected, i.e., you put your money in a wallet" S1NCCP rating (7)

" Money is what goes inside the wallet but it doesn't make them similar in any way. They 'go together' but are not alike" S11 NCRP rating (1)

The following protocols were for items from *things you find on a shelf*, 'with context' condition:

Newspaper/Telephone

"both may be found on a shelf. Because they are both found on a shelf they are therefore similar" S6WCCP rating (5)

Photograph/Clock

"So what if you find the two on a shelf, what makes them similar? Nothing! S3WCCP rating (1)

It seems that even when individuals notice a thematic link between the items, they are not always prepared to count it as the relevant kind of similarity. These protocols also indicate an awareness of differing sources of similarity.

It appears that individual differences exist in a) a tendency to perceive or report these links whilst undertaking this type of task and b) once perceived, a willingness to be influenced by such links in making a rating of similarity. Only 6/43 subjects in the 'no context' condition reported no links of this nature for Ad hoc items, but only 4 participants reported them in a third or more of their protocols (total of 18 protocols) with the highest percentage being 50% for just one participant. Most subjects reported such links in just one or two protocols for these items under this condition.

Of the four participants who made most use of these links, only one produced relatively high ratings for the items in question (mean rating 6.3) whereas the other three seemed comparatively uninfluenced in their ratings (mean ratings 1.0, 2.9 and 2.9). These three participants tended not to offer any other similarity basis for these items. The participant with the higher ratings produced longer protocols that elaborated the link in more detail.

The Relationship Between Protocols and Ratings

The content of the protocols can only be taken as revealing of the processes involved (even to a partial extent) in making similarity judgements if one believes that they are indeed the considerations made by the participants. A minimal requirement in establishing their validity must be that the reported considerations should at least be reflected in the ratings that they purport to explain. The following analysis attempted to address this issue.

A mean similarity rating was calculated for the protocols classified in the following way (see Table 3.10):

Sims only - protocol indicated only features the items shared and/or thematic links

Diffs only - the protocol indicated only features/properties that the items did not share.
Both – the protocol listed both properties/features that the items shared and others that they did not share
NI- the protocol was non-informative in as much as it was not possible to know what the participant was taking into consideration
Dclaim - the protocol consisted only of a claim that the items are totally different or not similar at all.
Don't know an item - the protocol indicated that the participant only had a vague idea about what one of the items was like.
No protocol - the participant had failed to write a protocol
SDclaim - the protocol consisted only of a claim that initially indicated a basis for similarity (they both…) but then went on to say that this did not make them similar
ClaimS - the protocol initially indicated that the participant thought that the items were totally different but then went on to find some similarity.
Can be other – the protocol consisted only of a claim that one item could be the other (e.g. a clock can be an ornament)

Table 3.10

Mean similarity ratings and standard deviations for the various classifications of protocols – ordered by similarity.

Protocol type	Mean	s.d.	N
Sims only	5.63	2.20	723
Both	4.77	2.12	965
Don't know an item	4.58	2.02	24
Can be other	4.32	2.28	22
Non-informative	3.18	2.10	175
SD claim	2.78	1.65	101
ClaimS	2.59	1.21	39
Diffs only	1.96	1.34	417
No protocol	1.31	.87	35
DClaim	1.24	.81	102
TOTAL			2603

It can be seen that this classification contains the three categories used in Tables 3.4 and 3.7 but also attempts to classify the 'other' category more finely. This latter classification was conducted by one judge only and is included only for information as to what other types of considerations constituted the 'other' category used earlier in this section. Since for this finer classification inter-rater reliability was not established and since it also gives rise to categories of very small size, all classifications apart from Sims only, Diffs only and Both were collapsed to give just four categories of protocols. The means for these can be seen in Table 3.11.

Table 3.11
Mean similarity ratings and standard deviations (in brackets) associated with protocols in each of the four categories.

Sims only (n=723)	Diffs only (n=417)	Both (n=965)	Other (n=482)
5.63 (2.20)	1.96 (1.34)	4.77 (2.12)	2.59 (2.45)

A one-way repeated measures Anova showed that there was a significant difference amongst these four means, Wilks Lambda $F\ (3,55) = 160.406$, $p < 0.01$, eta squared = 0.90. The ratings accompanied by protocols listing similarities only were significantly higher than the ratings against which participants had listed only differences between the items, $F\ (1,57) = 482.040$, $p<0.001$, eta squared = 0.894. The ratings for the protocols indicating both similarities and differences lay between the two values, somewhat nearer to, but still significantly different from the similarity ratings for similarities only $F, (1,57) = 77.963$, $p< 0.01$ eta squared = 0.578. This initial analysis gives cause for some confidence that the ratings were reflecting the reported considerations of the participants.

As a further test of the relationship between the protocols and the ratings, it was reasoned that the higher the percentage of similarity only protocols produced by a participant, the higher the overall mean rating should be. The percentage of each type of protocol (Sims only, Diffs only, Both and Other) was calculated for each participant.

The overall pattern of protocol types across all participants can be seen in Table 3.12.

Table 3.12

Mean percentage of Sims only, Diffs only, Both and Other protocols

Sims only	Diffs only	Both	Other
27.8	16.00	37.1	19.1

The percentage of similarity only protocols was found to be positively correlated with the overall mean rating, r (86) = 0.290, p= 0.0035, 1 tailed. The percentage of difference only protocols was found to be significantly negatively correlated with the overall mean ratings, r (86) = - 0.247, p = 0.011, 1 tailed.

The calculation of the percentage of each type of protocol produced by each participant enabled a mean profile to be produced across participants for both conditions of context and according to whether the context was instantiated or not, for both category types. This is illustrated in Figure 3.2.

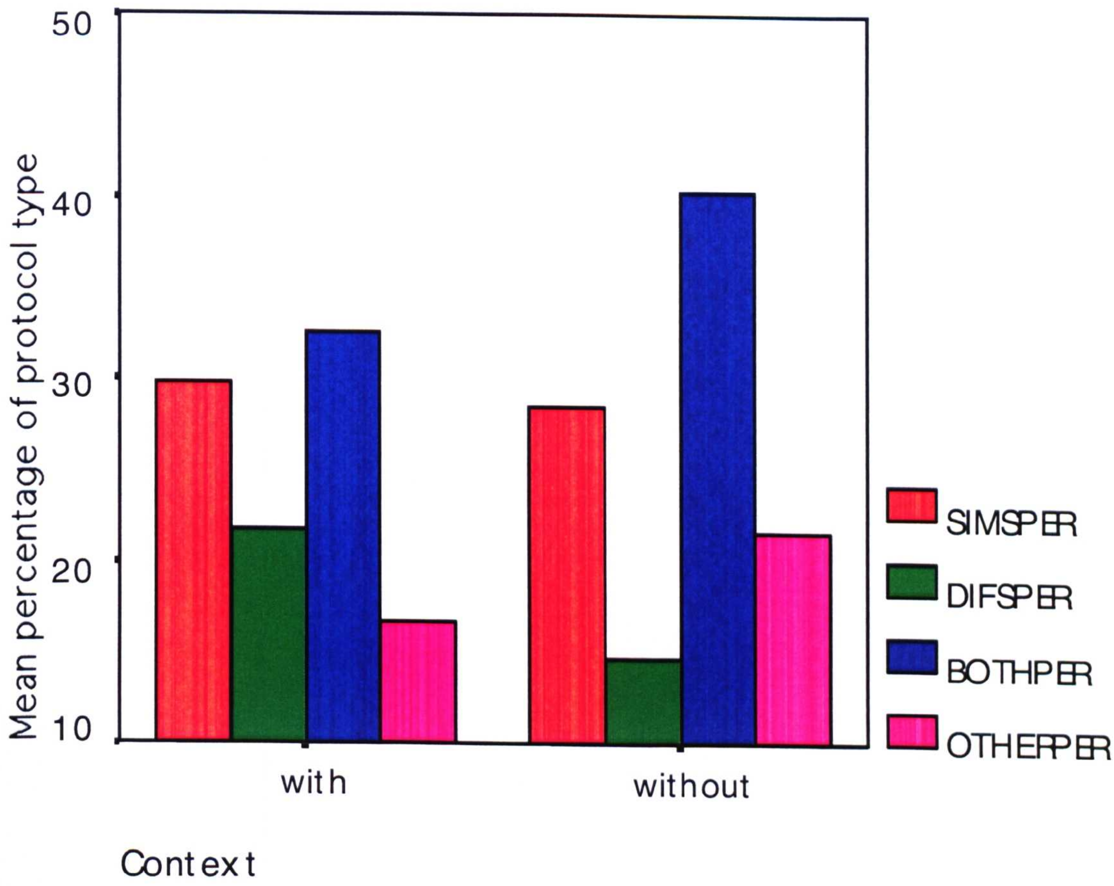


Figure 3.2 Mean percentage of each protocol type across participants with and without context

Despite the apparent variation in pattern of response across the context conditions, this was not found to be significant using this participant profile, $\chi^2 (3) = 1.074$, $p = 0.783$.

The final question to be addressed was whether the protocols that included thematic similarity resulted in higher ratings than did those without such considerations. It has already been argued that when participants in the 'no context' condition instantiated the ad hoc categories (or close to it), they were making thematic links between the perceptually dissimilar items. For example, when a participant says of dust and boxes that both can be found in attics, then they are forming a thematic association.

any other type of similarity the analysis is limited to 'similarity only' protocols for ad hoc category members where no context was provided.

Table 3.13
Mean ratings and standard deviations from ad hoc protocols (no context) referring to a thematic context and protocols making no such reference.

	Thematic context	No thematic context
Ad hoc	5.26 (2.04)	4.45 (2.12)

The ratings from the protocols where participants instantiated a thematic context for themselves when comparing ad hoc category members were significantly higher than the ratings from the protocols that showed no such context, $t\ (184) = 2.272$, $p=0.024$. Interestingly, if one adds to the thematic context group those protocols that contained novel thematic links, the mean rating for the group decreases ($M= 4.94$, $sd =2.15$) and the difference only approaches significance, $t\ (184) = 1.569$, $p = 0.059$. This suggests that the ratings given by participants for ad hoc items where there was no context provided, but the participant reported considering at thematic link between the items other than the original context (e.g. drugs and guns can both kill you) were somewhat less influenced by this link than the when the context instantiated was more specific.

General Discussion

The protocols collected for this study provided a rich source of qualitative data concerning how participants claim to have decided upon a scaled rating of similarity between items. A future analysis could attempt to catalogue the actual nature of the properties mentioned as being both shared and unshared, and differences could be classified as either alignable or non-alignable. For the purposes of this thesis, the analysis of the protocols was undertaken with two aims in mind:

- To see if they could help explain the quantitative data relating to the self-report condition – in particular the failure of context to enhance the ratings of the ad hoc item pairs which was evident in the NSR conditions.
- To look for evidence that knowledge of situations in which the items commonly occur was playing a part in the process of judging similarity

The success, or otherwise, in achieving these aims will be discussed in turn.

Explaining the Similarity Ratings

The first classification of the protocols was into those that instantiated a context and those that did not (when context was not provided) and those which made reference to a given context and those that did not (when context was provided). The data from the ratings showed that the provision of context for ad hoc items raised ratings in the no self report condition but in the self report condition the ratings appeared to be at much the same level as when context was not provided. It was speculated that either the effect of context was weakened by participants having more time in which to instantiate their own context when none was provided, or that the context was examined more critically when it was provided. There was some evidence that when faced with perceptually different items, a context was instantiated. Fourteen percent of the protocols in the 'no context' condition made reference to the very context provided in the context condition. Perhaps more importantly, 65% of the protocols where context was provided made no reference whatsoever to that context, suggesting that it did not serve to guide their judgements at all (see Table 3.3).

When the protocols were further classified into those that listed similarities only, differences only, both similarities and differences or some other expression altogether, it could be seen that the percentage of similarity only protocols was much the same overall for the 'with context' and 'no context' conditions. The highest percentage of difference only protocols was seen when context was provided but not mentioned and was lowest when context was not provided but was instantiated anyway by the participant. This suggests that there was an increased focus on differences in the 'with context' condition compared to 'no context' condition. It must be re-emphasised that the items in the ad hoc pairings were items that would not be expected to be viewed as similar at all under a traditional feature-matching model of similarity. Although all the ad hoc pairs had a mean similarity of less than 5, there was clearly felt to be some degree of similarity. The prevalence of similarities in the protocols for these items is a testament to participants' efforts to find similarity when faced with this task. It should also be noted that approximately a third of the protocols indicated a process of weighing of similarities and differences as implied by the Tversky (1977) model of similarity.

A similar breakdown for the taxonomic items failed to explain why the ratings for these items were higher under the self-report condition without context than with context. There was some suggestion that, by protocol, a greater proportion of differences were listed when context was provided but not mentioned; however, it is not clear that this in itself would explain the ratings for these items. A more detailed analysis of the specific properties mentioned for the items may assist in understanding the quantitative data but since this was not the main focus of the protocol collection, it was not deemed appropriate at this time.

Gricean Pragmatics

As was noted earlier (page142) the variability in the ratings (as indicated by the standard deviations) was notably less for all both category types and all pairtypes when no context was provided and there was no self-report. It seems then that there is the highest degree of consensus around the mean ratings when judgements are made fairly quickly and with no need to justify them and when the items are presented without any other information (i.e., the category label). This may reflect the various ways in which participants sought to make sense of the task with which they were presented. As previously mentioned, although participants seem to have little difficulty in coming to judgements about similarity between often unusual pairings of items – the requirement to do so, and to place a numerical rating on that judgement, is likely to occur only in experimental settings such as these. The task appears to have no goal other than to satisfy the (unspecified) requirements of the experimenter. Since the experimental setting can be seen as an interaction between the Experimenter and the participant, then the Gricean (1975) maxims of conversation may be expected to operate and participants are likely to seek to maximise the use of any information that is provided to them. In the case of this study, that information came in the form of the category label but also in the form of the protocol requirement. It is perhaps therefore somewhat surprising that less than 50% of the protocols for either the taxonomic or ad hoc pairings made reference to those labels. This is particularly so for the ad hoc pairings. The presentation of *Fruit* above the pairing of *orange* and *banana* may simply be understood as orienting the participant to the stimuli (acting almost as the word 'ready' in a lexical decision task). However, the ad hoc categories were perhaps less predictable and in some cases very unusual (e.g. Things sold on the Black market in Russia) and therefore one might have predicted that far more protocols would have mentioned this heading in their protocols. The comparatively high variability in ratings seen when protocols were required and when context was provided (even without protocols) suggest participants made use of this information in different ways. Although there were overall patterns of context and protocol effects, there was not a tight consensus amongst the samples on the way in which this affected the ratings that they decided upon. In the case of ad hoc items, it has been suggested that this, to a certain extent may reflect differences in individual susceptibility to note, and be influenced by, thematic relations between items but it is harder to explain the variability around ratings for taxonomic item pairs that seems to occur in all conditions except the relatively speeded, no context condition. The rather superfluous provision of the taxonomic category labels may act to 'confuse' participants wondering how to make use of it – this may also be the case when a protocol is required but a label is not, given the rather obvious perceptual basis for similarity that presents itself between two taxonomic. What does seem clear from the content of the protocols is that in a spirit of co-operation, most participants expend considerable effort in seeking similarity, as they seem to presume that this will meet the criteria for a successful interaction from the perspective of the experimenter.

Influence of Situational Knowledge

The main focus of the data collection was to see if participants who had been asked to judge similarity between items engaged in a much wider process than feature matching. Specifically, it was hypothesised that the type of situational knowledge exhibited in the strategies for the category member free-recall task reported in Chapter 2 would also influence the deliberations of the participants in the similarity task.

It could be considered that the taxonomic item pairs in the studies reported in this chapter generally correspond to the item pairs in Wisniewski and Bassok (1999) that shared attributes but no thematic association (A+T-) whilst the ad hoc items are more like A-T+ pairs, i.e., pairs of items that share few attributes but may share thematic links. The higher ratings of similarity for the taxonomic versus the ad hoc pairs overall would therefore be consistent with the Wisniewski and Bassok findings.

Since there was no systematic manipulation of items expected to evoke thematic similarity in this study, it is contended that it acts as a strong demonstration of the extent to which this type of similarity is spontaneously produced. It is true that in this study, a subgroup of the items provided to the 'no context' participants had been previously generated from an ad hoc category. In retrospect, it may be considered that the ad hoc category could, in itself, be seen as constituting a scenario/situation and that meant that these items were, á priori, thematically related. However, it could also be argued that the nature of some of these thematic associations (kept in a pocket, found on a shelf) would not have been expected to be as compelling as that between the items used in Wisniewski and Bassok's study (car and tow-truck, chisel and hammer). Some protocols in this study also exhibited a high level of creativity in establishing thematic links between dissimilar items.

During the protocol analysis, it became apparent that particular participants seemed to be predisposed to noticing or generating thematic similarity. Dunham and Dunham (1995) posited individual differences in perceptual-cognitive styles that manifest themselves in early conceptual development (3 yr. olds). A subsection of their participants showed a persistent thematic approach to categorisation tasks, the antecedents for which could be found in the use of a selective interest in relations among objects and the use of certain relational terms at earlier ages. The degree to which adults tend to make use of thematic organisation in such tasks as category –member generation, similarity judgements and categorisation (Lin and Murphy, 2001) may vary considerably between participants but may be a relatively stable characteristic across different tasks.

Interestingly, the intrusion of thematic relations into tasks that 'should' be based on taxonomic relations has been viewed by some as resulting solely from participants' misunderstanding of the instructions for the task. However, a variety of instruction manipulations have failed to eradicate the influence of thematic associations (Wisniewski and Bassok, 1999; Lin and Murphy 2001). Instead, Gentner and Brem (1999) have claimed that the discovery of these influences do not demand a radical review of similarity theories as suggested by Bassok and Medin (1997) and Wisniewski and Bassok (1999). Gentner and Brem proposed the confusability account in which thematic influences are not an essential part of the similarity process but that they influence the process when "participants have difficulty distinguishing between the mental output that arises from accessing associations as opposed to the output of a separate, independent comparison process" (p.180). The screening task of Gentner and Brem's second experiment required 702 participants to look at triads of object names that consisted of a target (e.g., dart) and two other items (e.g., bullseye and javelin) and asked to pick the one of the two that was most similar to the target. As a result of repeated trials they identified a group of participants (11%) who were "unable to distinguish similarity from association in over 90% of cases". A further group of 41% were 'correct' in over 90% of cases and the remainder of the participants was intermediate in performance. This is mentioned in connection with the protocol analysis reported here for two reasons. Firstly, it supports the notion of individual differences in predisposition to make use of thematic relations in "taxonomic" tasks. More importantly, the nature of a number of the protocols produced by the participants in the present study indicate a clear awareness of the existence of more than one 'type' of similarity and their inclination to take account of this in their ratings varies. Some participants clearly felt that the thematic relation was not relevant to the question of similarity whereas for others, the term similar clearly encompassed these relationships. There is little in the protocols collected here to support Gentner and Brem's claim that the participants were unable to distinguish between the two. Consider the following examples:

S9 No Context CP Wedding T/T Camera/Confetti rating 2
I think of a wedding and see that they both fit at a wedding and so they are similar but as objects themselves they are not.

S1 No Context RP Black Market A/A Magazine/Fashion Accessories rating 4
In terms of the materials they can be made can be quite different, however fashion accessories can be found in magazines.

S2 No Context RP Shelf A/A Book/Spider rating 2
They are not similar but are similar because it could be a book about spiders.

S9 No Context CP Wedding T/A Present/Hat rating 2
They both appear at weddings but the objects are completely different.

S2 No Context CP Attic T/A Dust/Spider rating 1
When you think of spiders you think of dust because spiders are usually found in dusty, dirty cellars and attics, you think of a spider web surrounded by dust but they are not very similar at all.

Summary of protocol analysis

The relationship between the nature of the protocols as classified (similarities only, differences only or both) and the ratings gives rise to a certain amount of confidence that the protocols are indeed informative of the process involved in the judgement of similarity between items under these experimental conditions. The protocols suggest that when people make direct assessments of similarity between items, with no specified goal other than that they have been asked to do so, they engage in a feature matching process but also generate other types of situational links between the items that subsequently influence ratings of similarity.

The discovery of the thematic similarity in the protocols may have been viewed with some caution were it not for the results of the studies in Chapter 2 and the subsequent publications by Bassok and Medin (1997) and Wisniewski and Bassok (1999). It can not be argued that the thematic similarity resulted purely from requiring protocols from the participants since Wisniewski and Bassok found no effect of protocol upon the ratings of items that were thematically associated. This suggests that the thematic integration of items took place even in the absence of self-report. However, in this study an interaction of self-report with context was found so self-report does seem to affect ratings under particular conditions. The low occurrence of reported thematic links in the protocols for the taxonomic pairing could be interpreted as suggesting that taxonomically based similarity may be preferred over thematic based similarity, where it can be utilised. However, it may also be the case that, as discussed in the introduction, the taxonomic category labels did not themselves suggest situations or contexts for the comparison in the same way that the ad hoc labels did and so may not have evoked situational links between the items.

The protocols did not show the same level of self-reference that was witnessed in the protocols from the category-member generation task but the similarity judgements did seem to draw on the same generic knowledge of situations/events in which the items can occur. It is the co-occurrence of the items in these contexts that seems to form a basis for thematic similarity.

Participants appeared to be aware of the difference between intrinsic feature-based comparison and thematic association and for some people these are considered simply to be different, but equally legitimate, sources of similarity. It has recently been found that educated adults can demonstrate a hitherto unexpected tendency to form categories on the basis of thematic relationships as opposed to taxonomic relations (Lin and Murphy, 2001; Murphy, 2001). If such relationships are seen as a valid basis of categorisation, and items that are categorised together are believed to be similar, then it seems reasonable that the same type of relationship is viewed as constituting similarity.

Both Bassok and Medin (1997) and Wisniewski and Bassok (1999) consider that 'abstract' attributes such as the thematic relation between items in familiar situations could be encompassed by traditional similarity models as another type of feature that serves to enhance or decrease similarity. It may be argued that participants are simply switching between the sharing of physical/functional properties and the sharing of situational scripts when they consider the respects in which items are seen as similar. However, the question remains as to why the situational scripts should exercise such an influence in adult cognition when it was previously thought that this organisation of items on the basis of how they co-occur in situations was a rather immature organisation, seen primarily in young infants (Markman, 1989) and believed to be superseded in development by a more sophisticated taxonomic organisation. This issue will be considered in detail in the concluding chapter.

As discussed in Chapter 2, it must be conceded that the content of the protocols can not be assumed to be providing a full or accurate description of the cognitive processes involved in completing the task set for participants. It is undoubtedly the case that many other considerations entered into their judgements other than those that they chose to report. This may account for the limited success in using the protocols to explain the effect of context upon the similarity ratings.

Summary of Results

The new factors included in the design of this study revealed the following:

- The effect of context on the ratings for ad hoc pairings shown by Barsalou (1982) was found only if the participants made unreflective judgements that they were not asked to explain or justify. The effect also seemed to apply most clearly to the pairs of typical items.

- An interaction of self-report and context suggested that the effects of context were influenced by time and effort.

- The content of the protocols suggests that context lead to a higher focus on differences between items.

- Protocols revealed substantial consideration of situation –based associations between items in addition to more standard feature matching.

- Where participants reported spontaneously instantiating the same context for ad hoc item pairs as the one from which they had been originally generated, similarity ratings were higher than when they did not make such a link between items. The situational similarity did, in these cases, therefore appear to increase the perceived similarity of the items to each other, as reflected by the ratings.

It also appeared from the protocols that the thematic link between some item pairs gave rise to a stronger feeling of similarity compared to other thematically associated pairs. The next chapter reports a study that set out to investigate the factors that affect how compelling participants find thematic similarity between perceptually dissimilar items.

CHAPTER FOUR

Factors that Determine Thematic Similarity

Overview

This chapter reports an investigation on whether thematic similarity depends upon the association between the items under consideration and familiar situations in which they commonly feature. Both item-to-event and event-to-item association are measured.

Background

In Chapter 3, the prevalence of 'thematic similarity' amongst the justifications for similarity ratings was demonstrated and discussed. There seems to be general agreement that this type of similarity arises when people think of items being 'linked' by virtue of appearing together in the same situation or event (Wisniewski and Bassok, 1999). It also seems to be clear that certain pairs of items are more likely to elicit feelings of such similarity than are others. In Wisniewski and Bassok (1999, exp.1), only half of the item pairs that shared both features and were thematically related were rated as more similar than the item pairs that shared features only. This suggests that some of the thematic relations did not enhance the similarity ratings. It seems that items vary in their thematic similarity in the same way that they vary in terms of their feature-based similarity. Whilst it is, arguably, easy to see why a sparrow and a robin may be viewed as more similar to each other than a penguin and a sparrow – it is perhaps harder to establish why a dog and bone may be viewed as more similar than a bride and a church. If, as has been suggested here, people draw on an alternative organisation of knowledge based on the situations in which items tend to be encountered, one might expect that the strength of association between individual items and these situations would be a factor in determining the relative thematic similarity between items which feature in the same situation. The studies reported in this chapter sought to explore this type of similarity.

The idea that contiguity and similarity are the two basic laws of association, is of long standing and has been attributed to, amongst others, J. S. Mill (1843/1930) who identified a law of similarity - "similar ideas tend to excite one another" - and of contiguity "when two impressions have been frequently experienced (or even thought of) either simultaneously or in immediate succession, then whenever either of these impressions or the idea of it recurs, it tends to excite the idea of the other". Lucariello and Nelson (1985) claim in their introduction, " Since Hume, two bases of association have been recognised, similarity and contiguity in space-time". These authors go on to say that "these are reflected in the presumed bases for category formation and formation of thematic groups (e.g. cup and milk) respectively"

There has been a tendency in cognitive psychology to separate similarity from association rather than to see similarity as a form of association. To understand thematic similarity it is helpful to see perceptual/functional similarity as only one, albeit important, way in which items may be associated together in thought. It seems that perceptually and functionally associated items form 'traditional' categories based on shared features. However, items can be associated in a variety of other ways that may also influence how we think about them.

If items are contiguous then they are adjoining, co-occurring or are in contact with each other. Items may therefore be spatially contiguous (literally occurring together in the same space) or temporally contiguous (occur together at the same point in time). The former clearly entails the latter but items, which are temporally contiguous, may vary in their degree of spatial contiguity. These types of contiguity should therefore be considered separately.

As mentioned in chapter 3, Lin and Murphy (2001) have reported a programme of work that is the most extensive to date to tackle thematic relations in adults' concepts. These authors conducted a series of experiments that showed that when adults were asked to choose which of two items would best form a category with a target item, a substantial proportion of choices were for the thematically linked option rather than the taxonomically matched item. This finding persisted across a range of presentation formats and instructions. Lin and Murphy explain thematic relations in the following manner " thematic relations are the external or complementary relations among objects, events, people and other entities that co-occur or interact together in space or time" . They proceeded to give examples of "some" thematic relations.

Spatial	– a roof is on top of a house
Functional	– chalk is used to write on a blackboard
Causal	– electricity causes a light bulb to glow
Temporal	– bills typically come after meals in restaurants.

This list is clearly not exhaustive. One could add 'episodic' relations, i.e., co-occurring in scripts or stories. One of the difficulties in attempting to unravel the nature of thematic similarities is the diverse and overlapping ways in which relations between entities may be described. Lin and Murphy point out that items may share more than one type of thematic relation (e.g. causal and spatial). This said, it seems that the four candidates referred to by Lin and Murphy would be relatively easy to identify and have been alluded to elsewhere in the literature. For example, Markman (1981) notes that " In naturally occurring situations, objects exist in spatial, temporal and causal contexts" and she also claims that "event-like meaningful structures might be a more spontaneous, natural way of organising information" (p.203). For the purposes of her thesis, Markman distinguishes between paradigmatic and syntagmatic associations where the former are based on similarity or inclusion and the latter arise from " some causal, temporal or thematic relation". (p.229) Note here that thematic is cited separately as a form of association from temporal or causal. Lucariello and Nelson (1985) also make reference to functional/syntagmatic relations as being those "things that appear together in the same time-space and bear a functional relation to one another, for example, cup and milk."

Wisniewski and Bassok's (1999) work is the only other paper to systematically examine the effect of thematic relations between items on similarity ratings. These authors assume that thematic groupings are considered to be plausible because they go together as " a unified scene or event". They go on to say that entities in thematic relations tend to each play a different role in that scene or event and therefore exhibit different role-appropriate properties which in turns makes them unalignable. That is to say they do not present themselves as suitable for comparison along shared dimensions. However, an examination of the item pairs used in their study shows that they all share a 'direct' relationship, i.e., one that brings them into direct contact with each other, such as car and towtruck, chisel and sculpture, ship and lifeboat. Seven out of 12 pairs in Wisniewski and Bassok's A-T+ category (pairs of items that do not share standard features but are thematically related) comprise agents who have use of, or contact with, the base object, e.g., cat and vet, chair and carpenter, car and mechanic. Other relations seem to be one of dependency through joint use e.g., hammer and chisel, kettle and cup. The nature of these relationships automatically suggests a scenario, though not necessarily a well-rehearsed event of the type to be used in the following study. This is also true of the materials used by Lin and Murphy.

There do not appear to be any item pairs in the Wisniewski and Bassok work that are thematically related only by virtue of playing some role in the same event or scenario – they all share a further relationship. Most of the item pairs in the current study, however, will be related purely by "theme" (e.g. *present* and *turkey*). This will then, in itself, constitute a considerable difference between this study and the previous work.

The Thematic Similarity Studies

As described above, there are a number of types of relationships that may exist between items and it is likely that some, if not all, of these affect the degree to which people are prepared to say that the items are similar to each other, even in the absence of shared perceptual and/or functional features. In the work reported here, it was decided to focus on the link that each item in the pair enjoys with a specified event/situation rather than on the direct link between the items. If thematic similarity reflects a feeling of "closeness" between the items arising from an organisation of knowledge around situations, then the extent to which each item invokes that situation, or is invoked by it, should predict the strength of thematic similarity as shown through ratings. It was necessary therefore to establish some measure or index of item-to-event association and event-to-item association. Both measures would be required since the association between an item and an event is not symmetrical. For example, the item 'confetti' may not be strongly elicited by the event 'wedding' since there are many other items that may be thought of first in connection with this situation (church, bride, vicar, ring etc). However, the presentation of the item 'confetti' is likely to reliably produce 'wedding' as the only associated event. It was a possibility that one "direction" of association would prove to be a better predictor of thematic similarity than the other- especially if item pairs are presented for a similarity judgement in the absence of context.

There are in existence a number of word association databases, but these did not result from a data collection that specifically sought to elicit the relationship between items and the events/situations in which they are regularly encountered. It was therefore determined that the putative explanatory factors of item-to-event and event-to-item strength of association should be measured by collecting data from similar samples to those who would be asked to provide the similarity ratings. There were consequently three stages of data collection. In stage 1 (eliciting item-to-event data) participants were asked to generate items that they associated with 15 different events/situations. The frequency of generation was taken to be a measure of association between the event and the item. In stage 2 (eliciting event-to-item data), a subset of the items produced in stage 1 were presented to a new group of participants and they were asked to list events/situations that they thought about in connection with that item. Once again, frequency of generation of the target event was taken as an indication of the strength of association between the item and the event. In stage 3, the items were paired together according to a manipulation based on the measures obtained in stage 1 and 2 and were presented to a third set of participants who were required to rate the pairs for similarity.

Stage 1: Eliciting Item-to-Event Data

Method

Participants

Twenty-eight cognitive science undergraduates (8 males and 20 females) took part in the pilot study for course credit. Four males were aged under 21 and 4 over 21. Eight females were under 21 and 10 were over 21. Two female participants failed to report their age group.

Materials

Fifteen situation/events were chosen by the experimenter (see Table 4.1) The events were such that most participants could be reasonably expected to have experienced them to varying degrees. An attempt was made to include both very specific cultural events (*Christmas Day*) to more generic situations (*getting ready for a special occasion*).

Table 4.1.

Events Used to Elicit Items.

A Wedding	Looking for Something in the Attic
Visit to the Dentist	Packing the Car to go on Holiday
A Day at the Beach	Preparing a Picnic
Having a Shower	Rescue at Sea
A Camping Trip	A Children's Party
A Trip to the Supermarket	Making a Cup of Coffee
Seeing a Film at the Cinema	Getting Ready for a Special Occasion
Christmas Day	

The stimuli were presented in a 16-page booklet of A4 typed sheets. The front page provided instructions and on each of the following pages the name of the event appeared in BOLD on the top of the page with the rest of the page lined to encourage listing. At the bottom of each page appeared the statement I **have been in this**

situation followed by the tick box options of Never, less than 10 times, more than 10 times, more than 50 times. This was included to provide some measure of the extent to which participants had personal experience of the events. The order of the stimuli was randomised for each participant (see Appendix J for instructions and sample materials).

Procedure

All participants provided data for all 15 events. The data were collected from all participants in one session. The booklets were distributed and the instructions shown on the front were read aloud by the experimenter as follows.

> **When you think of an event or situation, there are a variety of items that may come to mind. For example, when I think of a lecture I think of a whiteboard, an overhead projector, students, handouts etc. At the top of each of the following pages you will see the name of an event or situation. For each page you will be allowed 90 seconds to list as many items as you can which the event or situation bring to mind. At the bottom of each page I would like you to indicate how often you have been in this situation by ticking one of the following options:**
>
> **Never/less than 10 times/more than 10 times/more than 50 times**

You will be told when to start each page and when to turn over and start the next page. Please do not start each page until you are told to do so.

These written instructions were supplemented with the verbal instruction that in completing the question at the bottom of each page, participants were to understand that "more than 10 times" included "but less than 50 times". Participants were given the opportunity to ask questions before starting. Questions were raised about what sort of items would be appropriate and participants were encouraged to write down any sort of thing that came to mind when they thought about the event. A time limit of

90 seconds per event was imposed to limit the length of the session. This particular time constraint was used at it seemed, in Studies 1 and 2, to have proven to be sufficient time to allow for substantial data generation but not so long as to induce restlessness in participants. The procedure lasted 30 minutes in total.

Results

Fluency

An overall mean of 13.13 items were produced per event. The most items produced for any one event was 17.07 (*a day at the beach*) and least was 10.07 (*getting ready for a special occasion*). Across events, the highest mean number of items produced by any one participant was 17.73 items and the lowest mean number of items generated was 7.67. The mean number of items generated to each event can be seen in Table 4.2.

Table 4.2

Mean and standard deviation of number of items generated to each event.

Event	Mean	Standard deviation
A Day at the beach	17.07	4.32
A Wedding	15.64	4.33
A Camping Trip	14.71	4.40
A Trip to the Supermarket	14.43	4.61
A Children's Party	14.32	3.70
Christmas Day	13.50	4.00
Visit to the dentist	13.36	4.19
Preparing a Picnic	12.61	3.26
Seeing a Film at the Cinema	12.54	3.71
Having a Shower	12.29	3.58
Making a Cup of Coffee	12.29	3.73
Looking for Something in the Attic	11.86	4.37
Rescue at Sea	11.75	3.45
Packing the Car to go on Holiday	10.57	4.14
Getting Ready for a Special Occasion	10.07	3.77

Familiarity

Most of the events chosen were familiar to participants (see Table 4.3). Only one category had more than 5 responses in the Never category, that was *Rescue at Sea* (25 participants had never experienced this event) and 10/15 events had no responses at all in the "Never" experienced category. For 10/15 events, more than 70% of participants indicated that they had experienced these events at least more than 10 times. For 3 categories, more than 70% of participants claimed that they had experienced the event more than 50 times. *Making a cup of coffee, having a shower* and a *trip to the supermarket* appeared to be the most familiar in terms of experience. Participants had the least experience of *preparing a picnic, a camping trip* and *rescue at sea*. It should be noted that despite *rescue at sea* being experienced by only 2 participants, there appeared to be no difficulty in generating items for this event. Presumably participants were able to draw on second hand experience through film and fiction to complete the task.

Table 4.3.

Percentage of participants who indicated how often they had experienced the event.

Event	Never	<10 times	>10 times	>50 times
A Wedding	3.7	25.9	55.6	14.8
Visit to the Dentist	0	0	74.1	25.9
A Day at the Beach	0	0	70.4	29.6
Having a Shower	0	0	7.4	92.6
A Camping Trip	18.5	40.8	33.3	7.4
A Trip to the Supermarket	0	0	0	100
Seeing a Film at the Cinema	0	3.7	29.6	66.7
Christmas Day	0	0	74.1	25.9
Looking for Something in the Attic	0	37	48.2	14.8
Packing the Car to go on Holiday	0	33.3	55.6	8.1
Preparing a Picnic	11.1	44.4	40.8	3.7
Rescue at Sea	92.6	7.4	0	0
A Children's Party	0	29.6	48.2	22.2
Making a Cup of Coffee	0	3.7	18.5	77.8
Getting Ready for a Special Occasion	0	3.7	51.9	44.4

It should be noted that the for this breakdown only the percentage was calculated from 27 rather than 28 participants as one participant indicated that he/she had experienced Christmas Day more than 50 times despite having claimed to be under 21 – this cast some doubt on this aspect of their data and so their frequency of experience data was excluded.

Items Generated

No systematic classification was conducted on the items generated by the participants as it was not the focus of the study. It appears, however, that items tended to fall into the following fairly distinct categories: Objects, Food, Clothes, People, Activities, Weather, Emotions, Comments/personal associations. A full record of items and frequencies for each event can be seen at Appendix K. With reference to Appendix K, it should be noted that items mentioned by only one participant appear without any recorded frequency and if such items represent a qualified member of the category then they will not have been included in the total frequency for that item. For example, *warm fluffy towel* was mentioned by only one participant and was not included in the total frequency for *Towel*. This procedure was adopted in order to avoid any risk of over-inflating the frequencies.

Stage 2: Eliciting Event-to-Item data

The purpose of this stage of the study was to obtain some measure of the extent to which the items listed from the 15 events in the previous stage were associated with the events from which they were generated. A second group of participants were therefore presented with a subset of the 192 items and were asked to list events that they associate with those items.

Method

Participants

One hundred and twenty undergraduates and post-graduates from the University of Hertfordshire took part in this stage of the data collection. Some participants were awarded course credit, others were placed into a draw for a cash remuneration. These participants were drawn from a range of disciplines including law, nursing and psychology. The mean age of the group was 26.45 years. The range was 43 yrs (18yrs minimum, 61 yrs maximum). There were 89 female and 31 male participants.

Materials

Three event categories from stage1 were dropped, *rescue at sea* (least personal experience) *getting ready for a special occasion* (items overlapped to a large extent with *having a shower*) and *packing the car to go on holiday* (items overlapped with *day at the beach*). For each of the remaining 12 events, 8 items generated with a relative high frequency and 8 items generated with a low frequency were required for stage 2. The following constraints operated in choosing these items:

- only one non-concrete item was included from each event (e.g., noise, smiles)
- not more than one item from each event that made direct reference to the event was included (e.g., Christmas Pudding, Shower Curtain)
- Superordinates (e.g., food, clothes) were excluded altogether as they were considered too general and would cause difficulties in comparison across levels in the final stage of the study
- In order to ensure that similarity between these items would be mainly thematic and not perceptual, the inclusion of two items from the same superordinate category within each event was avoided unless it was considered that they shared few perceptual properties.
- Where an item had been generated for more than one event, it was included under the event for which it had been generated with the highest frequency.

Once these constraints had been satisfied, 8 items generated with the highest frequencies and 8 items generated with the lowest frequencies remaining in each category were identified. In some event categories (where frequency of generation dropped off quite quickly or where the above mentioned constraints had demanded removal of high frequency items)this meant that items generated by as few as 6/28 participants were classified as relatively high frequency. However, the progressive 'culling' nature of this procedure meant that it was desirable to include as many items as possible in the next stage. This gave rise to the set of items in Table 4.4.

Table 4.4

Items chosen from 12 events for inclusion in Stage 2 (Stage 1 frequency of generation shown after each item).

High Frequency	Low Frequency	High Frequency	Low Frequency
BEACH		ATTIC	
Sandcastle 24	Lifeguard 2	Ladder 20	Suitcases 3
Sea 23	Candyfloss 2	Boxes 16	Pipes 2
Sand 23	Dinghy 2	Dust 13	Photos 2
Spade 17	Sunhat 2	Cobwebs 9	Antiques 2
Bucket 16	Frisbee 2	Spiders 7	Light switch 2
Swimming costume 14	Seaweed 2	Junk 8	Bags 2
Suntan Lotion 8	Toes 2	Old Books 5	Damp 2
Shells 7	Wet 2	Xmas decorations 5	Aerial 2
PARTY		WEDDING	
Presents 20	Decorations 4	Bride 21	Invitations 4
Balloons 16	Bouncy Castle 3	Church 21	Hymns 3
Cake 14	Lemonade 2	Vicar 12	Bagpipes 2
Clown 13	Chocolate 2	Flowers 10	Shoes 2
Noise 11	Paper plates 2	Champagne 8	Cameraman 2
Jelly 10	Streamers 2	Ring 7	Bells 2
Hats 9	Toys 2	White dress 7	Jewellery 2
Party bags 8	Pretty dress 2	Confetti 6	Smiles 2
CAMPING		CINEMA	
Tent 27	Scouts 2	Popcorn 25	Ushers 4
Sleeping Bag 22	Discomfort 2	Dark 21	Coke 3
Rucksack 9	Fishing Rod 2	Ticket 12	Laughter 3
Tent pegs 8	Matches 2	Sweets 10	Toilet 3
Stove 8	Wellington Boots 2	Screen 7	Actresses 2
Compass 7	Lantern 2	Ice cream 6	Torches 2
Sausages 6	Fish 2	Seat 5	Nachos 2
Grass 6	Caravan 2	Adverts 5	Projector 2

Table 4.4 contd.			
COFFEE		PICNIC	
Milk 23	Fridge 4	Sandwiches 15	Rug 3
Sugar 23	Electricity 3	Blanket 13	Thermos 3
Kettle 18	Tap 2	Fruit 11	Kitchen Roll 2
Water 17	Sink 2	Crisps 10	Wet wipes 2
Cup 14	Filter 2	Plates 8	Coolbox 2
Coffee 14	Relaxing 2	Napkin 6	Fruit Squash 2
Spoon 13	Plug 2	Wine 6	Pork pies 2
Biscuits 8	Brandy 2	Sun 6	Plastic Cup 2
DENTIST		SHOWER	
Drill 19	Window 2	Shampoo 24	Flannel 3
Chair 17	Floss 2	Towel 20	Hair 3
Fillings 16	Posters 2	Soap 14	Bathrobe 3
Light 12	Small mirror 2	Shower Curtain 13	Talc 3
Pain 10	Wisdom Tooth 2	Conditioner 10	Moisturiser 2
Mouthwash 9	Gloves 2	Hot 8	Tiles 2
Dentist 9	Toothpaste 2	Steam 6	Bodywash 2
Needle 7	Dread 2	Razor 5	Body 2
SUPERMARKET		CHRISTMAS DAY	
Trolley 23	Magazines 4	Tree 23	Wrapping paper 4
Checkout 16	Videos 2	Turkey 18	Cold 3
Vegetables 15	Cashiers 2	Cards 10	Mince pies 3
Aisles 9	Cigarettes 2	Crackers 10	Happiness 2
Money 9	Cashpoint 2	Television 10	Carols 2
Basket 8	Tins 2	Queen's speech 9	Films 2
Shopping List 8	Expensive 2	Xmas Pudding 8	Dinner 2
Meat 7	Credit card 2	Snow 7	Candles 2

Procedure

The items were presented to participants in a 25-page booklet. The booklet consisted of a cover page of instructions and examples and then 24 pages with one item printed in bold at the top of each page (see Appendix L for instructions and sample materials). Hence each participant was required to generate events for 24 items. Eight different sets of materials were therefore constructed in order to collect data for each of the 192 items. These were constructed as follows:

One high and one low frequency item from each event were randomly allocated to each set. There was some concern that the co-presentation of certain items may inadvertently suggest events to participants and so two other versions of each of the 8 sets of materials were created. This gave rise to 24 different booklets. Five copies of each one was made and each copy was randomised according to one of five different item orders. The booklets were then completed by the 120 participants.

The following instructions were shown on the front page of the booklet and were also read aloud. "When you think of an object or item, there are a variety of events/situations which may come to mind. For example, when I think of a newspaper, I may think of being in a shop buying one, or reading one on the train, or someone using one to light a fire or putting them out in the rubbish. On each of the following pages, you will see a word that is the name of an object or item. Some of the words could be treated as verbs but I want you to think of them all as names of objects or items. For each page, you will be allowed 30 seconds to list as many events or situations as you can which the object/item brings to your mind. I will tell you when to start each page".

Participants were run individually or in small groups and participated for course credit or payment.

Results

Despite the specific nature of the instructions and the presentation of an example, many participants listed a number of associations other than events or situations. Nevertheless, situations and events were generated. The responses for each event were typed into a separate table. The situations that were mentioned by more than one participant were tabulated by the experimenter. Decisions had to be made regarding a) whether a response constituted a situation and b) whether two responses represented the same situation. A second judge looked at 20% (38) of the tables and independently tabulated the situations listed. The two judges then compared classifications and frequencies for this subset of the items. Where agreement varied, it tended to concern the subdivision of more 'inclusive' situations into smaller categories. After discussion, these smaller situation categories were agreed. On 12 of the items, agreement was 90%, on a further 20 agreement was over 70%. On only 2 items was the classification performed quite differently by the two judges. Although the level of agreement was not perfect, it was decided that it was sufficiently high to serve the purpose of the data collection. An example of the type of responses generated can be seen in Table 4.5.

Table 4.5

Example of responses from stage 2 (event-to-item generation) for item "BALLOONS".

Item No.	Subject No.	Data
1	1	Parties: Age: Different colours: Christmas: Helium: Logos and messages
2	9	Fun: Children: Bouncy: Fly away: Joke shop: Party: Go up in one(big)
3	17	Christmas: Parties: Birthdays :Colour: Names on balloons: Ages : price
4	25	I remember when I was a kid and I lost a balloon I was holding
5	33	Colourful balloons in birthday party: A small child holding a balloon
6	43	Party : Children: Food: Sweets: Music: Hats: plates
7	51	Party – birthday : Fun fair: Watching hot air balloons taking off

8	59	Going to birthday parties where I was always made to play horrible games which involved popping balloons and me ending in tears
9	67	Being at birthday parties: Balloons being let out of a net at a wedding: Clowns making objects with balloons
10	75	At a child's party: Hot air flying over on a clear day
11	84	Lots of colours: Blue ,red, etc.: Put up at parties: You can use helium: Hot air balloons : Fly in sky
12	92	Going to children's parties: Being scared of them popping The BBC hot air balloon on TV
13	100	With helium they fly: Children's b-day parties
14	108	Parties!: Summer- hot air balloon events
15	116	Being at a fairground: Children's parties : Flying in a balloon

The situations of principal interest were those from which the item had been originally generated. The situations generated for each of the 192 items were tabulated and these events were identified amongst the responses. The frequency with which each original event was generated in response to each item can be seen at Appendix M. On the basis of the frequency with which the event had been listed in response to the item, each item was classified as High or Low frequency. A table was created showing for each item, the number of participants who generated this item from the original event and who generated that original event in response to that item. Each item was then classified as having a high or low item-to-event frequency and a high or low event-to-item frequency using the following criteria:
For the item-to-event frequency, any item generated by more than 9 / 28 (32%) participants was classified as high frequency and an item generated by less than 9 /28 participants was classified as low frequency (scores of 9 were omitted). For the event-to-item frequency, any item which gave rise to the target event with a frequency of more than 5/15 (33%) was classified as high and an item which invoke the target event with a frequency of less than 5 /15 was classified as low frequency (scores of 5 were omitted).
Clearly, these criteria are not as stringent as would be desirable. However, imposing a more stringent criteria, such as requiring an item to have been generated by over 50% of participants to classify as high frequency would have resulted in an inadequate

number of stimuli in each category, bearing in mind the requirements for the final stage of the study. Since almost 150 participants had already been involved in collecting these data it was decided to run with these constraints. So, to take an example, 82% of the participants mentioned Beach or going to the Beach in response to SAND, giving SAND a high event-to-item frequency. Beach was only generated by 13% of participants in response to SPADE, and so SPADE has a low event-to-item frequency.

For the planned manipulation of materials in the final stage of the study, at least two items from each situation were required that fitted each of the following categories:

HH – item was generated from the situation with a high frequency in Stage 1 and in turn gave rise to that same situation with a high frequency in Stage 2.
HL - item was generated from the situation with a high frequency in Stage 1 but only gave rise to that same situation with a low frequency in stage 2.
LH - item was generated from the situation with a low frequency in Stage 1 but gave rise to that same situation with a high frequency in stage 2.
LL - item was generated from the situation with a low frequency in Stage 1 and also evoked that same situation with a low frequency in stage 2.

At the conclusion of the process described above, it was found that only 7 events had resulted in a sufficient number of each item type from the two stages of data collection to allow the planned manipulation in the final stage of the study. Five events were therefore dropped at this stage leaving the event/items seen in Table 4.6. It should be noted that there are items listed under particular events in Table 4.6 that did not appear under that same event in Table 4.4. This is due to the fact that if an item was generated for more than one event in Stage 1 then it was used only once in Stage 2. This enabled the collection of frequency data for additional items and avoided the repetition of any one item in Stage 2. So, for example, *candles* was generated in response to both Christmas day and Children's Party in Stage 1 but was presented only once in Stage 2. When *candles* was presented in Stage 2 it prompted mention of both Christmas day and Birthday party and so could be used for either event in the final stage of the study. Appendix M shows the additional events (from the original set) that were generated in response to this type of item.

Table 4.6

Items which could be potentially included in final stage of study.

	HH	HL	LH	LL
BEACH	SAND	SUN	FRISBEE	CANDYFLOSS
	SANDCASTLE	SPADE	LIFEGUARD	DINGHY
	BUCKET		SEAWEED	SUNTAN LOTION
	SWIMMING COSTUME		SHELLS	TOES
	SEA		SUNHAT	WET
COFFEE	KETTLE	COFFEE	BISCUITS	BRANDY
	CUP	MILK	FILTER	ELECTRICITY
		SUGAR		FRIDGE
		WATER		PLUG
				RELAXING
				SINK
				TAP
.				
SHOWER	HH	HL	LH	LL
	SHOWER CURTAIN	SHAMPOO	BODYWASH	BODY
	TOWEL	CONDITIONER	BATHROBE	HAIR
	WATER		FLANNEL	HOT
	SOAP			MOISTURISER
				RAZOR
				STEAM
				TALCUM POWDER
				TILES
CHRISTMAS	TURKEY	TREE	CAROLS	CANDLES
DAY	CARDS	TELEVISION	MINCE PIES	COLD
	CRACKERS		XMAS PUDDINGS	DINNER
	PRESENTS		WRAPPING PAPER	FILMS

			SNOW	HAPPINESS
			DECORATIONS	
CHILD PARTY	JELLY	CLOWN	PARTY BAG	BOUNCY CASTLE
	BALLOONS	CAKE	PAPER PLATES	CHOCOLATE
		PRESENTS		DECORATIONS
		NOISE		HATS
		CANDLES		LEMONADE
				PRETTY DRESS
				STREAMERS
				TOYS

S/MARKET	HH	HL	LH	LL
	CHECKOUT	VEGETABLES	BASKET	VIDEOS
	TROLLEY	FRUIT	CASHIERS	CASHPOINT
			CREDIT CARD	CIGARETTES
			SHOPPING LIST	MAGAZINES
			MONEY	MEAT
			TINS	
			EXPENSIVE	
WEDDING	BRIDE	FLOWERS	BELLS	BAGPIPES
	CHURCH	VICAR	CONFETTI	CAMERAMAN
			RING	HYMNS
			WHITE DRESS	JEWELLERY
			CHAMPAGNE	SHOES
			INVITATIONS	SMILE
			HATS	
			USHER	

Stage 3: Comparisons of Similarity Ratings

Following collection of the item-to-event and the event-to-item data, the items could be paired together in order that they might be judged for similarity.

Of the 70 pairs used for the final stage of the study (see Appendix N for full details of pairs), only 3 pairs clearly shared perceptual properties of the type that would be expected to underpin 'traditional' similarity, i.e., similarity based on shared features (fruit-vegetables, party bags-paper plates, shampoo –conditioner). It was therefore assumed that any similarity rated by participants for most of the pairs would be based on something else i.e., thematic similarity.

Since context had been shown to heighten similarity ratings (see Chapter 3) amongst items that shared few standard features, it was decided to present the pairs in this study both in the context of the situation to which they had been shown to be associated and also without that context. It was reasoned that any systematic effect of strength of association between the item and the situation on ratings should be most evident if participants were to be encouraged to focus on the situational organisation of their item knowledge by highlighting the situation in question. However, previous studies (e.g. Wisniewski and Bassok, 1999) did not provide any context for the comparisons and still thematic associations were seen to influence ratings of similarity. For this reason both a 'context' and a 'no context' condition were included in this study. It was also felt that a blocked presentation of the pairs relating to the same situation would further encourage participants to think about these items specifically as they related to the situation of origin whereas a randomised presentation of non – perceptually similar items would possibly weaken the effect, particularly if no context was provided either. Both conditions of presentation, blocked and random, were therefore prepared. It was hypothesised that items pairs with comparatively high item-to-event and event-to-item indices would result in higher similarity ratings than the pairs with low item-to-event and event-to-item indices.

Method

Participants

A further 264 participants took part in the experimental stage of the study. All participants, 53 males and 211 females were undergraduates on psychology and psychology related courses. The age range of the participants was 18 to 59 years and the mean age was 22 years. The mean number of years lived in this country was 20 years. 84% (n=222) of the participants had lived in this country for all of their lives. Only 6% of participants had lived in this country for less than 10 years.

Design

A five factor mixed design was utilised. The first between-subjects factor was order. There were two different orders of items as they were presented on each page. The second between-subjects factor was set. Three different sets of materials were prepared (see Materials and Appendix M). The third between-subjects factor was presentation. Item pairs were presented as either blocked (relating to the event with which they were associated) or random. The final between-subjects factor was context: participants saw the item pairs either in the presence of the event name (with context) or alone (no context).

The within participants factor was the paircode with ten possible pairtypes. The dependent measure was the ratings of similarity given on a scale of 1 – 9 where 1 indicated 'not similar at all' and 9 indicated 'very similar'.

Materials

The items identified through stages one and two of this study were combined into item pairs for the purpose of collecting similarity ratings between the two items. There were four possible item types (HH, LL,HL, and LH) and, without including all possible orders, ten combinations or "pairtypes" (see Table 4.7).

Table 4.7.

Pair combinations used in final study

HH	HH
HL	HL
LH	LH
LL	LL
HH	HL
HL	LH
LH	LL
LL	HH
HH	LH
HL	LL

The first letter of the coding for each word refers to the frequency with which the item was generated from the event (item-to-event index of association henceforth referred to as ie) – the second refers to the frequency with which the event was generated from the item (event-to-item index of association henceforth referred to as ei).

An attempt was made to order the items within the pairs so that each of the four types appeared first or second fairly evenly. Direction of comparison has been shown to affect the features taken into consideration when judging similarity of items (Goldstone, Gentner and Markman, 1993). It is possible that this is also the case for thematically related items (situations associated with the first item in the pair may be 'searched' for the second item). However, in the absence of empirical evidence to this effect, it was considered that a full combination of item types would constitute an unwieldy number of stimuli and the stimuli set was therefore limited to the ten combinations shown above. It was decided that participants could reasonably be asked to provide ratings for 35 pairs of items (five pairs from each event) without losing interest. In order to obtain ratings for all pairtypes and all events without participants seeing the same item repeated within the stimuli, it was necessary to create 3 different sets (A, B, C) of materials. Certain pairs were necessarily repeated across the sets but no item appeared more than once within the same set. A blocked and randomised version of each set, with and without context was produced, The stimuli were presented in an eight-page A4 booklet. The front page gave instructions and an example. On each of the remaining pages, five pairs of items were printed. Above each item pair appeared the question " How similar are the following items to each other?" Below each pair was a scale of 1 – 9 where the number 1 was labelled not similar at all and 9 was labelled very similar. In the blocked version, the five item pairs on each page were all related to one of the seven events. In the randomised version, the item pairs were presented randomly through the booklet. In the blocked 'with context' condition, the name of the event appeared in bold type at the top of the page (e.g. Having a Shower) and preceding each pair was the phrase " Think about: having a shower". In the blocked 'no context' condition the layout of the similarity question, the item pairs and the scale was preserved but there was no event mentioned anywhere on the page. In the 'randomised with context' condition, the phrase "Think about: followed by the relevant event" was included above each pair but this was removed for the 'randomised no context' condition. Finally a second order for each version was prepared. This was achieved by altering the order of the item pairs on each page for the blocked condition from 12345 in the first version to 34512 in the second version. For the randomised condition, a second randomised order was used. When the booklets were compiled, the order of the pages was reversed for half the materials. Instructions and sample pages for each of these conditions can be seen at Appendix O.

Procedure

The data collection was run in groups of various sizes (maximum 60 participants at any one time). Participants were asked to read the instructions on the first page and were given the opportunity to ask questions before starting the task. The standard demographic details were collected in addition to data concerning how long the participant had lived in this country. The reason for this was the culture specific nature of some of the events/situations. There was no time constraint placed upon the task.

Results

Analysis by Pairtype

The first analysis to be conducted was based on the original ten pairtypes. The mean ratings for each of the ten pairtypes were calculated (see Table 4.8).

Table 4.8

Mean similarity ratings and standard deviations for each pairtype. Means are displayed in ascending order.

	LL-LL	HL-LL	LH-LL	HL-LH	LL-HH	HH-LH	LH-LH	HH-HL	HL-HL	HH-HH
Mean	2.51	3.32	3.62	3.99	4.33	4.48	4.68	4.85	5.23	5.29
s.d.	(1.5)	(1.91)	(1.71)	(2.08)	(2.04)	(2.10)	(2.08)	(2.41)	(2.01)	(2.26)

The data were initially entered into a five factor (2*2*2*3*5) mixed ANOVA: order (2 orders, between group) by presentation (blocked or random, between group) by context (with or without, between group) by 3 set types (A, B, C, between group) by paircode (10 codes, repeated measures). All analyses were conducted at 95% confidence level. Wilks Lambda F values are reported throughout due to a violation of sphericity shown by a significant Mauchly's sphericity test, $p< 0.01$. There were

Figure 4.1 Mean similarity ratings as a function of pairtype for both conditions of context.

no main effects of order, set or presentation and so the data were collapsed across these factors and subjected to a 2*10 mixed ANOVA. This showed a main effect of pairtype $F(9,254) = 112.62$, $p< 0.001$, eta squared = 0.80 and of context, $F(1, 262) = 42.81$, $p< 0.001$, eta squared = 0.87) and a small interaction effect between pairtype and context $F(9,254) = 8.726$, $p< 0.0005$, eta squared = 0.24 (see figure 4.1). There was a significant linear polynomial contrast $F(1,262)= 659.85$ $p<0.001$, eta squared = 0.72 and contrasts were significant for all pairwise comparisons except LL-HH vs. HH-LH and HL-HL vs. HH-HH .

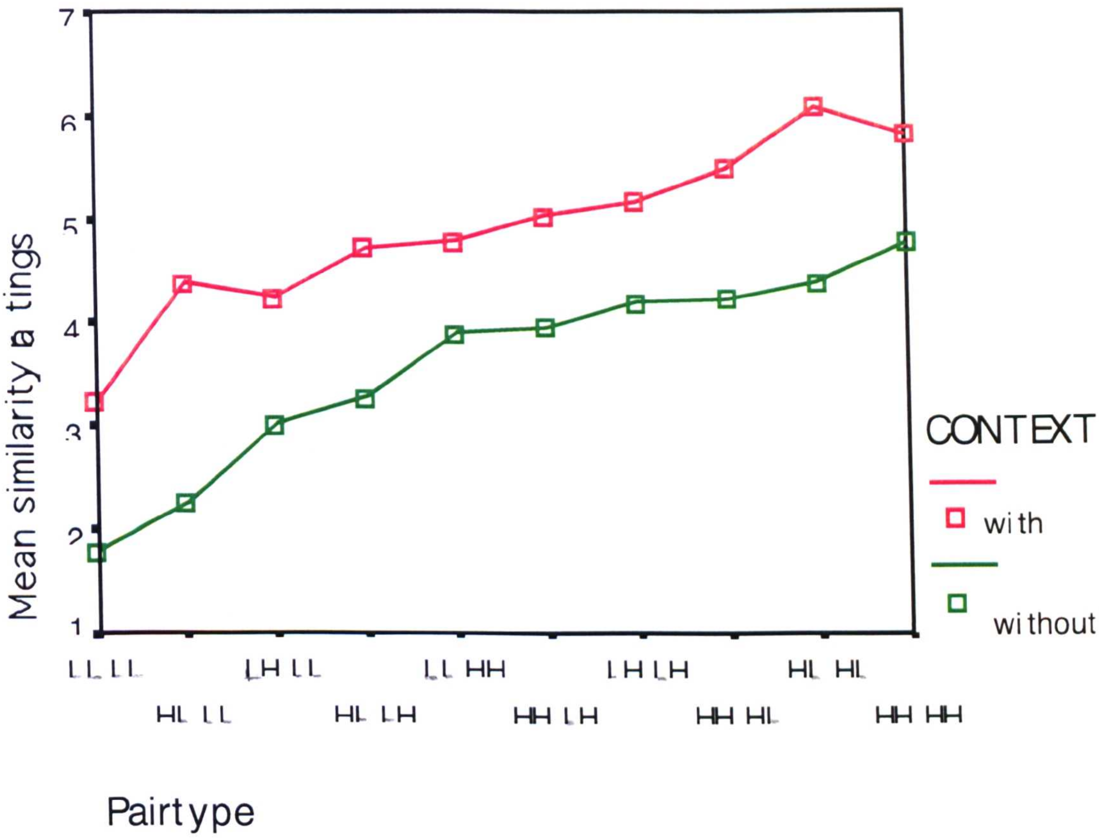


Figure 4.1 Mean similarity ratings as a function of pairtype for both conditions of context

Table 4.9

Marginal mean similarity ratings presented by position of item in pair and index type.

1st in pair				2nd in pair			
Item-to-event		Event –to-item		Item-to-event		Event-to-item	
L	H	L	H	L	H	L	H
3.78	**4.52**	**3.18**	**4.51**	**3.76**	**4.92**	**4.09**	**4.59**

In first pair, ratings are lowest when either ie or ei is low. Ratings are increased when ie or ei is high but it seems irrelevant which one is high. In second pair, the same story applies but ie seems to be most effective at raising ratings. The same means are re-arranged in ascending order in table 4.10.

Table 4.10

Marginal mean similarity ratings presented in ascending order

Position in pair	1st	2nd	1st	2nd	1st	1st	2nd	2nd
index	EI	IE	IE	EI	EI	IE	EI	IE
	L	L	L	L	H	H	H	H
	3.18	**3.76**	**3.78**	**4.09**	**4.51**	**4.52**	**4.59**	**4.92**

It appears, from the sub-set of combinations used in this study, that the position of the item in the pair does not affect the strength of the thematic similarity. Both items seem to be considered equally. It also seems to be the case that it does not matter whether it is the event-to-item or the item-to-event index that is high. It seems clear that if one of the indices is high in one of the pairs, the ratings will be raised compared to when that index is low.

Analysis based on Item-to-Event indices of Association Only

There are clearly alternative ways of grouping the pairtypes for analysis. It is possible that the component indices (item-to-event and event-to-item) may have exerted a differential effect upon the ratings. For this reason, two further sets of analyses were conducted based on these components. Firstly, the pairs were recoded according to whether the item-to- event indices for the pair were high. This gave rise to the following codings:

0= neither of the items in the pair were generated from the event with a high frequency

1= one of the items in the pair was generated from the event with a high frequency

2= both of the items in the pair were generated from the event with a high frequency

Table 4.11

Coding system used for distinguishing between item-to-event and event-to-item association

	Item-to-event	Event-to-item
LL-LL	0	0
LH-LL	0	1
HL-LL	1	0
LH-LH	0	2
HL-LH	1	1
LL-HH	1	1
HL-HL	2	0
HH-LH	1	2
HH-HL	2	1
HH-HH	2	2

Means ratings were calculated based on this new coding (see Table 4.12).

Table 4.12

Mean similarity ratings for each paircode under conditions of with and without context (sds in brackets) collapsed across mode of presentation.

	Paircode		
	0	1	2
With context	4.15 (1.59)	4.78 (1.90)	5.28 (2.00)
Without context	2.50 (0.99)	3.50 (1.55)	4.21 (1.84)
Overall	3.33 (1.56)	4.14 (1.84)	4.75 (1.99)

Table 4.13
Mean similarity ratings for each paircode under conditions of with and without context, blocked and random presentation (sds in brackets)

	Paircode 0	1	2
With Context			
Blocked	4.00 (1.71)	4.73 (2.08)	5.18 (2.14)
Random	4.30 (1.46)	4.83 (1.71)	5.39 (1.86)
Without Context			
Blocked	2.52 (1.12)	3.34 (1.69)	3.97 (1.92)
Random	2.49 (0.86)	3.67 (1.39)	4.45 (1.74)

The overall pattern of results with and without context was very similar to the pattern seen in the previous analysis. As in the previous analysis, effects of order and set were initially tested for by including them both as between factors in the initial ANOVA . Context and presentation were also between factors and the new coding was a repeated measures factor with three levels (0,1,2).

The ratings with context were higher than without context, $F(1,240) = 47.518$, $p< 0.01$, eta squared = 0.165 and the ratings differed with the paircode, $F(2,239) =230.89$, $p< 0.01$, eta squared = 0.659. Both with and without context, the ratings for the pairs with one item having a high item-to-event association were higher than for pairs where neither item had a high item-to-event association. Pairs where both members have a high item–to-event association were rated highest of all. There was a significant linear trend for paircode, $F(1,262) = 435.488$, $p< 0.01$ eta squared = 0.624. All pairwise comparisons were significant at $p< 0.001$.

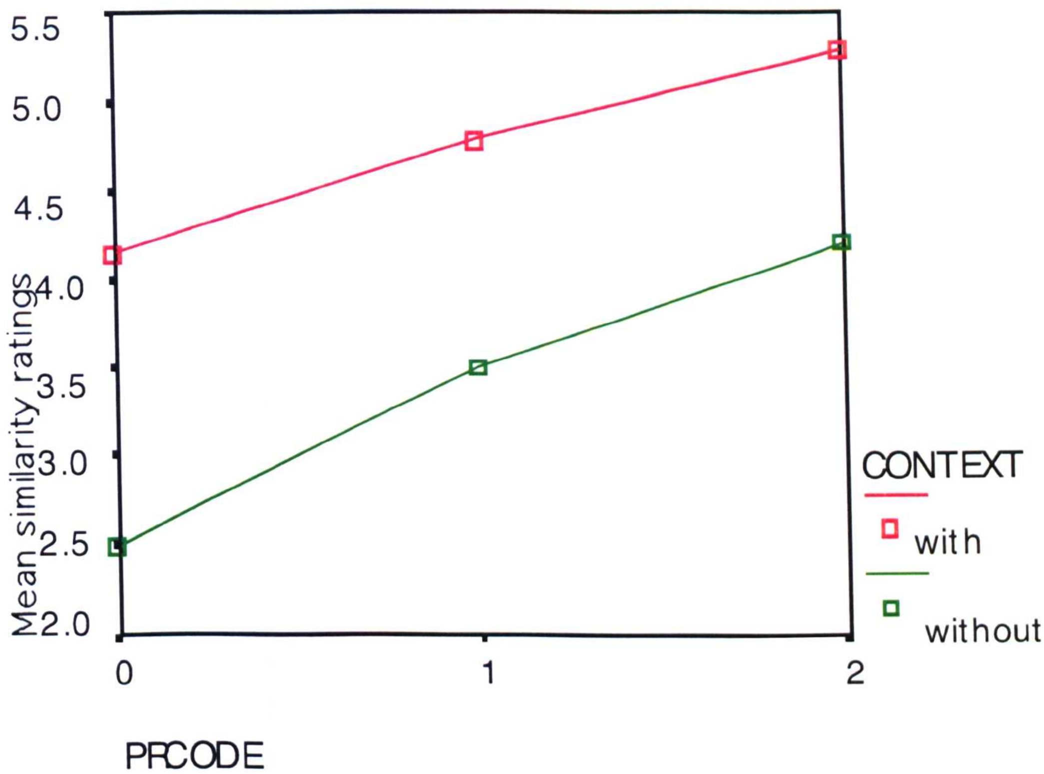


Figure 4.2 The effect of context and pair type on mean similarity ratings.

There were no other significant main effects. Paircode featured in a small interaction effect with context and presentation, F (2,239) = 3.466 p=0.033 eta squared = 0.028. Set featured in a small interaction effect with paircode , F (4,478) =13.21, p <0.001 eta squared =0.10. This latter effect appears to be the result of a slightly inflated pattern of increase in the ratings with code in set B but since the pattern of results is similar to elsewhere in the analysis, subsequent analysis collapsed the data across set and order.

The 2* 5 mixed ANOVA conducted on the data after removing order, set and presentation as factors confirmed the main effect of paircode, F(2,261) = 293.583 , p<.0005, eta squared = 0.625. Overall the pairs were rated significantly higher with context (M = 4.74) compared to without context (M = 3.4) F(1,262) = 46.632, p< 0.01 eta squared = 0.151. There was one other significant interaction, between paircode and context, F(2,261) = 8.926, p< 0.01, eta squared = 0.064.

One way repeated measures ANOVAs showed that the effect of paircode was present at both levels of context. With context $F(2,130) = 103.465$, $p < 0.01$ eta squared = 0.614, Without context $F(2,130) = 118.706$, $p < 0.01$ eta squared = 0.646.

.

Analysis Based on Event-to-Item Indices of Association Only

Finally , the pairs were coded once again, this time according to whether the event-to-item indices for the pair were high. This gave the following possible codings:

0 = neither of the items in the pair generated the target event with a high frequency

1 = one of the items in the pair generated the event with a high frequency

2= both of the items in the pair generated the event with a high frequency

See column 2 of Table 4.11 for the resulting pair groupings. Means ratings were re-calculated based on this new coding (see Table 4.14).

Table 4.14

Mean similarity ratings for each paircode under conditions of with and without context (sds in brackets) collapsed across mode of presentation.

	Paircode		
	0	1	2
With context	4.09 (1.70)	4.79 (1.85)	5.76 (2.03)
Without context	2.9 (1.19)	3.41 (1.44)	4.33 (1.78)
Overall	3.50 (1.58)	4.10 (1.79)	5.05 (2.04)

Table 4.15

Mean similarity ratings for each paircode under conditions of with and without context, blocked and random presentation (sds in brackets)

	Paircode		
	0	1	2
With Context			
Blocked	4.00 (1.85)	4.74 (2.06)	5.57 (2.17)
Random	4.10 (1.55)	4.83 (1.62)	5.96 (1.87)
Without Context			
Blocked	2.79 (1.31)	3.33 (1.57)	4.15 (1.98)
Random	3.01 (1.05)	3.49 (1.31)	4.51 (1.56)

The pattern of ratings seen in Table 4.18 is, unsurprisingly, similar to that seen in the previous two analyses. The provision of context inflated the similarity ratings compared to no context, $F(1,240) = 46.061$, $p < 0.01$, eta squared = 0.161, whilst presentation, blocked or random, exerted no effect overall. As before, all five factors were included in an initial ANOVA. There was a main effect of paircode, $F(2,239) = 300.711$, $p < 0.01$, eta squared = 0.716. If both items in the pair are items that are strongly associated with a common situation then the ratings are highest. If one item is strongly associated with the situation then the items are seen as less similar than if it is true of both items but are still perceived as more similar than if neither item is closely associated to the event. There was a significant linear trend of paircode, $F(1,262) = 610.221$, $p < 0.01$, eta squared = 0.7 and all pairwise comparisons were significant.

Set interacted significantly with paircode, $F(4,480) = 6.054$, $p < 0.01$, eta squared = 0.048. The interaction appeared to stem from the increased ratings for paircode 0 in set C. The interactions involving set led to an examination of the distribution of the data by set. Seven outliers were identified, cases 1,95,102,103,117,154 and 185. These were removed and the previous analyses were re-run but the principal findings remained unaltered. Since the interaction was associated with a very small effect size (0.062), it was decided to collapse the data across set.

The 2*3 mixed ANOVA conducted on the data collapsed across set, order and presentation confirmed a small main effect of context, $F(1,262) = 45.304$, $p< 0.01$, eta squared = 0.147 and a large effect of paircode, $F(2,261) = 303.962$, $p< 0.01$, eta squared = 0.7. This time there was no interaction between context and paircode.

Regression Analysis

It was decided that since the High/Low frequency classification could be subject to inconsistencies and relatively small levels of differentiation, a more useful analysis would be one that made use of frequency as a continuous variable. A stepwise multiple regression was therefore performed with the similarity ratings as the dependent measure and the frequencies associated with the four indices (first item in pair ie, first item in pair ei, second item in pair ie and second item in pair ei) as predictors. The best predictor of variance in the ratings was the ie index of the second item, explaining 12.5% of the variance, $F(1,68) = 10.867$, $p = 0.002$. The ie index of the first item was the next best predictor, explaining a further 8% of the variance $F(2,67) = 9.45$, $p< 0.001$ and the ei of the first item explained a further 5% of the variance $F(3,66) = 8.179$, $p< 0.001$. The total variance explained by the model was therefore 23.8%. The ei of the second item did not significantly add to the variance explained.

It was found that all four indices correlated positively and significantly with the ratings (all correlations were performed 2 –tailed):

1st item ie : $r(70) = 0.357$, $p = 0.002$
1st item ei : $r(70) = 0.321$, $p = 0.007$
2nd item ie: $r(70) = 0.371$, $p = 0.002$
2nd item ei: $r(70) = 0.286$, $p = 0.016$

The indices did not correlate significantly with each other with the exception of the 2nd item ie and ei indices $r(70) = 0.287$, $p = 0.016$. This suggests that by chance, the higher ie frequencies for the second item in the pairs tended to be linked with the higher ei ratings. An overall rating for the pair of items (calculated by averaging the four indices' frequencies) also correlated positively and significantly with the similarity ratings $r(70) = 0.527$, $p< 0.001$.

The indices of strength of association between the item and the event account for a significant amount of variance in the similarity ratings. This analysis gives a clearer picture of the relative contributions of the different indices and it seems that it is the ie indeces (indicating the likelihood of thinking of that item given a particular event) that have the greater influence on similarity ratings than the ei indeces though the chance correlation between the two indeces for the second item pair may have obscured any potential contribution from the ei index for the second item in the pair. There is still over 75 % of variance to be accounted for – other relationships discussed in the introduction to this chapter like spatial contiguity and temporal contiguity may well account for some of this unexplained variance.

Discussion

The empirical work reported in this chapter comprised three stages. The data from the first two stages were used to formulate the materials for the final stage. The first data collection consisted of requiring participants to list items that they associated with each of 15 events/situations. The relative frequency with which these items were generated was used to calculate an index (high or low) which was attached to that item as the item-to-event (ie) index. A subset of the items generated (192) were then presented to a different set of participants who were asked to list the events/situations that they associated with that item. The relative frequency with which the original event was listed for each item was then calculated and used to produce an index (high or low) of event-to-item (ei) frequency. Each item consequently had two indices. In the final data collection, these items were combined in pairs and presented to a final set of participants who gave scaled similarity judgements for the pairs. The pairs were presented with or without the event name (context) and blocked by event or randomised. It was predicted that pairs of items that were highly associated with the event would be rated as more similar than pairs of items that were weakly associated with the event and that the presence of context would inflate the ratings.

The following principal findings emerged from the analysis of the ratings in the final stage of the study:

- The strength of association between items and events was a major source of similarity when participants are asked to rate pairs of items under experimental conditions.
- A high strength of association between the item and the event seemed to have the more effect upon ratings of similarity as a high strength of association between the event and the item
- There was a strong effect of context on ratings of thematically related items – if even one index on one item was high, the mean ratings exceeded those given for any pairs without context. However, even without context , the ratings demonstrated sensitivity to the item/event association indices.

As mentioned in the introduction, previous studies that have documented the influence of thematic similarity have tended to use pairs of items that not only can be linked together through a common scenario or situation but also are linked directly to each other by spatial contiguity or functional relatedness. It is not surprising then that the co-presentation of those items give rise to a common situation in the mind of the participant. In the work reported here, most of the pairs of items were linked only by virtue of featuring in the "script' of the same situation. The first question then was whether this link would be sufficient reason for participants to find these items similar. Recall that the items generally shared few, if any, perceptual features. There was indeed a great deal of variability in the ratings, suggesting that participants are not all equally inclined to rate items as similar on this basis. However, the mean ratings indicate that this type of similarity is viewed as a legitimate basis for finding items similar in a standard similarity rating task. It could be argued that if participants are able to complete the phrase, 'they are similar because they both.........' , then the items are similar and in the case of these items, if they both feature in the same event/situation, then that is sufficient reason to rate them as being fairly similar to each other. Moreover, the results of the similarity ratings suggest that this type of similarity is sensitive, to a certain extent, to the strength of the association between each individual item and the event under consideration. The highest mean ratings were allocated to those items that were associated highly with the event on all indices whilst the lowest mean ratings were accorded to those items that scored low on all indices. Pairs coded in such a way as to represent some interim level of association with the event received interim ratings which gave rise to a significant linear trend.

It could be argued that the ratings of similarity are arising simply from the strength of word association between the two words presented rather than through the strength of their shared association to an event. Measures of word association are usually obtained from published norms and these norms, in turn, are usually an indication of the percentage of respondents who give a particular word as their first response in free association to a target word. It has been noted that these associations can themselves be categorised as particular types of relationships (see Hutchison, 2003) and so it is difficult to separate out the effect of these different types of association. No attempt was made in this study to provide this type of control and clearly a further study would be required to do so before associative links other than situational could be ruled out as an explanation for these results. However, a post-hoc search of the Nelson, McEvoy and Schreiber (1999) word association norms of the first item in each pair used in the study showed only 2 of the second words in the pair as an associate at all (shampoo-conditioner, 0.403 and vegetables-fruit, 0.208). A search of the second word in each pair as the target showed only 2 of the first words as an associate (conditioner-shampoo 0.455 and fruit-vegetables 0.082). This would suggest that there were few direct associations between the word pairs in this study. A number of the words used in this study did not appear as target words in these norms at all (n=27). Once it could be shown that the items used in this study are only weakly associated according to such norms, they could be compared against a set of wordpairs that are not situationally linked but are known to be associated by the standard of the association norms. If similarity ratings were to be as high for these items (which must not share perceptual or functional features either) as for the situationally linked items, then one could infer that thematic similarity is simply the reflection of <u>any</u> association between the words and is not actually thematic at all.

As predicted, ratings for pairs of items shown in the presence of the event were significantly higher than the ratings for pairs shown without the event name. Interestingly, the pattern described above was repeated both with and without context. This suggests that when participants are asked to judge the similarity of these apparently dissimilar items without any cue to the appropriate context, the context they instantiate is an event or situation brought to mind by one or both of the items and that their subsequent assessment of similarity reflects the level of association between the event and the item. Perhaps even more remarkably, this seems also to be true even when the items are presented in a randomised fashion. It appears then that the inclination to use a common situation or event as the basis of the judgement is strong and pervasive but it is also graded according to, at least in part, the strength of association between that situation and the items being assessed. Importantly, this suggests that the "link' between the items is not necessarily directly between them but is provoked by the link between each item and the event. This is analogous to the claim that taxonomically related items are related through their shared membership of a superordinate category.

The regression analysis conducted on the data suggested that the item-to-event indices were better predictors of variance in the ratings than the event-to-item indices. It might have been reasoned that, particularly in the absence of context, the extent to which the item brings to mind the event (event-to-item association) would exert a greater influence than the association in the other direction. This does not seem to have been the case. However, there are several factors that should be kept in mind when interpreting the data with regard to the different indices. Firstly, not all combinations of pairtypes were included in this study. The full range of possible combinations would be required to make more systematic comparisons. This could be achieved by using a smaller number of event categories or a larger participant group. One has to be mindful when conducting this type of study of the maximum number of judgements that can be made with the requisite degree of attention. Secondly, the frequencies on which the indices were based arose from a comparatively small number of participants (n= 28 for the item-to-event data and n = 15 for the event-to-item data). The frequencies often dropped off quite steeply and so the distinction between low and high frequency items was, in some cases, quite small. This may explain why the clearest difference in ratings based on this classification occurred when all indices were either high or low. A future study would focus on the final seven events used here and collect frequency data (in both directions) from a larger participant group. Having drawn attention to the limitations of this study, it could also be argued that the compelling and persistent pattern in the ratings across all conditions speaks to the robust nature of the effect.

As mentioned earlier, there are many other ways in which items may be linked together and only one factor has been considered here. There is clearly more explanation required for these ratings and further types of relationships between the item pairs to be investigated but the factor chosen does seem have been sufficient in explaining some variance in the rated similarity between these perceptually different items and perhaps more importantly, it illustrates that a direct link between the items (functional, causal, spatial) is not necessary for thematic similarity to be perceived. In the concluding chapter it will be argued that this constitutes further evidence that the organisation of object knowledge based around every day events and situations exerts a stronger influence on adult cognition than previously thought.

CHAPTER 5

General Discussion and Conclusions

Overview

The empirical evidence generated by this research programme suggests that a structured knowledge of the situations and events in which items commonly feature is sufficiently well-entrenched and important as to play a substantial role in the organisation of conceptual knowledge

In this chapter, the findings from the category generation task will first be summarised and will be discussed with reference to the two recent theories of conceptual knowledge, Barsalou (1999, 2002) and Funnell (2001), described in Chapter 1. The findings of the similarity studies will then be summarised and also discussed in the light of the same two models. Implications of the research for cognitive psychology more generally are then considered, leading on to a discussion of avenues for future investigation. The final summary highlights the main contribution of the research.

Generating Category Members in a Free Emission Task

In Chapter 2, two studies were reported in which participants were required to list members of common and ad hoc categories within a limited period of time. It was found that participants generated, on average, more items for the taxonomic than for the ad hoc categories. However, when the mean number of items for each category was plotted, there was no obvious discontinuity between the two category types, with *things that you keep in your pocket* resulting in the same level of output as *vehicle*. Similarly, when item dominance was examined for each category, there was no clear distinction between taxonomic and ad hoc categories. The categories with the highest degree of consensus on the top ten items generated were *fruit, vehicle, things that you* *keep in your pockets* and *things that dogs chase*. The item dominance distribution for these four categories did not differ significantly. This suggests that that in these respects, at least, common and ad hoc categories showed some unexpected overlap.

Most importantly for the perspective taken in this thesis, the protocols provided by participants explaining the strategies that they had used to complete the task revealed an unexpectedly high reliance on personal, generic experiences of situations in which the category members had been encountered. This was equally the case for common categories such as *fruit* and *birds* as for ad hoc categories such *as things people keep in their pockets*. It appeared that in order to retrieve category members from both types of category, participants cued themselves by instantiating environments in which they frequently come across the category, e.g. *fruit*; buying it, preparing it, keeping it in a bowl etc. What is the nature of these memories? They are not episodic in the sense that episodic memories are usually understood to refer to one specific experience. Rather they seem to encapsulate a number of experiences of the same type of episode (shopping in a supermarket, going to weddings etc.). Other researchers have noted these types of memory and have referred to them using varied terminology. Neisser (1986) has used the term "extendure" to refer to repetition of actions and to events that extend over a period (e.g. all the plane trips I have been on). Williams and Dritschel (1992) who noted the prevalence of this type of memory in studies designed to elicit specific autobiographical memories, refer to them as 'general memories'. Pillemer et al. (1986) described a general memory as one that "does not pin down a specific event, the narrative often goes beyond any single happening and presents a theme for numerous events " (p.112). These types of memory are most often mentioned in the literature appertaining to autobiographical memory where they are viewed as errors when the participant has been asked to produce a specific autobiographical memory. Williams and Dritschel further distinguish between general memories that refer to a category of events comprising of a number of specific episodes, e.g., drinking in pubs, and those general memories which cover an extended period of time (lasting more than one day). The first of these they call 'categoric' memories. It would seem that the participants in the free emission task were drawing on 'categoric' memories (as opposed to category knowledge) to aid their retrieval of semantic information.

Predictions from Models (Category Member Generation)

The Funnell (2001) model, described in Chapter 1, would have predicted a difference in strategy for the different category types. The generation of taxonomic category members might have been expected through the 'concepts' level of abstraction, given that this equates to most notions of semantic memory. On the other hand, since ad hoc categories are often more closely linked to events, then their members could be predicted to be generated via general event knowledge. If one assumes that participants are aware of this process then one might have expected to see more mention of events and situations in the reported strategies for these categories than for taxonomic. Looking at the results of the emission study in the light of the Funnell model, it would seem that participants searched their general event knowledge rather than their 'concepts' to produce category members for both types of category. Using the Funnell model, it could be argued that the level of abstraction that is utilised in achieving a cognitive task would be the one most suited to meeting the aim of that task. It is easy to see how generating members of an ad hoc category of *things people take to a wedding* could be achieved by consulting one's knowledge of this type of event in non-specific terms. However, it is harder to see why the level of general event knowledge was also the one used most frequently to generate members of taxonomic categories. A possible explanation for this may arise from the Gentner and Brem (1999) study mentioned in the discussion of Chapter 3. This showed that when participants were put under pressure of time to decide which of two items (e.g., bullseye or javelin) was most similar to a target item (e.g., dart), their performance times for picking the 'correct' match (the taxonomically related item) was slower in the presence of a themed alternative than in the presence of an unrelated foil. Gentner and Brem claim that "thematic influences rise with increased cognitive load" (p.179). This could imply that the most effortful access to object knowledge is at the 'decontextualised' level and accessing object knowledge via the general event level is relatively effortless. Perhaps this is why participants drew from the same event level knowledge base to produce members of taxonomic categories. In couching her model in terms of 'levels', Funnell has implied a hierarchical relationship between the different degrees of abstraction with 'concepts' being the highest level. This is presumably because this level would be the last to develop (information about regularities of objects would be drawn from repeated episodes of interaction with those objects, where the repeated episodes would themselves consist of a number of single episodes). This would also fit with the observation in dementia patients that the 'top' level is the first to be lost. However, this relationship requires testing against alternatives. Since non-clinical samples appear to draw on general event knowledge to perform tasks such as category generation for all types of category, then it would seem that the 'concepts' level is not necessarily privileged to deal with semantic tasks. Although the post hoc predictions have not been entirely supported, it could be considered that the results offer support for the need to recognise close links between general event knowledge and knowledge of objects.

The Barsalou position (1999) would predict that when listing category members, concepts should be simulated in settings and situations and one might therefore expect to see mention of these settings in the strategies reported by the participants. Barsalou's claims about situated concepts do not seem to distinguish between taxonomic and ad hoc representations and so the same strategy would be expected for both types of category. This was indeed what was found and Barsalou has interpreted the published version of results reported in Chapter 2 (Vallée-Tourangeau, Anthony & Austin, 1998) as demonstrating, amongst other studies, that categories are situated in background settings (Barsalou, 2003). He points out that work on visual object processing also demonstrates the strong links between objects and their settings. If Barsalou is on the right track then presumably the category name, e.g., *fruit* was simulated in a variety of settings, each of which was then searched for specific category members. This seems consistent with the strategies reported by participants - buying it, making fruit salad etc. Under Barsalou's theory, the simulations would also include information about the interactions of agents with the items (i.e., motor routines), emotional responses as well as visual, auditory and tactile information. This might explain why some participants employed such strategies as "fruit I like", "fruit my children like" and why so many of the strategies involved actions.

Similarity Judgements

The effect of context

In Chapter 3, the effects of type of category, context and typicality on ratings of similarity for pairs of everyday items were reported. Although the main purpose of this study was to collect protocols of the judgements in order to look for the influence of situational knowledge on a 'semantic task', the quantitative data showing the constraints upon the effect of context on similarity judgements was also of interest. It was found that although the provision of context (in the form of the category name) enhanced the similarity ratings for pairs of items from ad hoc categories pairs (as was the case in Barsalou, 1982), this effect applied only to the most typical items and only when participants were obliged to make relatively speeded judgements (the no self-report condition). These results will be re-visited in the light of both the Funnell and the Barsalou theories before considering thematic similarity more generally.

Barsalou (1982) interpreted the effect of context on the perceived similarity of ad hoc category members as a demonstration of context-dependent property activation. *Jewellery* and *pets* may not seem very similar in the absence of any context but the common property of being 'things people hold as precious' will render them more similar if thought of in the context *of things people save from their house in a fire*. It could be also argued that when the context, i.e., the ad hoc category title, is provided, it is activating knowledge at the general event level in Funnell's model. Consider again the 'contexts' provided in Chapter 3:

> *things you find in an attic, things you find on a shelf, things people take to a wedding, things that may be conveniently kept in pockets, things dogs chase and things sold on the black market in Russia.*

All of these constitute situations or settings in which items may or may not appear. It is certainly the case that certain situations will highlight different properties of an item (arranging flowers in a vase is unlikely to render 'roots' salient but planting flowers in a garden would be more likely to do so) but since the properties shared by many ad hoc category members tend to be few and very broad (*things dogs chase* may share the property of 'can move or be made to move'), one questions whether the highlighting of these characteristics would be sufficient to raise ratings of similarity by the observed amount. If one accepts that many ad hoc and goal derived categories <u>are</u>, or involve, situations then it may be that the realisation that one has, or can, or will come across both items in the situation under consideration that is in itself the basis for the feeling of similarity. Under this view, the sense of similarity arises from association at the general event level of memory rather than at the most abstract (concept level) of the model and is prompted by the provision of the name of the event itself (the category label).

The similarity ratings for the taxonomic items seemed to be unaffected by the provision of a category label. Why would this label not act in the same way as the ad hoc category name? The answer may lie in nature of the category labels themselves. Although both are presented in this study (and that of Barsalou) as providing 'context' - they can not be seen as equivalent if context in this instance is understood to be information about situations. The category names for taxonomic categories, unlike ad hoc category names, do not represent situations. Exemplars of *fruit, vehicle, furniture* and *bird* will indeed feature in general event knowledge, as shown in the first two studies, but the superordinate label itself would, at best, only be tied to extremely broad events (eating, travelling etc.). Any additional similarity to be felt due to their presence at the general event level of knowledge is therefore unlikely. Can this interpretation explain the rather surprising finding that in the self-report condition, the taxonomic pairs were rated as more similar in the <u>absence</u> of the category name? Possibly. It has just been proposed that the presence of the category superordinate (e.g., fruit) does not activate general event knowledge at a sufficiently specific level as to enhance similarity. Indeed, its diffuse association with a large range of events could in itself inhibit the activation of any more specific event knowledge. When participants perform the rating task quickly and unreflectively, the items are judged purely on the basis of their shared properties. However, when the participants have the chance to consider their judgement more carefully, it is possible that the co-presentation of individual fruits, for example, apples and bananas, may well activate a more specific situational script that then enhances the rating. This may be suppressed in the presence of the superordinate label. These explanations are seductive but it should be noted that no specific evidence for them was found in the protocols. It is of course possible that such considerations did take place but were not reported. Further protocol collection under different types of instruction may serve to elicit the evidence required to support these speculations.

Situated Conceptualisations

How do the effects of context and typicality, discussed in Chapter 3, fit into Barsalou's theory of concepts as simulators of situated conceptualisations? Under Barsalou's theory, the presentation of a concept name will evoke a situated conceptualisation. The comparison of two concept names would presumably involve a comparison of these conceptualisations. This could involve comparing the intrinsic features/properties of the items, comparing motor routines that drive use of the items and comparing the most common settings in which they have been experienced. Barsalou claims that the particular situated conceptualisation that is activated depends upon context. That would suggest that when the category label for ad hoc category pairs is provided, such as *things you keep on a shelf* alongside *photo* and *clock,* then the conceptualisation of those items in that particular setting is brought to mind and one does not therefore think of the setting of photos in an album or a clock beside your bed. Given that the same setting can reasonably surround each item, this forms the basis for similarity. However, why should the co-presentation of these items not suggest the same, shared, setting when not explicitly provided? The protocols suggested that for many participants this did happen and that may explain the loss of the effect in the self -report condition. Perhaps it can be inferred from the differing results in the no self-report condition that self-initiation of situational information takes some time and that under relatively speeded conditions, the comparison tends to be limited to the items themselves unless the situational information is already provided.

With regard to the ratings for the common category members, a similar story to the one drawn from Funnell's levels of meaning model could be told from the Barsalou perspective. The superordinate category may be associated with too many diverse settings for any one situated conceptualisation to apply and so comparison may be limited to intrinsic features of the objects. However, under slower, more reflective conditions, the presentation of the exemplars without the superordinate have the potential for allowing a comparison of situational information as well as of features.

Effect of Protocol (Time and Effort)

The protocols were collected to permit an investigation of whether or not participants considered situations and settings when judging similarity, However, it transpired that this condition also affected ratings, and in fact seemed to be responsible for eliminating the previously documented effect of context on ad hoc item pairs. It is not known to what extent participants felt that they were justifying their ratings but the lack of difference in ratings between the concurrent and retrospective protocol conditions suggests that the protocol effect arose from the increased time and depth of reflection that giving a protocol entails rather than from an element of post-hoc justification. Recall that the concurrent protocol was made prior to the rating. Interestingly, Wisniewski and Bassok (1999) also included an 'explanation' condition in which participants were instructed to write an explanation for their similarity ratings. In contrast to this study, they found that the pattern and magnitude of the ratings did not vary as a result of giving the protocol. This suggests that their participants' considerations did not differ with or without explanation. How is this finding compatible with the effect of protocol found here? It may be that the use of the word 'explanation' in the Wisniewski and Bassok instructions resulted in a justification of a fairly quick judgement rather than a consideration of new perspectives as has been argued to be the case here. The difference is also likely to be due to the nature of the stimuli. The Wisniewski and Bassok stimuli had been chosen specifically to meet the criteria of their manipulation. That is to say, the thematically linked items were chosen by the experimenter on the assumption that they were could clearly be integrated into a scenario (e.g., milk/cow, sailor/ship, cat/vet) whereas the items from the ad hoc categories in this study were less obviously linked and therefore additional time and effort afforded in the protocol condition could indeed allow a less obvious link to be made.

Thematic Similarity

The analysis of the protocols collected with the similarity ratings discussed in chapter 3 showed that judgements of similarity between two items, made under experimental conditions, often involve more than a weighting of shared and unshared perceptual or functional features. In many cases, other links and associations between the items are also considered to render them similar. Those links frequently result from the items appearing in a common location, situation or event. This did not occur in all cases, but it did occur in one form or another in approximately one third of the protocols collected and this is not easily dismissed. It should be noted that the situations instantiated by participants rarely included personal references of the sort seen in the protocols accompanying the category member generation task. Nevertheless, in many cases they seemed to make use of similar generic event or situational knowledge. The results of Chapter 3 add support to the findings of Wisniewski and Bassok (1999), published after the Chapter 3 data were collected. However, the Chapter 3 study differed from that published work in that the pairs of items were not specifically chosen to be 'thematically related' and no manipulation was grounded in that relationship. This study therefore provides an effective demonstration of the high degree to which thematic similarity is spontaneously produced during rating tasks. There can be little doubt that similarity is not a passive computation of matching and mis-matching features. Perhaps more importantly, the detailed nature of some of the protocols recorded in this study indicated clearly ;

- that participants were very aware of the different 'types' of similarity that were available to them when making their judgements
- that there was considerable individual difference in participants' willingness to base their ratings in this task on thematic similarity.

Neither of these observations arose from the Wisniewski and Bassok study and they may be important for the following reasons. Firstly, as mentioned previously, there have been attempts to portray thematic similarity as the results of unconscious interference from associative processes (Gentner and Brem, 1999). The introspection of some of the participants in the present study shows that there is nothing unconscious about the 'interference'. The conscious knowledge of two different types of reasons for finding things similar may speak to the awareness of two different organisations of object knowledge that may serve different goals. Secondly, the individual differences, which are rarely considered in such studies since the nature of the analysis is to collapse across these, may be predictive of differences in other domains. For example, if an individual is prone to high levels of influence from situational knowledge in this task (i.e., not only notes this similarity but allows it to enhance ratings of perceptually distinct items), then it is possible that this person may allow this conception of similarity to affect their choice of analogies when problem solving in this manner. It is contended then that this study makes a distinctive and useful contribution to the literature despite certain commonalities with the published work.

Both Funnell (2001) and Barsalou's (2003) models of conceptual knowledge could predict thematic similarity. Barsalou's (2003) proposal that any conceptual knowledge is likely to be situated would make it possible that this information could exert an influence on any cognitive operations involving concepts. As discussed earlier, a comparison of situated conceptualisations for the purpose of judging similarity could well include a comparison of those situations and this indeed appears to have been the case. Funnell's (2001) model, however, would predict thematic similarity only if judgements included accessing knowledge at the general event level. If the assessment of similarity took place only at the 'concepts' level, one would expect to see judgements were based only on feature comparison. The presence of thematic similarity therefore implies that this task taps both the proposed concepts level and the general event level.

It is easy to dismiss thematic similarity as a mistake that people make when being asked to judge similarity but this would be to risk overlooking important data. It seems more likely that when people tell you about thematic similarity or allow it to influence their ratings, they are not demonstrating their misunderstanding of the task, but are indicating a 'closeness' of these items in an organisation of knowledge that pivots around experiences of situations in which these items feature. The protocols of some participants in the Chapter 3 study indicate clearly that they are often aware that this is not the same type of similarity that arises from a closeness of items in an organisation based on overlapping perceptual and functional features, but it is, for many of them, similarity. Associations that are predicated on close positions in a hierarchical, feature-based structure of object memory have been privileged in the literature to such an extent that they are rarely referred to as 'associations' at all and this may have resulted in a demeaning of the role played by other types of association in cognition. New models of semantic memory or conceptual knowledge such as those considered here help to redress the balance and in doing so open up new directions for research in concepts.

Factors Determining Thematic Similarity

In the work described in Chapter 4, an attempt was made to start to isolate the factors that may determine the strength of thematic similarity. It is clear that when rating on the basis of thematic similarity, participants find the similarity much more compelling for some pairs than for others. Whilst both the Funnell and the Barsalou models can comfortably explain thematic similarity, neither would specifically predict these differences in ratings within a set of items said to be thematically related. In the introduction to Chapter 4, various types of relationships that can exist between entities were considered as candidates for influencing thematic similarity (spatial, temporal, and causal). The factor that was chosen for investigation was strength of association between the event (situation) and the item. Measures of this association, both item-to-event and event-to-item, were collected and used to manipulate the stimuli in the final study; these stimuli were paired and presented to participants to be rated for similarity to each other. It should be recalled that very few of the item pairs shared any perceptual or functional similarities and were all connected to one of seven events with either high or low frequency.

Despite the comparatively small participant groups providing the measures of strength of association and despite the sometimes small difference in criterion between high and low frequency, the results showed clearly that:

- item pairs that were highly associated to the event (in both directions of association) were rated the most similar and item pairs that were the most weakly associated to the event (in both directions of association) received the lowest ratings of similarity.
- Context (the name of the event) enhanced the ratings but the same effect was seen whether participants saw the items blocked by event or randomised.
- The pattern of ratings for item pairs representing intermediate combinations of levels of association was compelling but difficult to interpret due to the fact that only a subset of the total number of combinations available were used. Nevertheless, it is clearly not the case that any two items appearing in the same situation will receive the same rating of thematic similarity. It will depend upon the likelihood, in the experience of the rater, of them appearing in that particular event. In the terms of the Funnell model, this judgement would take place at the general event level of knowledge.

In view of the fact that the sensitivity to item-event strength was demonstrated even without context – why should providing a context result in overall higher similarity ratings? It can be speculated that perhaps the co-presentation of the event, whilst not suggesting a context that would not otherwise occur for the participant, has the pragmatic effect of confirming the participants' sense of similarity when judging perceptually dissimilar items and thus enhances the ratings.

Implications of the Research

There are a number of reasons why the investigation of concepts or semantic memory has tended to take place with comparatively little reference to other forms of memory. The requirement to identify items in our environment and to ensure a high degree of consensus amongst language users has led to a great deal of attention being paid to the information necessary for classification. Possibly, it is the phenomenology of generating such information that distinguishes it from the effort associated with 'remembering'. One feels that it is not necessary to *remember* that birds have beaks or that sparrows are birds - such propositions have the status of facts which are so common as to make little demand on memory, rather like recognising one's mother's face. It may be the repeated nature of human interaction with the extensions of concepts that gives the information we have about them a different 'flavour' to information which can be recalled from one experience or even from multiple events. Studies such as those of Rosch in which participants are requested to list features of everyday objects or to generate category members are not usually referred to as memory tasks. Instructions to participants tend to be phrased in terms of " list as many members of the category *bird* as you can" - the word "remember" is not included at the end of the sentence. If, on completion of such a generation task, omissions were to be pointed out to the participant, s/he may well explain this in terms of forgetting - "oh, yes...I forgot those". This does not indicate that the participant forgot that those members were birds but that they forgot to mention them in response to the experimenter's instruction. Nevertheless, knowledge beyond sensory experience is available to us only through the facility of memory and so one should be wary of investigating 'types' of memory without being constantly mindful of their inter-relationships.

A key issue in the interpretation of the results of this thesis is the extent to which the tasks used are believed to reveal underlying organisation in memory. For example, are the strategies that participants report using to achieve the free recall of category members indicative of a particular organisation? As stated in Chapter 2, there is precedent for inferring organisation from the nature of the clusters of outputs from free recall of learned lists (Barsalou and Sewell, 1985; Chase and Ericsson, 1981; Graesser and Mandler, 1975; Mandler, 1967; Reitman, 1976; Tulving, 1962,1964, 1966; Tulving and Pearlstone, 1966) and to a lesser extent from free recall of categories (Rips, Shoben and Smith 1973; Henley, 1969). It is also a plausible argument that it would be maximally effective to tap into an existing organisation to access information about category members rather than to deploy some other set of self-generated cues that are not reflected in that structure. It is interesting to recall that Walker and Kintsch (1985) documented a high incidence of 'experiential clusters' in the generation output of participants' recall of categories (77%) but when the same participants later sorted their own output, the tendency was to use semantic groupings (86%). This suggests different strategies for each task; is it the case that one of these strategies takes advantage of underlying knowledge structures and the other does not? This seems unlikely – in which case it must be that each task is drawing on a different organisation.

In this thesis, it has been shown that adults' knowledge of the everyday settings and situations that surround objects and entities in the world plays a substantial role in certain cognitive operations that might have been expected to depend only upon de-contextualised conceptual knowledge. The strong view that could be drawn from this evidence is that 'semantic' memory is inseparable from generalised event memory and that knowledge of objects and their features is never activated in isolation from a situation. This seems to be the position adopted in Barsalou's (2003) theory of situated conceptualisation. A weaker claim based on the same evidence would be that the organisation of object-knowledge around situations and events is an alternative organisation to that of objects based on shared-features and that adults switch between these knowledge bases according to the goal of the particular task. This would be more compatible with Funnell's (2001) model. The studies conducted here were not designed to test these alternatives and the results of the category generation and similarity tasks are open to explanation from either perspective. It would seem that the results from the final study, looking at the determination of thematic similarity strength suggest a further level of complexity of situational knowledge that is not specifically addressed by either of these models.

Critique of the empirical work

In both the category generation task and the similarity ratings task, the claims in this thesis concerning the role of situational knowledge flow directly from accepting that these studies have a) revealed something about how participants performed these tasks and b) that how they claim to have performed them reveals, in turn, something of import concerning the organisation of conceptual knowledge. Each of these premises is open to challenge. The argument concerning the reliance (or otherwise) that can be placed on protocols as windows upon cognitive processing was considered in some detail in chapter one and will not be re-run here. In reviewing a new edition of Ericsson and Simon's seminal work concerning verbal reports as data, Crutcher (1994) documents a growing body of work that views protocols as offering insight into cognitive processes and discusses the increasing level of credibility that such data is accorded. Crutcher also points out that thought sequences can be accurately reported by participants retrospectively if the task is brief and the protocol is recorded as soon as the task is completed. Both of these conditions were met by the design of the category generation task in chapter 2 and the similarity rating task reported in chapter 3.

A number of reasons were also provided in the discussion of chapter 2 as to why the protocol contents accompanying the category member generation task should be considered genuinely informative of cognitive processes and indeed, there is good precedent for believing that clusters in output from free recall can reflect underlying structure of memory. Nevertheless, this study was unfortunately not designed to permit validation of the strategies claimed in the protocols through examination of the output. Audio recording of the task performance would at least allow the identification of clusters of items through temporal pauses although it may still be difficult to subsequently confirm that any one cluster was indeed the result of a particular strategic search due to the personal nature of some of the strategies claimed. An improved approach may be to assume the validity of the strategies claimed and to require the task to be performed under a variety of conditions that one could predict to interfere with this method of retrieval. For example, the task could be accompanied by images of category members (unidentifiable birds flying in the distance during the generation of *Birds)* versus images of a compatible situation (e.g. a garden) versus images of incompatible situations (e.g. hospitals). If the presence of an incompatible situation image reduces fluency or results in a longer time to generate a comparable number of items compared to the other two conditions then one's confidence of the instantiation of situations as a means of retrieval may be increased.

It would also be desirable to collect the data in such a way as to be able to check the statistical differences between types of protocol more easily – the collection of nominal data with attendant difficulties of non-independence of cells leads to problems in analysis.

The requirement to provide a protocol for any task can potentially change the way in which the task is carried out; this may be because of the pragmatics of the situation i.e. some participants did feel comfortable with offering no real explanation of how they produced the category members (just wrote down the ones I knew – just popped into my mind) and so it could be argued that if that was what actually happened for all participants on most occasions then they could easily have said so. However, being asked to explain one's own performance repeatedly during the study may well result in participants feeling pressurised to provide 'better', more detailed, or even just more interesting explanations. Alternatively (or indeed additionally), the extra time afforded by the protocol and the attendant opportunity to reflect upon the task more carefully may actually result in the task being undertaken in a different manner altogether, as was demonstrated by Smith and Sloman (1994) and discussed in Chapter One. This indeed appeared to be the case in the study reported in Chapter 3. The effect of providing a context for the ad hoc category members, which enhanced similarity ratings under relatively speeded conditions, was not evident when the protocols were recorded. An examination of the protocols did not serve to entirely explain this finding.

The effect of (speak aloud) protocols in the Smith and Sloman (1994) study was to apparently encourage rule-based reasoning about whether an item belonged to particular category rather than making the same decision based purely on similarity. It could be argued that the participants in the similarity study reported here, although engaged in a different task to Smith and Sloman's participants (my participants were not required to make categorisation judgements at all), were drawn, through the protocols, into reasoning about the task in a way that would not have occurred in the no self-report conditions. That is to say, they possibly reasoned that they would not be asked about the similarity between apparently dissimilar items if there were not some similarity to be found and that their effort to find such similarity led them to make use of thematic similarity when they would not perhaps have done so otherwise. Whilst this interpretation seems plausible, it is not consistent with the findings of Wisniewski and Bassok (1999); an enhancing effect on similarity ratings for items that were thematically related could be seen in their study whether protocols were recorded or not.

The inclusion of typicality as a factor in study 3, whilst well-motivated, was not entirely successful in that the differential effects of context and protocol at the different levels of typicality that showed in analysis at the level of participant became less clear when examined at the level of item. Although the item effects did not swamp the effects of protocol and context, they do cast some doubt on the generalisability of the findings to other item pairs and much more work is needed on a wider range of items before more confidence can be placed in the conclusions. That said, the findings from the final study show that thematic similarity itself varies depending at least partly upon the strength of association between the items under consideration and the event/situation that into which they are being integrated. It may be the case then that for thematically related items, such item effects are inevitable if not systematically controlled.

To return to my second premise, even if the protocols in these studies truly reveal the thought processes of the participants- need this necessarily have any impact on the theories of organisation? The strategies reported for the category generation task may just be effective, context-dependent cues. The items could be contextualised at retrieval, having been retrieved from a context-free storage. It is difficult to argue conclusively that these interpretations are mistaken and it is challenging to devise experiments that will help to decide upon these issues but let us imagine that context-free organization of conceptual knowledge is at some time in the future shown to be the case- if conceptual information is consistently contextualised at retrieval then it will still be necessary for the use of this situational information to be carefully documented as to its role in various cognitive operations.

The final studies did not rely on protocols but other problems arose. The frequency of generation from event-to-item and item-to-event was a fairly crude measure of item/event association. The position of the item in the list generated was not used as an as indicator of association strength, neither was any account made of the extent to which any one item might be associated with other events (*bucket* may reliably evoke beach but may equally reliably evoke window cleaning). Other, more complex measures that may include these possibilities should be developed since the effect of the measure used may have underestimated the role of item/event association in influencing thematic similarity. Possibly picture based measures that assess levels of surprise at the presence of items in a scene could be utilised. Arguably, pictures could be considered more appropriate in capturing situational aspects of an item than word cues in generation tasks.

As noted in the discussion to Chapter 4, a larger corpus of normative data for item/event association is required.
This would allow for greater flexibility in testing a greater combination of item pairs and would also allow for word association to be controlled for a priori and items chosen on this basis rather than checking post –hoc as was the case here. It would have been preferable if the event -to–item data could have been collected from the same number of participants as the item-to event data but the large number of items required in the former data collection necessitated a considerable number of participants even in order to obtain data from a sample of 15 for each item.

There are clearly alternative interpretations to be made of the data presented here that would not support the argument for situational information as an organising factor in conceptual knowledge. In studies 1 and 2, the retrieval strategies used by participants may be just that – retrieval strategies and may not necessarily indicate anything new about the structure of conceptual knowledge. The presence of context in Study 3 may act to render particular properties more salient as suggested by Barsalou's (1982) paper; the participants producing thematic similarity in Study 4 may simply be misguided in their understanding of the term similarity and the results from Study 5 may simply reflect associations between the pairs of words themselves that undoubtedly exist. However, it is contended that taken together, and viewed in conjunction with other recent empirical work (Wisniewski and Bassok 1999; Lin and Murphy, 2001; Murphy 2001) these studies should be seen as having demonstrated that situational knowledge of objects plays a role in adult cognition that has been, at worst, ignored and, at best, under-specified until quite recently. The degree to which this limitation may have constrained our understanding of other cognitive processes that rely upon concepts, such as reasoning and problem solving, should now be established.

Future Research

Other Directions for Research

If one accepts that information that was not previously seen as part of a 'core concept' consistently plays a role in cognitive tasks, then it seems unhelpful to cling to separate notions of decontextualised concepts and (unspecified) associated information as separate knowledge organisations. It would seem more elegant to accept that a concept comprises both the abstracted feature descriptions and the information of background settings, motor routines etc. Similarly, it seems to be time to extend our understanding of semantic memory along the same lines. These ultimately may turn out to be arguments about terminology and in my view, such arguments need not bear on the direction of empirical research which should continue to garner further evidence for, or against, the 'situated concepts' position. However, I believe that the research should also now move on to examine and test the implications of such a view by addressing the following questions:

- Which subsets of the full range of conceptual information available to a person become activated by which tasks and for which purposes? Are there any tasks in which only decontextualised object knowledge is utilised?

- How does conceptual information about items in situations affect other cognitive processes believed to rely upon concepts? If it affects perception of similarity, does it affect reasoning by analogy or inductive reasoning?

- Are there individual differences in the extent to which possibly inappropriate or unhelpful conceptual knowledge is permitted to enter into reasoning about objects/items?

- Can this new view of conceptual knowledge shed light on cases of semantic impairment? Funnell has already suggested that a broader conception of semantic memory can be useful in predicting and understanding patterns of performance and behaviour in semantic dementia.

Ongoing Work

Through the supervision of undergraduate and postgraduate work, I have started to pursue some of the issues raised above. We have made use of insights into thematic similarity to re-interpret over-inclusive and under-inclusive categories formed by patients suffering from schizophrenia. Performance on sorting tasks has often been judged as 'incorrect' if it is not based on standard taxonomic relationships between the items. However, the patient samples have been shown to make use of thematic similarity in sorting tasks and to note the same types of thematic similarity as normal participants in protocols that accompany a similarity judgement task (Green, 2002). It has also been demonstrated that schizophrenia patients tend to allow thematic similarity to influence their ratings to a higher degree than do normal participants and current work suggests that they may make use of thematic organisation even when it is not considered appropriate to the task by normal controls. This interpretation of data from schizophrenic performance on semantic tasks suggests that discrepancies between this and the data obtained from control groups may not indicate a difference in semantic structure and category boundaries, as is usually suggested, but may reflect their use of an event-based organisation of knowledge in tasks that controls tend to use a taxonomic organisation to tackle.

A recent undergraduate project examined the effect of a heightened perception of thematic similarity on problem solving through example. Participants were classified on the basis of their tendency to make thematic similarity judgements. All participants were trained on probability principles using a training example and then asked to solve new problems. The new problems could have the same or different story line to the training example and could also have the same or different object correspondence. It was found that all participants performed worse on the problems with reversed object correspondences but that the thematic raters performed even worse when this condition was combined with a dissimilar story line. This finding can be explained in terms of the thematic raters being less able to see the similarity between the test and training problem when the story line differs. It seems that differences in perception of thematic similarity may well affect performance on other tasks where similarity plays a role.

Follow-on Studies

The overlapping nature of many of the relationships that can exist between two items makes it difficult to extricate any one particular relationship and to argue for its unique influence on organising conceptual knowledge and hence of cognitive processes. The sharing of perceptual and functional features has been the most successful relationship to date in providing explanations of the way in which people seem to think about and use concepts and so structures based on these relationships have enjoyed a virtually unrivalled supremacy in many models of semantic memory.

It has been argued here that the demonstration of the influence of the relationship that results from items regularly co-occurring in common situation and events on simple cognitive tasks requires that we look more closely at this type of association. The last study reported here attempted to start to break down the source of thematic similarity by seeing whether it could be explained by the strength of item-to-event and event-to-item association. As mentioned in the discussion of Chapter 4, there are a number of ways that this study could be extended to consolidate the findings. Firstly, a manipulation based on situational association versus word association is required to rule out the possibility that the findings were the due to the latter rather than the former. Secondly, additional data could be collected from a larger group of participants for a small range of events in order to obtain a full set of combinations of the event-to-item and item-to-event indices. This may enable a clearer picture to emerge of the ratings for the items that have intermediate strength of association to the event and to thus reveal the extent of participants' sensitivity to less dramatic differences in the item-to-event and event-to-item strength of association.

Further studies to collect other types of data that may be used to model thematic similarity more precisely are also called for. I am in the process of collecting data concerning the temporal contiguity of items within an event that I suspect will also contribute to their perceived similarity. For example, *kettle* and *sugar* may be viewed as less similar than *kettle* and *tap* as the former occur at the beginning and end of a 'making a cup of coffee' scenario whereas the latter both occur near the beginning. There is a problem of separating this relationship from that of spatial contiguity and reliable methods of measuring spatial contiguity are under deliberation. Since the judgement of similarity is proving to be such a complex process, this fine-grained level of examination of the factors that affect is deemed to be necessary.

It is also desirable that alternative methodologies should be utilised to further establish the extent of influence exerted by a situational organisation of conceptual knowledge. One such possibility is the use of a priming paradigm.

Priming Studies

Priming studies have a long history of being used to test the strength of association between items. The task most frequently used in connection with this paradigm is the lexical decision task whereby a participant has to identify a letter string as a real word or a non-word. If the task is facilitated by the pre-presentation of a different word (the prime) then it is inferred that the two words are associated within the lexicon. Priming effects have been found to arise from a variety of relationships between the prime and the target such as phrasal associations (bee- hive), synonyms and category co-ordinates (bee-wasp). Considerable attention has been paid to the priming effects found for shared category membership since this has been taken as further evidence of the status of semantic (taxonomic) categories as organising structures in semantic memory. Different manipulations have been adopted in an attempt to establish whether the facilitation from this type of relationship provides support for a holistic model of semantic memory (such as Collins and Loftus, 1975) or whether is more consistent with a distributed model. The difference between these two models is that nodes in holistic models represent entire lexical units whereas in distributed models the nodes of the network are individual features (e.g. Moss, Hare, Day and Tyler, 1994). By attempting to separate the prime/target stimuli pairs into those that share features and are additionally associated, those that share features and are not additionally associated and those that are associated but have no overlapping features, the applicability of each of these models can be tested. If priming occurs for 'overlapping feature only' pairs but not for 'associative only' pairs then this is taken as evidence for the distributed model. In reviewing some of the studies that claim to have demonstrated either 'pure' semantic priming or 'pure' associative priming, Hutchison (2003) points out a consistent difficulty with the creation of stimuli for studies of this type, namely that of discriminating between feature overlap and associative strength. As mentioned in the previous chapter, for the purposes of this line of research, 'strength of association' between a pair of words is usually measured by the percentage of respondents who give one of the words as their first response in free association to the other word. However, Hutchison (2003) suggests that the assumption that such association norms arise principally from phrasal contiguity between the items may be mistaken since a close examination of the norms reveal a variety of other associations between the items including natural category membership, functional relations and script relations. These relationships could all be reasonably classed as semantic, even though this term is frequently reserved for relationships based on shared features. It follows then that researchers using these norms as a measure of non-semantic association in order to create stimuli to compare against semantically related items will be confounding the manipulation. Hutchinson contends that many priming studies have suffered from this problem. A clear demonstration of the priming potential of pairs that are truly not associated or are only weakly associated according to association norms is required before one can suggest that their relationship has particular significance within semantic memory or the lexicon. Since associated items are invariably related semantically as well as associatively, one would expect their potential for priming to be the result of both components of their relationship (with semantic relations being additionally 'boosted' through association). Items that are only semantically linked but are not highly associated should show priming due only to their semantic relationship which should be less than that of the associated pairs. This was convincingly demonstrated in a study by Moss, Ostrin, Tyler and Marslen-Wilson (1995) that clearly distinguished between semantic relatedness and association (based on 'first-word - generated' norms) and also distinguished between three types of semantic relatedness, category-coordinates, functional relatedness and instrumental relatedness. Moss et al.

found that across three experiments using different modes of presentation, pairs that shared 'instrumental' relationships (e.g. *hammer-nail, broom-floor*) showed robust priming effects even when they were not associated. When compared to the priming effects found for artifact category coordinates, this finding leads the authors to claim that functional information about artifactual items may be more easily accessed than category information and a case is made for raising the status of this type of relationship. Since this thesis seeks to make to make a similar plea on behalf of items related 'situationally', it follows that a similar demonstration of this relationship in priming studies is called for. Interestingly, the Moss et al. study also included script-related pairs (such as *restaurant-wine* and *circus-lion*). These were found to show priming effects when presented as an auditory stimulus but not when presented visually (even when associated). Moss and her colleagues argue that this may mean that there is no automatic priming from these relationships or that the time course of access to this type of relationship is slower than for the functional information. There are several comments to make upon this result in relation to this thesis.

- In the Moss et al. study the item pairs consisted of the situation itself and atypical item that would appear in a script for that situation and not of two items that share a common script (such as the pairs used in Chapter 4). Although one might expect *restaurant* to prime *wine* in the same way that *fruit* will prime *apple* –it may be that the relationship between event names and items that feature in them do not share the same type of superordinate relations.
- Many of the primes in this subset of materials were locations rather than events. Whilst it may be argued that events and places are invariably entwined, it is possible that naming a location such as 'hospital' does not evoke the kind of specific simulation of a situation that is required to link the items together in the way that would be necessary to demonstrate the organising role of situations.
- When the instrumentally linked stimuli are examined, it can be seen that many of these pairs also suggest situations, e.g., razor-leg (shaving), drill-teeth (at the dentist), spoon-dessert (having a meal), hammer –nail (fixing or making), pub-beer (going for a drink), kitchen-sink (washing up), butcher-meat (shopping), kettle –tea (making a drink). Indeed, these are very similar pairs as those used by Wisniewski et al. (1999). One would expect these pairs to be viewed as 'thematically similar' and as has been repeatedly claimed throughout this thesis,

the basis of thematic similarity is the co-occurrence of these items in a situation or scenario. The nature of the relationship between these items is complex and bears further scrutiny.

It is suggested, then that further investigation of the priming potential of items that are 'situationally' related may serve to test whether co-occurrence in everyday situations provide links that 'more than associative'. If 'pure' semantic priming raises the status of feature based organisations, and priming from instrumentally linked pairs can raise the status of functional relationships then a convincing demonstration of priming from purely situationally related items could be used to argue for the status of situational organisation.

Summary of Main Findings

The presentation of this body of empirical work has been set in the context of recent changes in the way some researchers wish to think about the ways in which we store and access information about the things that surround us in our environment. I have argued that the findings are in line with a broader view of semantic memory than as a warehouse of facts about what objects are like. However, there are also new and specific findings arising from the data that have extended our knowledge about the two tasks utilised, category member generation and judgements of similarity under experimental conditions:

- The category generation study was the first to provide an in-depth examination of the strategies that participants claim to use when generating members of a wide range of categories (Ch.2).

- The protocol analysis for the above uncovered the use of everyday situations to cue retrieval of category members (Ch.2).

- Taxonomic and ad hoc categories vary less than might have been expected in terms of participants' ability to generate members; the degree of consensus on the items generated and the strategies that participants report having used to generate their members (Ch.2).

- The quantitative data from the similarity judgement study showed that important qualifications should be made to a much cited study (Barsalou, 1982) concerning the effect of context on similarity of ad hoc item pairs, namely that this effect appears to manifest itself only when the speed of judgements allows little chance for reflection and that it is strongest when items are comparatively typical of the ad hoc category (Ch.3).

- The protocols accompanying the similarity judgements showed evidence of participants spontaneously and consciously including situational links between the items in their consideration of similarity (Ch.3).

- There was evidence of individual differences in the extent to which the above links were viewed as constituting 'similarity' (Ch.3).

- The first known attempt to examine the factors that govern the strength of thematic similarity was made via a manipulation based on item-to-event and event-to-item association (Ch.4).

- Previous suppositions that thematic similarity is based on items being integrated into a common scenario or situation were empirically supported through the manipulation based on item-to-event and event-to-item association (Ch.4).

- If treated with caution, written protocols may be used to shed light on the way in which participants achieve certain simple tasks.

Furthermore, the studies have raised new research questions that may ultimately help to test recent theoretical positions regarding conceptual knowledge.

References

Ahn, W., Kalish, C, Gelman, S.A.,Medin, D.L.,Luhmann, C.,Atran, S.,Coley, J.D., & Shafto, P. (2001). Why essences are essential in the psychology of concepts. *Cognition, 82*, 59-69.

Allen, S.W., & Brooks, L.R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3-19.

Anderson, J.R., & Ross, B.H. (1980). Evidence against a semantic-episodic distinction. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 441 – 465.

Anderson, J.R., Kline, P.J.,& Beasley. C. M., Jr. (1979). A general learning theory and its application to schema abstraction. In G. R. Bower (Ed.). *The psychology of learning and motivation* (Vol 13. pp 277-318). San Diego.CA: Academic Press.

Armstrong, S.L.,Gleitm,an, L.R.,& Gleitman, H. (1983). On what some concepts might not be. *Cognition, 13*, 263-308.

Barsalou, L.W. (1981). *Ad hoc categories and cross classification.* Unpublished manuscript

Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition, 10*, 82-93.

Barsalou, L.W. (1983). Ad Hoc categories. *Memory and Cognition, 11*, 211-227.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Barsalou, L.W. (1987). Theinstability of graded structure in concepts. In U.Neisser (Ed.), *Concepts and conceptual development* (pp 101-140). Cambridge: Cambridge University Press

Barsalou, L.W. (1991). Deriving categories to achieve goals. In G.H.Bower (Ed.) , *The psychology of learning and motivation: Advances in research and theory* (Vol.27, pp. 1-64). San Diego, CA: Academic Press

Barsalou, L. W. (1992). *Cognitive psychology* (ch. 7). Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L.W. (1993). Flexibility ,Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional system of perceptual symbols. In A.C. Collins, S.E. Gathercole, M.A. Conway & P.E.M. Morris (Eds.), *Theories of memory* (pp. 29-101). London: Lawrence Erlbaum Associates.

Barsalou, L.W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 633-660.

Barsalou, L.W. (2003). Situated Simulation in the Human Conceptual System. *Language and Cognitive Processes*, 18, 513-562.

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

Barsalou, L. W., Yeh, W., Luka, B. J., Olseth, K. L., Mix, K. S., & Wu, L.-L. (1993). Concepts and meaning. In K. Beals, G. Cooke, D. Kathman, S. Kita, K.-E. McCullough, & D. Testen (Eds.), *Chicago Linguistic Society 29, Vol. 2*: Papers from the parasession on the correspondence of conceptual, semantic and grammatical representations (pp. 23-61).

Bassok, M., & Medin, D. L. (1997). Birds of a feather flock together: Similarity judgments with semantically rich stimuli. *Journal of Memory and Language, 36,* 311-336.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80* (3, Pt. 2).

Bousfield, W. A, (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49,* 229-240.

Bousfield, W. A., & Sedgewick, H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology, 30,* 149-165.

Brooks, L.R. (1987). Decentralized control of categorization: the role of prior processing episodes. In U.Neisser (Ed.), *Concepts and conceptual development* (pp 141-174). Cambridge: Cambridge University Press

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

Chase, W.G., & Ericsson, A. (1981). Skilled memory. In J.R.Anderson (Ed.),*Cognitive skills and their acquisition* (pp.141-190). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cofer, C.N. (1965). On some factors in the organizational characteristics of free recall. *American Psychologist, 20,* 261-272

Cofer, C.N., Bruce, D.R., & Reicher, G.M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology, 71,* 858-866.

Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240-247.

Collins, A.M., & Quillian, M.R. (1972). Experiments on semantic memory and language comprehension. In L.W. Gregg (Ed.), *Cognition in learning and memory*. Oxford, England: John Wiley & Sons.

Collins, A.M., & Loftus, E.F. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82*, 407-428.

Conrad, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology, 92*,149-154.

Conway, M. A. (1987). Verifying autobiographical facts. *Cognition, 25*, 39-58.

Conway, M. A. (1990a). Associations between autobiographical memories and concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 799-811.

Conway, M. A. (1990b). *Autobiographical memory*. Milton Keynes, UK: Open University Press.

Conway, M. A. (1990c). Conceptual representation of emotions: The role of autobiographical memories. In K. J. Gilhooly, M. T. G. Keane, R. H. Logie, and G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought, Vol. 2*. (pp. 133-143). Chichester: John Wiley & Sons.

Crowder, R.G. (1976). *Principles of learning and memory*. Hillsdale NJ: Lawrence Erlbaum Associates.

Crutcher, R.J. (1994). Telling us what we know: The use of verbal report methodologies in psychological research. *Psychological Science, 5*, 241 – 245.

Dunham, P.,& Dunham,F. (1995). Developmental antecedents of taxonomic and thematic strategies at 3 yrs of age. *Developmental Psychology, 31*, 483-493

Ericcson, K.A., & Simon, H.A. (1980). Verbal reports as data. *Psychological Review, 87*, 215 – 251.

Farah, M.J., & McClelland, J.L. (1991). A computational model of semantic memory inpairment: Modality specificity and emergent category specificity. *Psychological Review, 120*, 339-357.

Fodor, J.A. (2004). Having concepts: a brief refutation of the twentieth century. *Mind and Language,19*, 29-47.

Fodor, J.A. (1998). *Concepts: Where Cognitive Science went wrong*. Oxford: Clarendon Press.

Fodor, J.A.,& LePore, E. (1996). The Red Herring and the Pet Fish: Why concepts still can't be prototypes. *Cognition, 58*, 253-270.

Friendly, M. L. (1977). In search of the M-gram: The structure of organization in free recall. *Cognitive Psychology, 9,* 188-249.

Franks, B., & Braisby, N.(1997). Concepts in Action: The Evolutionary Role of Concepts and Similarity. *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (*SimCat97*) (pp.91-97) . Edinburgh: Department of Artificial Intelligence Edinburgh University.

Franks, J.J., & Bransford, J.D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology, 90,* 65 –74.

Funnell, E. (1996). Semantic Dementia: *A role for goals and scripts in the memory for object concepts in every-day situations*. Paper presented to the British Neuropsychological Society Annual General Meeting, London.

Funnell, E. (2001). Evidence for scripts in semantic dementia: Implications for theories of semantic memory. *Cognitive Neuropsychology*, *18*, 323-341.

Garrard, P., Ralph, M.A.L.,Hodges, J.R., & Patterson, K. (2001). Prototypicality,distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and non-living concepts. *Cognitive Neuropsychology, 18*, 125-174.

Gelman, S.A., & Wellman, H.M. (1991). Insides and essences: early understandings of the non-obvious. *Cognition, 38*, 213-244.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*, 155 –170.

Gentner, D. (1989). The mechanisms of analogical learning. In Ortony, A.,& Vosniadou, S. (Eds.), *Similarity and analogical reasoning* (pp. 199-241). New York, NY: Cambridge University Press.

Gentner, D., & Brem, S.K. (1999) Is snow really like a shovel? Distinguishing similarity from semantic relatedness . Proceedings of the 21st Cognitive science Annual Conference. NJ, US: Lawrence Erlbaum Associates.

Glenberg, A.M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1-55

Goethals, G.R., & Reckman, R.F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology, 9*, 491-501.

Goldstone, R. L. (1994). The role of similarity in categorization: providing a groundwork. *Cognition, 52*, 125-157.

Goldstone, R. L. (1995). Mainstream and avant-garde similarity. *Indiana University Cognitive science Technical Report 132*

Goldstone, R. L., & Medin, D. L. (1994). The time course of comparison. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 29-50.

Goldstone, R.L.,& Barsalou, L.W. (1998). Reuniting perception and conception. *Cognition, 65*, 231-262.

Goodman, N. (1970). Seven strictures on similarity. In L. Foster & J.W. Swanson (Eds.), *Experience and Theory*. Amherst: University of Massachusetts Press.

Graesser, A. II, & Mandler, G. (1978). Limited processing capacity constrains the storage of unrelated set of words and retrieval from natural categories. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 86-100.

Green, N.E.A. (2002) *Semantic memory anomalies in patients with schizophrenia: A cognitive psychological approach*. Unpublished doctoral dissertation.

Grice, H.P. (1975). Logic and Conversation in P. Cole and J. L. Morgan (Eds.), *Syntax and semantics III: Speech Acts*. New York: Seminar Press

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12*, 550-561.

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 225-240.

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

Hassin, R. (1997). Similarity and difference oriented similarity comparisons. *Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation (SimCat97)*. Edinburgh. Department of Artificial Intelligence Edinburgh University.

Henley, N. M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior, 8*, 176-184.

Herz, R.S. (1997). The effects of cue distinctiveness on odor-based context-dependent memory. *Memory and Cognition, 25*, 375-380.

Hewitt, K. (1973). Context effects in memory: a review. Cambridge University Psychological Laboratory

Hilton, D. J. (1995). The social context of reasoning: Conversational influence and rational judgment. *Psychological Bulletin, 118*, 248-271.

Hutchison, K.A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychnomic Bulletin and Review, 10*, 785-813.

Jackendoff, R. (1987) *Consciousness and the Computational Mind*. Cambridge, MA: MIT Press

Jenkins, J.J., & Russell, W.A. (1952). Associative clustering during recall. *Journal of Abnormal and Social Psychology, 47*, 818-821.

Kaplan, A.S., & Murphy, G.L. (2000). Category learning with minimal prior knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26*, 829-846.

Keil, F.C. (1989). *Concepts, kinds and cognitive development*. Cambridge MA, US:MIT Press

Kendler, H.H. (1966). coding: associationistic or organizational? *Journal of verbal learning and Verbal Behavior, 5*,198-200.

Kelly, M. H., & Keil, F. C.(1987). Metaphor comprehension and knowledge of semantic domains. *Metaphor and Symbolic Activity*, 2, 33-51.

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

Kripke, S. (1972). Naming and Necessity. In D.Davidson and G. Harman (Eds.), *Semantics of natural language*. Doedrecht, Holland: Reidel

Kucera, H.K., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Labov, W. (1973). The boundaries of words and their meanings. In C.-J.N.bailey and R.W. Shuy (Eds.), New ways of anlaysing variations in English. Washington: Georgetown University Press.

Lancaster, J.S., & Barsalou, L.W. (1997). Multiple organisations of events in memory. *Memory, 5*, 569-599.

Laurence, S. & Margolis, E. (1999). Concepts and cognitive science in E. Margolis, & S.Laurence (Eds,), *Concepts: Core readings*(pp. 3- 81). Cambridge, MA, US: MIT Press

Lin, E.L, & Murphy, G.L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General, 130*, 3-28

Lucariello, J. & Nelson, K. (1985). Slot-filler categories as memory organizers for young children. *Developmental Psychology*, 2, 272-282.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior, Research Methods, Instruments, & Computers, 28*, 203-208.

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

McKoon, G., Ratcliff, R., & Dell, G.S. (1986). A critical evaluation of the semantic-episodic distinction. *Journal of Experimental Psychology: Learning, Memory and Cognition, 12,* 295-306.

McRae, K.,de Sa, V.R., & Seidenberg, M.S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126,* 99-130.

Malt, B. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology, 29,* 85-148.

Mandler, G. (1967). Organization and memory. In K.W. Spence & J.T. Spence (Eds.), *The psychology of learning and motivation* Vol 1. (pp. 327-372). New York: Academic Press

Mandler, G., Worden, P.E., & Graesser, A.C. (1974). Subjective disorganization: search for locus of list organization. *Journal of verbal learning and Verbal Behavior, 13,* 220-235.

Markman, E.M. (1981). Two Different Principles of Conceptual Organization. In M.E. Lamb, & A.L. Brown (Eds.), *Advances in developmental psychology* Vol.1. (pp. 199-236). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Markman, E.M. (1989). *Categorization and Naming in Children: Problems of Induction.* Cambridge,MA: MIT Press

Markman, A.B., & Gentner, D. (1993a). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32,* 517-535

Marshall, G.R. (1967). Stimulus characteristics contributing to organization in free recall. *Journal of Verbal learning and Verbal Behavior, 6,* 364-374.

Medin , D.L. (1989). Concepts and conceptual structure. *American Psychologist, 44,* 1469-1481.

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., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology, 20* 158-190

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254-278

Medin, D.L., Altom. M.W., Edelson., S.M., & Freko. D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: learning, Memory and Cognition, 8,* 37-50.

Medin, D.L., Lynch, E.B.,& Solomon, K.O. (2000). Are there kinds of concepts? *Annual Review of Psychology, 51*, 121-147.

Millikan, R.G. (1998). A common structure for concepts of individuals, stuffs and real kinds: More mama, more milk and more mouse. *Behavioral and Brain Sciences, 21*, 55 –100.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519-533.

Morris, M.W., & Murphy, G.L. (1990). Converging operations on a basic level in event taxonomies. *Memory & Cognition*, 18, 407-418.

Moss, H.E., Hare, M.L., Day, P. & Tyler, L.K. (1994). A distributed memory model of the associative boost in semantic priming. *Connection Science, 6*, 413-427.

Moss, H.E., Ostrin, R.K., Tyler, L.K., & Marslen-Wilson, W.D. (1995) Accessing different types of lexical semantic information: Evidence from priming. *Journal of Experimental Psychology: Learning, Memory and Cognition 21*, 863-883.

Murphy, G.L. (2001). Causes of taxonomic sorting by adults: A test of the thematic-to-taxonomic shift. *Psychonomic Bulletin and Review. 8*, 834-839.

Murphy, G.L. (2002). *The Big Book of Concepts*. Cambridge,MA: MIT Press

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

Nelson, D.L., McEvoy, C.L., & Schreiber, T. (1999). University of South Florida word association, rhyme and word fragments norms. Available at http://cyber.acomp.usf.edu/Free Asoociation/.

Neisser, U. (1986). Nested structure in autobiographical memory. In D.C. Rubin (Ed.), Autobiographical Memory (pp.77-81). Cambridge and New York: Cambridge University Press.

Nelson, K., & Gruendel, J. (1981). Generalized event representations: Basic building blocks of cognitive development. In M.E.Lamb & A.L. Brown (Eds.), *Advances in Developmental Psychology*, Vol 1. (pp. 131 – 158). Hillsdale NJ: Lawrence Erlbaum Associates.

Nisbett, R.E., & Schachter, S. (1966). Cognitive manipulation of pain. *Journal of Experimental Social Psychology, 2*, 227 – 236.

Nisbett, R.E., & Wilson. T.D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review, 84*, 231 – 259.

Osherson, D.N., Smith, E.E.,Wilkie,O., Lopez A., & Shafir, E. (1990). Category-based induction. *Psychological Review, 2*, 185-200.

Pillemer, D.B., Rhinehart, E.D., & White, S.H. (1986). Memories of life transitions: The first year in college. *Human Learning, 5*, 109-123.

Pointer, S.C., & Bond, N.W. (1998). Context-dependent memory: Colour versus odour. *Chemical senses, 23*,359-362.

Pollio, H.R., Kasschau, R.A., & DeNise, H.E. (1968). Associative structure and the temporal characteristics of free recall. *Journal of Experimental Psychology, 76*, 190-197.

Pollio, H.R., Richards, S.,& Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal learning and Verbal Behavior, 8*, 529-536.

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

Puff, C.R. (1970). Role of clustering in free recall. *Journal of Experimental Psychology, 86*, 384-386.

Putnam, H. (1975). The meaning of meaning. In K.Gunderson (Ed.), Language, mind and knowledge. Minnesota Studies in the philosophy of science (Vol.7). Minneapolis: University of Minneapolis Press.

Quillian, M.R. (1968). Semantic memory. In M.Minsky (Ed.), *Semantic Information processing*. Cambridge, MA: MIT Press

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93-134.

Reed, S.K. (1972). Pattern recognition and categorization. *Cognitive psychology, 3*, 382-407.

Reisberg, D. (1997). Concepts and generic knowledge. In Reisberg, D. Cognition: exploring the science of the mind (pp.304-345). New York: W.W.Norton.

Reitman, J.S. (1976). Skilled performance in Go: Deducing memory structures from inter-response times. *Cognitive Psychology, 8*, 336 – 356.

Rips, L.J. (1989). Similarity, typicality and categorization. In S.Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning*. Cambridge: Cambridge University Press

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.). *Cognitive development and the acquisition of language*. New York: Academic Press

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General, 104,* 192-233.

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

Ross, B.H.,& Murphy, G.L. (1999). Food for thought: cross classification and category organization in a complex real world domain. *Cognitive Psychology, 38,* 495-553.

Roth, E.M., & Shoben, E.J. (1983). The effect of context on the structure of categories. *Cognitive Psychology, 15,* 346-378.

Rubin, D.C. & Olson, M.J. (1980). Recall of semantic domains. *Memory and Cognition, 8,* 354-366.

Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J.Spiro, B.C.Bruce. & W.F. Brewer (Eds.). *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale.NJ: Erlbaum.

Schank, R.C. (1982). *Dynamic Memory.* Cambridge: Cambridge University Press.

Schank, R.C., & Abelson, R. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Sjöberg, L. (1972). A cognitive theory of similarity. *Goteborg Psychological Reports,* 2 (10).

Sjöberg, L. & Thorslund, C. (1979). A classificatory theory of similarity. *Psychological Research, 40,* 223-247.

Smiley, S.S.., & Brown, A.L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology, 28,* 249-257.

Smith, E.E., & Sloman, S.A. (1994). Similarity versus rule-based categorization. *Memory and Cognition, 22,* 377-386.

Smith, E.E., Patalano, A.L., & Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65,* 167-196.

Smith, J. D., & Kemler Nelson, D. G. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General* , *113,* 137-159.

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

Smith, L.B.,& Samuelson, L.K. (1997). Perceiving and remembering; category stability, variability and development. In D.R. Shanks, & K.Lamberts (Eds.), (pp. 161-195). *Concepts and Categories*. Cambridge MA: MIT Press

Snowden, J, Griffiths, H, & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology, 11*, 265-288.

Snowden, J, Griffiths, H, & Neary, D. (1995). Autobiographical experience and word meaning. *Memory, 3*, 225-246.

Strevens, M. (2001). Only causation matters: reply to Ahn et al. *Cognition, 82*, 71-76.

Suzuki, H., Ohnishi, H., & Shigemasu, K. (1992). Goal directed processes in similarity judgment. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Thompson, C.P., Hamlin, V.J., & Roenker, D.L. (1972). A comment on the role of clustering in free recall. *Journal of Experimental Psychology, 94*,108-109.

Tulving, E. (1962). Subjective organization in free recall of "unrelated' words. *Psychological Review, 69*, 344-354.

Tulving, E. (1964). Intratrial and intertrial retention: Notes towards a theory of free recall verbal learning. *Psychological Review, 71*, 219-237.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free recall learning. *Journal of Verbal learning and Verbal Behavior, 5*, 193-197.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving, and W. Donaldson (Eds.), *Organization of memory* (pp. 381-403). London: Academic Press.

Tulving, E. (1984). Relations among components and processes of memory. *Behavioral and Brain Sciences, 7*, 257-268.

Tulving, E. (1986). What kind of a hypothesis is the distinction between episodic and semantic memory? *Journal of Experimental Psychology: Learning, Memory and Cognition 12, 307-311.*

Tulving , E. & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of verbal learning and Verbal Behavior, 5*, 381-391.

Tulving, E., & Thompson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

Tversky, A. (1977). Features of similarity, *Psychological Review, 84*, 327-352.

Vallée-Tourangeau, F., Anthony, S. H., & Austin N. G. (1998). Strategies for generating multiple instances of common and ad hoc categories. *Memory,6*, 555-592.

Van Gelder, T. (1995). Mind as motion: Explorations in the dynamics of cognition. Cambridge, MA, US : The MIT Press

Walker, W. H., & Kintsch, W. (1985). Automatic and strategic aspects of knowledge retrieval. *Cognitive Science, 9*, 261-283.

Ward, T. B. (1983). Response tempo and separable-integral responding: Evidence for integral-to-separable processing sequence in visual perception. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 103-112.

Warrington, E.K., & Shallice, T. (1984). Category-specific semantic impairment. *Brain, 107*, 829-854.

Whittlesea, B.W.A. (1987). Preservation of specific experiences in the representation of general knowledge. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 3-17

Whittlesea, B.W.A. (1997). The Representation of general and particular knowledge. In K.Lamberts & D.R. Shanks (Eds.), *Knowledge, concepts and categories. Studies in Cognition* (pp. 335-370). Cambridge, MA: The MIT Press.

Whittlesea, B.W.A., & Brooks, L.R. (1988). Critical influence of particular Experiences in the perception of letters, words, and phrases. *Memory and Cognition, 16*, 387-399.

Williams, J.M.G., & Dritschel, B.H. (1992). Categoric and extended autobiographical Memories. In Conway, M.A.,Rubin, D.C., Spinnler, H., & Wagenaar, W.A (Eds.), Theoretical perspectives on autobiographical memory, Netherlands: Kluwer

Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very long-term memory. *Cognitive Science, 5*, 87-119.

Wilkinson, L. (1990). SYSTAT: the system for statisitics. Evanston, IL: SYSTAT Inc.

Wisniewski, E.J., & Bassok, M. (1999). What makes a man similar to a tie? Stimulus compatibility with comparison and integration. *Cognitive Psychology, 39*, 208-238.

Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Blackwell.

Wu, L. & Barasalou, L.W. (in press). Grounding concepts in perceptual simulation: Evidence from property generation.

Yeh, W., & Barsalou, L. W. (1996). The role of situations in concept learning. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 469-474). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

APPENDIX A

Published article: Vallée-Tourangeau, F., Anthony, S. H., & Austin N. G. (1998)Strategies for Generating multiple instances of common and ad hoc categories. *Memory,6,* 555-592.

PAGE
NUMBERING
AS
ORIGINAL

MEMORY, 1998, *6* (5), 555–592

Strategies for Generating Multiple Instances of Common and Ad Hoc Categories

Frédéric Vallée-Tourangeau, Susan H. Anthony, and Neville G. Austin

University of Hertfordshire, UK

In a free-emission procedure participants were asked to generate instances of a given category and to report, retrospectively, the strategies that they were aware of using in retrieving instances. In two studies reported here, participants generated instances for common categories (e.g. *fruit*) and for ad hoc categories (e.g. *things people keep in their pockets*) for 90 seconds and for each category described how they had proceeded in doing so. Analysis of the protocols identified three broad classes of strategy: (1) *experiential*, where memories of specific or generic personal experiences involving interactions with the category instances acted as cues; (2) *semantic*, where a consideration of abstract conceptual characteristics of a category were employed to retrieve category exemplars; (3) *unmediated*, where instances were effortlessly retrieved without mediating cognitions of which subjects were aware. Experiential strategies outnumbered semantic strategies (on average 4 to 1) not only for ad hoc categories but also for common categories. This pattern was noticeably reversed for ad hoc categories that subjects were unlikely to have experienced personally (e.g. *things sold on the black market in Russia*). Whereas more traditional accounts of semantic memory have favoured decontextualised abstract representations of category knowledge, to the extent that mode of access informs us of knowledge structures, our data suggest that category knowledge is significantly grounded in terms of everyday contexts where category instances are encountered.

INTRODUCTION

The studies reported here concern the nature of the strategies people spontaneously use when asked to retrieve instances for a given category. Item retrieval during this so-called "free emission" procedure appears to reflect

Requests for reprints should be sent to Frédéric Vallée-Tourangeau, Department of Psychology, University of Hertfordshire, Hatfield, Hertfordshire, UK, AL10 9AB. E-mail: psyqfv@herts.ac.uk

A partial report of these data was presented at the BPS Cognitive Section XII Annual Conference, Bristol, September 1995.

certain organisational processes in memory. Bousfield and Sedgewick (1944, p. 153), for example, remarked about the nature and order of the instances retrieved for *animals* that "various types of contiguity are apparent (...) we may observe groups of domesticated animals, commonly exhibited species, and various zoölogical phyla." It soon became apparent that delineating these clusters was an important thing to do, albeit a difficult one (Kausler, 1974).

Research on the nature of the items and their clustering, as well as on the nature of the underlying retrieval process was neglected in favour of efforts aimed at quantifying various aspects of the participants' output. Retrieval norms were produced (e.g. Battig & Montague's 1969 *Category norms for verbal items in 56 categories*) and algorithms were developed to identify the size and number of clusters. Thus, Graesser and Mandler (1978, Study 2) obtained item generation data using a six-minute free-emission procedure and determined that clusters averaged about five items. Gruenewald and Lockhead (1980) reported that in their free-emission data 74% of the clusters were smaller than three items, and that an hyperbolic function fitted the latency between clusters of items during item retrieval. Kail and Nippold (1984) assessed developmental differences in the size and number of clusters generated and concluded that hyperbolic functions modelled adequately the cumulative output of participants from different age groups. Gronlund and Shiffrin (1986) investigated the modelling performance of the search of associative memory model of Raaijmakers and Shiffrin (1981) for cumulative output and item dominance curves produced under different retrieval conditions, for example by providing cues or by having subjects use a particular strategy (e.g. listing instances of *fish* by cycling through the letters of the alphabet).

In most of these studies, the strategies that participants use to generate category instances received scant attention. Graesser and Mandler (1978, p. 94) spoke briefly about them suggesting that subjects accessed "conceptual dimensions" which guided item generation. Kail and Nippold (1984, pp. 950–951) stated that "[our] account of retrieval is distinctly 'nonstrategic' in nature"; they assumed that retrieval was determined by the associative strength between a category name and the category exemplars. Gronlund and Shiffrin (1986) observed that subjects instructed to use an alphabetic strategy (or one based on the size of exemplars; see also Barsalou & Sewell, 1985) produced fewer items than subjects who were free to retrieve instances as they pleased; the nature of these "idiosyncratic" strategies was not addressed however. In contrast, Gruenewald and Lockhead (1980) ventured the broad outlines of a mechanism underlying item production. They proposed a two-process model: The first process searches for a "semantic field" and a second retrieves items associated with that field. Gruenewald and Lockhead sometimes take semantic fields to be taxonomic or sub-taxonomic groups (e.g. *pets*, p.239), but in other places their notion seems to correspond to a much broader class of things, including autobiographical facts and personal experiences of varying specificity. For example, they speculated about the exploration of the following semantic fields during the generation of items for *food* (pp. 229–230): "Consider recalling food words. Perhaps you would consider that supermarkets have foods and attempt to search aisles in your imagination; then you might recall foods on the dining room table; perhaps next to vegetables, then to fruits, to foods that are disliked, and so on; to foods of different countries, then foods from memorable gourmet meals, then (aha) wines, cheeses and so forth."

Two studies, however, sought to characterise the retrieval strategies in greater detail: Williams and Hollan (1981) and Walker and Kintsch (1985). Williams and Hollan studied a small group of subjects who were asked to retrieve, over many sessions, names of high school classmates, in some cases 16 years after graduation. Subjects were asked to comment on their retrieval strategies as they were producing the names. On the basis of the subjects' protocols, an iterative "retrieval cycle" was proposed, consisting of three stages: (1) establishing a retrieval context, (2) searching that context, and (3) verifying candidate to-be-recalled items. Many of these retrieval contexts were essentially experiential in that they were contexts that specified places, situations, and activities where classmates were encountered.

Walker and Kitsch (1985) used a 12-minute free-emission procedure with three categories, *automobiles*, *soups*, and *detergents*. Half of their subjects concurrently described their retrieval strategies and the other half retrospectively described how they retrieved instances. Walker and Kintsch reported that 77% of the retrieval strategies or cues were of an episodic memory character. Thus, for the category *automobiles*, subjects used episodic retrieval strategies like "*my cars*, *friends' cars*, *cars I have wrecked*, *cars I have seen on TV* [in contrast to semantic cues like] *small cars*, *foreign cars*, *GM cars*, *luxury cars*" (Walker & Kintsch, 1985, p.269). Thus, the retrieval process seemed to depend considerably on memory of personal experiences of varying specificity. Walker and Kintsch also noted that the nature of the strategies reported in concurrent or retrospective protocols did not differ, although the number of reported strategies was greater with the concurrent protocols (we shall return to this point later).

The nature of the subjects' introspection in Williams and Hollan (1980) and Walker and Kintsch (1985) suggests that exemplar generation is mediated by a reflection on personal experience or the re-creation of contexts in which subjects interacted with these objects as members of the given category. However, the range of categories was extremely narrow. The studies reported here sought to document over a wider range of categories and in greater detail the retrieval strategies subjects use spontaneously as reported in their retrospective protocols. We chose different kinds of categories to encourage a broad range of retrieval strategies. Some of the categories were well established *common* categories (taken from Rosch & Mervis, 1975) like *fruit* and *furniture*, others were more *ad* *hoc*[1] (Barsalou, 1983). The ad hoc categories were chosen as a function of their value along two dimensions, namely a concreteness dimension, where concreteness characterises the physical nature of the instances, and a familiarity dimension, or the (likely) degree to which subjects would be familiar with the category (even if they had rarely, if ever, instantiated the category as such). Thus, we compared the retrieval strategies underlying item production for *things people keep in their pockets* (high concrete, high familiar) with *animals found on the Galapagos* (high concrete, low familiar) with *excuses for arriving somewhere late* (low concrete, high familiar); we did not construct a low concrete, low familiar ad hoc category.

STUDY 1

The participants in this study generated instances for each category for a 90-second period. After each category the participants wrote down the strategies they had used in the process. One of our aims was to assess the extent of the prevalence of experiential strategies documented in Walker and Kintsch (1985) across a more representative range of categories, and in particular to assess their usage for common categories. Previous research has established a close connection between autobiographical memories and ad hoc categories (Conway, 1990a) and we expected that instance retrieval from "familiar" ad hoc categories would be mediated to a greater extent by experiential strategies. We believed that instance retrieval for categories that are better established in long-term memory, for example, the common category *fruit*, might not be mediated by experiential cues to the same degree as familiar ad hoc categories: for example, as suggested in Barsalou (1983), the strong category–instance associations might render such mediations superfluous. In addition, if experiential cues are used spontaneously for familiar ad hoc categories, it is

[1] The scope of the class of categories that should be designated "ad hoc" is not entirely clear. Barsalou (1983) proposed that such categories are "not well established in memory" (p.211), "are created spontaneously for use in specialized contexts" (p.211), "to achieve goals" (p.214). Barsalou (1985, p.632) later proposed a more general class of "goal-derived categories" which "include both ad hoc categories and better established categories that were once ad hoc." We use the term to denote sets of things, events, or ideas which we presume (i) are not generally well rehearsed as sets, (ii) do not have a commonly used conventional name, and (iii) have as instances or members entities that satisfy requirements stated in the phrase describing the category. The category description may state characteristics of the instances that would enable category members to be recognised independent of a context (e.g. things made mostly of plastic) or may mention a situation or context into which instances must fit but do not state explicitly what characteristics of the instances themselves give the fit to the situation (e.g. things people hate when they are ill—see Study 2; "food to eat on a diet"). Common categories (e.g. fruit, etc.) are presumably well rehearsed, clearly do have conventional names, but are silent on the criteria for membership.

of interest to investigate what cues subjects employ when they have had little or no experience with a category (e.g. *animals found on the Galapagos*).

Method

Participants. A total of 50 undergraduates from the University of Hertfordshire volunteered for this study.

Categories. Ten categories were selected, two common categories taken from Rosch and Mervis (1975), namely *vehicle* and *fruit*, and eight ad hoc categories. These eight ad hoc categories were made up of three subgroups of categories. The first group was one for which the category instances were concrete objects whose grouping was likely to be familiar to the participants, namely *things people keep in their pockets*, *things dogs chase*, *things people take to a wedding*, *things made mostly of plastic*. For ease of exposition we will refer to these four categories as *pockets*, *dogs chase*, *wedding*, and *plastic* respectively. The second group of ad hoc categories was one for which the category instances would cover a different ontological range including, for example, moods, feelings, life events, causal explanations, although participants were unlikely to think the groupings odd. These were *reasons for going on a holiday*, and *excuses for arriving somewhere late*. For convenience these will be referred to as *holiday* and *late*. The final group of ad hoc categories was made up of concrete entities whose groupings were unlikely to be familiar to the participants. These were *animals found on the Galapagos*, and *things sold on the black market in Russia*[2] (these two categories will be referred to as *Galapagos* and *black market*).

Design and Procedure. All participants received all 10 category names in one of 10 randomised orders. After generating category members for each category, participants were asked to rate the difficulty of the retrieval process and to estimate how many more items they could have produced if they were given an extra 10 minutes (as data from these two questions were not the focus of this investigation, they will not be reported in the main body of the Results section). The participants were then asked to answer in writing the question "how did you go about thinking of items for this category?". Answers to this question provided the data on the nature of the retrieval strategies.

Test materials were assembled in a booklet. Single category names were printed on the top of separate lined sheets. After each category sheet, a sheet with the three questions was inserted. The front page of the booklet informed the

[2] Our better informed colleagues told us that it was more appropriate to speak of a black economy in Russia than a black market as such. We suspect, however, that our inaccurate and stereotypical view of the Russian economy did not unduly concern our subjects.

participants about the nature of the task and that for each category they would be given 90 seconds to list the names of as many items they could think of.

Participants were run in groups ranging in size from five to ten. Once the experimenter had answered any queries, participants were instructed to turn over the cover page and start listing items for the first category. After 90 seconds had elapsed, participants were instructed to turn over the page; they were given another 90 seconds to answer the three post-retrieval questions. This procedure was repeated for the remaining nine categories. We chose to segment the experimental procedure into those time intervals in order that test booklets would take no longer than 30 minutes to complete.

Measures. Three measures are reported. These are, (1) the number of items produced for each category (*output fluency*); (2) the proportion of participants who generated each of the 10 most generated items for each category, a measure of consensus termed *item dominance* (more comprehensive production frequencies for the items in each of the 10 categories are reported in the Appendix); and (3) the segmentation and classification of subjects' answers to the question "how did you go about thinking of items for this category?". These answers were first segmented into parts, where each segment appeared to correspond to a distinct idea or approach to item generation. (In many cases the answers were naturally segmented by the participants with the use of terms such as "and" and "then".) Two investigators independently segmented all protocols and met to compare their resulting segmentations and resolve disagreements. The number of protocol segments ranged from 93 (*holiday*) to 117 (*pockets*), with a mean of 102 segments per category (s.d. = 9.5).

All protocol segments were then transcribed onto index cards labelled with a subject number and the category name. Triplicates of the cards were produced and a set given to each of the three authors. The cards were independently sorted into groups. In a first classificatory stage, segments were sorted largely on the basis of their surface form, that is on a minimal interpretation of their meaning. For example, for *fruit*, protocol segments like "fruit I eat" and "fruit I like" were sorted initially in different groups. In a second phase we sought to identify groupings of protocol segments that appeared to reflect similar approaches to item generation. All three authors agreed the final classification of the segments. Three broad classes of strategies were identified: (1) *experiential mediation*, (2) *semantic mediation*, and (3) *unmediated retrieval.* Of the protocol segments, 88% could be classified under these headings (8% resisted such groups and 4% were uninterpretable). Table 1a reports all the subgroups that were identified in each of the three categories for *fruit* from the protocol segments of Study 1. Protocol segments classified within the experiential mediation category indicated that the generation of category instances was based on autobiographical facts; or mediated by the recreation of specific or generic personal experiences during which these instances are encountered. This latter distinction is based on the one made in Conway (1990c, p. 134) between specifically dateable experienced events and "experienced events (...) which subjects could only date to a lifetime period and which were abstracted from specific experiences." Barsalou et al. (1993, p. 38) similarly argue that "(...) an episodic situation represents a single event that occurred at a specific time, [and that] a generic situation generalized over related episodic situations." Thus, our subjects reported using specific personal experiences to aid retrieval, such as, for *wedding*, "the last wedding I attended" or more generic experiences, such as for *dogs chase*, "dogs I've seen in the park". Finally, strategies that described secondhand personal information, that is, of being told directly of the experience (e.g. by friends, via the television) were also included in the experiential mediation category. Segments classified within the semantic mediation category suggested that instance retrieval was mediated by access to subtaxonomic groupings; or by the consideration of the abstract characteristics of the category; or by analogy to related indices ("I just thought of the type of things which are black market in other places"); or by using a distinctive reasoning strategy, like a recipe for a fruit salad (or thinking of the contents of yoghurt pots). The unmediated retrieval category grouped together protocols that implied that instances were not retrieved via an intermediate cue of which subjects were aware at the time of report or which implied an appreciation of the graded structure of the category (i.e. that some instances were "common"). Segments classified in that category specified that instances "popped into my mind", or "came unbidden", or were "common knowledge".

Table 1b reports representative examples of actual protocol segments for *fruit* in Study 1. The protocol segments classified in the experiential mediation

TABLE 1A
Subgroups of Protocol Segments for Fruit Within Each of the Three Classes of Retrieval Strategies in Study 1

Experiential	*Semantic*	*Unmediated*
1. Fruit I eat 2. Fruit I like or don't like 3. Fruit I buy 4. Fruit I've seen 5. Fruit seen in supermarkets 6. Fruit bowl at home 7. Specific memories of eating a fruit 8. Fruits I've heard of 9. Media 10. What I know other people eat	1. Consideration of subcategories: citrus fruits, exotic fruits, garden fruits 2. Reasoning about features (sweet vs. bitter) 3. Recipe 4. Cued by other products, e.g. types of yoghurt, fruit juices 5. Linguistic cues like "berry"	1. Popped into head 2. General knowledge ("I just knew them") 3. Notion of typicality 4. Inter-item cueing

category ranged from autobiographical facts (e.g. "what I hate"), reflections on specific life episodes where fruits were encountered (e.g. "the fruit salad my friend made the other day"), to generic episodes (e.g. "fruits I see in the supermarket"), to experiences of others (e.g. "fruits my children like"). The protocol segments classified in the semantic mediation category indicated that instances were generated on the basis of strategies that made no reference to personal experience, such as a reflection on taxonomic organisation (e.g. "summer fruits"); a consideration of where they grow (e.g. "garden") or how they are utilised (e.g. "yoghurt pots"). The unmediated retrieval category was more liberally constructed, including any protocol segments that did not refer to a deliberate strategy and/or that indicated that the category instances were retrieved automatically. As will become apparent later, because the proportion of protocol segments fitting this category was quite small, adopting such a liberal classification criterion did not artificially inflate the importance of these protocol segments.

TABLE 1B
Actual Protocol Segments that Described Instance Retrieval Strategies for Fruit in Study 1

Experiential Mediation
"What I hate"
"What I eat"
"First thinking of all my favourite fruits."
"Thought about the fruits I bought"
"I thought of the fruit my friend made the other day"
"My auntie's fruit bowl"
"What I have growing in my garden"
"I took the easy way out and wrote what I see in the supermarket"
"Fruits that my children like or dislike"
"Visualised media adverts"

Semantic Mediation
"I thought of categories, e.g. summer fruits"
"Tropics, orchard, garden"
"I thought of fruit which you can grow in the garden"
"Fruits to make fruit salad"
"Yoghurt pots"

Unmediated Retrieval
"Fruits—all different kinds are common knowledge"
"General knowledge"
"By the word association i.e. one fruit brought back recollection of the next"
"Whatever came to mind"
"I listed the fruit I knew of"

Results

Output Fluency. The mean number of items generated for each category and each category type are reported in the second column (Study 1) of Table 2. Subjects generated most items for the common categories (mean of 12.98) and fewest for concrete unfamiliar ad hoc categories (mean of 5.87)[3] To illustrate the differences and similarities both between and within category types, a number of pairwise comparisons were carried out. Thus, within the common categories, fluency for *fruit* was significantly greater than for *vehicle*, $t(49) = 4.86$, $P < .0001$ (all tests two-tailed). The means for *vehicle* and *pockets* were identical, but within the concrete familiar ad hoc type, fluency for *plastic* was significantly smaller than for *pockets*, $t(49) = -10.3$, $P < .0001$, and across ad hoc category type the mean for *plastic* was smaller than for *holiday*, $t(49) = -3.44$, $P < .001$. Finally, the fluency for *late* was significantly greater than the fluency for *black market*, $t(49) = 2.23$, $P < .03$.

Item Dominance. Figure 1 plots the proportion of subjects who generated the *i*th most generated items (*i* ranged from 1 to 10) for each category (right panel) as well as the average by category type (left panel). The picture conveyed by the left panel is clearer: item dominance distributions for common categories showed the most consensus while the least consensus was observed for the unfamiliar ad hoc categories. The right panel portrays a fuzzier picture where the item dominance distributions for each category overlap a great deal. On the basis of Kolmogorov–Smirnov analyses (Wilkinson, 1990), the item dominance distribution for *fruit*, which reflected the highest consensus, was significantly different from the distributions for most categories, except *vehicle*, *pockets*, and *dogs*. In turn, the item dominance distribution for *plastic*, which reflected the least consensus, was significantly different from most categories except *Galapagos* and *black market*. For the categories between these two extremes, the item dominance distributions did not differ significantly.

Retrieval Strategies. The percentage of experiential, semantic and unmediated protocol segments for each category as well as averages for each

[3] Participants were also asked to rate, on a scale from 1 to 10 (1 = "not difficult at all" and 10 = "very difficult"), the difficulty of the retrieval process for each category, as well as to predict the number of items they could produce had they been given 10 more minutes. Both these measures were systematically related to output fluency. Producing instances for the common categories was rated the easiest (mean of 2.19) and for the unfamiliar ad hoc type the hardest (mean of 7.38). The correlation between output fluency and output difficulty ratings was strongly negative, $r = -.62$, $P < .0001$. Estimates of output fluency forecast varied a lot across subjects but were nonetheless significantly correlated both with output fluency, $r = .26$, $P < .001$, and difficulty ratings, $r = -.26$, $P < .001$.

TABLE 2
Mean Output Fluency for the 10 Categories in Study 1 and the 9 Categories in Each of the Three Conditions of Study 2

Category	Study 1	Study 2		
		A	B	C
Common Categories				
Fruit[1,2]	14.80	13.65	13.97	14.39
	(0.63)	(0.52)	(0.43)	(0.64)
Vehicle[1,2]	11.14	8.23	8.40	10.04
	(0.61)	(0.82)	(0.93)	(0.57)
Bird[2]		9.19	10.83	12.52
		(1.05)	(0.97)	(1.00)
Furniture[2]		11.74	12.57	10.75
		(0.54)	(0.57)	(0.40)
Mean	12.98	10.70	11.44	11.96
	(0.48)	(0.43)	(0.42)	(0.37)
Concrete Familiar Ad Hoc Categories				
Pockets[1,2]	11.14	11.97	11.43	9.03
	(0.40)	(0.59)	(0.71)	(0.64)
Dogs chase[1,2]	9.10	11.43	10.50	9.77
	(0.30)	(0.58)	(0.61)	(0.48)
Wedding[1]	8.04			
	(0.41)			
Plastic[1]	6.78			
	(0.35)			
Walls[2]		11.94	11.07	9.40
		(0.53)	(0.57)	(0.31)
Mean	8.77	11.79	11.00	9.40
	(0.21)	(0.32)	(0.36)	(0.31)
Abstract Familiar Ad Hoc Categories				
Holiday[1]	8.20			
	(0.36)			
Late[1]	8.18			
	(0.36)			
Ill[2]		9.71	7.67	6.60
		(0.66)	(0.52)	(0.41)
Mean	8.19	9.71	7.67	6.60
	(0.26)	(0.66)	(0.52)	(0.41)
Unfamiliar Ad Hoc Categories				
Galapagos[1]	5.00			
	(0.37)			
Black market[1,2]	6.74	9.39	8.03	7.50
	(0.53)	(0.68)	(0.78)	(0.45)
Mean	5.87	9.39	8.03	7.50
	(0.34)	(0.68)	(0.78)	(0.45)

Standard error in parentheses.
1 = included in Study 1; 2 = included in Study 2.

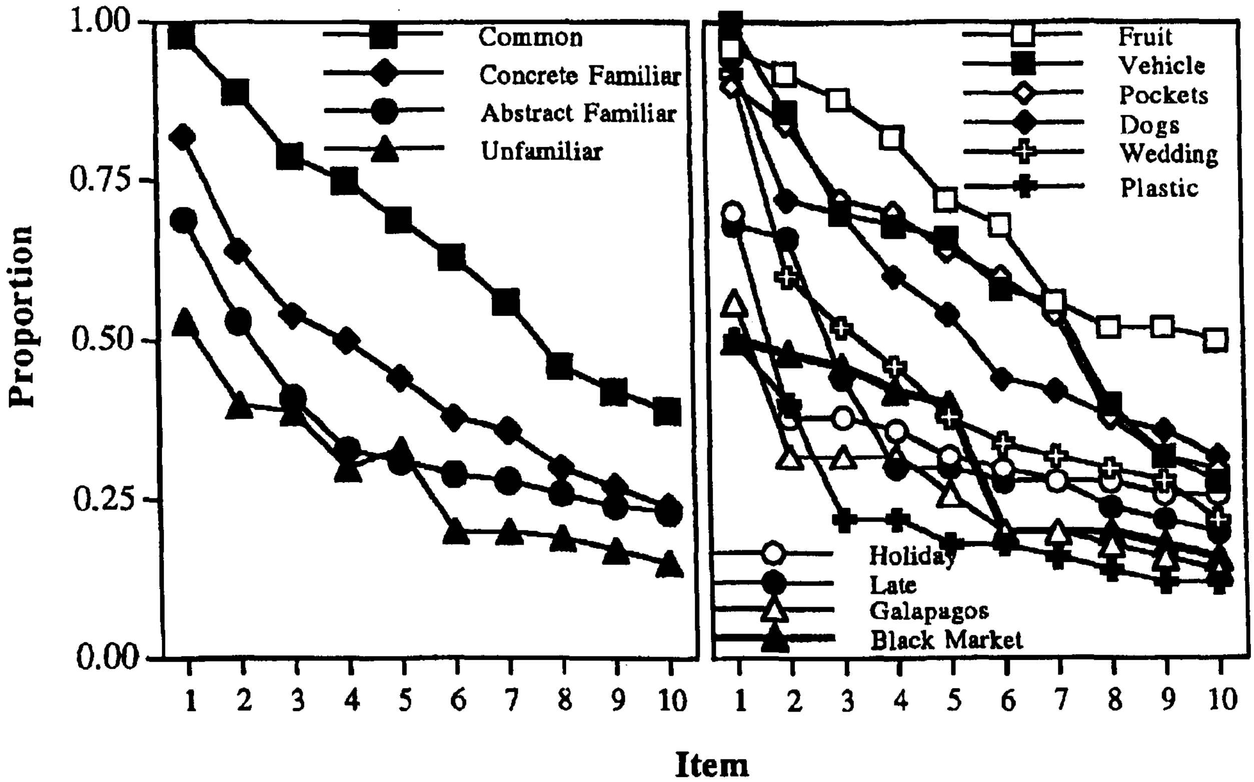


FIG. 1. Proportion of participants generating each of the 10 most generated items for the 10 categories in Study 1 (right panel); averaged proportions for each category type (left panel).

category type are reported in Table 3. As most subjects reported more than one strategy, protocol segments were not independently produced. Consequently, no inferential statistics are reported. However, clear patterns in the segment distributions can be observed. For seven of the eight categories, experiential strategies greatly outnumbered both semantic strategies and reports of unmediated retrieval. In terms of individual categories, the percentage of experiential protocols outnumbered semantic protocols for *fruit*, *pockets*, *dogs chase*, *wedding*, *plastic*, holiday, and *late*; for *vehicle* the percentages were nearly even. Striking reversals of this pattern were observed for *black market* and *Galapagos*.

Discussion

The data on output fluency and retrieval consensus will be discussed first, followed by a discussion of the data on retrieval strategies. Participants found it easier to generate items for the common than for the ad hoc categories, and within ad hoc categories, they found it easier to generate items for the familiar types. These results are in general agreement with those reported in Barsalou (1983) where the average output fluency for nine common categories and the average for nine ad hoc categories differed reliably. These statistical differences may suggest a clear dichotomy between common and ad hoc categories. However, if the mean output fluency for each category is plotted in rank order,

TABLE 3
Percentage of Protocol Segments Classified as Indicating Experiential Mediation, Semantic Mediation, and Unmediated Retrieval for Each Category and Each Category Type in Study 1

	Experiential	Semantic	Unmediated
Common Categories			
Fruit	67	16	17
Vehicle	36	40	25
Mean	52	28	21
Concrete Familiar Ad Hoc Categories			
Pockets	83	10	7
Dogs chase	54	31	15
Wedding	62	25	14
Plastic	74	12	14
Mean	68	19	13
Abstract Familiar Ad Hoc Categories			
Holiday	60	27	13
Late	69	10	22
Mean	65	19	18
Unfamiliar Ad Hoc Categories			
Galapagos	24	67	9
Black market	32	65	4
Mean	28	66	7

as in Fig. 2, a more subtle picture emerges: while *fruit* and *Galapagos* anchor the two extremes of the graph, the smooth transition within these two poles suggests no important discontinuities corresponding to a processing boundary between common and ad hoc categories (nor between the different types of ad hoc category). In this respect, it is interesting to note that our item dominance distributions overlapped considerably (see right panel of Fig. 1) as they did, in fact, in Barsalou (1985).[4]

The data on the reported strategies (see Table 3) suggest that the mediators that participants often used were more closely related to personal experiences that we believe establish, sustain, and tune the meanings of everyday concepts than to the decontextualised semantics in which theories of everyday concepts are generally framed. For seven out of the ten categories in this study, the

[4] The item dominance distributions for Barsalou's (1985) common and ad hoc categories were calculated for the top 10 most generated items from the raw data provided in the appendix of his paper. Kolmogorov–Smirnov analyses of these distributions revealed that only *weapons* and *personality characteristics in others that prevent you from being friends with them* differed reliably from the other categories in terms of item dominance; the remaining categories showed item dominance distributions that did not differ statistically.

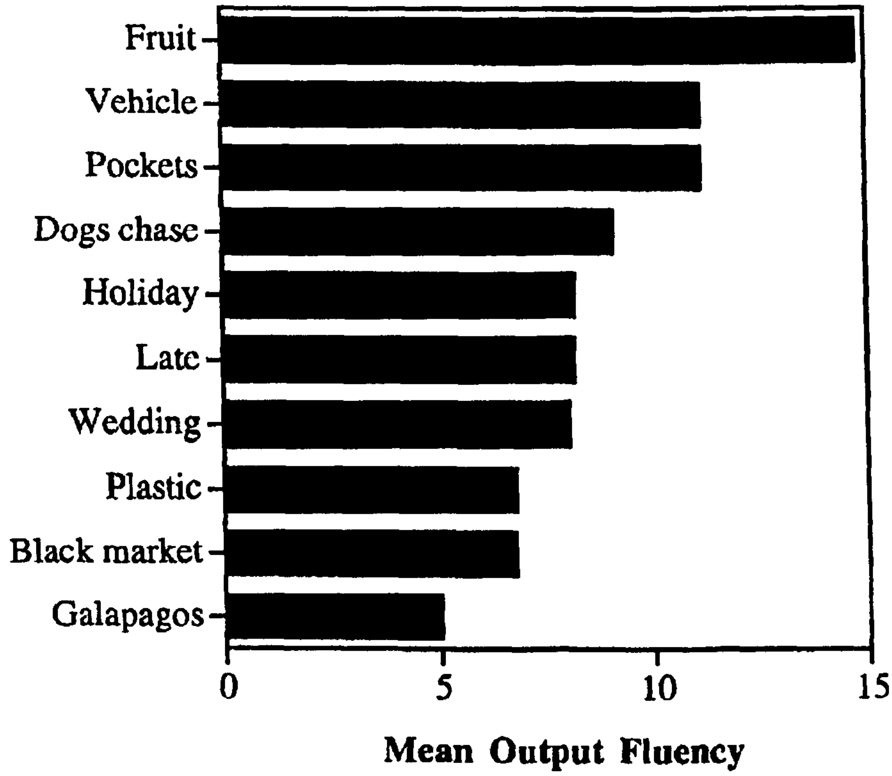


FIG. 2. Rank order of the mean output fluency for each of the 10 categories in Study 1.

proportion of protocol segments indicating an experiential strategy was much greater than those indicating a semantic strategy. This pattern was noticeably reversed for *Galapagos* and *black market*. In these two categories, less than a third of the protocol segments indicated an experiential strategy.[5] Participants were forced to reason explicitly about the membership criteria for these two unfamiliar categories. Furthermore, the distribution of experiential/semantic strategies for these two categories indicates that the question "how did you go about thinking of items for this category?" did not bias the retrospective description of the strategies in a way that would artificially favour experiential over semantic strategies.

Semantic strategies were about as frequent as experiential strategies for *vehicle*. It appears that despite the common usage of this category label, in both everyday discourse and in concept research, a number of subjects experienced some uncertainty in defining the scope of category membership. For example, one participant wrote: "Thinking that vehicle includes a moving means of transport on 4 wheels. Is it true? I'm not sure" and another wrote: "I think I may

[5] Lest the reader thinks that subjects reporting an experiential strategy for these categories were well travelled, *all* the protocol segments classified as experiential for *Galapagos* described television programmes on exotic islands or the Galapagos Archipelago. For *black market*, descriptions of media sources accounted for a third of the experiential strategies (in comparison, for the remaining eight categories, the average percentage of segments describing a media source was 7%); one subject reported using her last visit to Russia to generate items; the remaining protocol segments classified as experiential mediation described information heard from other sources such as friends and relatives.

have got a bit carried away and put things which would come under transport rather than vehicles ...". This uncertainty over category membership seemed to result in some subjects seeking to establish a definition of the category in terms of properties, for example, "Transportation over distance for 1 or more human beings which requires a source of power...". Protocols such as these were classified as semantic. This was the only category that appeared to cause widespread "boundary disputes".

The percentage of protocols indicating unmediated retrieval was lower than 20% in eight of the ten categories. It is likely that only the production of the first few items for any category was unmediated (Walker & Kintsch, 1985): once the strong associates of the category labels have been listed, the production of the remaining items was mediated by elaborative retrieval strategies.

Certain considerations might have led us to expect a greater proportion of semantic strategies for both common *and* ad hoc categories. In some traditional accounts, the long-term memory representation of categories is characterised as being structured along conceptual dimensions (e.g. Henley, 1969; Rips, Shoben, & Smith, 1973; Lund & Burgess, 1996; Rubin & Olson, 1980). These dimensions are uncovered through analyses of item proximity in subjects' free recall output; or on the basis of pairwise similarity ratings; or on the basis of lexical co-occurrence in text. Matrices of item proximity are cluster analysed (e.g. Friendly, 1977) or subjected to a multidimensional scaling analysis that yields "interpretable" dimensions along which items vary. For example, Rubin and Olson (1980) report that size and ferocity are dimensions that underlie *mammals*. Given that subjects' output for a common category can be characterised by a small set of semantic dimensions, it is perhaps surprising that subjects referred to them so rarely when generating instances. It has been generally assumed that the nature of the cues that provide the most successful access to category exemplars shares important similarities with the way these exemplars are represented in memory (e.g. Barsalou & Sewell, 1985, p. 651; Morris, Bransford, & Franks, 1977; Tulving & Thompson, 1973). The protocol segments from Study 1 suggest that experiential aspects of the context in which objects as instances of a category are encountered are an intrinsic part of the category representation.

As Barsalou (1991) has demonstrated, an explicit consideration of the ideals defining an ad hoc category often guides the generation of instances during planning. For example, in planning a holiday, people need to consider the category *vacation location*. Barsalou reported data indicating that people tailor the candidate instances of *vacation locations* along certain ideals, reflecting their current goals and constraints. Furthermore, the graded structure of an ad hoc category is determined in part by the extent to which its instances satisfy the ideal or goal associated with the category (Barsalou, 1985). In this respect, one might surmise that ad hoc categories are perhaps to some extent even more amenable to a conscious consideration of the semantic dimensions that define them than are common categories. Thus, in the case of *pockets*, it may not have been unreasonable to expect subjects' protocols to display an appreciation of the practical nature of the objects carried in pockets in terms of quotidian needs and goals. Yet such considerations very rarely surfaced in the subjects' protocols, in stark contrast with the panoply of personal experience descriptions.

Before we discuss the implications of the prevalence of experiential strategies in our subjects' protocols, two possible methodological artefacts must be addressed. The first concerns the way in which ad hoc categories were defined to the participants. For example, the formulation of *things people keep in their pockets* might be interpreted to mean "things people I know—myself included—keep in their pockets". Instead of focusing on the nature of the instances that belong to the category, participants might have been encouraged to retrieve personally relevant memories of things kept in pockets. It is possible to rephrase ad hoc categories in a way that focuses more clearly on the nature of the instances; e.g. *things that may conveniently be kept in pockets*. This formulation not only eliminates reference to "people" but also emphasises the nature (and perhaps even the function) of the category instances.

The second potential artefact concerns the relative proportion of ad hoc and common categories in Study 1 (4 to 1), and their order of presentation. On average, across all the randomised orders of presentation created for Study 1, participants were much more likely to have generated instances for ad hoc categories before generating items for the common categories. If the wording of the ad hoc categories encouraged an experiential retrieval strategy or otherwise did not sufficiently stress the conceptual dimensions that characterise the category members, the participants might have simply continued using experiential strategies for common categories. Study 2 sought to remedy these shortcomings by (1) varying the wordings of ad hoc categories, (2) reducing the ratio of ad hoc to common categories, and (3) by fixing the order of presentation such that some participants generated items for common categories before they were presented with the ad hoc categories.

STUDY 2

Method

Participants. A total of 92 undergraduates from the University of Hertfordshire volunteered for this study.

Categories. Nine categories were used, four common and five ad hoc. The common categories, taken from Rosch and Mervis (1975), were *fruit*, *vehicle*, *bird*, and *furniture*. Three of the five ad hoc categories were of the concrete familiar type, namely *things people keep in their pockets* (or the alternative wording, *things that may conveniently be kept in pockets*), *things dogs chase* (or *things that might be chased by dogs*), and a new category *things people put on* *walls* (or *things that can be put on walls*). The fourth one was a new abstract familiar ad hoc category, namely *things people hate when they are ill* (or *things people would hate when they are ill*), and the fifth one was the unfamiliar *things sold on the black market in Russia* (or *things that could be sold on the black market in Russia*). For convenience the five ad hoc categories will be referred to as *pockets*, *dogs chase*, *walls*, *ill*, and *black market*.

Design and Procedure. The order of presentation of the categories and the wordings of the ad hoc categories were manipulated in three conditions. In all three conditions, the categories were blocked such that the participants were presented with the four common or the five ad hoc categories in consecutive order. Four different randomised orders of presentation were created within each block of categories.

Condition A. The four common categories were presented first, followed by the five ad hoc categories defined with the original wording of Study 1.

Condition B. The five ad hoc categories, defined with the original wording, were presented first, followed by the four common categories.

Condition C. The same presentation order as in B was used. However, the ad hoc categories were now defined with an alternative wording that might encourage the participants to consider the nature and/or the function of the category members, for example *things people keep in their pockets* was presented as *things that may conveniently be kept in pockets*.

The same experimental procedures as in Study 1 were employed. There were 31 participants in Condition A, 31 in Condition B, and 30 in Condition C. Across all three conditions, 83% of the protocol segments were classified within one of the three retrieval strategy categories (experiential, semantic, or unmediated), 9% resisted the classification scheme, and 8% were uninterpretable. The average number of classified protocol segments, across categories, varied little within each condition: means of 41.2, 41.0, and 40.7 for Conditions A, B, and C respectively.

Results

Output Fluency and Item Dominance. The mean output fluency for each category and the average output fluency by category type in the three conditions of Study 2 are reported in the last three columns of Table 2. Highest fluency was observed for some of the common categories (e.g. *fruit*).[6] For the common categories, fluency seemed relatively stable across condition (if anything the

[6] As in Study 1, difficulty ratings were negatively correlated with output fluency in all three experimental conditions, smallest $r = -.35$, $P < .0001$ (in Condition A), and output forecast was positively correlated with output fluency in all three conditions, smallest $r = .22$, $P < .0001$ (Condition C), and negatively correlated with difficulty ratings in all three conditions, smallest $r = -.23$, $P < .0001$ (Condition A).

means slightly increased from A to C). In contrast, for the three types of ad hoc categories, output fluency seemed to decrease across the conditions. A series of ANOVA for each category type revealed that output fluency did not differ reliably across conditions for common categories [$F(2, 243) = 1.69$] nor for the unfamiliar ad hoc category [$F(2,90) = 2.27$]. However, output fluency decreased reliably across conditions in the abstract familiar [$F(2,90) = 8.52$, $P < .0004$], and concrete familiar ad hoc category types [$F(2, 176) = 6.86$, $P < .002$]. The difference between Conditions B and C, which reflects the effect of rewording the ad hoc categories, was reliable for two of the five ad hoc categories, namely *pockets* ($P < .01$ using Fisher's post hoc least significant difference) and *walls* ($P < .03$).

The experimental manipulations had little effect on the item dominance distributions: within each category, item dominance distributions did not differ significantly (on the basis of Kolmogorov–Smirnov analyses); production frequencies for the categories used in Study 2 are reported in the Appendix. As in Study 1, there were few salient differences between categories. Thus *fruit* showed the highest consensus and *ill* the lowest. Among the common categories, consensus for *bird* and *vehicle* were the lowest, and the item dominance distributions for *bird* differed significantly from the ones for *fruit* in all three conditions. The three familiar concrete ad hoc categories showed very similar distributions across conditions which, generally, did not differ significantly from the distributions for the common categories. Finally, the item dominance distributions for *ill* differed significantly from the item dominance distributions from all other categories with the exception of *bird*, *black market*, and *walls* in Condition A, and *black market* in Condition C.

Retrieval Strategies. The percentages of protocol segments coded as reflecting an experiential strategy, a semantic strategy, or no strategy (unmediated retrieval) are shown in Fig. 3 for each category in all three conditions. To facilitate the comparison with the percentages observed for the categories that were also used in Study 1 (*fruit*, *vehicle*, *dogs chase*, *pockets*, and *black market*) they are plotted alongside the percentages observed in Study 2. The experimental manipulations appeared to have had no systematic effect on the relative distribution of the protocol segments: as in Study 1, for all but the unfamiliar ad hoc category *black market*, experiential strategies outnumbered semantic strategies in all conditions, including the two new ad hoc categories, *ill* and *walls*, and, importantly, the two new common categories, *bird* and *furniture*. The sole exception to this pattern was for *vehicle* in Condition A. As in Study 1, retrieving instances for *black market* was done mostly on the basis of semantic strategies.

The influence of the prior presentation of the ad hoc categories (using the wording of Study 1) on the percentage of experiential strategies in common categories can be assessed by examining the data for the four common

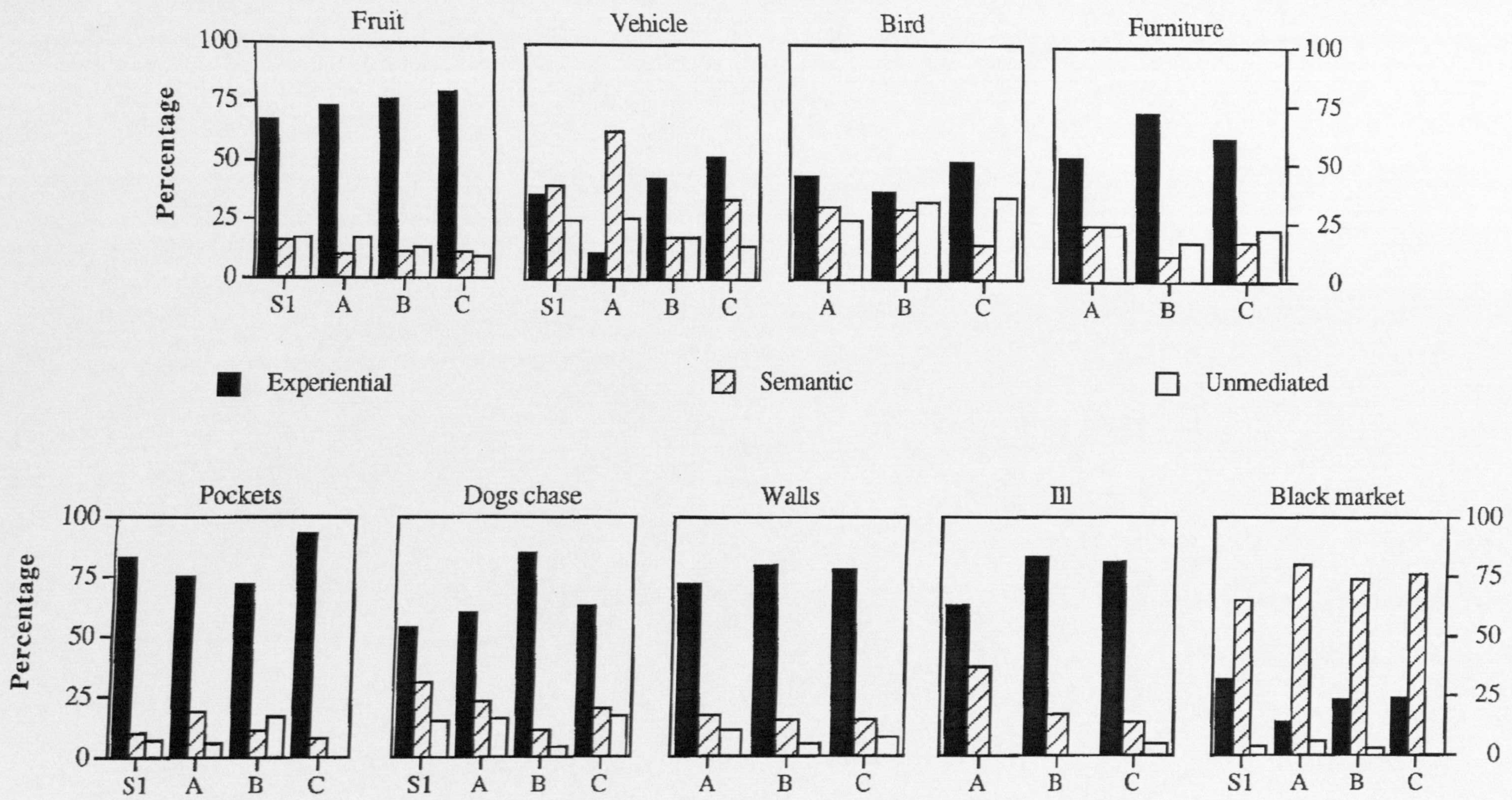


FIG. 3. Percentage of experiential, semantic, and unmediated protocol segments in Conditions A, B, and C of Study 2. For the categories *fruit*, *vehicle*, *pockets*, *dogs chase*, and *black market*, the percentages of each segment type from Study 1 are also plotted.

categories between Condition A, in which they were presented first, and Condition B, in which they followed the ad hoc categories. *Vehicle* registered an important difference in that experiential strategies were used much more after the presentation of the ad hoc categories (Condition B and also in Condition C) than when the common categories preceded the ad hoc ones (Condition A). For the remaining three common categories, *fruit*, *bird*, and *furniture*, the presentation order seemed to have little influenced the extent to which experiential strategies were employed.

A comparison of Conditions B and C for the ad hoc categories assesses the effect of rewording the ad hoc concepts in a way that emphasised object-centred properties. As Fig. 3 indicates, the preponderance of Experiential strategies did not vary across Conditions B and C for *walls* and *ill*; it increased slightly for *pockets*, but decreased slightly for *dogs chase*.

Discussion

The data on output fluency and retrieval consensus essentially painted the same picture produced by the data in Study 1. *Fruit* was the best established category with the largest output fluency and the greatest output consensus. The new abstract ad hoc category *ill*, in turn, showed the smallest output fluency and the lowest output consensus. However, in between these two extremes, neither the output fluency data nor the item dominance distributions could produce a clear and systematic wedge between common and ad hoc categories.

The data produced by the first study appeared to indicate that people, when asked to produce a list of items that are members of some target category, will do so primarily by retrieving memories of personal experiences in which they interacted with those items *qua* members of the target category. Importantly, this kind of retrieval strategy was employed as much for common as for ad hoc categories. The categories for which experiential strategies were not used where the ones for which the participants were unlikely to have experienced the objects as members of the group specified by the concept. For example, even if the participants in this study were likely to be very familiar with the items they listed for *black market* (e.g. alcohol, jeans, cigarettes) they did not construct personally relevant contexts in which they encounter these objects. Rather, as the category is instantiated in Russia and all our participants (except one) had never set foot there, they were forced to *reason* about what they knew or intuited about the Russian black market.

The prevalence of experiential strategies in the first study could have resulted from certain methodological limitations which Study 2 sought to remedy. Thus, the wordings of the ad hoc categories were modified to emphasise the characteristics of the category instances, to see if this would then encourage a consideration of those characteristics as reported in the retrospective protocols. The change in wordings had some impact on the participants, as revealed by the lower output fluency for all ad hoc category types. Yet as Fig. 3 reveals clearly, the prevalence of experiential strategies in ad hoc categories was not affected by the change in wordings (compare Conditions B and C). The category *black market* was still the only ad hoc category for which semantic strategies outnumbered experiential strategies.

Another important aim of the second study was to assess whether the prevalence of experiential strategies in common categories such as *fruit* (and to a lesser extent *vehicle*) could be attributed to the greater proportion of ad hoc categories in the test booklets. To that end the ratio of common to ad hoc categories was modified from 1:4 to 4:5, and, importantly, in one condition (A), the four common categories were presented in a block before the presentation of the ad hoc categories. Using this procedure, experiential strategies were dominant in both Conditions A and B in three of the four common categories replicating the findings for *fruit*, and extending them to *bird* and *furniture*. In Condition A, however, only 11% of the protocol segments indicated an experiential strategy for *vehicle*. Although *vehicle* was the only familiar category in Study 1 that showed a large proportion of semantic protocols, the distribution of strategies for *vehicle* in Condition A is difficult to interpret in light of the fact that the distribution of retrieval strategies for the other three common categories was stable across the three experimental conditions. This category was, however, the only one whose scope appeared to be ambiguous to our subjects.

GENERAL DISCUSSION

We have been fairly sanguine about the informativeness of the retrospective protocols obtained in these studies. It could be argued that the protocols merely reflect a reconstructive process ancillary to the one that determined the kind and number of instances generated, and that the predominance of experiential retrieval strategies is an artefact of that process. A number of considerations, however, strongly suggest that the strategies reported by the participants did reflect the ways in which they generated instances. These are, (1) the specificity of some of the protocol segments, (2) the presence of clusters in the generated item list that corresponded to the reported strategies, (3) the tendency to group items into semantic, but not experiential, categories when people sort a list of items (Walker & Kintsch, 1985), and (4) the lack of difference in the nature of the strategies reported in concurrent and retrospective protocols (Walker & Kintsch, 1985). Let us examine each of these points in turn.

First, the participants often reported strategies that were very specific. This specificity casts doubt on the hypothesis that the participants invented strategies in order to answer the question "how did you go about thinking of items for this category?". For example, for *fruit*, one participant wrote:

> If you go to Tesco's in Royston you meet the fruit and vegetable section first of all. I imagined I was shopping and can remember the layout and most of the fruit.

Another reported the following strategy for *bird*:

> I do a lot of sailing and many boats in the type of sailing boats I sail in are named after birds; so I went through my memory of the opposition's boats' names.

It seems unlikely that subjects would have put in the effort to confabulate these detailed and idiosyncratic accounts as opposed to reporting the strategies they used spontaneously.

The procedure employed for the studies reported here did not lend itself to a systematic analysis of the correspondence between reported strategies and instances generated. This is due, in part, to the fact that most strategies were too idiosyncratic to permit a plausible mapping of generated items to strategies. For example, a subject might report the strategies "what I see at the supermarket" and "fruits I like" for *fruit*, but it is then hard to determine whether the production of any particular item was done on the basis of the first or the second strategy. However, when the reported strategies were based on norms or knowledge shared by both participants and experimenters (e.g. arrangement of bedroom furniture) analysis of the protocols revealed some striking correspondences. Table 4 lists items generated by three participants (from Study 2), shown in the left portion of each column, along with their retrospective protocols, shown in the right portion of each column (the back-slashes indicate how these were segmented). The protocol from 65c accounts for all the items generated, although the series from *blue tit* to *house martin* could have included smaller clusters produced by local strategies or processes. The protocol from 82a matches the clusters only to some extent: presumably *furniture van* was triggered by *van*, and *scooter* is a member of a previous category. Finally, the protocol from 62c, while perfectly interpretable, is not specific enough to suggest clusters, and without knowing the arrangement of this participant's house, it is hard to be confident that he is being quite as systematic as he says.

Third, Walker and Kintsch (1985) asked subjects in one session to sort items they had themselves generated in a previous free-emission task. The subjects were then asked to sort the same items according to the strategy that produced them. The two groupings hardly overlapped: 85% of the clusters in the first sorting task were semantic whereas 77% of the clusters in the second sorting task were episodic. This dissociation suggests that if our participants had no introspective access to their retrieval strategies and compensated by looking back on the items they had generated to identify meaningful groupings, they would have reported, mostly, descriptions of semantic groupings. In contrast, our subjects converged in great numbers on experiential strategies, replicating the findings of Walker and Kintsch.

TABLE 4
Protocols of Three Subjects

65c bird	
eagle owl — blue tit robin chaffinch sparrow blackbird thrush swift house martin — ostrich — pigeon dove seagull	I thought of big birds/then of birds I would see in the garden/then in a zoo./ When I thought of pigeon the next two came automatically to mind.
82a vehicle	
car lorry bus train coach — motorbike bicycle — tractor Land Rover JCB crane van — furniture van — scooter	The modes of transport seen on the road./ Followed by two wheeled variety/and then the building trade. My husbands works in this field.
62c furniture	
table chair dresser piano sofa easy chair coffee table wardrobe bed dressing table hi-fi unit stool ottoman desk	My own house— going room to room

Fourth, the distribution of experiential and semantic strategies is unlikely to have been biased by the retrospective nature of the protocol. In Walker and Kintsch (1985) half of the participants described their retrieval strategies as they were generating instances and half were asked to do so retrospectively. Although the concurrent protocols yielded a greater number of strategies, an analysis of the kinds of strategies reported with both procedures revealed that the relative distribution of semantic and episodic strategies was the same (see Walker & Kintsch, 1985, p. 277).

Experiential Contexts as Effective Retrieval Strategies: Implications for Theories of Concepts

The analysis of the protocols generated in the present studies suggests that output in a free-emission procedure is governed to a considerable extent by the nature of the participant's personal experiences. Some of the examples of semantic fields provided by Gruenewald and Lockhead (1980) correspond to what we have called experiential strategies (e.g. memorable gourmet meals, visits to the local supermarket) and these "semantic fields" might be better characterised as event frames (Barsalou, 1992) where memories for objects, actions, participants, and situations are woven in a rich fabric of thematic connections.

Barsalou and his colleagues have offered empirical evidence supporting the idea that aspects of the situations in which objects are encountered become integrated in the knowledge of objects and categories. In a concept learning experiment, Yeh and Barsalou (1996) presented their subjects with instances from the same category but in different situations. Each situation was correlated with a different set of relevant properties. If the category representation that develops as subjects experience different instances across situations is principally decontextualised, reflecting an abstraction of properties across contexts, then the recognition of situation-relevant properties should not be faster when it takes place in that situation. Yet, situations primed the recognition of situation-relevant properties. Such results suggest that aspects of the situations in which category instances are encountered form an integral part of their representation.

The feature list that people report when cued by the name of a category is presumably a reflection of their mental intension of that category, their concept of it. Such intensions are known to vary with the cultural point of view the person takes, they are context-dependent. Similarly context-dependent are the decisions people make as to the extensional scope of category names, presumably because the concepts they are instantiating differ with context. Concrete instantiations of general concepts also vary with context. (See Barsalou, 1993, for a review). Each of these phenomena can be understood on the supposition that people are responding on the basis of having retrieved a restricted set of individual or generic instances associated with the explicit, implied, or an analogous context in the stored representation of that concept in memory. Our own subjects' retrieval efforts consisted mostly in the re-creation of situations where objects were encountered and experienced.

Barsalou et al. (1993, p. 26) sketch the outlines of a broad ranging theory of conceptualisation and meaning in which "the fundamental conceptual representations in the human cognitive system are schematic perceptual images extracted from all modes of experience" (see also Barsalou, 1993). More relevant to understanding why experiential context frequently mediates retrieval in our experiments is their equally radical proposal (p. 38) that "concepts are neither context independent nor universal but are situated and local." Mental concepts of both particular individuals and generic types of individual, are always established within and include the perspective of the experiential scenarios in which they are encountered or imagined. Rather than having an essential meaning, or even a univocal core, according to Barsalou et al. (1993, p. 47), they are "collections of all specialised models for a particular type of individual together with their associated generic situations." Context and context effects are not an optional extra to context-free theories of concepts and categorisation but are intrinsic to the mental representations constituting concepts.

That (some) fruit can be bought from a particular display counter in a supermarket might ordinarily be regarded as encyclopaedic information about fruit rather than being a part of what "fruit" really *means*. An alternative view, however, is that "fruit" has varying usages and does not have a singular real meaning; rather, "fruit" extensionally refers to just those (partially overlapping) sets of things, and intensionally to the represented characteristics of the things in each set that have historically become referred to as "fruit" in particular experiential contexts (cf. Bloom, 1996). On this view, the concept underlying "fruit" may be conceived as comprising a probably fuzzy collection of *combined concepts*[7]—e.g. fresh fruit salad fruit, garden fruit, exotic fruit, cooked desert fruit, forest tree fruit, botanical textbook fruit. The prototypical regularities that have been found in analyses of taxonomic concepts (e.g. Rosch & Mervis, 1975) may reflect a pragmatic default of reporting about sets of instances in common contexts in the absence of any specific directive otherwise—subjects respond cooperatively to what sense they can make of

[7] The term *combined concept* (or related terms such as *conjunctive concept* or *conjunctive category*) is commonly taken as designating concepts whose verbal label consists of more than a single word. We do not address the question of cases in which the components of a *combined concept* may combine in a compositional manner. As used here the term refers to phrases that we see as acting in the manner of pointers to prestored experientially related bodies of information. Thus the term "garden fruit" may be thought of as a verbal shorthand for "fruit that people grow in their gardens" and not one whose meaning is to be computed from the separate meanings of "garden" and of "fruit". We suggest later that monolexemic terms may behave in a similar fashion.

the investigator's intent. The commonalities that we found in our subjects' strategies for retrieving instances of common taxonomic categories and a variety of ad hoc categories would be consistent with this view. Some recent discussions of the role played by the retrieval of stored instances in the comprehension of combined concepts are compatible with this view (e.g. Gray & Smith, 1995; Hampton, 1997).

Barsalou (1983, 1985) noted a variety of measures that, to varying extents, were correlated with judgements of goodness of category membership. Similarity to a category ideal and frequency of instantiation (experiencing potential instances as members of the named category) correlated with membership judgements in both his ad hoc and in common taxonomic categories. Ad hoc and established goal-oriented categories, *ex hypothesi*, do not cut nature at the joints and nor do conventional categories of artefacts (e.g. furniture, vehicles, tools, weapons, clothing) or social roles (e.g. professions). In fact membership in both artefact and social role categories conforms to the sort of satisfaction criteria that apply to explicitly goal-oriented categories—hence *good* members are close to some ideal(s) and are likely to be frequently encountered instances. Is there good reason to suppose that so-called natural kind categories are different? Our retrieval data suggest not, as do Barsalou's aforementioned findings. If not, then there may be a crucial difference between the technical use of terms and the folk uses (as in the classic arguments about whether tomatoes should be *called* fruit[8] or penguins *called* birds) which we assume that experimental subjects generally employ.

There is in fact good reason to doubt whether, even in their technical scientific sense, many (most?) natural kind categories are as clear-cut as seems to be commonly assumed. Referring to "species" in his exposition of the Darwinian legacy Dennett (1995, pp. 93–94) remarks that "As Darwin pointed out, if it weren't for the separations that time and the extinction of the intermediate stepping-stones has created (...) we could not put them into a 'natural classification'—we need the biggish gaps between *extant* forms to form

[8] The technical definition of "fruit" is "the matured ovulary of angiosperms" (Greulach & Adams, 1976). It is interesting to note that this is a reference to historical origin rather than to characteristics of an entity taken out of (developmental) context. On this definition pea, bean, peanut, walnut, olive, pumpkin, and cucumber are fruit, as well, of course, as tomato. "Vegetable" seems to be a culinary rather than biological category. Standard dictionaries seem to offer meanings for the vernacular use of terms, and for "fruit" describe a variable, albeit qualified mix of culinary, physical, developmental, and sensory characteristics, together with instances, exceptions and contrast categories, e.g. "An edible part of a plant, generally sweet, acid and juicy, esp. a part that contains the seed, but sometimes extended to include other parts (e.g. the leaf stalk in rhubarb), and popularly distinguished from the vegetable, its savoury, firm-fleshed counterpart." (The Chambers dictionary, 1994); "Something that you can eat that grows on a tree or bush. It has soft or firm flesh, and contains seeds or a stone. Oranges, bananas and grapes are fruit. Fruit is eaten raw or cooked, usually as a dessert." (Collins COBUILD dictionary, 1987). The vagueness and contradictions and differences in emphasis of these two accounts are striking.

the 'boundaries' of any such classes' (emphasis in the original). Addressing higher levels of taxonomic classification Morris (1995, p. 662) remarks that "Even now there are several very peculiar animals ... which apparently defy placement in known metazoan groups." The extent to which natural kinds are experienced as distinct kinds (Malt, 1995) may depend on experiencing only a limited range of variation in a particular ecological range. There is also clear evidence that folk uses of many purportedly biological natural kind terms differ from the uses of the same terms in a scientific biological sense in quite radical ways (Dupré, 1981; Gould, 1990), subdividing and crosscutting scientifically differentiated kinds.

Perhaps one of the most puzzling questions about conceptual representation is similar to one concerning beliefs about the paranormal. With respect to the latter it is not so much whether paranormal phenomena occur or not, as why there appears to be such a widespread belief that they do. Similar, perhaps, is the puzzle as to why there appears to have been a widespread view that vernacular concepts of the sorts discussed here are most usefully characterised by reference to a singular, integrated, coherent, and context-free meaning at their core. The view sketched here suggests that common folk concepts of natural, social, and artefact categories may profitably be viewed as collections of representations of (partially overlapping) contextually focused *combined categories* with substantial autobiographical reference. The independent status of the head concept in such combinations is interesting. Consider *furniture*, *bedroom furniture*, and *schoolroom furniture*. It would seem that, for most people, instances of *bedroom furniture* are also instances of *furniture*, but at least some instances of *schoolroom furniture* are not (Hampton, 1982, 1988). From the current perspective this apparent anomaly may arise because a common but restricted set of *furniture* (household furniture perhaps) is being queried for the membership of, say, chalkboard. Naming a superordinate category with a common monolexemic designator does not mean that all the category is accessed. The category may be instantiated by an experientially bound but limited subset of the general category. In this sense a concept is created on the fly.

Our retrieval data, together with Barsalou et al.'s (1993) theoretical framework, suggest that at least for concepts of common superordinate categories there is an important type of constituent conceptual organisation distinct from that of a taxonomic organisation of subordinates. This level is organised on a foundation of generic autobiographical experiences and consists of a corresponding variety of combined concepts whose experiential perspectives integrate the specific and generic individuals falling within their range. Such experientially based combined concepts would have a coherence that derived precisely from the points of view or standpoints or goals that characterised those experiences and in this sense they may share organisational characteristics with ad hoc or explicitly goal-oriented concepts.

Autobiographical Memory and the Meaning of Objects

The literature on the conceptual representation of categories makes little systematic reference to the circumstances in which everyday concepts are applied and which sustain their meaning. There is evidence, however, that supports the importance of autobiographical memory in underpinning meaning. Work with neuropsychological patients with semantic dementia reported by Snowden, Griffiths, and Neary (1994; see also 1995) has shown that even if these patients' performance on general tests of word comprehension is extremely poor, their understanding of objects, places, and people that are encountered on a regular basis is relatively intact. Snowden et al. (1994, p. 287) argued that it is "the experiential input into the meaning system" that sustains their meaning. The meaning of objects, places, and people that were part of the patients' lives some time in the past is lost. This difference "highlights the role of autobiographical experience in the maintenance of meaning, and underlies the interrelationship between episodic and semantic memory" (Snowden et al., 1994, p. 265).

Conway's research has also pointed to a close interdependence between autobiographical memory and semantics. Conway (1987) asked subjects to verify true or false semantic facts and autobiographical facts. Semantic verifications were of the form "is an apple a fruit?", while autobiographical verifications took the form "are apples your favourite fruit?". These questions were preceded either by a neutral word or by a category-name prime, for example "fruit". Semantic verifications were, as expected, faster when preceded by the presentations of the relevant category-name primes. Importantly, autobiographical verifications were also primed by the presentation of a relevant category name, e.g. "fruit" speeded responses to "are apples your favourite fruit?". Conway (1990b, p. 176) suggested that "frequent and current autobiographical knowledge might be stored in memory with semantic knowledge and so help 'customize' the semantic system."

Conway (1990c) provides evidence that memories of specific and generic personal experiences are closely connected to representation of common and ad hoc categories. The general procedure employed in that study involved giving subjects names of categories and asking them to "bring to mind an image of whatever [you take] the word or phrase to refer" (Conway, 1990c, p. 134). Conway supplied subjects with three predetermined categories in which to classify their images: (1) specifically dateable autobiographical memory images; (2) loosely dateable generic images derived from specific experiences; and (3) images not based on specific or generic experiences which Conway labelled "semantic" images. The majority of images elicited by ad hoc categories (e.g. *things to take on holiday*) were dateable autobiographical memories (53%; Experiment 1) and 23.5% were generic images. In the classification of the protocol segments in our studies, the specific and generic images categories were fused into the experiential mediation category. On average, 72.6% of the strategies for the (familiar) ad hoc categories were experiential, a figure that closely corresponds to the sum of the two percentages just quoted (76.5%). This convergence is important because, unlike in Conway (1990c), we did not provide our participants with predetermined categories that would help them identify their retrieval strategies. Furthermore, the sum of the percentage of the specific (33.8%) and generic (36.8%) images generated by common category labels summed to approximately the same value, namely 70.6% (note the correspondence with the figure reported in Walker & Kintsch, 1985, of 77% episodic retrieval strategies). Thus, Conway's studies provide further evidence of the close connections between memory of personal life events and concepts.

The traditional characterisation of semantic and episodic memory (e.g. Tulving, 1972) attributes language comprehension to semantic memory and the storage of dateable events to episodic memory. As the basic building blocks of knowledge are specific episodes, one challenge has been cast as documenting the process by which more abstract conceptual knowledge emerges out of individual experiences (e.g. Lucariello & Nelson, 1985). The intimate connection between personal experience and category instances, as demonstrated by the former's ability to retrieve the latter, should also encourage the investigation of the role of specific personal experiences in grounding meaning and governing reasoning.

Manuscript received 8 August 1996
Manuscript accepted 3 June 1997

REFERENCES

Barsalou, L.W. (1983). Ad hoc categories. *Memory & Cognition, 11*, 211–227.

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.

Barsalou, L.W. (1991). Deriving categories to achieve goals. In G.H. Bower (Ed.), *The psychology of learning and motivation, 27* (pp.1–64). London: Academic Press.

Barsalou, L.W. (1992). *Cognitive psychology* (Ch. 7). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Barsalou, L.W. (1993). Flexibility, structure, and linguistic vagary in concepts: Manifestations of a compositional system of perceptual symbols. In A.F. Collins, S.E. Gathercole, & M.A. Conway (Eds.), *Theories of memory* (pp.29–101). Hove, UK: Lawrence Elbaum Associates Ltd.

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

Barsalou, L.W., Yeh, W., Luka, B.J., Olseth, K.L., Mix, K.S., & Wu, L.-L. (1993). Concepts and meaning. In K. Beals, G. Cooke, D. Kathman, S. Kita, K.-E. McCullough, & D. Testen (Eds.), *Chicago Linguistic Society 29, Vol. 2*: Papers from the parasession on the correspondence of conceptual, semantic and grammatical representations (pp.23–61). Chicago: Chicago Linguistic Society.

Battig, W.F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80* (3, Pt. 2).

Bloom, P. (1996). Intention, history, and artefact concepts. *Cognition*, *60*, 1–29.

Bousfield, W.A., & Sedgewick, H.W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, *30*, 149–165.

Collins COBUILD English language dictionary (1987). London: Collins.

Conway, M.A. (1987). Verifying autobiographical facts. *Cognition*, *25*, 39–58.

Conway, M.A. (1990a). Associations between autobiographical memories and concepts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 799–811.

Conway, M.A. (1990b). *Autobiographical memory.* Milton Keynes, UK: Open University Press.

Conway, M.A. (1990c). Conceptual representation of emotions: The role of autobiographical memories. In K.J. Gilhooly, M.T.G. Keane, R.H. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought, Vol. 2* (pp.133–143). Chichester, UK: John Wiley & Sons.

Dennett, D.C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* Harmondsworth, UK: Penguin Books.

Dupré, J. (1981). Natural kinds and biological taxa. *The Philosophical Review*, *XC*, 66–90.

Friendly, M.L. (1977). In search of the M-gram: The structure of organization in free recall. *Cognitive Psychology*, *9*, 188–249.

Gould, S.J. (1990). What, if anything, is a zebra? In *Hen's teeth and horse's toes.* Harmondsworth, UK: Penguin Books.

Graesser, A. II, & Mandler, G. (1978). Limited processing capacity constrains the storage of unrelated set of words and retrieval from natural categories. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 86–100.

Gray, K.C., & Smith, E.E. (1995). The role of instance retrieval in understanding complex concepts. *Memory and Cognition*, *23*, 665–674.

Greulach, V.A., & Adams, J.E. (1976). *Plants: An introduction to modern botany* (*3rd Edn.*). New York: Wiley.

Gronlund, S.D., & Shiffrin, R.M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *12*, 550–561.

Gruenewald, P.J., & Lockhead, G.R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 225–240.

Hampton, J.A. (1982). A demonstration of intransitivity in natural categories. *Cognition*, *12*, 151–164.

Hampton, J.A. (1988). Overextension of conjunctive concepts: Evidence for a unitary model of concept typicality and class inclusion. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 12–32.

Hampton, J.A. (1997). Emergent attributes in combined concepts. In T.B. Ward, S.M. Smith, & J. Viad (Eds.), *Conceptual structures and processes: Emergence, discovery and change* (pp.83–110). Washington DC: American Psychological Association.

Henley, N.M. (1969). A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behaviour*, *8*, 176–184.

Kail, R., & Nippold, M.A. (1984). Unconstrained retrieval from semantic memory. *Child Development*, *55*, 944–951.

Kausler, D.H. (1974). *Psychology of verbal learning and memory.* London: Academic Press.

Lucariello, J., & Nelson, K. (1985). Slot-filler categories as memory organizers for young children. *Developmental Psychology*, *21*, 272–282.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-orrence. *Behavior, Research Methods, Instruments, & Computers*, *28*, 203–208.

Malt, B. (1995). Category coherence in cross-cultural perspective. *Cognitive Psychology*, *29*, 85–148.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.

Morris, S.C. (1995). A new phylum from the lobster's lips. *Nature*, *378*, 661–662.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Rips, L.J., Shoben, E.J., & Smith, E.E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

Rosch, E., & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*, 563–605.

Rubin, D.C., & Olson, M.J. (1980). Recall of semantic domains. *Memory & Cognition*, *8*, 354–366.

Snowden, J., Griffiths, H., & Neary, D. (1994). Semantic dementia: Autobiographical contribution to preservation of meaning. *Cognitive Neuropsychology*, *11*, 265–288.

Snowden, J., Griffiths, H., & Neary, D. (1995). Autobiographical experience and word meaning. *Memory*, *3*, 225–246.

The Chambers dictionary (1994). Edinburgh, UK: Chambers Harrap.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp.381–403). London: Academic Press.

Tulving, E., & Thompson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.

Walker, W.H., & Kintsch, W. (1985). Automatic and strategic aspects of knowledge retrieval. *Cognitive Science*, *9*, 261–283.

Wilkinson, L. (1990). *SYSTAT: The system for statistics*. Evanston, IL: SYSTAT Inc.

Williams, M.D., & Hollan, J.D. (1981). The process of retrieval from very long-term memory. *Cognitive Science*, *5*, 87–119.

Yeh, W., & Barsalou, L.W. (1996). The role of situations in concept learning. In G. W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp.469–474). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

APPENDIX 1

Production frequencies for the categories used in Study 1 and Study 2 (the categories appear in alphabetical order). An item is included if a minimum of 10% of the subject sample produced it (i.e. minimum N = 14 when the category was used in both studies, minimum N = 5 when the category was used in Study 1 only, and minimum N = 9 when the category was used in Study 2 only). In Study 2 the ad hoc category descriptions were altered in Condition C (see text), and these are marked by an asterisk.

Animals found on the Galapagos

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
BIRDS	26	26			
INSECTS	17	17			
TURTLES	16	16			
LIZARDS	15	15			
TORTOISES	12	12			
MONKEYS	10	10			
SNAKES	10	10			
FISH	9	9			
CRABS	7	7			
PARROTS	7	7			
SEALS	7	7			
IGUANAS	6	6			
PEOPLE	6	6			
REPTILES	5	5			
SEAGULLS	5	5			

Bird

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
EAGLE	50		17	16	17
ROBIN	49		17	12	20
SPARROW	44		14	17	13
BLACKBIRD	43		14	14	15
BLUE TIT	40		13	14	13
PIGEON	28		7	9	12
THRUSH	28		7	12	9
PARROT	26		8	10	8
DUCK	24		9	10	5
MAGPIE	24		5	11	8
STARLING	22		4	9	9
OWL	21		7	4	10
CROW	20		4	6	10
SEAGULL	16		5	5	6
CHICKEN	15		8	1	6

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
OSTRICH	15		4	4	7
WOODPECKER	14		6	4	4
BUDGIE	14		5	5	4
CANARY	14		5	6	3
HAWK	13		6	4	3
SWAN	13		4	8	1
GREAT TIT	12		3	6	3
SWALLOW	12		3	7	2
TURKEY	11		5	4	2
FINCH	11		4	4	3
GOOSE	11		4	4	3
RAVEN	11		4	3	4
PHEASANT	11		3	4	4
PEACOCK	10		4	1	5
KESTREL	9		3	2	4
WREN	9		3	1	5
CUCKOO	9		2	4	3

Excuses for arriving somewhere late

Item	*Overall Frequency*	*1*	*1.A*	*2.B*	*2.C*
CAR BREAKDOWN	34	34			
HEAVY TRAFFIC	22	22			
OVERSLEPT	21	21			
BUS LATE	15	15			
LOST WAY	14	14			
PHONE RANG AS I WAS LEAVING	14	14			
DELAY ON TRAIN	13	13			
MISSED BUS	12	12			
ALARM FAILED	11	11			
TRAFFIC JAM	11	11			
FORGOT	10	10			
ACCIDENT—ROAD	9	9			
COULDN'T FIND KEYS	8	8			
ILLNESS (SELF)	8	8			
WATCH STOPPED	8	8			
WEATHER CONDITIONS	8	8			
COULDN'T FIND THE PLACE	6	6			
THOUGHT YOU SAID ANOTHER TIME	6	6			
SOMEONE CALLED AS I WAS LEAVING	6	6			
CLOCK STOPPED	5	5			
EMERGENCY	5	5			
FORGOT SOMETHING	5	5			
TRANSPORT FAILURE	5	5			

Fruit

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
APPLE	134	49	29	29	27
BANANA	131	46	28	28	29
ORANGE	127	44	29	26	28
PEAR	101	36	21	25	19
GRAPE	91	41	15	17	18
STRAWBERRY	86	22	24	20	20
KIWI	83	34	17	18	14
MELON	77	26	20	17	14
PINEAPPLE	76	26	17	15	18
GRAPEFRUIT	68	30	16	12	10
PEACH	68	24	11	18	15
PLUM	63	25	7	13	18
MANGO	61	21	11	13	16
RASPBERRY	60	17	18	13	12
CHERRY	52	18	10	11	13
LEMON	50	19	12	10	9
TOMATO	47	25	7	10	5
BLACKBERRY	45	15	11	12	7
NECTARINE	40	16	3	11	10
LIME	34	13	7	7	7
PASSION FRUIT	30	11	7	5	7
STAR FRUIT	29	12	8	5	4
APRICOT	26	10	7	3	6
GOOSEBERRY	26	9	8	6	3
BLACKCURRANT	25	8	7	5	5
BLUEBERRY	24	8	7	6	3
SATSUMA	23	13	5	1	4
TANGERINE	23	11	4	2	6
LYCHEE	23	9	6	4	4
GUAVA	19	8	3	3	5
AVOCADO	17	6	3	4	4
POMEGRANATE	16	8	3	2	3
CLEMENTINE	15	14	0	0	1

Furniture

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
CHAIR	83		28	26	29
TABLE	81		27	27	27
BED	68		22	24	22
WARDROBE	53		16	14	23
SOFA	43		7	18	18
DESK	42		14	16	12
STOOL	30		13	13	4

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
CUPBOARD	30		9	13	8
CHEST OF DRAWERS	25		8	8	9
CABINET	24		12	5	7
COFFEE TABLE	24		9	10	5
SETTEE	24		7	11	6
DRESSING TABLE	23		7	8	8
SIDEBOARD	23		7	8	8
ARMCHAIR	17		6	5	6
DINING TABLE	16		5	8	3
LAMP	14		8	3	3

Reasons for going on holiday

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
TO RELAX	26	26			
TO GET A TAN	16	16			
FOR FUN	14	14			
SUN	14	14			
TO REST	13	13			
BREAK	11	11			
VISIT FAMILY	11	11			
VISIT FRIEND	10	10			
MEET NEW PEOPLE	9	9			
TO SEE NEW PLACES	9	9			
SPEND TIME WITH FAMILY	8	8			
STRESS	8	8			
CHANGE OF SCENERY	7	7			
LEARN ABOUT DIFFERENT CULTURES	7	7			
WIN A HOLIDAY	7	7			
CHANGE OF CLIMATE	6	6			
FOR PLEASURE	6	6			
SEA	6	6			
SEE THE WORLD	6	6			
SPEND TIME WITH FRIENDS	6	6			
TO GET AWAY	6	6			
FOR A CHANGE	5	5			
LEARN ABOUT DIFFERENT CUSTOMS	5	5			
SPORTS	5	5			
TO TRY DIFFERENT FOODS	5	5			
WARMTH	5	5			

Things dogs chase

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C**
CATS	137	46	31	30	30
BALLS	94	37	18	20	19
OTHER DOGS	90	26	23	24	17
STICKS	75	34	12	13	16
CARS	67	22	17	15	13
POSTMEN	67	17	18	16	16
PEOPLE	66	30	11	16	9
RABBITS	66	19	14	15	18
BIRDS	56	17	17	10	12
OWN TAIL	49	21	12	12	4
CHILDREN	40	8	13	10	9
BICYCLES	29	9	10	5	5
FLIES	24	9	7	6	2
BONES	22	6	2	6	8
SHEEP	20	6	3	8	3
FOXES	19	9	3	3	4
FRISBEES	18	9	8	0	1
TOYS	18	8	4	5	1
BURGLARS	18	6	6	4	2
LEAVES	15	6	1	2	6
SQUIRRELS	14	6	3	2	3

Things made mostly of plastic

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
PENS	19	19			
COMPUTERS	15	15			
TOYS	15	15			
BAGS	11	11			
CARRIER BAGS	10	10			
CHAIRS	7	7			
FOOD CONTAINERS	7	7			
BIN LINERS	6	6			
BIROS	6	6			
BOWLS	5	5			
CONTAINERS	5	5			
DRINK BOTTLES	5	5			
TUPPERWARE	5	5			
WASHING UP BOWLS	5	5			

Things people hate when they are ill

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C**
NOISE	38		10	15	13
FOOD	30		13	10	7
TAKING MEDICINE	28		10	11	7
BEING TOO HOT	19		8	9	2
VOMITING	18		8	4	6
PAIN	17		6	9	2
HOSPITALS	15		8	7	0
DOCTORS	15		6	6	3
HAVING TO WORK	15		5	5	5
BEING TOO COLD	14		9	4	1

Things people keep in their pockets

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C**
KEYS	117	42	22	27	26
MONEY	106	39	17	27	23
PENS	93	33	23	18	19
WALLET	90	35	18	23	14
HANDKERCHIEF	73	29	18	14	12
SWEETS	66	29	8	14	15
TISSUES	62	27	12	13	10
CREDIT CARD	42	19	8	9	6
LOOSE CHANGE (COINS)	40	12	15	6	7
CIGARETTES	37	15	8	12	2
PURSE	32	8	10	9	5
BITS OF PAPER	31	9	14	2	6
COMB	29	13	10	4	2
LIGHTER	27	12	5	8	2
CHEWING GUM	25	11	7	4	3
DIARY	22	10	4	4	4
PENCILS	22	5	5	7	5
ID CARD	21	8	5	5	3
LIPSTICK	21	6	5	5	5
CONDOMS	20	4	8	4	4
MATCHES	19	12	1	5	1
WATCH	19	6	4	3	6
GLOVES	14	8	2	3	1
RECEIPT	14	7	3	3	1

Things people put on walls

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
PICTURES	76		25	25	26
WALLPAPER	66		22	24	20
PAINT	53		18	18	17
POSTERS	43		17	16	10
SHELVES	40		13	15	12
PHOTOS	34		13	14	7
MIRRORS	27		8	9	10
LAMPS/LIGHTS	25		7	12	6
PAINTINGS	21		8	8	5
CUPBOARDS	18		6	6	6
GRAFFITI	16		6	7	3
CLOCKS	16		5	7	4
HOOKS	16		6	4	6
PLATES	14		6	4	4
NOTICEBOARDS	12		5	5	2
(LIGHT) SWITCHES	12		3	6	3
TAPESTRY	10		5	3	2
CALENDARS	10		4	4	2
PLASTER	10		4	4	2
TELEPHONES	10		4	4	2
ORNAMENTS	10		3	4	3
BLACKBOARDS	9		3	2	4
NAILS	9		3	2	4

Things people take to a wedding

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
PRESENTS	45	45			
CAMERA	30	30			
CONFETTI	25	25			
FLOWERS	21	21			
CARD	19	19			
HAT	18	18			
CARS	14	14			
RINGS	14	14			
MONEY	11	11			
CHILDREN	10	10			
INVITATION	9	9			
HANDBAG	7	7			
RICE	7	7			
VIDEO CAMERA	7	7			
HORSESHOE	6	6			
THEMSELVES	6	6			
CAKE	5	5			
FRIEND	5	5			
RELATIVES	5	5			

Things sold on the black market in Russia

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
DRUGS	74	23	18	16	17
FOOD	73	24	15	19	15
CLOTHES	59	19	13	12	15
JEANS	52	20	12	11	9
ALCOHOL	40	16	7	11	6
FOREIGN MONEY	35	14	7	6	8
CIGARETTES	34	7	13	8	6
MEAT	25	10	9	3	3
CARS	24	7	5	6	6
JEWELLERY	22	4	6	7	5
ELECTRICAL GOODS	19	8	3	6	2
GUNS	19	6	3	5	5
BREAD	18	3	8	6	1
TV	17	7	4	3	3
WATCHES	17	4	6	4	3
MEDICINES	14	3	3	7	1
PASSPORTS	14	7	3	2	2
DOLLARS	14	5	3	3	3

Vehicle

Item	*Overall Frequency*	*1*	*2.A*	*2.B*	*2.C*
CAR	126	47	29	22	28
BICYCLE	92	34	22	17	19
BUS	80	33	19	12	16
MOTORCYCLE	78	34	10	14	20
LORRY	75	31	15	13	16
TRAIN	66	29	16	11	10
PLANE	65	29	11	12	13
VAN	58	20	15	12	11
BOAT	36	16	3	9	8
COACH	33	12	10	6	5
TRACTOR	31	9	5	10	7
TRAM	27	11	7	6	3
TRUCK	25	12	6	6	1
HELICOPTER	21	13	2	3	3
MOPED	21	7	4	3	7
SHIP	21	6	4	3	8
HOVERCRAFT	17	12	2	2	1
TRICYCLE	17	8	3	3	3
SCOOTER	16	5	4	1	6
FERRY	14	13	0	0	1

PAGE
NUMBERING
AS
ORIGINAL

APPENDIX B

INSTRUCTIONS AND SAMPLE OF MATERIALS
CATEGORY MEMBER GENERATION STUDY1

Thank you for taking part in this study

INSTRUCTIONS

This booklet consists of 10 exercises. At the top of the first page of each exercise you will see the name of a category, for example,

THINGS THAT HAVE TO BE WASHED OFTEN

You are asked to list the names of items which you think belong in this category. You will be allowed 90 seconds in which to list as many items as possible. On the next page you will be asked to answer three written questions about the list you have just produced. A further 90 seconds will be allowed to complete your answers.

You should repeat this procedure for each of the 10 exercises.

You will be told when to start each exercise and when the 90 seconds allowed for each section is up.

There are no right or wrong items to list. All that is required is that you list the items which occur to you personally.

THINGS PEOPLE TAKE TO A WEDDING

Please complete the following questions

1) **Place a mark on the scale below to indicate how difficult it was to think of items for this category?**

1	2	3	4	5	6	7	8	9	10
Not difficult at all									Very difficult

2) **If you were given another 10 minutes, how many more items do you think you would be able to produce? (give a number)**

____

3) **How did you go about thinking of items for this category?**

____

____

____

____

____

____

____

____

APPENDIX C

INSTRUCTIONS AND SAMPLE OF MATERIALS CATEGORY MEMBER GENERATION STUDY2

Thank you for taking part in this study

INSTRUCTIONS

This booklet consists of 9 exercises. At the top of the first page of each exercise you will see the name of a category, for example,

THINGS THAT HAVE TO BE WASHED OFTEN

Or, perhaps

VEGETABLE

You are asked to list the names of items which you think belong in this category. You will be allowed 90 seconds in which to list as many items as possible. On the next page you will be asked to answer three written questions about the list you have just produced. A further 90 seconds will be allowed to complete your answers.

You should repeat this procedure for each of the 9 exercises.

You will be told when to start each exercise and when the 90 seconds allowed for each section is up.

There are no right or wrong items to list. All that is required is that you list the items which occur to you personally.

VEHICLE

Please complete the following questions

1) **Place a mark on the scale below to indicate how difficult it was to think of items for this category?**

	1	2	3	4	5	6	7	8	9	10	
Not difficult at all											**Very difficult**

2) **If you were given another 10 minutes, how many more items do you think you would be able to produce? (give a number)**

____

3) **How did you go about thinking of items for this category?**

APPENDIX D

INSTRUCTIONS AND SAMPLE OF MATERIALS
TYPICALITY RATINGS

INSTRUCTIONS

Groups of objects or entities tend to be grouped together in **Categories.** There are some categories, such as Vegetable or Sport which are very familiar to you and for which you would easily be able to think of examples or **members.** There are other categories which may seem more unusual, such as *Things people leave behind in Theatres* or *Things that need painting regularly* - you may find it a little harder to think of members of these categories. What we have done in this study is provide you with the names of some categories (some familiar, some less so) and a list of things which are members. What we would like you to do is consider these members and make decisions about how good an example of the category any particular member is. For example, if you think of the category **SPORT,** you may consider that the member **Football** is a very good example of this category, that **volley-ball** is quite a good example but not as good as football and that **Chess** is a very poor example.

THE TASK YOU ARE REQUIRED TO DO HAS TWO STAGES.

Stage 1

At the top of each of page in this booklet you will see the name of a category. Below that will be the names of 15 members of that category. What we would like you to do first is to arrange those members in your own **PERSONAL** order of "goodness of example" for the category name given. For example:-

VEGETABLE	**My Order of "Goodness of Example"**
Carrot	1.*Potato*
Potato	2.*Lettuce*
Pea	3.*Pea*
Aubergine	4.*Onion*
Lettuce	5.etc.
Onion	6.
Cauliflower	7.
Cabbage	8.
Cucumber	9.
Sweetcorn	10.
Radish	11.
Runner bean	12.
Broccoli	13.
Artichoke	14.
Mushroom	15.

If you thought Potato was the best example of the category VEGETABLE you would put that in position 1 - if you thought Lettuce was the next best example then that would occupy position 2 etc. down to the least good example of the category which you would put in position 15.

Stage 2.

When you have done that for the first category, we want you to go back to each item in the order you have produced and give it a rating of "goodness of example" by circling the appropriate number on the rating scale beside each member. The scales look like this:-

1	2	3	4	5	6	7	8	9
Poor Example							**Excellent Example**	

So if you were considering **Lettuce** as an example of the category **Vegetable** you may think this a good example but not the best possible example and so circle a rating of say, 7. Remember, you have only been given a selection of members for each category so you may not necessarily consider any of them to be an 'Excellent' example of that category, not even the one you put in first position on your ordering. In the same way, the member that came last in your ordering need not necessarily be a poor member - it may be quite a good member but the worst of the selection you were given.

If you wish, you may give the same rating to more than one member.

Overall Procedure

When you have completed the ratings for all members of that category then turn to the next category and repeat the procedure i.e. order the members first in order of goodness-of-example- then give each member a goodness-of-example rating, and so on until you have completed all categories in the booklet. Take your time, we'd like you to think about this exercise carefully.

There are no right or wrong responses, it is your personal orderings and ratings which we are interested in.

It's quite difficult to decide how a number of items should be ordered and so you have been provided with scrap paper so you can have a go at trying different orders until you find one which you are happy with and that should be given as your "final ordering". Your final ordering and ratings should be given in the answer booklet provided.

Sample Page:

CATEGORY: FRUIT

ORANGE

GRAPE

PEAR

MELON

APPLE

KIWI

PINEAPPLE

GRAPEFRUIT

PEACH

CHERRY

LEMON

BLACKBERRY

BANANA

MANGO

CLEMENTINE

APPENDIX E

STIMULI USED TO OBTAIN TYPICALITY RATINGS

Stimuli for Typicality ratings

BIRD	FURNITURE	SHELF	DOGS CHASE
Eagle	Chair	CDs	Cats
Robin	Table	Newspapers	Other Dogs
Blackbird	Piano	Photographs	Balls
Sparrow	Sofa	Plants	Sticks
Blue Tit	Bed	Bottles	Bicycles
Starling	Wardrobe	Stereo	Postman
Hen	Desk	Pictures	Birds
Gull	Cabinet	Videos	Cars
Thrush	Cupboard	Screws	Rabbit
Vulture	Stool	Telephone	Their own Tails
Emu	Sideboard	Vases	Burglars
Pelican	Footstool	Clock	Squirrels
Owl	Dressing Table	Ornaments	Their owners
Crow	Cushion	Books	Strangers
Wren	Bookshelves	Plates	Cows

ATTIC	WEDDING	BLACK MARKET	VEHICLE
Xmas Decorations	Flowers	Food	Car
Memories	Confetti	Drugs	Bicycle
Old Clothes	Present	Jeans	Lorry/Truck
Books	Card	Clothes	Bus
Spiders	Champagne	Cigarettes	Train
Dust	Cake	Cars	Plane
Suitcases	Rings	Currency	Van
Toys	Video Camera	Guns/Weapons	Coach
Furniture	Food	Electrical goods	Boat
Old letters	Suit	Jewellery	Tram
Antiques	Hat	Video Recorders	Submarine
Cobwebs	Family	T.Vs	Rickshaw
Water Tank	Cars	Magazines	Yacht
Animal Nest	Rice	Furniture	Trailer
Boxes	Camera	Fashion Accessories	Moped

Stimuli for Typicality ratings contd.

FRUIT	POCKETS	
Apple	Keys	
Orange	Money	
Banana	Pens	
Pear	Wallet	
Grape	Handkerchief	
Melon	Tissues	
Grapefruit	Scraps of paper	
Lemon	Sweets	
Pineapple	Credit cards	
Kiwi Fruit	Cigarettes	
Pomegranate	Chewing gum	
Apricot	Receipt	
Peach	Lipstick	
Clementine	Condoms	
Avocado	Buttons	

APPENDIX F

TYPICALITY RATINGS: ANSWER BOOK SAMPLE PAGE

FRUIT

ORDER OF "GOODNESS OF EXAMPLE" **RATING OF "GOODNESS OF EXAMPLE"**

1.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent example)
2.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
3.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
4.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
5.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
6.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
7.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
8.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
9.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
10.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
11.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
12.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
13.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
14.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)
15.__________	1 2 3 4 5 6 7 8 9 (Poor … Excellent)

APPENDIX G

TYPICALITY RATINGS: MEANS AND STANDARD DEVIATIONS
(Items chosen for use in similarity study shown in bold)

TYPICALITY RATINGS: MEANS AND STANDARD DEVIATIONS

VEHICLE			FRUIT		
CAR	9	0	APPLE	8.53	1.61
BUS	8.02	1	ORANGE	8.51	1.45
LORRY	7.2	1.64	BANANA	7.98	1.73
TRAIN	7.02	1.82	PEAR	7.3	1.63
COACH	7.02	1.61	PEACH	6.37	1.7
VAN	7	1.79	GRAPE	6.33	2.07
PLANE	5.98	2.21	PINEAPPLE	5.91	1.95
BICYCLE	5.89	2.37	MELON	5.53	1.97
BOAT	5.41	2.23	LEMON	5.14	2.4
MOPED	4.84	1.94	CLEMENTINE	5.14	2.26
TRAM	4.25	1.88	APRICOT	5.02	2.02
YACHT	3.23	1.82	GRAPEFRUIT	5.02	2.01
TRAILER	3.11	1.93	KIWI	4.74	2.18
RICKSHAW	2.55	2.02	POMEGRANITE	3.14	2.36
SUBMARINE	2.41	1.7	AVOCADO	2.4	1.94
WEDDING			DOGS		
RINGS	6.66	2.48	CATS	8.14	1.09
PRESENT	7.5	1.72	STICKS	8.07	1.47
CAMERA	7.3	1.49	BALLS	7.52	1.7
CONFETTII	7.11	1.85	OTHER DOGS	6.61	1.99
FAMILY	6.7	2.54	POSTMENT	6.07	2.49
SUIT	6.43	2	RABBITS	5.66	2.37
FLOWERS	6.3	1.94	OWN TAILS	5.55	2.37
VIDEO CAMERA	6.16	1.95	BIRDS	5.45	2.26
CARD	5.89	2.33	BURGLARS	5.05	2.28
CARS	5.48	2.42	STRANGERS	5.02	2.2
CHAMPAGNE	5.11	2.13	SQUIRRELS	4.51	2.4
HAT	5.09	2.25	CARS	3.93	2.28
CAKE	4.8	2.52	BICYCLES	3.86	2.03
FOOD	3.91	1.84	OWNERS	3.82	2.3
RICE	2.93	2.24	COWS	1.84	1.35

TYPICALITY RATINGS: MEANS AND STANDARD DEVIATIONS Contd.

B.MARKET			ATTIC		
DRUGS	**7.41**	1.81	**BOXES**	**8.00**	**1.00**
GUNS	**7.02**	1.97	**WATER TANK**	**7.55**	**1.63**
ELECTRICAL GOODS	**6.52**	1.7	**DUST**	**7.09**	**2.63**
CIGARETTES	5.95	2.31	SUITCASES	6.73	2.10
CURRENCY	5.93	2.4	TOYS	6.64	1.75
FOOD	5.86	3.02	XMAS DECOS	6.64	2.16
VIDEOS	5.84	2.08	MEMORIES	6.36	2.46
TVS	5.82	1.79	COBWEBS	6.27	2.41
JEWELLERY	5.77	2.04	OLD CLOTHES	6.18	1.47
CLOTHES	5.14	1.9	OLD LETTERS	6.00	2.32
JEANS	4.91	2.09	**SPIDERS**	**5.55**	**2.16**
CARS	4.91	2.25	FURNITURE	5.55	1.81
FASHION ACCESSORIES	**3.64**	5.82	**ANTIQUES**	**5.27**	**1.74**
FURNITURE	**3.2**	1.86	**BOOKS**	**5.18**	**2.18**
MAGAZINES	**2.48**	1.64	ANIMAL NESTS	4.80	3.05
SHELF			**POCKETS**		
ORNAMENTS	**8.64**	**0.67**	**MONEY**	**8.43**	**0.9**
BOOKS	**7.91**	**2.12**	**WALLET**	**8.27**	**1.11**
PHOTOS	**7.73**	**1.27**	**KEYS**	**8.2**	**1.07**
VASES	6.45	1.86	TISSUE	6.75	1.81
CLOCKS	6.36	1.91	CIGARETTES	6.55	1.89
PLANTS	6.27	1.56	CHEWING GUM	6.36	1.91
VIDEOS	6	2.24	CREDIT CARD	6.11	2.21
CD	5.82	2.4	HANKY	5.75	2.14
BOTTLES	5.45	2.11	RECEIPT	5.39	2.01
PLATES	5.09	2.51	SWEETS	5	1.8
STEREO	4.91	2.34	PAPER	4.86	2.39
PICTURES	4.55	2.5	CONDOM	4.7	2.02
TELEPHONE	**3.64**	**1.57**	**PEN**	**4.61**	**2.06**
NEWSPAPER	**2.45**	**1.51**	**LIPSTICK**	**4.14**	**2.06**
SCREWS	**2**	**1.55**	**BUTTONS**	**2.41**	**2.09**

TYPICALITY RATINGS: MEANS AND STANDARD DEVIATIONS Contd.

FURNITURE			**BIRD**		
CHAIR	**8.45**	0.52	**BLUE TIT**	**8.5**	**0.71**
WARDROBE	**7.82**	1.17	**SPARROW**	**8.5**	**0.71**
TABLE	**7.82**	1.66	**ROBIN**	**8.2**	**1.87**
BED	7.55	1.29	BLACKBIRD	8.2	1.32
CHEST OF DRAWERS	7.36	1.21	THRUSH	8.1	0.99
SOFA	7.09	1.45	CROW	7.8	1.23
DESK	6.91	1.64	STARLING	7.56	2.07
CUPBOARD	6.55	1.86	EAGLE	7.1	1.6
DRESSING TABLE	6.36	2.42	GULL	6.7	2.1
STOOL	6.36	2.29	WREN	6.33	2.12
SIDEBOARD	5.82	2.93	OWL	6.2	2.26
BOOKSHELVES	5.64	1.75	VULTURE	5.9	2.18
FOOTSTOOL	**5.27**	2.49	**PELICAN**	**5.7**	**2.31**
PIANO	**2.82**	1.4	**HEN**	**4.6**	**2.27**
CUSHION	**2.55**	1.97	**EMU**	**4.2**	**2.66**

APPENDIX H

ITEM PAIRINGS FOR SIMILARITY STUDY

Item Pairings for similarity rating study

	Typical - Typical	Typical-atypical	Atypical-atypical
FRUIT			
	Apple-Orange	Apple-Avocado	Avocado-Pomegranate
	Orange-Banana	Apple -Pomegranate	Pomegranate-Kiwi
	Apple -Banana	Apple-Kiwi	Avocado-Kiwi
		Orange-Avocado	
		Orange-Pomegranate	
		Orange-Kiwi	
		Banana-Avocado	
		Banana-Pomegranate	
		Banana-Kiwi	
FURNITURE			
	Chair-Table	Chair-Cushion	Cushion-Piano
	Table-Wardrobe	Chair-Piano	Piano-Footstool
	Chair-Wardrobe	Chair-Footstool	Cushion-Footstool
		Table-Cushion	
		Table-Piano	
		Table-Footstool	
		Wardrobe-Cushion	
		Wardrobe-Piano	
		Wardrobe-Footstool	
BIRD			
	Sparrow-Blue Tit	Sparrow-Emu	Emu-Hen
	Blue-Tit - Robin	Sparrow-Hen	Hen-Pelican
	Sparrow-Robin	Sparrow-Pelican	Emu-Pelican
		Blue-Tit - Emu	
		Blue-Tit - Hen	
		Blue-Tit - Pelican	
		Robin - Emu	
		Robin-Hen	
		Robin-Pelican	
VEHICLE			
	Car-Bus	Car-Submarine	Submarine-Rickshaw
	Bus-Lorry	Car-Rickshaw	Submarine-Yacht
	Car-Lorry	Car-Yacht	Rickshaw-Yacht
		Bus-Submarine	
		Bus-Rickshaw	
		Bus-Yacht	
		Lorry-Submarine	
		Lorry-Rickshaw	
		Lorry-Yacht	

	Typical - Typical	Typical-atypical	Atypical-atypical
THINGS DOG CHASE			
	Cat-Stick	Cat-Cow	Cow-Owner
	Stick-Ball	Cat-Owner	Owner-Bicycle
	Cat-Ball	Cat-Bicycle	Cow-Bicycle
		Stick-Cow	
		Stick-Owner	
		Stick-Bicycle	
		Ball-Cow	
		Ball-Owner	
		Ball-Bicycle	
THINGS YOU FIND IN AN ATTIC			
	Box-Water Tank	Box-Book	Book-Antique
	Water Tank-Dust	Box-Antique	Antique-Spider
	Box-Dust	Box-Spider	Book-Spider
		Water Tank-Book	
		Water Tank-Antique	
		Water Tank-Spider	
		Dust-Book	
		Dust-Antique	
		Dust-Spider	
THINGS PEOPLE TAKE TO A WEDDING			
	Present-Camera	Present-Rice	Rice-Cake
	Camera-Confetti	Present-Cake	Cake-Hat
	Present-Confetti	Present-Hat	Rice-Hat
		Camera-Rice	
		Camera-Cake	
		Camera-Hat	
		Confetti-Rice	
		Confetti-Cake	
		Confetti-Hat	
		Photograph-Telephone	
		Clock-Screw	
		Clock- Newspaper	
		Clock- Telephone	

	Typical - Typical	Typical-atypical	Atypical-atypical
THINGS YOU FIND ON A SHELF			
	Ornament-Photograph	Ornament-Screw	Screw-Newspaper
	Ornament-Clock	Ornament-Newspaper	Newspaper-Telephone
	Photograph-Clock	Ornament-Telephone	Screw-Telephone
		Photograph- Screw	
		Photograph- Newspaper	
		Photograph- Telephone	
		Clock-Screw	
		Clock- Newspaper	
		Clock- Telephone	
THINGS SOLD ON THE BLACK MARKET IN RUSSIA			
	Drugs-Gun	Drugs- Magazine	Magazine -Furniture
	Gun-Electrical Goods	Drugs- Furniture	Furniture-Fashion Accessories
	Drugs-Electrical Goods	Drugs- Fashion Accessories	Magazine- Fashion Accessories
		Gun- Magazine	
		Gun- Furniture	
		Gun- Fashion Accessories	
		Electrical Goods-Magazine	
		Electrical Goods-Furniture	
		Electrical Goods-Fashion Accessories	
THINGS THAT MAY BE CONVENIENTLY KEPT IN POCKETS			
	Money-Wallet	Money-Button	Button-Lipstick
	Wallet-Key	Money-Lipstick	Lipstick-Pen
	Money-Key	Money-Pen	Button-Pen
		Wallet-Button	
		Wallet-Lipstick	
		Wallet-Pen	
		Key-Button	
		Key-Lipstick	
		Key-Pen	

APPENDIX I

INSTRUCTIONS AND SAMPLE PAGES OF MATERIALS FOR SIMILARITY STUDY

(No Context Retrospective Protocol)
(With Context Retrospective Protocol)
(No Context Concurrent Protocol)
(With Context Retrospective Protocol)
(No Context No Protocol)
(With Context No Protocol)

INSTRUCTIONS

Thank you for agreeing to take part in this study.

On each of the following pages you will see names of items presented together in pairs, for example:

SHIRT **SHOE**

You will be asked to consider how similar you think these items are to each other and to mark your judgement on a scale like this:

1	2	3	4	5	6	7	8	9
not similar at all								very similar

You indicate your judgement by circling the appropriate number on the scale. For example, if you felt that the items were quite dissimilar but did have some similarity to each other then you might want to circle the 3 or 4. If you felt that they were very similar you might want to circle the 7,8 or 9 depending on the strength of your feeling.

You will be required to write down the things that you considered when you made your judgement. Try to give as full account as possible.

You will be given 90 seconds to complete each page. You will be told when to start each page and when the 90 seconds allowed for each page is up.

OWNER ***BICYCLE***

HOW SIMILAR ARE AN **OWNER** AND A **BICYCLE**?

Mark your judgement about the similarity of an **OWNER** and a **BICYCLE** on the scale below

1	2	3	4	5	6	7	8	9
not similar at all								very similar

NOW WRITE DOWN THE THINGS YOU ARE CONSIDERING IN MAKING YOUR JUDGEMENT OF SIMILARITY.

INSTRUCTIONS

Thank you for agreeing to take part in this study.

On each of the following pages you will see the name of a category and below that the names of two members of that category presented together in pairs, for example:

THINGS THAT PEOPLE WEAR

SHIRT ***SHOE***

You will be asked to consider how similar you think these items are to each other and to mark your judgement on a scale like this:

1 2 3 4 5 6 7 8 9

not similar at all **very similar**

You indicate your judgement by circling the appropriate number on the scale. For example, if you felt that the items were quite dissimilar but did have some similarity to each other then you might want to circle the 3 or 4. If you felt that they were very similar you might want to circle the 7,8 or 9 depending on the strength of your feeling.

You will be required to write down the things that you considered when you made your judgement. Try to give as full account as possible.

You will be given 90 seconds to complete each page. You will be told when to start each page and when the 90 seconds allowed for each page is up.

THINGS DOGS CHASE

OWNER ***BICYCLE***

HOW SIMILAR ARE AN **OWNER** AND A **BICYCLE**?

Mark your judgement about the similarity of an **OWNER** and a **BICYCLE** on the scale below

1 2 3 4 5 6 7 8 9

not similar at all very similar

NOW WRITE DOWN THE THINGS YOU ARE CONSIDERING IN MAKING YOUR JUDGEMENT OF SIMILARITY.

INSTRUCTIONS

Thank you for agreeing to take part in this study.

On each of the following pages you will see names of items presented together in pairs, for example:

SHIRT ***SHOE***

You will be asked to consider how similar you think these items are to each other and to write down the things you are considering in making a judgement of similarity. Try to give as full an account as possible. When you have done this, you will be required to mark your judgement on a scale like this:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	
not similar at all										**very similar**

You indicate your judgement by circling the appropriate number on the scale. For example, if you felt that the items were quite dissimilar but did have some similarity to each other then you might want to circle the 3 or 4. If you felt that they were very similar you might want to circle the 7,8 or 9 depending on the strength of your feeling.

You will be given 90 seconds to complete each page. You will be told when to start each page and when the 90 seconds allowed for each page is up.

OWNER *BICYCLE*

HOW SIMILAR ARE AN **OWNER** AND A **BICYCLE**?

WRITE DOWN THE THINGS YOU ARE CONSIDERING IN MAKING YOUR JUDGEMENT OF SIMILARITY.

Now mark your judgement about the similarity of an **OWNER** and a **BICYCLE** on the scale below

1 2 3 4 5 6 7 8 9

not similar at all very similar

INSTRUCTIONS

Thank you for agreeing to take part in this study.

On each of the following pages you will see the name of a category and below that the names of two members of that category presented together in pairs, for example:

THINGS THAT PEOPLE WEAR

SHIRT ***SHOE***

You will be asked to consider how similar you think these items are to each other and to write down the things you are considering in making a judgement of similarity. Try to give as full an account as possible. When you have done this, you will be required to mark your judgement on a scale like this:

1 2 3 4 5 6 7 8 9

not similar at all **very similar**

You indicate your judgement by circling the appropriate number on the scale. For example, if you felt that the items were quite dissimilar but did have some similarity to each other then you might want to circle the 3 or 4. If you felt that they were very similar you might want to circle the 7,8 or 9 depending on the strength of your feeling.

You will be given 90 seconds to complete each page. You will be told when to start each page and when the 90 seconds allowed for each page is up.

THINGS DOGS CHASE

OWNER *BICYCLE*

HOW SIMILAR ARE AN **OWNER** AND A **BICYCLE**?

WRITE DOWN THE THINGS YOU ARE CONSIDERING IN MAKING YOUR JUDGEMENT OF SIMILARITY.

Now mark your judgement about the similarity of an **OWNER** and a **BICYCLE** on the scale below

1 2 3 4 5 6 7 8 9

not similar at all ... very similar

INSTRUCTIONS

Thank you for agreeing to take part in this study.

On each of the following pages you will see the names of items presented together in pairs, for example:

SHIRT ***SHOE***

You will be asked to consider how similar you think these items are to each other and to mark your judgement on a scale like this:

1 2 3 4 5 6 7 8 9

not similar at all **very similar**

You indicate your judgement by circling the appropriate number on the scale. For example, if you felt that the items were quite dissimilar but did have some similarity to each other then you might want to circle the 3 or 4. If you felt that they were very similar you might want to circle the 7,8 or 9 depending on the strength of your feeling. I would like you to give each pair and each rating *some* thought but ask you not to spend too long on each pair.

OWNER *BICYCLE*

HOW SIMILAR ARE AN **OWNER** AND A **BICYCLE**?

Mark your judgement about the similarity of an **OWNER** and a **BICYCLE** on the scale below

1	2	3	4	5	6	7	8	9
not similar at all								very similar

INSTRUCTIONS

Thank you for agreeing to take part in this study.

On each of the following pages you will see the name of a category and below that the names of two members of that category presented together in pairs, for example:

THINGS THAT PEOPLE WEAR

SHIRT	***SHOE***

You will be asked to consider how similar you think these items are to each other and to mark your judgement on a scale like this:

1	2	3	4	5	6	7	8	9
not similar at all								**very similar**

You indicate your judgement by circling the appropriate number on the scale. For example, if you felt that the items were quite dissimilar but did have some similarity to each other then you might want to circle the 3 or 4. If you felt that they were very similar you might want to circle the 7,8 or 9 depending on the strength of your feeling. I would like you to give each pair and each rating *some* thought but ask you not to spend too long on each pair.

THINGS DOGS CHASE

OWNER *BICYCLE*

HOW SIMILAR ARE AN **OWNER** AND A **BICYCLE**?

Mark your judgement about the similarity of an **OWNER** and a **BICYCLE** on the scale below

1 2 3 4 5 6 7 8 9

not similar at all very similar

APPENDIX J

INSTRUCTIONS AND SAMPLE PAGE OF MATERIALS FOR COLLECTION OF ITEM-TO-EVENT DATA

INSTRUCTIONS

Thank you for agreeing to take part in this study.

Please read the following instructions very carefully and ensure that you fully understand the task that is required of you.

When you think of an event or situation, there are a variety of items that may come to mind. For example, when I think of a lecture I think of a white board, an overhead projector, students, handouts etc. At the top of each of the following pages, you will see the name of an event or situation. For each page you will be allowed 90 seconds to list as many items as you can which the event or situation bring to mind. At the bottom of each page I would like you to indicate how often you have been in this situation by ticking one of the following options:

Never ☐ less than 10 times ☐ more than 10 times ☐ more than 50 times ☐

You will be told when to start each page and when to turn over and start the next page. Please do not start each page until you are told to do so.

HAVING A SHOWER

I have been in this situation

Never	less than 10 times	more than 10 times	more than 50 times
☐	☐	☐	☐

APPENDIX K

ITEMS GENERATED TO EVENTS (FREQUENCIES)

SHOWER		GOING OUT	3	DENTIST	9
SHAMPOO	24	BODYWASH	2	NEEDLES	7
FAVOURITE SHAMPOO		BATH CRÈME		WAITING ROOM	7
TOWEL	20	RADOX		DENTIST'S NURSE	7
WARM FLUFFY TOWEL		BODYSCRUB		WHITE COAT	6
WARM TOWELS		FACE WASH	2	LARGE WHITE COAT	
SHOWER GEL	16	MOISTURISER	2	MAGAZINES	6
SOAP	14	TILES	2	TOOTHBRUSH	6
NICE STRAWBERRY SOAP		BUF PUF	2	INJECTION	5
LIQUID SOAP		BATH CRÈME	2	RECEPTION	5
SHOWER CURTAIN	13	MIRROR	2	RECEPTION ROOM	
WATER	12	TOOTHPASTE	2	WAITING AREA	
HOT WATER		BODY	2	RECEPTION AREA	
COLD WATER		TEMPERATURE	2	RECEPTIONIST	5
CONDITIONER	10	COLD WATER	2	FACE MASK	5
FAVOURITE CONDITIONER		WARM	2	TEETH	5
HOT	8	REFRESHING	2	BRACES	4
HOT WATER	6	WAKE UP	2	MIRROR	4
STEAM	6	BATH	2	FEAR	4
SHAVE/SHAVING	6	SLIPPERY	2	WAITING	4
RAZOR	5	WASH	2	WAITING AND WAITING	
COLD	5	SCRUB	2	PEOPLE WAITING IN QUEUES	
CLEAN	5	SINGING	2	QUEUE	
WET	5	**DENTIST**		PLASTIC/RUBBER GLOVES	4
TAP(S)	4	DRILL	19	INSTRUMENTS	3
HOT TAP		(DENTIST) CHAIR	17	METAL INSTRUMENTS	
COLD TAP		RECLINING CHAIR		PICK	
SPONGE	4	BIG CHAIR		LONG METAL INSTRUMENTS	
BATHMAT	4	SEAT		SCALPEL LIKE ITEMS	
RELAXING	4	FILLINGS	16	SILVER INSTRUMENTS	
RELAX		LIGHT(S)	12	SMELL(AT DENTIST)	4
DEODERANT	3	BIG LIGHT IN YOUR FACE		SMELL OF DENTIST OFFICE	
TALC	3	LIGHT PROJECTOR		SMELL	
SHOWER HEAD	3	SWIVEL LAMP		DENTAL ASSISTANT	3
SHOWER UNIT	3	PAIN	10	RINSE	3
LOOFAH	3	PAINFUL		PULLING TEETH OUT	3
FLANNEL	3	PAIN IN MY MOUTH		HYGIENIST	3
HAIR	3	HURT		ANAESTHAETIC	2
BATHROBE	3	SORE		SMALL MIRROR	2
SHOWERHAT/CAP	3	MORE PAIN		POSTERS ON CEILING	2
WARMTH	3	MOUTHWASH	9	PICTURES ON CEILING	
MORNING	3	DRINK(PINK LIQUID)		POSTERS	2

DENTIST Contd.		XMAS PUDDNG	8	COLD	3
SPITBOWL	2	FOOD	8	LAUGHTER	3
WATER	2	BREAKFAST		GRAVY	2
GLOVES	2	BACON SANDWICH FOR BKFAST		DINNER	2
WISDOM TOOTH	2	SNACKS		ROAST DINNER	
TOOTHPASTE	2	CRISPS		FILMS	2
ORAL B		DECORATIONS	7	OLD FILMS	
COLGATE		(XMAS) CAKE	7	GOOD FILMS	
MONEY	2	DRINK(S)	7	JESUS CHRIST	2
CREDIT CARD		FIZZY DRINKS		CANDLES	2
FLOSS	2	SNOW	7	CAROLS	2
NERVOUSNESS	2	FRIENDS	7	HYMNS	
APPREHENSION		CHOCOLATE	6	PEOPLE	
DREAD	2	STOCKINGS	5	MEETING PEOPLE	
PANIC	2	FATHER XMAS	5	EARLY RISING	2
ROOT CANAL	2	SANTA		EARLY RISING	
TOOLS	2	(ROAST) POTATOES	5	VERY EARLY	
NUMB MOUTH	2	SPROUTS	5	EXCITEMENT	2
NUMB	2	STUFFING	5	HAPPINESS	2
NASTY SMELL2	2	CHILDREN	5	HAPPY	
WINDOW	2	WRAPPING PAPER	4	JOY	
OPEN MOUTH	2	LIGHTS	4	JOYFUL	
XMAS DAY		CHURCH	4	FUN	2
PRESENTS	23	GAMES	4	MIDNIGHT MASS	2
TREE	23	BOARD GAMES		PHONE CALLS	2
TURKEY	18	TRIVIAL PURSUITS		COOKING	2
FAMILY	15	BLOODY TRIVIAL PURSUITS		MESS	2
FAMILY GATHERING	2	MUSIC	4	**BEACH**	
BEING WITH FAMILY		SONGS		SANDCASTLES	24
RELATIVES		XMAS SONGS		SAND	23
FAMILY DAY		ROAST DINNER		GOLDEN SAND	
FAMILY FUN		LUNCH		SANDY	
HAVING ALL FAMILY OVER		XMAS DINNER	3	SEA	23
DAD		TRADITIONAL XMAS DINNER		SEASIDE	
SISTER-IN-LAW		MINCE PIES	3	SPADE	17
CARDS	10	SWEETS	3	BUCKET	16
CRACKERS	10	ALCOHOL	3	TOWEL	15
TELEVISION	10	SHERRY		LARGE TOWELS	
LATE NIGHT FILMS		LOTS OF ALCOHOL		SWIMMING COSTUME	14
OLD FILMS		WINE	3	SUN	13
GOOD FILMS		MIDNIGHT MASS	3	ICE CREAM	12
QUEEN"S SPEECH	9	JESUS'BIRTHDAY	3	SWIMMING	12
QUEEN 3PM		FIRE	3	SWIM	

BEACH Contd.		CLIFFS	2	RED FLAG	
CHILDREN	10	ROCKPOOL	2	TOES	2
SUNTAN LOTION	8	SEAWEED	2	PUNCH AND JUDY	4
(COLD) DRINKS	8	BEACHBALLS	2	**ATTIC**	
BEER		FOOTBALL	2	LADDER	20
SHELLS	7	FRISBEE	2	DARK	17
WATER	6	SUNHAT	2	BOXES	16
SALTWATER		BEACH TOWEL	2	BIG BOXES	
FAMILY	6	SUNBLOCK	2	DUST	13
HUSBAND		ICE LOLLIES	2	COBWEBS	9
PICNIC	6	FISH AND CHIPS	2	(OLD) TOYS	8
HOT	6	CANDY FLOSS	2	OLD TEDDY BEARS	2
BOATS	6	(PICNIC) BLANKET	2	OLD GAMES	
ROCKS	5	FAIRGROUND	2	JUNK	8
PIER	5	AMUSE,MENTS		SPIDERS	7
CRAB	5	AMUSEMENT ARCADES		RUBBISH	6
DECKCHAIR	5	AMUSEMENT PARK		DUSTY	6
FISH	4	ARCADES		(OLD) BOOKS	5
PEBBLES	4	FAIR		OLD CLOTHES	5
BALL	4	PEDALOES	2	CLOTHES	
JELLYFISH	4	DINGHY	2	XMAS DECORATIONS	5
SUNTAN OIL	4	GAMES	2	XMAS ITEMS	
FOOD TO TAKE	4	TOES	2	TORCH	5
FOOD		WET	2	WATER TANK	5
HOT DOGS		(EDIBLE ROCK)	1	COLD	5
BURGERS		SUNBATHING	2	INSULATION	4
SNACKS		LIFEGUARDS	2	INSULATOR	
CHIPS		PEOPLE	2	SUITCASES	3
SAUSAGE ROLLS		TOURISTS		CASES	
SEAFOOD		MEN		BABY CLOTHES	3
WINDBREAK	4	BLOKES		BLACK BAGS	3
SUNGLASSES	4	OLD MEN		LIGHT SWITCH	3
PARASOLS	4	SURFERS		LIGHTS	3
RUBBER RINGS	4	SUNBATHERS		TILES	3
INFLATABLES		SWIMMERS		SLATES	
LILO		SANDWICHES	2	ROOFTILES	
PUNCH AND JUDY	4	SOGGY SANDWICHES		FLOORBOARDS	3
WAVES	3	BEACH MAT	2	WOODENFLOORBOARD S	
BIKINI	3	FLAGS		PAPERS	2
HATS	3	YELLOW FLAG		BOARD GAMES	2
SHORTS	3	GREEN FLAG		PICTURES	2
SUNBURN	3	RED FLAG		ANTIQUES	2
VOLLEYBALL	3	GREEN FLAG		DUVETS	2

ATTIC Contd.		BICUITS	8	HOT WATER	2
BAGS	2	NESCAFE	5	COFFEE GRANUL;ES	2
BIN BAGS		KENCO		BRANDY	2
PLASTIC BAGS		GOLD BLEND		STEAM	2
PHOTOS	2	DECAFF	5	RELAXING	2
PHOTO ALBUMS		CAFETIERE	4	WATCH TV	2
MESS	2	REGULAR		BOILING WATER	2
LIGHT SWITCH	2	LATTE		WAITING	2
SLANTED ROOF	2	TURKISH COFFEE		STIR	3
ROOF		CAPPUCINO		WAKE UP	2
JOISTS		EXPRESSO		CHAT	2
BEAMS	2	FREEZE DRIED		CONVERSATION	
RAFTERS		IRISH COFFEE		**CHILD PARTY**	
WOODEN BEAMS		INSTANT		PRESENTS	20
PIPES	2	FILTER COFFEE		GIFTS	
WATER PIPES		HOT	4	BALLOONS	16
AERIAL	2	KITCHEN	4	GAMES	15
OLD ITEMS	2	FRIDGE	4	CAKE	14
OLD STUFF		CREAM	4	CLOWN	13
OLD THINGS		FRIENDS	4	NOISE	11
DIRTY	2	MOTHER		CANDLES	10
DIRT		PEOPLE		JELLY	10
DAMP	2	RELATIVES		CRISPS	9
STORAGE	2	BREAKFAST	4	HATS	9
SMELLY	2	SAUCER	3	PASS THE PARCEL	8
QUIET	2	TEASPOON	3	MUSICAL CHAIRS	8
SCARY	2	ELECTRICTY	3	PARTY BAGS	8
CREAKY	2	ELECTRIC		TAKE HOME BAGS	2
COFFEE		BOILING WATER	3	MUSIC	8
MILK	23	COFFEE BEANS	3	CHILDREN	7
SUGAR	23	TEA	3	LOTS OF CHILDREN	
KETTLE	18	TEABAGS		CAKES	7
KETTLE FULL OF BOILING WATER		CAFFEINE	3	PARENTS	6
WHITE ELECTRIC KETTLE		COFFEE MAKER	2	ENTERTAINERS	6
WATER	17	PERCOLATOR		FOOD	6
CUP	14	COFFEE JAR	2	SAUSAGE ROLLS	
SPECIAL CUP		FILTER	2	PARTY FOOD	
BIG CUP		SINK	2	BREAD STICKS	
SMALL CUPS		TAP	2	ICE CREAM	6
COFFEE	14	COLD TAP		JELLY AND ICE CREAM	
SPOON	13	PLUG	2	SWEETS	6
MUG	11	ELECTRIC PLUG		DRINKS	6
BIG MUG				WINE	

PARTY Contd.		EXCITEMENT	2	TOWELS	3
MAGICIAN	5	MESS	2	KEYS	3
MUSICAL CHAIRS	5	DISCO	2	MONEY	3
PARTY GAMES	4	THEMES	2	FIRST AID BOX	3
DONKEY		**PACKING CAR**		FAMILY	3
PIN THE TAIL ON THE DONKEY		(SUIT)CASES	16	CHILDDREN	
PARTY DRESSES	4	FOOD	9	DAD	
PARTY HATS	4	FOOD FOR JOURNEY		NOISY KIDS	
SILLY HATS		FOOD AND DRINK FOR JOURNEY		TYRES-PRESSUR AND TREAD	3
DECORATIONS	4	DRINKS		CDS	3
LAUGHTER	4	SNACKS		EXCITEMENT	3
SCREAMING	3	SOUP		LUGGAGE	2
CRYING	3	CLOTHES	9	HAND LUGGAGE	
MOTHERS	3	TROUSERS		BOXES	2
RELATIVES		SHIRTS		STRESS	2
FRUSTRATED PARENTS		TRACKSUIT BOTTOMS		FRIENDS	2
GRANDPARENTS		SOCKS		SWEETS	2
FAMILY		BOXER SHORTS		LAST MINUTE SWEETS	
BOUNCY CASTLE	3	SHORTS		SWIMWEAR	2
SANDWICHES	3	JEANS		SWIMMING TRUNKS	
HALL	3	BAGS	8	SUNHAT	2
FRIENDS	2	CARRIER BAGS		BIKE RACK	2
BIRTHDAY CAKE	2	BIG BAGS		PETROL	2
SAUSAGES	2	SMALL BAGS		SPARE TYRE	2
SAUSAGE ON STICKS	2	HEAVY BAGS		PILLOWS	2
SNACKS	2	HOLDALLS		TOO MUICH STUFF	2
CHOCOLATE	2	RUCKSACK		GAMES	2
BISCUITS	2	GIRLS' BAGS		CROSSWORDS	
PLASTIC CUPS	2	PASSPORT	6	SUNGLASSES	2
PAPER PLATES	2	ROOF RACK	5	CAMERA	2
PLATES		CASSETTE TAPES	5	VIDEO	2
FIZZY DRINKS2	2	MUSIC		SUNTAN CREAM	2
COKE		MUSIC TO LISTEN TO		SUNTAN OIL	
LEMONADE	2	BLANKETS	4	COOLER BAG	2
PARTY CLOTHES	2	TICKETS	4	LISTS	2
PRETTY DRESSES	2	DRINKS	4	**PICNIC**	
FANCY DRESS	2	BOOKS	4	DRINKS(S)	19
BIRTHDAY	2	SHOES	3	BASKET	16
STREAMERS	2	FLIP FPLOS		PICNIC BASKET	
TOYS	2	TRAINERS		SANDWICHES	15
CARDS	2	BOOT	3	FRESH SANDWICHES	
BIRTHDAY CARDS		CAR	3	CAKE(S)	13
DRESSING UP	2	TYRES- PRESSURE AND TREAD	3	BLANKET	13

PICNIC Contd.		WEATHER FORECAST	2	MAGAZINES	4
PICNIC BLANKET		PLASTIC CUPS	2	TINS	2
FRUIT	11	WET WIPES	2	PASTA	3
CRISPS	10	KITCHEN ROLL	2	TESCOS	3
FOOD	9	TOWEL	2	SAINSBURY	
BREAD	9	LARGE TOWEL		WAITROSE	
FRENCH BREAD		COOLBOX	2	(COLD) FREEZERS	3
ROLLS		COOLBAG		ALCOHOL	2
CUTLERY	8	BOX	2	REWARD CARD	2
PLATES	8	BOXES		STORE CARD	
CUPS	7	SUNAN LOTION	2	CREDIT CARD	2
BUTTER	6	MUSIC	2	BARGAINS	2
MARGE		MAKING SANDWICHES	2	2FOR1DEALS	2
CHEESE	6	**SUPERMARKET**		DELICOUNTER	2
NAPKIN	6	TROLLEY	23	CIGARETTES	2
WINE	6	CHECKOUT	16	CASHPOINT CARD	2
SUN	6	VEGETABLES	15	EGGS	2
SUNNY		FOOD	15	EXPENSIVE	2
SUNNY DAY		FOODS		SALES ASSISTANTS	2
HAMPER	4	FRUIT	13	CASHIERS	2
CHOCOLATE	4	MONEY	9	SCREAMING CHILDREN	2
SALAD	4	AISLES	9	CHILDREN	2
FORKS	3	SHOPPING LIST	8	WINE	2
KNIVES	3	BASKET	8	VIDEOS	2
SPOONS	3	SHOPPING BASKET		SHELF STACKERS	2
SAUSAGE ROLLS	3	MEAT	7	SHOPPING BAGS	2
THERMOS	3	BREAD	6	CARRIEIR BAGS	
HAM	3	PEOPLE	6	BOXES	2
FAMILY	3	DRINK	6	FROZEN FOOD	2
FRIENDS	3	DRINKS	5	TOILETRIES	2
BISCUITS	3	BOOZE		CAKES	2
RUG	3	TILL(S)	6	CHOCOLATES	2
JAM	2	CAR	5	RESTAURANT	2
PORK PIES	2	OFFERS	5	DOUGHNUTS	2
CRACKERS	2	CAR PAR	5	QUEUE	2
PIES	2	QUEUE	5	LONG QUEUE	3
SWEETS	2	MILK	5	EXPENSIVE	2
COKE	2	MUSIC CDS/TAPES	5	**CAMPING TRIP**	
FRUIT SQUASH	2	BAKERY	5	TENT	27
COFFEE	2	CLOTHES	4	SLEEPING BAGS	22
COUNTRYSIDE	2	DELICATESSANT	4	FIRES	10
PARK	2	BISCUITS	4	RUCKSACK	9
GRASS	2	SHELVES	4	FOOD	9

CAMPING Contd.		POLES	2	TEMPLE	
DRY FOOD		GAS STOVE	2	FAMILY	14
TINNED FOOD		CAMPING GAZ		MOTHER OF BRIDE	
TENT PEGS	8	CALOR GAS STOVE		MOTHER IN LAW	
COLD	7	CAMPING STOVE		FATHER OF BRIDE	
COMPASS	7	STOVE		FATHER IN LAW	
GRASS	6	PRIMUS STOVE		BRIDESMAIDS	14
SAUSAGES	6	GAS HOB		FOOD	14
TREES	6	TOWEL	2	VICAR(PRIEST)	12
TORCH	6	FAMILY	2	MINISTER	
RAIN	5	MUM AND DAD		RECEPTION	11
MARSHMALLOWS	5	DAMP	2	DANCING	11
MUD	5	CANVAS	2	FRIENDS	10
MUDDY		CARAVAN	2	FLOWERS	10
GROUND SHEET	4	TOILET BLOCK	2	DRINK(S)	9
CAR	4	SHOWER BLOCK		WINE	
CLOTHES	4	TOLIET PAPER	2	CHAMPAGNE	8
OLD CLOTHES		RAINCOAT	2	WHITE DRESS	7
WARM CLOTHES		STARS	2	CAR(S)	7
CHANGE OF CLOTHES		FOREST	2	POSH CARS	
CLEAN UNDERWEAR		WOOD		WEDDING CAR	
WATERPROOF		WOODS		BIG EXPENSIVE CARS	
WATERPROOF CLOTHING		LAKES	2	BEST MAN	7
ANORAK		FISH	2	SPEECH	7
BEANS	4	FISHING ROD	2	RING(S)	7
MAP	4	FRIENDS	2	WEDDING RING	
PEOPLE	4	BUGS	2	SUITS	6
CAMP FIRE	4	SPIDERS		HIRING A SUIT	6
FIRESIDE		FLIES		CAKE	6
WIND	3	LANTERNS	2	WEDDING CAKE	5
WINDY		WELLINGTON BOOTS	2	TIERED CAKE	
WET	3	DRINK(S)	2	PAGE BOY	6
TORRENTIAL RAIN		POLES	2	PHOTOGRAPHS	6
THUNDERSTORMS		HIKING	2	CONFETTI	6
FLOOD		MATCHES	2	GUESTS	5
FIELDS	3	DISCOMFORT	2	EXPENSIVE	4
RIVERS	3	COOKING	2	CHILDREN	4
INSECTS	3	SCOUTS	2	DRESS	4
BAGS	3	**WEDDING**		REGISTRY OFFICE	4
PILLOW	3	BRIDE	21	INVITATIONS	4
WATER	3	GROOM	21	MUSIC	4
WALKING	3	CHURCH	21	PARTY	4
FUN	3	SYNAGOGUE		HATS	3

WEDDING Contd.		WAVES	5	SELECT NICE SHOES	
TOP HAT		BIG WAVES	2	SPECIAL SHOES	
WEDDING DRESS	3	FEAR	4	BATH	6
BEAUTIFUL DRESS		DANGER	4	HAIR DO/STLE	5
WHITE	3	DANGERS		WASH HAIR	5
HONEYMOON	3	PANIC	4	CLOTHES	5
BOUQUETS	3	PEOPLE	4	FRIENDS	5
CEREMONY	3	ROCKS	3	PERFUME	4
HYMNS	3	DROWNING	3	DEODERANT	4
PRESENTS	3	HELP	3	MONEY	4
BALLOONS	2	FLARE	3	MUSIC	4
SERVICE	2	LIFE RING	3	PRESENT	4
PEOPLE	2	WHISTLE	3	TROUSERS	4
GIFTS	2	COASTGUARD	3	TIE	4
USHERS	2	BIG WAVES	2	POSSBLY TIE	4
BAG PIPES	2	BEACH	2	JACKET	3
CAMERAMAN	2	TITANIC	2	SHIRT	3
PHOTOGRAPHER	2	RNLA	2	SHAVE	3
BELLS	2	FAMILY	2	HAIRDRYER	3
SHOES	2	LIFE BELT	2	IRONING CLOTHES	3
JEWELLERY	2	CHOPPY SEAS	2	PERFUME	2
MOTHER IN LAW	2	WATER	2	HAIR	2
IN LAWS		SUNKEN SHIPS	2	FIX HAIR	
SMILES	2	SHIP	2	SORT HAIR OUT	
HEN NIGHT	2	TOWELS	2	BRUSH HAIR	
I DO	2	TV	2	DOING HAIR	
SPECIAL	2	AFRAID	2	NEW CLOTHES	2
COMMITMENT	2	DEATH	2	NAILS	2
RESCUE AT SEA		NOISE	2	DRESS	2
HELICOPTER	23	NAVY	2	DRESSING	2
LIFEBOAT	19	ARM BANDS	2	NICE CLOTHES	
LIFE JACKET	13	RESUSSITATION	2	SMART DRESS	
BOAT(S)	8	SINKING	2	SPECIAL DRESS	
FERRY		**OCCASION**		DRESS UP	2
YACHT		MAKEUP	12	SORTING OUT CLOTHES	2
LIFEGUARD	8	DOING MAKEUP		SMART	2
COLD	8	FACE POWDER		TIME	2
STORM(S)	6	CONCEALER		RUSH	2
WET	6	LIPSTICK		EXCITEMENT	2
WIND	5	EYESHADOW		NERVOUS	2
ROPE	5	MASCARA		ORGANISING TRANSPORT	2
SEA	5	SHOWER	11	BABYSITTER	2
DINGHIES	5	SHOES	7	WINE	2

OCCASION Cond.		USHERS	4		
HANDBAG	2	EXCITEMENT	4		
SELECT NICE BAG		COMEDY	4		
SUIT	2	ROMANCE	3		
TIGHTS	2	LAUGHTER	3		
AFTERSHAVE	2	COKE	3		
WEDDINGS	2	SURROUND SOUND	3		
BIRTHDAY		FRIENDS	3		
PARTIES		TOILET	3		
ANNIVERSARIES		NACHOS	3		
EATING OUT		ACTORS	2		
MEAL		STAR	2		
BABYSITTER	2	ACTRESSES	2		
FAMILY	2	FIZZYDRINK	2		
OUTFIT	2	WIDE SCREEEN	2		
CINEMA		UNCOMFORTABLE	2		
POPCORN	25	MONEY	2		
DARK	21	TALK	2		
THE DARK		LOTS OF PEOPLE	2		
DARK ROOM		CROWDS			
DRINKS	16	CHOCOLATE	2		
COLD DRINKS		TORCHES	2		
BIG DRINKS WITH STRAWS		NACHOS	2		
TICKET	12	PROJECTOR	2		
SWEETS	10	BUY TICKETS	2		
HUGE SCREEN	7	CRYING	2		
SCREEN	5	HORROR	2		
FILM PREVIEWS	6	SCIENCE FICTION	2		
ICE CREAM	6	AIR CONDITIONING	2		
PEOPLE	6	FALL ASLEEP	2		
SEAT	5				
RED SEATS					
FOLD DOWN SEATS					
BIG SEATS					
SEATING					
COMFY SEATS	3				
CHAIRS	3				
CHAIRS IN ROWS	2				
ADVERTS	5				
LOUD	5				
NOISE	5				
FILM	4				
HOT DOGS	4				

APPENDIX L

INSTRUCTIONS AND SAMPLE MATERIALS FOR COLLECTING EVENT-TO-ITEM DATA

INSTRUCTIONS

Thank you for agreeing to take part in this study.

Please read the following instructions very carefully and ensure that you fully understand the task that is required of you.

When you think of an object or item, there are a variety of events/situations which may come to mind. For example, when I think of a newspaper, I may think of being in a shop buying one, or reading one on the train, or someone using one to light a fire or putting them out in the rubbish. On each of the following pages, you will see a word which is the name of an object or item. Some of the words could be treated as verbs but I want you to think of them all as names of objects or items. For each page, you will be allowed 30 seconds to list as many events or situations as you can which the object/item brings to your mind. I will tell you when to start each page.

Please indicate your age, gender and course code below:

Age..........................
Gender......................
Course code................

SAND

APPENDIX M

EVENTS GENERATED TO ITEMS (FREQUENCIES)
(PLUS OTHER EVENTS FROM ORIGINAL SET GENERATED TO SAME ITEMS)

ITEM	TARGET EVENT	FREQUENCY (N=15)	ITEM	TARGET EVENT	FREQUENCY (N=15)
BUCKET	BEACH	7	CARAVAN	CAMPING	2
CANDYFLOSS	BEACH	0	COMPASS	CAMPING	5
DINGHY	BEACH	0	DISCOMFORT	CAMPING	0
FRISBEE	BEACH	10	FISH	CAMPING	0
LIFEGUARD	BEACH	10	FISHING ROD	CAMPING	0
SAND	BEACH	9	GRASS	CAMPING	0
SANDCASTLE	BEACH	12	LANTERN	CAMPING	3
SEA	BEACH	6	MATCHES	CAMPING	0
SEAWEED	BEACH	9	RUCKSACK	CAMPING	4
SHELLS	BEACH	10	SAUSAGES	CAMPING	0
SPADE	BEACH	2	SCOUTS	CAMPING	9
SUNHAT	BEACH	6	SLEEPING BAG	CAMPING	14
SUNTAN LOTION	BEACH	4	STOVE	CAMPING	0
SWIMMING COSTUME	BEACH	7	TENT PEGS	CAMPING	10
TOES	BEACH	3	TENT	CAMPING	11
WET	BEACH	0	WELLINGTON BOOTS	CAMPING	0
BISCUITS	COFFEE	6	CHAIR	DENTIST	0
BRANDY	COFFEE	0	DENTIST	DENTIST	9
COFFEE	COFFEE	0	DREAD	DENTIST	0
CUP	COFFEE	6	DRILL	DENTIST	0
ELECTRICITY	COFFEE	0	FILLINGS	DENTIST	11
FILTER	COFFEE	6	FLOSS	DENTIST	6
FRIDGE	COFFEE	0	GLOVES	DENTIST	0
KETTLE	COFFEE	13	LIGHT	DENTIST	0
MILK	COFFEE	4	MOUTHWASH	DENTIST	5
PLUG	COFFEE	0	NEEDLE	DENTIST	0
RELAXING	COFFEE	0	PAIN	DENTIST	0
SINK	COFFEE	0	POSTERS	DENTIST	0
SPOON	COFFEE	5	SMALL MIRRORS	DENTIST	0
SUGAR	COFFEE	0	TOOTHPASTE	DENTIST	2
TAP	COFFEE	0	WINDOW	DENTIST	0
WATER	COFFEE	0	WISDOM TOOTH	DENTIST	8

ITEM	TARGET EVENT	FREQUENCY (N=15)	ITEM	TARGET EVENT	FREQUENCY (N=15)
BLANKET	PICNIC	0	BATHROBE	SHOWER	7
COOLBOX	PICNIC	11	BODY	SHOWER	0
CRISPS	PICNIC	0	BODYWASH	SHOWER	10
FRUIT SQUASH	PICNIC	0	CONDITIONER	SHOWER	3
FRUIT	PICNIC	0	FLANNEL	SHOWER	10
KITCHEN ROLL	PICNIC	0	HAIR	SHOWER	2
NAPKIN	PICNIC	0	HOT	SHOWER	0
PLASTIC CUP	PICNIC	4	MOISTURISER	SHOWER	0
PLATES	PICNIC	0	RAZOR	SHOWER	2
PORK PIES	PICNIC	3	SHAMPOO	SHOWER	4
RUG	PICNIC	0	SHOWER CURTAIN	SHOWER	10
SANDWICHES	PICNIC	5	SOAP	SHOWER	6
SUN	PICNIC	0	STEAM	SHOWER	0
THERMOS	PICNIC	5	TALCUM POWDER	SHOWER	2
WETWIPES	PICNIC	3	TILES	SHOWER	0
WINE	PICNIC	0	TOWEL	SHOWER	10
CANDLES	XMAS DAY	2	AERIAL	ATTIC	0
CARDS	XMAS DAY	9	ANTIQUES	ATTIC	0
CAROLS	XMAS DAY	10	BAGS	ATTIC	0
COLD	XMAS DAY	0	BOXES	ATTIC	0
CRACKERS	XMAS DAY	13	COBWEBS	ATTIC	0
DINNER	XMAS DAY	0	DAMP	ATTIC	0
FILMS	XMAS DAY	0	DUST	ATTIC	0
HAPPINESS	XMAS DAY	0	JUNK	ATTIC	0
MINCE PIES	XMAS DAY	13	LADDER	ATTIC	3
QUEENS SPEECH	XMAS DAY	9	LIGHT SWITCH	ATTIC	0
SNOW	XMAS DAY	7	OLD BOOKS	ATTIC	0
TELEVISION	XMAS DAY	0	PHOTOS	ATTIC	0
TREE	XMAS DAY	0	PIPES	ATTIC	0
TURKEY	XMAS DAY	13	SPIDERS	ATTIC	2
WRAPPING PAPER	XMAS DAY	13	SUITCASES	ATTIC	0
XMAS PUDDINGS	XMAS DAY	6	XMAS DECORATIONS	ATTIC	0

ITEM	TARGET EVENT	FREQUENCY (N=15)	ITEM	TARGET EVENT	FREQUENCY (N=15)
BALLOONS	CHILDS PARTY	*14*	AISLES	SUPERMARKET	*6*
BOUNCY CASTLE	CHILDS PARTY	6	BASKET	SUPERMARKET	9
CAKE	CHILDS PARTY	0	CASHIERS	SUPERMARKET	8
CHOCOLATE	CHILDS PARTY	0	CASHPOINT	SUPERMARKET	3
CLOWN	CHILDS PARTY	3	CHECKOUT	SUPERMARKET	9
DECORATIONS	CHILDS PARTY	0	CIGARETTES	SUPERMARKET	0
HATS	CHILDS PARTY	0	CREDIT CARD	SUPERMARKET	9
JELLY	CHILDS PARTY	10	EXPENSIVE	SUPERMARKET	6
LEMONADE	CHILDS PARTY	0	MAGAZINES	SUPERMARKET	0
NOISE	CHILDS PARTY	0	MEAT	SUPERMARKET	0
PAPER PLATES	CHILDS PARTY	6	MONEY	SUPERMARKET	7
PARTY BAG	CHILDS PARTY	7	SHOPPING LIST	SUPERMARKET	13
PRESENTS	CHILDS PARTY	0	TINS	SUPERMARKET	7
PRETTY DRESS	CHILDS PARTY	0	TROLLEY	SUPERMARKET	13
STREAMERS	CHILDS PARTY	0	VEGETABLES	SUPERMARKET	0
TOYS	CHILDS PARTY	0	VIDEOS	SUPERMARKET	0
BAGPIPES	WEDDING	*4*	ACTRESS	CINEMA	*1*
BELLS	WEDDING	8	ADVERTS	CINEMA	0
BRIDE	WEDDING	14	COKE	CINEMA	0
CAMERAMAN	WEDDING	2	DARK	CINEMA	0
CHAMPAGNE	WEDDING	6	ICE CREAM	CINEMA	0
CHURCH	WEDDING	6	LAUGHTER	CINEMA	0
CONFETTI	WEDDING	15	NACHOS	CINEMA	4
FLOWERS	WEDDING	3	POPCORN	CINEMA	14
HYMNS	WEDDING	3	PROJECTOR	CINEMA	3
INVITATIONS	WEDDING	7	SCREEN	CINEMA	8
JEWELLERY	WEDDING	0	SEAT	CINEMA	0
RING	WEDDING	9	SWEETS	CINEMA	0
SHOES	WEDDING	0	TICKET	CINEMA	3
SMILE	WEDDING	0	TOILET	CINEMA	0
VICAR	WEDDING	4	TORCH	CINEMA	0
WHITE DRESS	WEDDING	15	USHER	CINEMA	4

OTHER TARGET EVENTS GENERATED FROM ITEMS

LANTERN	CHRISTMAS	3
BRANDY	CHRISTMAS	2
WATER	SHOWER	4
	BATH	8
COOLBOX	HOLIDAYS	5
	BEACH	3
FRUIT SQUASH	SUPERMARKET	2
WINE	SUPERMARKET	2
HOT	HOLIDAYS	6
	BEACH	2
STEAM	BOILING KETTLE	5
	MAKING TEA	2
TOWEL	BEACH	6
CANDLES	BIRTHDAYS	7
CARDS	BIRTHDAYS	12
FILMS	CINEMA	12
DAMP	SHOWER	3
PHOTOS	HOLIDAYS	6
SUITCASES	HOLIDAYS	11
XMAS DECORATIONS	DECORRATING XMAS TREE	9
	CHRISTMAS	3
CAKE	SUPERMARKET	4
	CHRISTMAS	5
CHOCOLATE	CHRISTMAS	2
DECORATIONS	CHRISTMAS	15
PRETTY DRESS	WEDDING	4
STREAMERS	CHRISTMAS	2
AISLES	WEDDING	8
BASKET	PICNICS	3
HYMNS	CHRISTMAS	4
SWEETS	DENTIST	3
TORCH	CAMPING	4
USHER	WEDDING	12

CANDYFLOSS	FUNFAIR	12
LIFEGUARD	SWIMMING POOL	9
SPADE	GARDENING	10
WET	RAINING	9
SWIMMING COSTUME	SWIMMING POOL	8
STOVE	COOKING	13
SINK	WASHING UP	11
POSTERS	BEDROOMS	8
WISDOM TOOTH	IN PAIN	11
NEEDLE	SEWING	9
BLANKET	KEEPING WARM	8
KITCHEN ROLL	MOPPING UP SPILLS	11
NAPKIN	EATING OUT	8
PLATES	WASHING UP	9
FLANNEL	WASHING SELF	10
WET WIPES	FEEDING BABY	8
SANDWICHES	HAVING LUNCH	8
SHAMPOO	WASHING HAIR	10
STEAM	SEEING STEAM TRAINS	8
CANDLES	GOING TO CHURCH	8

APPENDIX N

STIMULUS PAIRS USED IN FINAL STUDY

Stimulus Pairs used in final study.

BEACH		
1	HH-LH	Bucket - Frisbee
2	LL-HH	Suntan Lotion - Sandcastle
3	HL-LL	Spade -Dinghy
4	LH-LH	Lifeguard-Seaweed
5	HH-HH	Sandcastle-swimming Costume
6	HL-LH	Sun-Shells
7	LL-LL	Toes-Suntan Lotion
8	HH-HL	Sea-Spade
9	HL-HL	Spade-Sun
10	LH-LL	Frisbee-candyfloss

COFFEE		
1	HH-LH	Cup-Biscuits
2	LL-HH	Fridge - Kettle
3	HL-LL	Coffee - Tap
4	LH-LH	Biscuits - Filter
5	HH-HH	Kettle - Cup
6	HL-LH	Milk-Filter
7	LL-LL	Brandy-Electricity
8	HH-HL	Kettle - Sugar
9	HL-HL	Water - Sugar
10	LH-LL	Biscuits - Sink

SHOWER		
1	HH-LH	Soap-bathrobe
2	LL-HH	Razor-Shower Curtain
3	HL-LL	Conditioner –Talcum Powder
4	LH-LH	Bodywash-Bathrobe
5	HH-HH	Soap-Shower Curtain
6	HL-LH	Shampoo-Flannel
7	LL-LL	Hair -Tiles
8	HH-HL	Towel-Conditioner
9	HL-HL	Shampoo-Conditioner
10	LH-LL	Flannel-Steam

XMAS DAY		
1	HH-LH	Presents – Mince Pies
2	LL-HH	Candles - Cards
3	HL-LL	Tree - Dinner
4	LH-LH	Wrapping Paper-Xmas Puddings
5	HH-HH	Turkey- Cards
6	HL-LH	Television - Snow
7	LL-LL	Films - Cold
8	HH-HL	Crackers - Tree
9	HL-HL	Tree - Television
10	LH-LL	Snow - Dinner

SUPERMARKET		
1	HH-LH	Checkout -Tins
2	LL-HH	Money -Trolley
3	HL-LL	Vegetables - Cigarettes
4	LH-LH	Basket - Cashiers
5	HH-HH	Checkout - Trolley
6	HL-LH	Fruit – Credit Card
7	LL-LL	Meat - Videos
8	HH-HL	Trolley - Vegetables
9	HL-HL	Vegetables - Fruit
10	LH-LL	Shopping List - Magazines

CHILDS PARTY		
1	HH-LH	Balloons Paper Plates
2	LL-HH	Streamers - Jelly
3	HL-LL	Noise - Hats
4	LH-LH	Party Bag – Paper Plates
5	HH-HH	Jelly - Balloons
6	HL-LH	Cake – Party Bags
7	LL-LL	Bouncy Castle - Chocolate
8	HH-HL	Jelly - Clown
9	HL-HL	Cake - Clown
10	LH-LL	Paper Plates - Lemonade

WEDDING		
1	HH-LH	Church Champagne
2	LL-HH	Jewellery - Church
3	HL-LL	Flowers - Bagpipes
4	LH-LH	Confetti - Cameraman
5	HH-HH	Bride - Church
6	HL-LH	Vicar - Ring
7	LL-LL	Shoes - Smile
8	HH-HL	Bride - Flowers
9	HL-HL	Vicar - Flowers
10	LH-LL	White dress - hymns

APPENDIX O

INSTRUCTIONS AND SAMPLE MATERIALS FOR FINAL STUDY
(With context blocked)
(With context random)
(No context blocked)
(No context random)

Thank you for agreeing to participate in this study.

On the following pages you will see lists of pairs of items. At the top of each page, you will see the name of an event or situation in which these items can be found. I would like you to look at each pair of items and think about them as they relate to that event or situation and then give a rating of how similar you think the items are to each other on the scale provided. The scale looks like this:

1 2 3 4 5 6 7 8 9

not similar at all — very similar

Please circle the number that indicates your rating. So, if you think that the items are not similar at all you would circle 1 but if you feel that they are very similar then you would circle 9. Use the numbers in between the extremes of the scale to circle intermediate ratings of similarity.

There is no time limit. Take your own time.

There are no right or wrong answers. I am interested in your own reactions. Please make sure that you have given a rating for every pair.

Gender..................

Age.....................

Course code.................

No. of years lived in Britain ....

A DAY AT THE BEACH

Think about : **A DAY AT THE BEACH**
How similar are the following items to each other?
SANDCASTLE SWIMMING COSTUME

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A DAY AT THE BEACH**
How similar are the following items to each other?
SPADE SUN

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A DAY AT THE BEACH**

How similar are the following items to each other?
BUCKET FRISBEE

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A DAY AT THE BEACH**
How similar are the following items to each other?
TOES SUNTAN LOTION

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A DAY AT THE BEACH**
How similar are the following items to each other?
LIFEGUARD SEAWEED

1 2 3 4 5 6 7 8 9

not similar at all very similar

Thank you for agreeing to participate in this study.

On the following pages you will see lists of pairs of items. Just above the pairs of items, you will see the name of an event or situation in which these items can be found. I would like you to look at each pair of items and think about them as they relate to that event or situation and then give a rating of how similar you think the items are to each other on the scale provided. The scale looks like this:

1 2 3 4 5 6 7 8 9

not similar at all ... very similar

Please circle the number that indicates your rating. So, if you think that the items are not similar at all you would circle 1 but if you feel that they are very similar then you would circle 9. Use the numbers in between the extremes of the scale to circle intermediate ratings of similarity.

There is no time limit. Take your own time.

There are no right or wrong answers. I am interested in your own reactions. Please make sure that you have given a rating for every pair.

Gender..................

Age......................

Course code.................

No. of years lived in Britain ....

Think about : **MAKING A CUP OF COFFEE**
How similar are the following items to each other?
WATER SUGAR

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A DAY AT THE BEACH**
How similar are the following items to each other?
SPADE SUN

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A TRIP TO THE SUPERMARKET**
How similar are the following items to each other?
CHECKOUT TINS

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A CHILDREN'S PARTY**
How similar are the following items to each other?
BOUNCY CASTLE CHOCOLATE

1 2 3 4 5 6 7 8 9

not similar at all very similar

Think about : **A DAY AT THE BEACH**
How similar are the following items to each other?
TOES SUNTAN LOTION

1 2 3 4 5 6 7 8 9

not similar at all very similar

Thank you for agreeing to participate in this study.

On the following pages you will see lists of pairs of items. I would like you to look at **each pair of items** and then give a rating of how similar you think the items are to each other on the scale provided. The scale looks like this:

1 2 3 4 5 6 7 8 9

not similar at all ... very similar

Please circle the number that indicates your rating. So, if you think that the items are not similar at all you would circle 1 but if you feel that they are very similar then you would circle 9. Use the numbers in between the extremes of the scale to circle intermediate ratings of similarity.

There is no time limit. Take your own time.

There are no right or wrong answers. I am interested in your own reactions. Please make sure that you have given a rating for every pair.

Gender..................

Age.....................

Course code.................

No. of years lived in Britain ....

How similar are the following items to each other?
SANDCASTLE SWIMMING COSTUME

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
SPADE SUN

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
BUCKET FRISBEE

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
TOES SUNTAN LOTION

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
LIFEGUARD SEAWEED

1 2 3 4 5 6 7 8 9

not similar at all very similar

Thank you for agreeing to participate in this study.

On the following pages you will see lists of pairs of items. I would like you to look at **each pair of items** and then give a rating of how similar you think the items are to each other on the scale provided. The scale looks like this:

1 2 3 4 5 6 7 8 9

not similar at all — very similar

Please circle the number that indicates your rating. So, if you think that the items are not similar at all you would circle 1 but if you feel that they are very similar then you would circle 9. Use the numbers in between the extremes of the scale to circle intermediate ratings of similarity.

There is no time limit. Take your own time.

There are no right or wrong answers. I am interested in your own reactions. Please make sure that you have given a rating for every pair.

Gender..................

Age.....................

Course code.................

No. of years lived in Britain ....

How similar are the following items to each other?
WATER SUGAR

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
SPADE SUN

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
CHECKOUT TINS

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
BOUNCY CASTLE CHOCOLATE

1 2 3 4 5 6 7 8 9

not similar at all very similar

How similar are the following items to each other?
TOES SUNTAN LOTION

1 2 3 4 5 6 7 8 9

not similar at all very similar