CHAPTER FOUR

EXPERIMENT TWO

4.00 CHAPTER OVERVIEW

The main focus of this chapter is a detailed description of the rationale and method underlying Experiment 2. Section 4.01 sets out the question and issues which are tested in Experiment 2. Section 4.02 describes the theoretical background of some of the issues tested in Experiment 2, including Neisser's (1967) description of categories as either well- or ill-defined. This view of categories has given rise to the unitary approach in researchers' interpretation of their data, with regard to how categories are represented, and their exemplars categorized and structured.

Section 4.03 sets out the unitary approach to the categorization process, then presents an alternative account of categorization processes which takes the form of a two-stage process of comparison. Two assumptions made by the unitary approach about processing of exemplars are tested in Experiment 2. Section 4.04 deals with the unitary approach to membership structure, and Section 4.05 deals with the mental representation of the category. Evidence which contradicts these descriptions is set out. Two assumptions made by the unitary view of structure and representation are tested in this experiment, and described in this section.

Section 4.06 summarises the four assumptions about representation, structure and process underlying the unitary approach, which were tested in Experiment 2. Briefly, the issue for each of the three category-types is not only whether it possesses a membership structure based upon degrees-of-typicality, but also whether its structure is derived from a one-stage process of similarity comparison. Section 4.07 develops a rationale for the tasks chosen as measures of representation of categories, their membership structure and of the categorization process. Section 4 08 is a description of the method and procedures by which Experiment 2 was conducted.

Throughout Experiment 2, data analyses used a participants-based format, rather than an items-based format, so that individuals might be considered. The probability value of .05 is accepted as the alpha level of significance. Section 4.09 deals with analyses of the participants' categorization processes, where the first two assumptions of the unitary approach were tested. Section 4.10 reports on tests of the unitary view's assumptions concerning the internal structure of membership structure of categories and their definedness. The data are analyzed with regard to whether typicality can predict membership structure, and also whether people agree about category boundaries. The results are summarised in Results Summary One. Section 4.11 deals with the data analysis concerning the kind of knowledge which constitutes concept representations. The implications of these results for a single or dual representation of each category-type are listed in Results Summary Two.

Section 4.12 begins with interpretations of the data analyses, discussing the experiment's main findings in turn with regard to process, structure and representation of categories. Finally, possible implications for a theory of concept representation are discussed in Section 4.13.

A unitary approach to people's understanding of categories would claim that people represent natural superordinate category-types as fuzzy prototypes, and that other category-types, such as property types, are represented as rules or definitions. Briefly stated, the general aim of Experiment 2 was to examine whether people hold different representations for different types of categories (for example, natural superordinate versus property types). An alternative possibility, not tested in this experiment, is that people can hold different representations for the same concept.

4.01 QUESTION AND ISSUES OF EXPERIMENT 2

Experiment 1 confirmed that the production frequency distributions in each of the three category-types (superordinates, properties, and ad hoc) did reflect an underlying organized structure of some kind. Since the data-analysis was concerned with patterns of agreement across subjects, it might safely be assumed that the kind of structure was one of normative knowledge. Rosch and her colleagues would claim that these gradients were represented as structures of similarity (that is, prototypes) (Mervis & Rosch, 1981; Rosch, 1975b). The significant interaction between category-type and production frequency levels, however, suggested that these underlying normative structures differed across the three category-types, but the nature of the data did not allow a valid conclusion to be drawn about how they might differ. Consequently, the main question asked in Experiment 2 is whether different category-types might be represented differently. For example, a category might have one representation for the person's concept of the category (i.e., its intension), and another representation for the actual items which constitute that category (i.e., its extension).

One issue raised in the discussion of Experiment 1 concerned subject performance and normative knowledge. If people represent category knowledge in an idiosyncratic organization, yet understand normative structures also, the idiosyncratically based structure might be more detailed with regard to information than the normative structure. For example, the influence of idiosyncratic information upon an individual's category behaviour might be of relevance if a more distinctive account of an individual's use of categories is to be made (Barsalou, 1987; Freyd, 1983; Kahnemann & Miller, 1986; Kahnemann & Tversky, 1972)).

Throughout the presentation of Experiment 2, a distinction will be made between idiosyncratic and normative stimuli. Each task will be carried out twice: once using idiosyncratic stimuli and once with normative stimuli. In this way, a subjective organization of a category's representation, structure and process can be compared with the objective organization of the category, and analyzed as to cognitive efficiency. The experiment compares people's performance on tasks like order of exemplar-generation and production frequency, membership decision, typicality ranking, and an "experience" measure: frequency-of-instantiation ranking. A description of these tasks is found in Section 4.08 and in Appendix E. The results should allow conclusions to be drawn about suitable category-models, perhaps with a different representational model being appropriate for each category-type.

4.02 DEFINEDNESS AND THE UNITARY APPROACH

Neisser's (1967) notion of definedness is concerned with a category's boundaries. "Definedness" concerns whether an object's or creature's membership/non-membership in a category involves a fuzzy or absolute categorization decision. Neisser saw category structure as a dichotomy of definedness, with well-defined categories at one pole being contrasted with ill-defined categories at the opposite pole. He did not, however, describe definedness of categories as a continuum. Neisser (1987) assumed that the nature of categories was such that they were either well-defined or ill-defined, with the conditions for membership being either a precise rule or a fuzzy prototype.

Category boundaries are considered to be well-defined and clear-cut if the conditions for membership state some precise rule, so that category instances are either members of the category or not. According to this assumption, well-defined categories have: firstly, members which must all contain a critical, invariant feature (for example, three lines); a limited variability in membership; and underlying category-dimensions (for example, triangular shape) which can be specified precisely. In short, membership depends upon a potential instance having the necessary and sufficient features to define the category. Most people have strong intuitions that any category possesses a clear definition of its membership conditions, but for the most part, they cannot specify these exactly (Armstrong, Gleitman & Gleitman, 1983; McNamara & Sternberg, 1983). The property categories used in Experiment 1 seem to involve the kind of decisions necessary for well-defined types. For example, in producing exemplars for *Red Things*, the rule "the object must be coloured red" will apply.

On the other hand, ad hoc categories are the unknown quantity, since their exemplars do not share a physical similarity, or an immediately discernible categorization rule. When the conditions for membership are imprecise, then the category's boundaries will be ill-defined, with no clear way of distinguishing membership from non-membership. Decisions about membership cannot be absolute, and are accompanied by uncertainty to a greater or lesser degree. The acceptance or rejection of new items as members depends upon an instance's *degree* of resemblance to the prototype, which means only a fuzzy idea of the conditions for belonging to a category. A category can be said to be ill-defined when its members are infinitely variable, have varying degrees of membership status, possess obscure category dimensions, and some members are better examples than others, including nonmembers. The natural superordinate categories used in Experiment 1 are examples of ill-defined types, and their representation, structure and process as tested in this experiment should reflect the features of ill-definedness.

Accounts which describe categories as ill-defined or well-defined types have a common ground and approach to category behaviour, being derived from theories of how environmental structure constrains and determines our mental representations of categories. Consequently, both accounts adopt a unitary approach to categorization behaviour, claiming a one-to-one relationship between the categories in the environment and people's representation of them. No allowance is made in either theory for active interpretation of the environment, through imposition of personal meaning upon it. Category knowledge is either normative or objective. In Experiment 2, the unitary approach is tested mainly with regard to its assumptions underlying prototype theories of ill-definedness.

The unitary view originated in a study by Rosch and Mervis (1975), who investigated what they termed natural categories (both biological and artifactual kinds). Their description of natural categories consisted of lists of independent features characteristic of the category. The principle of family resemblance is said to govern the category structure, in that membership is the result of attribute matching between the physical features of a potential instance in the world and the independent feature list of a category representation. The degree of overlap of physically similar features equals the representativeness of the instance, or how good an example it is of its category. Degrees of representativeness result in a graded structure of the items belonging to the category, so that instead of being equivalent (as in well-defined categories), the members of the category vary in typicality, depending upon the degree of overlap (Mervis & Rosch, 1981; Rips, Shoben & Smith, 1973). Rosch and Mervis (1975) interpreted this graded structure as reflecting not only typicality, but also the membership structure of the instances in the category.

In her early research, Rosch had refused to nominate any one of the three prototypes described in chapter 2 - feature lists, abstract composites, or exemplar instances - as being the "true" prototype representation for category-models. By its very nature, a fuzzy prototype cannot be precisely defined, except to say that it *is* undefined. At the same time, she interpreted the correlations between the gradience in goodness-of-example ratings and categorization response times as directly reflecting the internal structure of the category-members (Rosch, 1975a; Rosch & Mervis, 1975). In short, she was providing a unitary account of natural or semantic categories: their representation, internal graded structure and categorization processes all were said to be due to the same underlying cause - prototypicality.

At other times, however, as in her 1973 article "On the internal structure of perceptual and semantic categories", she wrote that the *formal* definition of categories does not fully correspond to the category as a psychological unit. Finally, in 1978, she said that the notion of prototypicality should be used to explain certain phenomena concerned with variation in typicality, but its role should not be confused with cognitive processing, representation, and formation of categories:

> ... the empirical findings about prototypicality have been confused with theories of processing - that is, there has been a failure to distinguish the structure of categories from theories concerning the use of that structure in processing. (Rosch, 1978, p.36).

Since the late 1970's, Rosch herself has consistently distinguished between what her experimental results show, and any theories that might account for those results. She has reached the conclusion that the graded structure of physical appearance, of the sort produced by typicality ratings, does not have a one-to-one correspondence with people's mental representation or concept of the category (Rosch, 1983). Prototype effects are superficial phenomena which may have many sources other than degree of membership (Lakoff, 1987).

Rosch's early unitary approach has given rise to a certain interpretation of the data, which in turn has generated a number of category-models. One example of these is Hampton's polythetic model (1979), which involved a specific unitary theory based upon the work of Rosch and Mervis, (1975). His studies showed that the usefulness of a distinction between defining and characteristic features was doubtful. Consequently, he developed the notion of a polymorphous/polythetic concept which he described as one in which "an instance is categorized as belonging to a certain class, if and only if, it possesses at least a certain number of a set of features, none of which need be necessary or sufficient" (1979, pp. 450-451). His theory also allowed for differential weighting of the features for their cue validity (that is, their relevance and importance), resulting in degrees of category membership for the category extension.

The unitary approach to category representation is an influential one, and any study about category-types needs to take it into consideration. Hampton (1979; 1987; 1988) is just one of many researchers who have continued to base the interpretation of their research results upon the assumptions inherent in a unitary approach - that prototype effects reflect something *direct* about the nature of human categorization and mental representation of categories (see Hampton & Dubois, 1993; Medin & Smith, 1984). In spite of this, the more recent research articles in the literature show a growing consensus that, though such interpretations might not be mistaken as far as they go, the unitary approach provides an *in*sufficient account of categories and concepts. If so, the structures derived from goodness-of-example rating tasks (that is, typicality measures) *under*determine mental representations, limiting the kind of knowledge concepts might contain. This study examines in turn the unitary approach's four main assumptions concerning categorization processes, structure and representation, which are described in more detail below.

4.03 CATEGORIZATION OF MEMBERS

The unitary approach makes two main assumptions about the categorization process. Firstly, the only important information used for categorizing a potential member is said to be either normative knowledge (in the case of prototype theories), or objective knowledge (in the case of classical models). Secondly, processing of potential members would involve a comparison of an object's features with those of its category, in a one-stage process of categorization.

> Assumption One: People represent their categories in a normative organization of information, and so the use of normative stimulus items in categorization tasks will elicit the best subject performance, as measured by speed of response and level of accuracy.

Concerning the kinds of information used to decide membership, Hampton (1979) takes the view that the only useful information about categories concerns normative knowledge. His study compared his polythetic concept theory with that of Smith, Shoben, and Rips' (1974) two-stage model (described below). He argued that their dual process model was superior only so long as it could motivate a functional distinction between characteristic and defining features. However, their model was claimed to fail in this, because specifying the precise defining features of a concept is well-nigh impossible, and the distinction of functions becomes blurred. Hampton states that his unitary theory of category process, structure and representation is superior upon the ground of parsimony. His polymorphous / polythetic concept consists of both kinds of features, the crucial factor during categorization being that an object possess a certain number of a set of features before it could be judged to be a member (Hampton, 1979, 1988; 1993).

Studies of category tasks which adopt the unitary view often use norms as word-stimuli for their experiments, yet Freyd (1983) noted that normative knowledge becomes less detailed than subjective knowledge in the process of being shared. At least two researchers, then, claim that normative knowledge is less detailed and more prone to distortion and inaccuracy, than is knowledge organized according to concrete individual experience (Bartlett, 1932; Freyd, 1983). It is suggested that, in order for category behaviour to be most efficiently realized, an autobiographical or emotively-based organization of category members should be used, rather than an organization based upon normative or objective facts alone.

One example of the power of subjectively organized knowledge is the self-reference effect upon recall and recognition performance. If a word is encoded with reference to oneself, both accuracy and speed of recall are higher than if encoded in a neutral context (Rogers, Kulper, & Kirker, 1977). Furthermore, empirical evidence supports the notion that there is an advantage for memory of material that is self-generated, and this advantage has been found in many memory tasks, such as cued recall (Begg & Snider, 1987; Greene, 1989; Slamecka & Graf, 1978). The general consensus amongst researchers seems to be that an active generation of materials gains its advantageous effects by inducing participants to devote greater rehearsal (Slamecka & Katsaiti, 1987) or mental processing (Begg & Snider, 1987; McFarland, Frey & Rhodes, 1980) to generated material than to passively copied or read material.

It is argued here that, if subjective knowledge is shown to have such a strong effect in these cognitive tasks, then it is feasible that it will also be of relevance in categorization processes, which are said to be similar to those of cued recall and recognition (Miller & Johnson-Laird, 1976). This experiment gathered idiosyncratic stimuli via an order of generation task, with each participant using his/her own individually generated exemplars in a membership decision task. One aim of Experiment 2 was to compare the differing effects (if any) of the use of normative and idiosyncratic stimuli upon categorization performance in three category-types.

Prediction One: It is predicted that categorization performance in the three category-types will show better accuracy and speed of recall when an individual participant's idiosyncratic stimulus items are used in a membership decision task, than when normative stimulus items are used.

Assumption Two: Membership decisions are carried out in a one-stage process of direct retrieval of instance-category relations, which are fixed in memory, and reflect the structure of the outside environment.

Concerning the view that categorization is a one-stage process, the unitary view assumes that the way we perceive category relations of objects in the world (e.g., instance-category memberships) is constrained by the similarity structures which already exist in the environment (Rosch, 1978). The unitary approach assumes that people encode such instance-category relations directly in memory, and any active analysis or constructive interpretation of the input is not possible during categorization (Lakoff, 1987). Such views include feature models which consist of a single computational process of similar feature comparison between instance and prototype (Hampton, 1979; 1984; McCloskey & Glucksberg, 1979; Rosch & Mervis, 1975) and discrete algorithmic models (Anderson, 1991a).

An alternative to accounts of single-stage categorization are the dual processing models, which assume that category relations are not represented as a direct reflection of category structures in the environment. The two-stage process of feature comparison is an example of an inferential, categorization process where the relationship between instance and category is neither stored directly in memory, nor is it always computed simply in one stage. Here, categorization occurs indirectly, with different kinds of information being used in separate stages, in order to reach a membership decision (Smith, 1978).

In dual processing models (for example, Smith, Shoben & Rips, 1974), the meaning of a word is said to be represented not as a whole unit, but as a set of semantic features. Each concept has both defining/essential and characteristic/typical features associated with it, each with separate functions. Defining features provide the necessary and sufficient criteria for deciding whether any word is a category member, and characteristic features both determine the typicality or centrality of category members, and the semantic relatedness of nonmembers. The two kinds of features are distinguished by setting a cut-off point on a continuum of *definingness* along which the features are assumed to be ordered, with the characteristic features being lower on the continuum than the defining ones. The features are computed in a two-stage process of categorization, with each stage being processed independently of the other, since different kinds of information are being used (Smith, 1978).

In the first stage, features are compared in a rapid parallel process of comparison, with the more features there are, the faster the membership is confirmed or contradicted. Initially, categorization consists of comparing the features of the potential member item with those of a proposed category. In this first stage, both defining and characteristic features are considered in an estimate of how many similar features are shared by the item and category, involving an *overall* holistic similarity computation. If this estimate exceeds an upper-level criterion where the item-category pair are seen to be highly related, then a fast "yes" response is given. Similarly, a fast "no" response is given if the elements of the item-pair are unrelated, falling beneath a lower level criterion.

However, when the computation of overall similarity or relatedness falls between these two criteria, the evidence is not conclusive, and initiation of the second, more *analytic*, stage becomes necessary, where only the defining features are considered. If each defining feature in the item matches a defining feature in the category, *and* a sufficient number of defining features are present to reach criterion level, a positive decision is made, or otherwise a negative response is given. The first stage of parallel processing may be all that is needed, if typicality is assessed as very high or very low. If this fast estimate is inconclusive, only then is the exemplar analyzed for the presence of sufficient defining features which will settle the uncertainty. In brief, then, serial processing of features takes place sequentially, with each feature comparison in the item-category relationship being completed before the next feature comparison can begin. In parallel processing, all the relevant features of the word item are perceived simultaneously, being processed in the one holistic comparison with their category.

In order to test Assumption Two, it will be necessary to compare serial/additive and one-stage/parallel models of categorization. Townsend (1990) claims that this distinction is important because by so doing, conclusions can be drawn about the complexity of the input material being processed, and the mental architecture which has the limited/unlimited capacity to do the processing. Townsend describes a number of ways to do this which involve the use of various experimental strategies based on mathematical demonstrations, including the method based on reaction times used in this experiment. He describes this method as looking for "factorial interactions with selective influence of cognitive subprocesses" (1990, p. 48). The strategy is based on the assumption that experimental factors can be found that affect separate cognitive subprocesses, that is, stages. Reaction time is measured under all combinations of the various factor levels and then the data-analysis is examined for the presence of interactions or the lack thereof. The presence of an interaction in an analysis of variance has been mathematically demonstrated to be an indication that parallel processing is occurring, that is, that features are being computed in a multiplicative fashion rather than an additive one (Ashby & Townsend, 1980; Donders, 1969; Sternberg, 1969a; 1969b; 1975).

In the method of factorial interactions, an experimental factor need merely affect the processing time of a stage. A lack of interaction is termed *additivity* because the two factors are affecting reaction time in a separately additive fashion. In other words, the effect of Factor X is the same, whatever the level of the second Factor Y. Also, both factors should lead to significant main effects, and a lack of interaction between them (Townsend, 1990).

The two factors used in this experiment are those of category-types and level of generation order (or production frequency in the case of normative word-stimuli). This method was used by Hampton (1984) in his comparison of superordinate/property category-types, where he argued for an ongoing parallel processing of instance-category relationships. He assumed that differences in response times for category-types might occur because each category-type reflects a different level of feature-abstraction from the physical environment. If his assumption is applied to the independent variables used here, then natural superordinates should involve purely concrete physical features since they reflect the physical appearance of objects. Property types require a slightly more abstract level of categorization, since they involve the presence of only a single physical feature from the physical environment. Ad hoc types are at the highest level of abstraction, since concrete physical features are of no relevance in their category-formation. The second factor is levels of representativeness in the various category items produced by either order of generation or production frequency. High, middle and low levels of generation order reflect the degrees of representativeness found for the items at each level.

Sternberg (1969a; 1969b; 1975) argued that, when two independent variables affect the same stage of an information-processing sequence, then their effects on overall time will interact. Hence, an interaction would indicate that categorization is carried out in a single process of access and retrieval of category relations. If an interaction is found between the two factors, that would be an indication that membership decision in all three category-types involves the same single process of access and retrieval of category relations, regardless of the feature-complexity of the various items at the different levels of representativeness. Alternatively, the lack of an interaction but the presence of two significant main effects would indicate additivity between stages of processing (McClelland, 1979; Schweickert, 1978; 1983).

Prediction Two: It is predicted that, because participants are using more information than that involved in a simple one-stage computation of similarity, their membership decisions will show evidence of a two-stage process.

4.04 INTERNAL MEMBERSHIP STRUCTURE

> Assumption Three: The unitary approach assumes that a category's membership structure is based upon its typicality structure. Consequently, it might be expected that natural superordinates will have ill-defined category boundaries, whilst rule-based categories such as property types will have well-defined category boundaries.

In early studies, Rosch claimed that the goodness of example (or typicality) ratings by subjects were a direct measure of that category's internal membership structure (Rosch, 1973, 1975a, 1975b). Once this was accepted, it was concluded that the varying degrees of typicality derived from goodness of example tasks could occur *only if* category membership criteria themselves were fuzzy or uncertain (rather than absolutely either-or).

More recently, Hampton (1979; 1988) makes the claim which supporters of the unitary approach also make throughout the literature. An item's status as a member of its category (in Hampton's case, specified in terms of his derived scale) will also predict a number of category phenomena. These include the rated typicality of category members, the membership in the category of any item, the response times taken to make a membership decision, the existence of borderline cases of membership and the relatedness of nonmembers. In other words, it is described as a unitary approach because all these task performances are based upon the one underlying principle - that it is the typicality of the instances which determines their membership status.

According to Neisser's (1967; 1987) dichotomy of definedness, the well- or ill-defined conditions for membership govern the nature of the category's boundaries with other categories: they can be either fuzzy *or* precise. A fuzzy boundary is accompanied by a gradient structure in its category-items. This has been well-attested in the literature. From this, the inference was drawn that the opposite would also be true: a precise and clearcut category boundary would not generate a gradience in its exemplars, because they are all of equal status as members or non-members (Mervis & Rosch, 1981; Rosch & Mervis, 1975).

However, research studies have shown that these well-defined categories *also* demonstrated typicality gradience effects in their members. It was shown in a number of studies that the presence of gradience in a category's structure does *not* exclude the possibility that such categories have clear boundaries and well-defined memberships. Graded typicality structure has been found in a number of categories besides fuzzy natural ones, including ad hoc types (Barsalou 1983, 1985); abstract types (Hampton, 1981); property types (Barsalou 1985; Barsalou & Ross 1986); and classical types (Armstrong, Gleitman & Gleitman 1983). Most of these types (except the natural categories) have specifically defined membership criteria which should have ensured that they had clear well-defined boundaries separating category members from nonmembers. For a well-defined category, any given object or event may be unambiguously and nonarbitrarily classified as a member or nonmember of the category (McCloskey & Glucksberg, 1978).

People might believe that the boundaries for most everyday categories are clearly demarcated, but these studies showed that their participants did not put these beliefs into practice. Armstrong, Gleitman and Gleitman (1983) found that typicality judgments could consistently be given to concepts that are not fuzzy at all, for example *Even numbers, Odd numbers* or *Male, Female.* Ill-defined internal structure was found to exist even in these category-types, where a formal rule *could* be clearly specified and defined (for example, "Odd number = an integer which cannot be divided by 2"). When judging members' typicality in such categories, participants still stated that instances had degrees of typicality. This demonstration of typicality variation amongst members of equal status threw doubt on the supposed common mechanism underlying typicality gradience and membership decisions by dissociating the two phenomena.

Osherson and Smith (1981) provided evidence of typicality gradients in well-defined categories, focusing on the area of conjunctive combinations. For example, they found that *guppy* was a highly typical example of *Pet Fish,* but a very atypical example of *Fish* and of *Pet.* This runs contrary to the predictions of the unitary view's fuzzy logic (Zadeh, 1965; 1978). Osherson and Smith (1981) suggest that prototype approaches might be reflecting identification procedures in which objects fit the similarity heuristics to a greater or lesser degree, yet are not concerned with the conceptual core, where membership must be all-or-none and is presumably represented by necessary and sufficient features. They concluded that fuzzy set notions are appropriate only to judgments of typicality gradients whereas functions concerned with acceptance/rejection of membership (for example, class inclusion, conjunction, disjunction and negation) follow classical set logic (i.e., well-defined rules). Further evidence from research on conjunctive concepts contradicts a unitary view of structure and representation (Medin & Shoben, 1988; Murphy, 1988; but cf. Hampton, 1988). The general consensus is one very like Putnam's notion of a "division of linguistic labour" (Putnam, 1975a; 1975b).

If both well-defined (for example, property types) and ill-defined categories (for example, superordinate types) have internal graded structures, then typicality cannot be regarded as a reliable measure of membership status. It is an indicator of gradient structure, but that does not necessarily mean that the *membership* of the items must also be graded. In other words, the typicality structure and the membership structure might not be one-and-the-same structures, but belong to different representations.

Rosch herself (1983), in her dual representation model of reference point reasoning (see chapter 2), came to the conclusion that category membership can be well-defined, yet the members vary in their degrees of typicality at the same time. She had hinted at this view earlier, but did not elaborate upon it in later papers:

> The attributes which define an instance as a category member and the attributes which define it as a good or less good member may be different. (Rosch, 1973, p.141)

Assumption 3 is tested by asking whether it is possible to find a high level of agreement amongst study participants about where category boundaries should fall in ill-defined categories. Theoretically, according to the unitary approach, high levels of participant agreement might be expected where the membership rule is clearly defined and the criteria are precise. If category boundaries are fuzzy, however, then participants would be expected to agree less amongst themselves about where precisely the boundary might fall on the continuum of gradient structure Note that gradient structure, according to ill-defined views of categories, involves a continuum which includes members and non-members. If typicality is not a reliable indicator of a category's definedness, that is, whether its membership structure should be ill-defined or well-defined, then an alternative measure is subject agreement about category boundaries.

Participant agreement is a cleaner indicator of the nature of a category's boundary than are typicality ratings (Barsalou, 1983; McCloskey & Glucksberg, 1978; Rosch, 1978). Instead of typicality rating tasks, McCloskey and Glucksberg (1978) used *disagreement* amongst participants about membership or nonmembership as an indicator of ill-definedness in a category's membership criteria. They claimed that well-definedness should be accompanied by high levels of agreement amongst participants in the acceptance or rejection of items. Their stimuli consisted of instances from common categories like *Fish, Ship,* and even *Natural Earth Formation* and *Weather Phenomenon.*

They reasoned that the presence of typicality gradience need not be an automatic indication of an ill-defined category, and that a well-defined category could still have a typicality gradient in its members. From their results, they concluded that natural categories are ill-defined, with no clear boundaries separating category members from nonmembers. They briefly considered the possibility of multiple types of representations for information, but rejected it for reasons of parsimony. This rejected solution was a forerunner to proposals later made by Rosch (1983) and Armstrong, Gleitman and Gleitman (1983) for dual representational models.

In summary, the main consequence of the unitary view is that physical similarity is taken to be the underlying basis for both the internal structure of a category, and for its external relations with other categories. Prototype accounts (Rosch, 1973;1975a; 1975b) have a view of graded structure which describes it as including items both internal *and* external to the category - that is, information about membership status *and* typicality of the items in the one representation. Whilst internal typicality gradients are obviously ubiquitous to all category-types, using them to explain a category's concept of what constitutes membership results in an impoverished account of how we understand and represent categories.

<u>Prediction Three:</u> Firstly, it is predicted that typicality will predict the order of members (as in the criterion Ranks variable) in both superordinates and property types. Secondly, it is predicted that participant disagreement (i.e., ill-definedness) about an item's positive or negative membership will be found in property category-types, as well as those category-types whose members vary in typicality (e.g., natural superordinates). Overall, the results will show a combination of typicality structures and ill-definedness of membership criteria in property types, as well as in natural superordinates.

4.05 MENTAL REPRESENTATION OF CATEGORIES

It is not the concern of the present experiment to contrast network and feature-based models of semantic meaning representation (Collins & Loftus, 1975; Glass & Holyoak, 1975). A growing number of researchers have suggested that both network and feature-based models make the same predictions (Anderson, 1978; Hollan, 1975; Johnson-Laird, Hermann, & Chaffin, 1984; Shanon, 1988a; Smith, 1978; but cf. Lucas, 1991). Consequently, a network model or feature list model is not really an important theoretical difference, since the two have been shown to be structurally the same (see Hollan, 1975), with both generating the same empirical predictions (Eysenck & Keane, 1990).

> Assumption Four: The unitary approach assumes that the mental representation of a category is as a prototype, and hence consists of typicality information only.

With respect to the representation of a category, Hampton's (1988; 1993) polythetic concept is essentially equivalent to Rosch and Mervis' (1975) prototype of independent features. He claims the difference is one of focus: their prototype has an extensional emphasis in that it describes a structural similarity relationship among object members, making no assumptions about how the category of objects as a whole might be represented mentally. On the other hand, Hampton's (1988) polythetic concept has an intensional emphasis, because it describes a relationship between objects and a conceptual representation of the category (Komatsu, 1992). The unitary approach assumes that not only are the category members (or extension) said to be internally structured according to their degree of typicality, but their representative prototype (or intension) is said to consist of this structure (Mervis & Rosch, 1981; Rosch, 1975b). Thus, the unitary approach makes no distinction between a category's intension and its extension, with categories being represented as structural concepts.

Because of the unitary view's assumption that membership structure is the same as typicality structure, then it is also assumed that the category as a whole is represented by a prototype. This interpretation of the data in Rosch's early studies sees the prototype as a structural representation which might involve a specific object or animal which is the best exemplar of its category; or it could be a schema or feature bundle, consisting of similar features abstracted across a number of exemplars. This assumption can be tested *indirectly* by looking at what kind of information underlies the organization of category exemplars in memory, whether it be direct experience of an object (frequency-of-instantiation task); physical similarity (goodness-of-example task); or membership criteria (categorization task).

4.05.1 Typicality as a basis for internal graded structure

Goodness-of-example judgment tasks have been used by a number of researchers to measure typicality of items: the more similar an item is to the other members of its category, and the less similar it is to members of contrast categories, the higher will be its rating as a good example of its category (Rosch & Mervis, 1975). Typicality effects are said to reflect an underlying pattern of family resemblance which is said to organize the category in memory.

Typicality might be relevant information for *natural* category-types, but not for ad hoc category-types where members seem to share no overlapping of physically similar, correlated features. Consequently, goodness-of-example typicality ratings might be unsuitable for measuring representativeness in ad hoc category types, where one instances's similarity to another is irrelevant. For their part, property category-types are defined by, and necessarily share, one similar feature. Typicality effects have been found in well-defined categories like property types (Barr & Caplan, 1987; Barsalou & Ross, 1986), so it may be that typicality information, as measured here by the goodness-of-example task, will be found to be relevant in these category-types also.

However, representativeness measures such as goodness-of-example have been shown by Vandierendonck (1988,1990) to be insufficient. He used artificial categories constructed of dots and squares, finding that frequency-of-instantiation rating tasks were poor predictors of internal gradience, whilst goodness-of-example rating tasks were only moderate predictors. He concluded that category objects and the category's membership criteria were represented in separate, dual representations. In contrast, Barsalou (1985) has found measures of direct experience (frequency-of-instantiation) to be strong determinants of category representation.

4.05.2 Experience as a basis for internal graded structure

If goodness-of-example is a measure of objective typicality in that it is biased towards the use of semantic/theoretical knowledge (Malt & Smith, 1982), then a subjective measure of direct experience is necessary, because categorization is concerned with "the coding of experience" (Smith, 1990, pp 34-

35). Frequency-of-instantiation judgments would tap into information of personal experience in the context of everyday life, as the subject judges how often, in his/her personal experience, an object has been experienced in a particular category-context.

Frequency-of-instantiation judgments need not necessarily measure typicality. Supporters of typicality as the quintessential information which governs category behaviour deny that frequency has any influence upon people's judgments of typicality gradience, as the latter is a structural concept rather than one based upon learning. For example, Rosch, Simpson and Miller (1976) compared typicality and repetition in a production frequency task, and found repetition of items to have no effect upon generation of items. However, the tests used to detect repetitive frequency were not very sensitive, and the experimental design was not capable of detecting *simultaneous* effects of feature comparison and frequency of repetition. If frequency had a weaker but still significant effect, it was not detected. Also, there are a number of different kinds of frequency (see Nosofsky, 1988, 1989).

Barsalou (1985) distinguished between the overall *objective* frequency with which a person has experienced an object (familiarity), and subjectively experiencing an object *in a particular category-context* (frequency-of-instantiation). Barsalou has described the task as being ".....someone's subjective estimate of how often they have experienced an entity as a member of a particular category. Where familiarity is a category-independent measure of frequency, frequency-of-instantiation is a category-specific measure of frequency" (1985, p. 631). How the two variables differ can be illustrated as follows: people generally appear more familiar with *chair* than with *log,* having experienced the former more often across all contexts. However, people have probably experienced *log* more often than *chair* as an instantiation of *Things which can be used for firewood.* Thus, people are more familiar with *chair,* having experienced it frequently across a number of contexts, but they have experienced *log* more often than *chair* in the context of *Firewood.* As a result, the semantic association *log-Firewood* would be stronger than *chair-Firewood,* even though *chair* might be the more familiar concept.

The empirical basis for the frequency-of-instantiation task is Wilkins' (1971) and Conrad's (1972) research on conjoint frequency. Here, gradience effects are explained as the various degrees of associative learning for each item as a member of its category. When two events are repeatedly presented together and the participant receives feedback as to whether the two events are relevant, the association between exemplar and its category is strengthened. If the two events are irrelevant to one another, then the association is weakened by lack of feedback. The theoretical difference between the two tasks lies in their explanation for category structure: according to the frequency-of-instantiation measure, category structure can be developed through learning, whilst the typicality measure claims structure is *perceived* because it is inherent in the environment. To conclude this section, it is to be noted that Rosch herself finally rejected the notion of prototypicality as the basis for structure and representation of categories:

> To speak of a prototype at all is simply a convenient grammatical fiction; what is really referred to are judgments of degree of prototypicality ...... Prototypes do not constitute a theory of representation for categories. (Rosch, 1978, p.40)

<u>Prediction Four:</u> It is claimed that the informational content of a concept involves more than typicality, and it is predicted that the best determinant of representation would be frequency-of-instantiation or a rule-like membership criteria in all three category-types.

4.06 AIMS OF EXPERIMENT 2

To summarise, the general aim of Experiment 2 is to test the validity of the unitary view of categories, and the four assumptions its supporters make when interpreting their results concerning representation and structure of categories, and categorization of members. Three more specific aims are set out below.

(1) The experiment will investigate the nature of the categorization process; whether it consists of a simple direct retrieval or one-stage computation of similarity relationships; or whether it requires two stages of processing. If the latter, then the implication is that the categories we form are not a mirror reflection of the environmental structure.

(2) The experiment will compare tasks to discover which one best reflects internal gradient structure, the unitary view being that typicality would best predict membership structure. If this is so, then the implication is that participants would not agree about the typicality-based category-types, thus providing evidence for their ill-defined boundaries. If, however, agreement amongst participants is found to be high in fuzzy (typicality-based) categories, this pattern is inconsistent with the unitary view's description of one

representation for both typicality and membership criteria. Such a pattern suggests that a category might be represented in more than one way.

(3) The experiment will compare tasks to discover which one best predicts mental representation of each category-type. The independent criterion for mental representation will be the order of items from the generation task (idiosyncratic); and the order of items from the production frequency task (normative). Since each predictor task is considered to draw upon different kinds of information, the results will indicate whether a category can be represented in more than one way.

4.07 RATIONALE FOR TASKS CHOSEN AS MEASURES OF REPRESENTATION, STRUCTURE AND PROCESS

4.07.1 Categorization process

Categorization of positive and negative items was measured by reaction times in a membership decision task (see Appendix E). Categorization reaction times (measured in milliseconds) are considered to be a direct reflection of people's application of the membership criteria, to decide whether an exemplar is a member or non-member of a category. As the process is taking place speedily, at a close to automatic level of information processing, response times provide a measure of internal membership structure.

In other words, the claim is that reaction times are a "cleaner" measure in terms of reflecting the order of members in a category than are some other measures, such as typicality ratings (Smith, 1978). Rating judgments and generation tasks generally require considered thought about an item, thereby leaving room for confounding factors such as frequency of experience over and above the basic dynamics of the 'system".

4.07.2 Internal membership structure

The membership response times were transformed into the Ranks variable, with the fastest item in a category being ranked 1, and the slowest in the category being ranked 6. The transformation of the membership decision times into Ranks was carried out to enable comparison of internal membership structure with the other ranking tasks described in Appendix E, such as goodness-of-example and frequency-of-instantiation. Internal membership structure is the result of whatever criteria are being used by subjects during decision-making, whether that be schematic information, definitional rules, or similarity heuristics.

4.07.3 Representation of the category in memory

The experiment had two measures of category representation: production frequency and order of generation. Both tasks reflect the probability of a participant in the study producing an exemplar when asked to generate members of a particular category. The two measures differ in their explanations for that probability. Production frequency data is based upon the frequency of mention of an item across a large subject population, from which are derived norms such as those collected by Battig and Montague (1969) or Loftus and Scheff (1971). The analysis of Experiment 1 suggests that such data reflect normative and cultural knowledge.

Order of generation performance reflects the availability of category-items in *an individual's memory*. In this task, the order in which items are generated in response to a category label is assumed to reflect their subjective organization in memory, with item availability being governed by any number of possibilities, including autobiographical knowledge.

The two measures also differ in their ways of organizing the data. The order of generation variable reflects each individual participant's data, whilst the production frequency variable reflects the same task performance, but this time averaged across a large participant population. The essential difference between these two variables is that the order of generation task captures *the individual's* experience with the members of a category; while the production frequency data, once averaged across participants, represents *the norm* for category-members across a large population.

Both production frequency and generation order have been used as valid measures of a category's mental representation (Rips, Shoben & Smith, 1973; Rosch, 1973). For this experiment, it was considered that the two tasks would achieve an *atheoretical* measure of category representation, as they simply reflect the order of items in memory, without prescribing how the categorical knowledge *should* be organized in memory. On the other hand, measures of typicality and of experiential frequency both assume that some sort of organization is basic to the category. As a result, these tasks were not considered suitable as atheoretical measures of mental representation.

A distinction should be made between the Ranks variable (internal membership structure), and the production frequency/order of generation variable (mental representation of the category). The Ranks variable reflects whether an item is a member, or not. An item is accepted or rejected as a member during a speeded categorization task, during a "quick and hasty" decision which might involve overall similarity heuristics, rather than the considered application of diagnostic information (Medin & Ortony, 1989). As such, Ranks reflects the organization of the members in a category extension.

On the other hand, production frequency and order of generation both reflect the category intension, that is, the diagnostic information which produced exemplars. Production (or generation of items) only took place after considered thought about a number of possbile items which might be merely related with the concept (for example, category *Vegetables* - exemplar *shopping list?*) but were not necessarily genuine instances. Separate tasks to measure the two structures are necessary because, whilst some category theories consider intensional and extensional structure to be one and the same (Hampton, 1979;1988;1993; Rosch, 1973; Rosch, 1975a; Rosch & Mervis, 1975), other theories consider them to carry different information and/or be the product of separate cognitive functions (Gleitman, Armstrong, & Gleitman, 1983; Putnam, 1975a, 1975b; Osherson & Smith, 1981).

4.08 METHOD

Use of both idiosyncratic and normative item-stimuli

Some of the tasks (membership decision, frequency of instantiation ranking, and goodness of example ranking tasks) used in the experiment were carried out *twice*, once using idiosyncratic stimulus items and once using normative stimulus items. It was expected that tasks using an individual participant's own exemplars might prove to be more sensitive measures in reflecting both category models of representation, and bases of internal gradience. The normative and idiosyncratic stimulus items were derived from the production frequency norms and order of generation task. The norms collected across 100 subjects in Experiment 1 (see Appendix D) were also used as the stimuli-items in a membership decision task, a goodness-of-example task, and a frequency-of-instantiation task. The exemplars collected from each individual in the order of generation task were later used as idiosyncratic stimulus items in another membership decision task, a goodness-of-example task, and a frequency-of-instantiation task.

Participants

Seventeen university students were employed as participants in the study, but the computer data of four was unusable since they did not take sufficient care during the membership decision task, hitting keys at random. The data of thirteen people was used in this study, seven females and six males, with ages ranging from eighteen to fifty years. Students were paid $25.00 each for four hours, returning to the experimental laboratory for one hour every week for a total of four weeks. Each student carried out the four tasks which constituted the study.

Materials

The materials used can be divided into two groups: normative items and idiosyncratic items. The idiosyncratic items for the experiment were generated by an order-of-generation task carried out by the participants in this experiment The normative items were collected from a production frequency task performed by those participants used in Experiment 1 (see Appendix D). Both tasks involved the generation of twelve exemplar items in response to a category-label.

The normative items were derived from the data produced by one hundred participants in Experiment 1. Their task was to generate eight exemplars for each of the fifteen category-labels. These labels consisted of five superordinate categories, five property categories, and five ad hoc categories, the actual categories being listed in Appendix D. The responses of these people were then collated into production frequency rankings, where the items are ordered (in each category) according to the number of times they were produced as an exemplar by the overall group.

The ranked order of the normative data was derived as follows. The item mentioned by the highest number of 100 participants was ranked first. The item next most frequently mentioned was ranked second and so on, until the top twelve items that were most frequently produced were collected. Of these twelve, six items were selected for use as stimulus materials in Experiment 2: item ranks 1 and 2; 6 and 7; and 11 and12. These particular rankings were chosen in order to provide a sufficient range of differences between the items, reflecting likelihood of production. The same stimulus materials were used in the subsequent three tasks for all participants.

The idiosyncratic material was taken from a generation task given to the people who were to participate in this experiment. Participants generated twelve items in response to the same fifteen category-labels that were used for generation of the normative items. Of these twelve items, six were chosen for an individual's idiosyncratic stimuli, in the same rankings used for the normative data: 1 and 2; 6 and 7; and 11 and 12.

Rating or ranking of items

There were a number of reasons for using ranking of items in the frequency of instantiation and the goodness of example tasks, instead of the more common rating tasks. The traditional means by which typicality is measured in such tasks as goodness of example involves rating of items. On the other hand, frequency of instantiation tasks have more commonly been assessed through ranking of items (Barr & Caplan, 1987; Barsalou 1985); and as a comparison of performance on the different tasks was planned, it was necessary to choose either rating or ranking of items.

The membership decision reaction times tipped the balance in favour of ranking of items, as these data in their natural form did not allow comparison with the other three tasks: order of generation/production frequency, frequency-of-instantiation, and goodness-of-example. Consequently, the millisecond data (response times) for each person in each category were transformed into ranks (fastest reaction time = 1, to slowest = 6).

Stimuli

Participants were to be exposed to both idiosyncratic and normative item-sets, carrying out some of the tasks twice: once with the normative data as stimuli and once with the items he/she had individually generated. These included the membership decision, frequency-of-instantiation ranking, and goodness-of-example ranking tasks. For the frequency-of-instantiation and goodness-of-example tasks, the six normative items in each of the fifteen categories were presented together under their category-label. The same was done for the ninety idiosyncratic items, thus totalling one hundred and eighty experimental items for each participant in each of these two ranking tasks. The stimuli for the two membership decision tasks consisted of the same ninety idiosyncratic and ninety normative stimulus items described above.

The same ninety distractor items were used for both membership decision tasks. For each six positive items representative of a category, there were six distractor items which did not belong to that category, and to which the correct response would be negative. The distractor items were not generated by participants, being constructed by the experimenter, with particular properties in mind. Forty-five items were disjointed (that is, unrelated) negative (for example, *Vegetables - car*) and forty-five items were related negative (for example, *Vegetables - cashews*). Disjointed negatives had no relationship at all with the category whilst the related negatives had some distant association (other than category membership), for example, *cashews* and *Vegetables* can both be eaten. A property category example would be *Red Things - fete* as a disjoint negative, and *Red Things - snow* with the latter having the relationship of salient (though different) colour features.

Procedure

Participants were tested individually or in pairs, except for the membership decision tasks which were executed on the computer individually. There were four types of tasks in all, and the instructions given are found in Appendix E. The tasks were carried out in separate sessions, so that people would not remember details about their answers, and consciously give the same one in all tasks.

Week 1: the initial task was generation of exemplars, which were also used as idiosyncratic items in the next three tasks.

Week 2: the membership decision task was carried out on computer, using both idiosyncratic and normative items.

Week 3: the two frequency-of-instantiation ranking tasks (based on idiosyncratic and normative items) were carried out.

Week 4: the two goodness-of-example ranking tasks, also using idiosyncratic and normative items, were completed.

The first task given was always the generation task as it provided the idiosyncratic data for their subsequent tasks. Production frequency had comparable status and had already been carried out by a population of one hundred participants. In the order of generation task, participants were given an eight-page booklet which contained an instruction page and fifteen category-labels. They read aloud the instructions on the first page, a copy of which is attached in Appendix E. Any questions were discussed, and then the task carried out. The task consisted of generating twelve exemplars and writing these down on the twelve lines under each category-label.

The second task, membership decision, was carried out a week to ten days after the generation task, to avoid possible priming effects on response times. Subjects were seated in front of a computer, and read aloud the task instructions, a copy of which is included in Appendix E. They were instructed to respond as quickly and as accurately as possible. Reaction times and accuracy were recorded on computer. Any questions were discussed.

Items in the membership decision task were presented in one of three blocks, according to category-type. Each block was preceded by seven practice items, both negative and positive exemplars of that category-type. The test items in each block consisted of thirty positive instances, and thirty negative instances. The thirty negative instances consisted of fifteen related, and fifteen unrelated, negatives.

There were two membership decision tasks, those whose positive items were based upon individually generated items, and those whose positive items were derived from a production frequency task administered to one hundred subjects. Both membership decision tasks were carried out at the same sitting, with a ten-minute break in between. Order of presentation of the two tasks was varied across the thirteen subjects.

The remaining two tasks (goodness-of-example and frequency-of-instantiation) took place on separate days and their priority was randomly varied across participants to control for order effects. These tasks took place after the membership decision tasks, again to avoid priming effects on response times.

For the goodness-of-example tasks, participants were given the goodness-of-example booklet, and read aloud the instructions, a copy of which is included in Appendix E. Any questions were discussed, and then the task carried out. The booklet consisted of: person's name and task instructions on the first page; three practice categories each with six items on the second page; then six pages containing fifteen category labels, each label with six items to be ranked according to their degree of typicality in that category. The ad hoc category labels also specified that category's goal. At the head of each page was a key to ranking, setting out degree of typicality on a scale of 1 to 6, with 1 signifying "best example" to 6 being "poorest example".

For the goodness-of-example task based on idiosyncratic stimuli, the experimenter took the items generated by that individual previously and used them to fill in the exemplars to be ranked for typicality. For the task based on normative stimuli, the items were the same for all subjects and were taken from a production frequency task.

The frequency-of-instantiation tasks used the same procedure as the goodness-of-example task, and was based upon the same items, except that the participant was requested to rank the items according to his/her frequency of experience of that item in that category-context. The booklet differed (from the goodness of example booklet) in that the ranking key at the beginning of each page consisted of a scale on which 1 signified "most frequently"; 3 was "frequently"; and 6 was "least frequently". A copy of the instructions on the booklet's front page is placed in Appendix E.

Participants were carrying out two tasks in each session, one with idiosyncratic item ranks; and the same task again, only this time with normative item ranks. The order of presentation of normative- or idiosyncratic-based tasks varied randomly across persons, so as to control for possible practice or fatigue effects.

4.09 RESULTS: CATEGORIZATION PROCESS

Prediction One: Idiosyncratic stimulus items will elicit faster and more accurate responses during decision-making than normative stimulus items.

The first analysis looked at whether participants were sensitive to the idiosyncratic/ normative nature of the word-stimuli used in the experiment, and as a consequence, whether their processing of the three category-types would differ with regard to response times and error rates. To assess this, a 3 x 2 ANOVA (repeated measures) was used to analyze the response times (msecs) for membership decisions in positive responses. The variables used in the ANOVA were category-type (superordinate, property, ad hoc) and stimulus-type (idiosyncratic, normative). An additional 3 x 2 ANOVA (repeated measures) was used on percentage errors. The descriptive data are set out in Table 3.

Table 3: Membership Decision: Participant means for response times (msecs) and subject means for percentage errors*

	Category - types			
Stimuli-types	Superords	Property	Ad Hocs	Means
Idiosyncratic	874.15 *(4.1)*	1035.67 *(9.2)*	1060.49 *(5.4)*	990.10 *(6.2)*
Normative	828.78 *(3.3)*	903.55 *(9.7)*	871.34 *(6.7)*	867.89 *(6.6)*
Mean	851.46 *(3.7)*	969.61 *(9.5)*	965.91 *(6.1)*	928.99 *(6.4)*

* *percentage errors in italicized parentheses.*

Response Time

A x B Interaction

There was no significant interaction between the category-type and stimulus-type effects (F (2,24) = 2.869, $p > .05$).

Type of Stimuli

Referring to Table 3, although idiosyncratic items elicited a longer mean response time to make membership decisions (990.10msecs) than did the normative items (867.89msecs), this main effect was not significant (F (1,12) = 3.170, $p > .05$).

Type of Category

Significant differences were found amongst the three category-types (F (2,24) = 6.253, $p < .01$). A series of post-hoc comparison tests were conducted on the means for the three category-types (critical HSD Tukey = 106.70). The superordinate categories were found to be significantly faster than the property categories by 118.15 msecs; and also significantly faster than the ad hoc categories by about 114.45 msecs. The ad hoc and property types could not be said to be different from one another.

Error Rates

A x B Interaction

No significant interaction was found between stimulus-type and category-type (F (2,24) = 0.082, $p > .05$).

Type of Stimulus

Referring to Table 3, the means show that membership decision using general stimuli did not elicit a significantly higher percentage of errors than did the same task using idiosyncratic stimuli, with $F(1,12) = 0.06$, $p > .05$. Thus, the hypothesis that use of idiosyncratic stimuli would elicit a more accurate performance than when normative stimuli were used, was not supported.

Type of Category

The range of error rates varied across the three category-types. Regardless of stimulus-type, superordinate categories had the lowest error rates (3.7), property categories had the highest (9.5), and ad hoc types (6.1) fell in between the two. This difference in error percentages for each category-type is significant where $F(2,24) = 10.41$, $p < 0.005$

A Tukeys HSD post-hoc test (critical difference = 3.18) showed that variance in error rates for category-types was found mostly between natural superordinates and property categories (mean difference = 5.775), with the superordinates eliciting significantly more accurate decisions. Properties also elicited significantly more errors than ad hoc stimuli (3.47), whilst superordinates and ad hoc did not significantly differ in error rates.

These results do not support Assumption One set out in Chapter Four, that normative items elicit the best subject performance in categorization tasks (as measured by speed of response and accuracy), than do idiosyncratic stimulus items. The direction of the means is consistent with this assumption, a weak effect can be found here, but the analysis lacks power. Neither do the results support Prediction One as set out in Chapter Four, where it was predicted that idiosyncratic item-stimuli would elicit the best subject performance during categorization.

Overall, the lack of differences between idiosyncratic and normative stimuli in both data-analyses (response times and error percentages) supports the possibility that subjective representation of category knowledge is as well-established as a normative representation (Rosch & Mervis, 1975; Hampton, 1984; 1993) or an objective representation (Collins & Loftus, 1975; Collins & Quillian, 1969). The main conclusion is that categorization might involve the subjective coding of experience, *as well as* the organization of normative knowledge, but converging evidence will be required from further analyses before a firm conclusion can be reached.

Also of interest is the superior ease-of-processing (as measured by speed and accuracy) which was found in the natural superordinate category-types, as compared to property and ad hoc types. The lack of a significant interaction in both the response time analysis and the error analysis indicate that, whilst participants' response times were significantly influenced by the type of category about which they had to make decisions, the pattern did not vary *as a consequence* of the two types of stimulus-items.

Prediction Two: Membership decisions will be carried out in two stages of processing, involving both diagnostic feature comparison and inference of category relations, rather than a one stage process which retrieves instance-category relations directly from memory.

To assess this prediction, two 3 x 3 ANOVAs (repeated measures) were used to analyze the response times (msecs) from a membership decision task which was based on positive items.

Membership decision task based on idiosyncratic stimuli

Table 4 shows descriptive data (based on idiosyncratic stimuli derived from the order of generation task), gained from the analysis results for 3 category-types (superordinate, property, ad hocs) by 3 levels of generation order (items at high, middle and low levels of generation).

Table 4 Membership Decision for Idiosyncratic items
Participant Means for response times (msecs)*

	Category - Types			
Genorder**	Superords	Property	Ad Hoc	Means
First Order	803 (316)	952 (460)	949 (433)	901 (409)
Middle Order	876 (302)	1033 (405)	1029 (407)	979 (377)
Last Order	943 (322)	1122 (464)	1203 (506)	1089 (446)
Mean	874 (314)	1036 (443)	1060 (457)	990

* Standard deviations in parenthesis
** Order of generation of exemplars (idiosyncratic stimuli)

A x B Interaction

No significant interaction was found between the category-type and order of generation, where $F(4,48) = 0.624$, $p > .05$.

Order of Generation

Referring to Table 4, the main effect for generation order is significant, where $F(2,24) = 18.513$, $p < .001$.

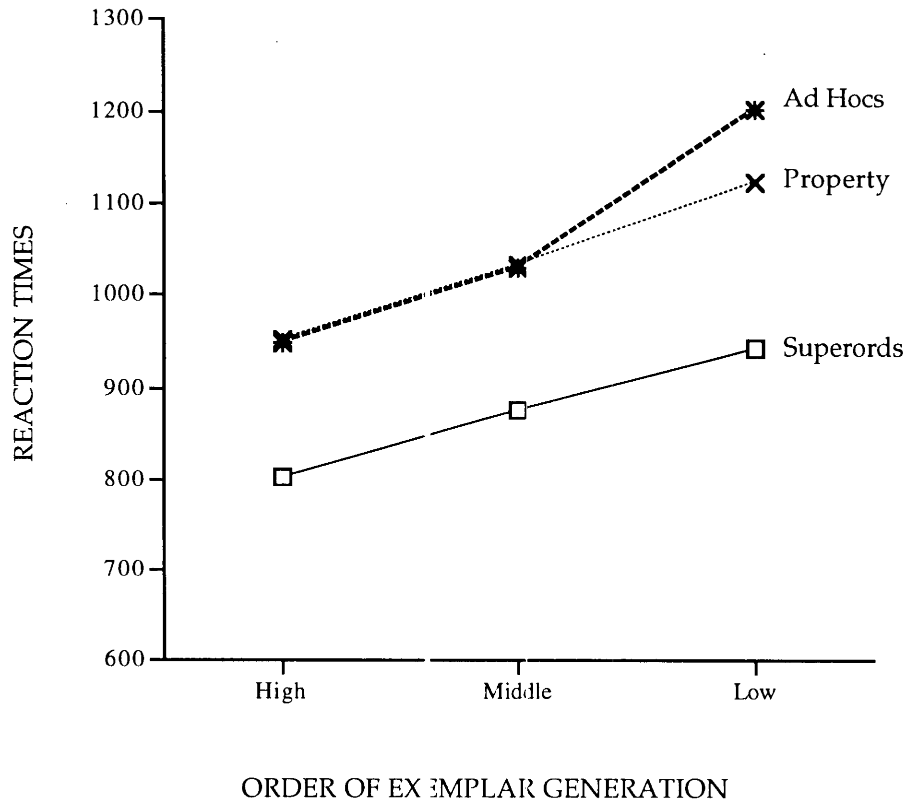
Post-hoc comparisons were carried out on the means from the high, middle and low generation levels, using Tukey's HSD critical difference of 77.52 msecs. From this, significant differences were found between high and middle generation levels (78 msecs); between middle and low generation levels (110 msecs), and between high and low generation levels (188 msecs).

Type of Category

A significant main effect was found for category-type, at $F(2,24) = 9.509$, $p < 0.001$. Tukey's post-hoc comparison of means was carried out, using the critical difference of 115.83 msecs. The post-hoc comparison tests showed, and it can be seen in Figure 2, that the superordinate categories were significantly different from both the property types (162 msecs) and the ad hoc types (186 msecs). The means of these latter two were not significantly different (24 msecs). Comparing the category-types, the faster means of the superordinates might be explained as being due to these categories being experienced and used more often than either of the property and ad hoc category-types. The differences between the superordinate and the property / ad hoc types are set out in Figure 2.

In summary, the pattern of the data-analysis (presence of significant main effects for the two independent variables, but lack of a significant interaction between them) supports the interpretation that categorization was carried out in a two-stage, serial process (Townsend, 1990). The data derive from idiosyncratic stimuli which reflect a subjective or autobiographical organization of category items, so any additive model used to explain the results would need to incorporate this kind of information.

Figure 2.



Membership Decision for Idiosyncratic Items
Subject Means for response times (msecs)

Membership decision task, based on normative stimuli

Table 5 shows descriptive data (based on normative stimuli derived from the production frequency task), gained from the analysis on the 3 category-types by 3 levels of production frequency (items at high, middle and low frequency of exemplar-production).

Table 5: Membership Decision for Normative items
Participant Means for response times (msecs)*

	Category - Types			
Product Freq**	Superords	Property	Ad Hoc	Means
High Frequency	740 (467)	783 (367)	831 (389)	785 (407)
Middle Frequency	842 (414)	940 (413)	855 (431)	879 (416)
Low Frequency	905 (573)	988 (493)	927 (390)	940 (486)
Mean	829 (488	904 (431)	871 (401)	868

* Standard deviations in parentheses
** Frequency of exemplar production (normative stimuli)

A x B Interaction

No significant interaction was found between the category-type and levels-of-frequency effects F (4,43) = 1.138, $p > .05$.

Level of Production Frequency

Referring to Table 5, the mean for frequency level of production is 868 msecs, the range being from high frequency normative items at 785 msecs, through middle frequency at 879 msecs, to low frequency items at 940 msecs. The high level had the lowest standard deviation. Collapsing across the category factor, level of production frequency is significant, where F (2,24) = 10.321, $p < .001$.

Tukeys HSD (critical difference = 86.037) post-hoc comparison of means was carried out, and the high level means of production frequency were found to be significantly different from the middle level means (94 msecs); and from the low level ones (155 msecs). There was no significant difference between the middle and low levels of production frequency.

Type of Category

The means for category-type range from the means for superordinates (829 msecs) and properties (904 msecs) to ad hoc categories (871 msecs). A significant main effect was not found, with F (2,24) = 1.099, $p > .05$.

The problem with these results is that only one main effect was found, that of stimulus-type. To infer additivity, as is described in the Smith, Shoben and Rips (1974) model, both factors should have shown main effects (Townsend, 1990). However, the results do allow the logical classical model to be eliminated. The presence of significant gradience in response times as indicated by significant differences between the levels rules out the possibility of the discrete processing proposed by the logical classical model (Collins & Quillian, 1969). Since classical models claim that all exemplars of a category have equal membership status, support for this model would require that no significant differences be found between the levels of production frequency.

In summary, comparing the two analyses based on the idiosyncratic and normative data, it can be seen that data derived from the categorization task based on normative stimulus items yielded far fewer significant results than the one based on idiosyncratic items. The normative stimuli materials elicited only two significant differences for the levels factor and none at all for the category-type factor. In contrast, the idiosyncratic data-analysis could pinpoint differences between the category-types, as well as all three levels of generation order. It is tentatively concluded that the use of normative stimuli in tasks lead to a less finely-tuned performance, than that gained when stimuli idiosyncratic to the participant are used.

Concerning the single or dual stage processing of members in a categorization task, the results on the data elicited from the normative stimuli were inconclusive as to a processing model. The data analysis on response times from the idiosyncratic stimuli indicates that serial, additive processing was being employed by subjects; and that a dual stage processing, featural model, would fit the results.

4.10 RESULTS: INTERNAL MEMBERSHIP STRUCTURE

Prediction Three: Unitary assumption three was that the order of exemplars in a category (i.e., their membership structure) is based upon their degrees of typicality and similarity. Consequently, the boundaries of categories with such graded structure is ill-defined. Also, the unitary assumption claims that categories whose membership criteria is not based upon fuzzy similarity (such as property types) cannot be expected to have ill-defined boundaries (Neisser, 1967; 1987). In contrast, it is predicted here that *all* category-types (whether superordinates or properties) will show typicality gradience in the membership order of their instances, and consequently, will also show ill-defined boundaries. In short, it is being predicted that participants will disagree about category boundaries in all category-types (Armstrong, Gleitman, & Gleitman, 1983).

The Ranks variable was derived from the membership decision task, which consisted of both positive and negative items. It consists of the membership response times ranked in order of magnitude. Each category had six positive exemplars tested during the membership task. Rank 1 denotes the fastest time (of the six items) and Rank 6 signifies the slowest time. The Ranks variable could thus be said to mirror the internal membership structure of items in a category. It was considered that this method of reflecting membership structure would be the least biased in terms of theoretical assumptions.

Table 6: Six linear regression analyses, each involving 6 ranks in goodness-of-example (predictor) regressing upon Ranks (dependent variable) as the criterion.(Number of items = 78).

Analysis on:	Std.Error	BetaCoeff	T-Statistic	Prob.*	Adj R-Sq	R
Idiosyncratic:						
Superordin**	1.617	0.356	3.326	0.001	0.116	.356
Property	1.656	0.264	2.367	0.020	0.057	.264
Ad Hocs	1.689	0.218	1.943	0.056	0.035	.218
Normative:						
Superordin**	1.621	0.349	3.250	0.002	0.110	.349
Property	1.625	0.324	2.964	0.004	0.093	.324
Ad Hocs	1.689	0.217	1.936	0.057	0.034	.217

* Probability, two-tailed
** Natural superordinate category-types

The first part of the assumption under test here is concerned with whether membership structure (as reflected in the Ranks variable) can be predicted by typicality effects (as measured by the goodness-of-example task). Six simple linear regression analyses analysed performances separately for idiosyncratic and normative items, on each of the three category-types. In each of the analyses, the Ranks variable was used as the criterion, and goodness-of-example as the sole predictor variable. The results are summarised in Table 6.

As indicated by the standardized error of the estimate (for natural superordinates) in the first column of Table 6, the value for idiosyncratic items (1.617) did not differ substantially from the value for normative items (1.621), suggesting the two kinds of word stimuli elicited roughly the same prediction error. The beta co-efficients in the second column indicate that the slope in the 6 ranks of the superordinates was closest to the membership structure (Ranks variable) of that category, since those were the categories which resulted in the highest beta coefficients (idiosyncratic = 0.356; and normative = 0.349). The property category-types were not far behind, with smaller beta coefficients (idiosyncratic = 0.264; and normative = 0.324). In both these category-types, goodness-of-example was found to be a significant predictor of membership structure, as indicated by their probability values, which were less than .05. In other words, the typicality gradience of the goodness-of-example task (as measured by the six ranking values for the items) was the best predictor of the membership structure (as reflected in the six ranking values for the items in the Ranks variable) of these two category-types.

The same cannot be said for the ad hoc category-types which, whether based upon idiosyncratic or normative items, were found to have small T-statistics with a probability greater than .05. This lack of prediction by goodness-of-example for the membership structure of ad hocs is not surprising, considering that degree-of-typicality (theoretically) is derived from physical similarity of features. The examples of ad hoc categories are not necessarily grouped together on the basis of similarity of appearance, as is demonstrated by the exemplars *wallet, dog, documents, children* for the category *Things to save from a burning home.*

As can be seen from the adjusted R-squared, typicality showed the lowest accountability for variance in the ad hoc types (3.5% and 3.4%). Typicality was highest in accountability for variation in the natural superordinates (11.6% and 11.0%).

Finally, if this analysis alone were to be considered, the assumption made by the unitary theory about membership structure would be supported, at least for the superordinate and property category-types. Both these category-types show that goodness-of-example (that is, typicality) can predict internal membership structure. However, it is to be remembered that Armstrong, Gleitman, and Gleitman (1983) showed that it is possible to have fuzzy typicality structure in a category which has well-defined boundaries. In other words, it is possible to have both a typicality structure *and* a well-defined membership structure, thus implying dual representations for a category. The next step is to determine which (if any) of these six category-types have well-defined boundaries, in spite of their members' structure being linked to typicality gradience in the previous analysis.

The second part of assumption three being tested here concerns the definedness of category boundaries, whether it is precise or fuzzy. If a category is well-defined, then subjects should agree amongst themselves about items' membership (or non-membership), even when the items concerned are unusual or atypical. To assess this, the data on subjects' error rates in the membership decision task had to be re-assessed, and was used to calculate subjects' rates of correct proportion of decisions. An error in categorization was taken to signify subject disagreement, whilst an accurate decision (whether it involved a negative or positive item) was taken to indicate subject agreement. Thus, judges were being ranked on their "degree of agreement" with an accurate decision, regardless of whether the item was positive or negative.

Each participant's response was graded as to its accuracy, regardless of whether items were positive or negative. Each category used in the task had 3 positive members and 2 negative members to be decided upon. Looked at in another way, each category had five levels or degrees-of-membership on a continuum. They included high, middle and low degrees-of-membership (the three positive items); and related or unrelated non-membership (the two negative items). In each category-type, each participant had ten items at each positive level (30 items); and fifteen items at each negative level (30 items). Each of the five levels in a category was scored for proportions of accuracy across these ten-item sets (for positives) and fifteen-item sets (for negatives). Complete accuracy for a ten-item set was defined as having a value of 1.00, so that each correct item in that positive set was worth .1.

In the negative item-sets, each correct item in that set was worth .067 (one-fifteenth of 1.00 = complete accuracy). The difference in values between a negative item and a positive item was due to their difference in numbers of items at each level. Thus, if a person answered only 8 out of 10 items correctly at the positive second level (middle degree of membership), his or her score would be

.8. Similarly, if only 11 out of 15 items were answered correctly at the negative fifth level (disjoint nonmembership), the score would be .74.

The dependent variable consisted of rankings allocated to each level's accuracy score. Rank 5 was given to the level with the lowest accuracy, Rank 4 to the level with the next lowest accuracy ratio, and so on until Rank 1 for the highest accuracy ratio was reached. If two levels had the same accuracy score, their Ranks were added and divided by two. Similarly, if three levels had the same accuracy score, their Ranks were added and divided by three.

Kendall's coefficient of concordance (W) was used to calculate agreement among participants. The rationale underlying Kendall's measure is that, if all judges agree upon the rankings to be allocated to each level, the column totals for each level would show considerable variability. On the other hand, if judges show maximal disagreement, each column should have some high ranks and some low ranks assigned to it by different judges, so that the column totals would be roughly equal. Thus, the variability of the column totals, given disagreement (or random behaviour) among judges, should be low. Kendall's coefficient was estimated for each category-type: superordinate, property and ad hoc. Refer to Table 7 for participants' agreement on levels within category-types based upon idiosyncratic items, and to Table 8 for participants' agreement about rankings for different levels within category-types based upon normative items.

Table 7: Rank Sums for five levels in Idiosyncratic Categories:
(in parentheses: analysis run on levels 3 and 4 alone)

Levels	Superord		Property		Ad Hoc	
1. High Genord*	30.00		36.00		35.00	
2. Middle Genord	40.00		42.50		43.00	
3. Low Genord	35.50	(17.50)	38.00	(16.50)	39.00	(19.00)
4. Related Negative	50.50	(21.50)	50.50	(22.50)	46.00	(20.00)
5. Disjoint Negative	39.00		28.00		32.00	
Kendall's Coeffic.	0.134	(0.095)	0.163	(0.213)	0.077	(0.006)
Chi-square probab	0.138	**(0.267)**	0.076	**(0.096)**	0.406	**(0.782)**

* Order of generation. The sequence of items on which the membership decision task was based.
In bold: Chi-square probability (df = 1) for levels 3 and 4, alpha level of significance < .05.

Kendall's coefficient reflects the variance of column totals, in the form of a percentage of the total maximum variance. Thus, the coefficients for each category-type in Table 7 indicate that the column totals for each of the five levels reflected a low percentage of the maximum possible variance (13.4% for natural superordinates; 16.3% for property types; and 7.7% for ad hoc types). This signifies that people disagreed strongly in all three category-types about which items should be allocated to which level. The chi-square statistic is a test of the null hypothesis that there is no agreement among judges. As can be seen, the probability of agreement amongst judges did not reach significance level (i.e., $p >$ 0 .05).

Since level 3 is very weak membership and level 4 is very weak nonmembership (i.e., related negatives), the area between these two levels on a continuum might safely be assumed to represent the category boundary. Consequently, agreement results about the items in levels three and four are of especial interest, since they indicate whether category structure is ill or well-defined. A high agreement among participants about which rankings to allocate to these two levels would be an indication of the category-type's well-definedness of category boundaries. Low agreement would indicate an imprecise or fuzzy notion of where exactly the boundary lies between membership and nonmembership.

Idiosyncratic category-types

As can be seen from Table 7, there is low variability between rank sums, a sign of high *disagreement* amongst participants about the level in a category to which an item belongs. Thus, all three category-types which were based upon idiosyncratic item-stimuli show disagreement amongst participants. As can be seen from Table 7, none of the three Kendall's coefficients of concordance showed a significant agreement, since their chi-squares all had probability values over .05. It is not surprising that all three category-types seemed to elicit so much disagreement amongst judges as to the levels to which certain item-sets should belong. People might agree as to which items belong/do not belong to a category, but the exemplars themselves might be organized in an idiosyncratic fashion. One participant might think that *pets* are a highly desirable item to save for the category *Things to save from a burning home,* and place that item in Level 1 (high degree of membership). Another participant might think that *pets* are not so important as *cash, jewellery* and *investment bonds* and relegate *pets* to Level 3 (very low degree); or even Level 4 (related negative non-member).

Consequently, to gain a more precise result, an analysis of the data for Levels 3 and 4 only was carried out. These levels reflect low membership and strongly related non-membership respectively, and as such, are testing for the ill-definedness of a fuzzy category boundary; or the well-definedness of a category boundary based upon an abstract rule. As can be seen from the values in parentheses for each category-type in Table 7, nowhere was chi-square's probability value below .05, signifying that agreement amongst participants did not reach significance level (Howell, 1987). It must be concluded that participants' disagreement about the category boundaries was a reflection of an ill-defined structure, which was found across all category-types.

Table 8: Rank Sums for five levels in Normative Categories: (in parentheses: analysis run on levels 3 and 4 alone)

Levels	Superord		Property		Ad Hoc	
1. High ProdFreq*	28.50		32.00		27.50	
2. Middle ProdFreq	34.50		41.00		37.00	
3. Low ProdFreq	39.00	(15.00)	56.50	(23.00)	46.50	(18.50)
4. Related Negative	58.00	(24.00)	41.50	(16.00)	50.00	(20.50)
5. Disjoint Negative	35.00		24.00		34.00	
Kendall's Coeffic.	0.30	(0.48)	0.35	(0.29)	0.20	(0.02)
Chi-square probab	0.004	**(0.013)**	0.001	**(0.052)**	0.034	**(0.579)**

* Production Frequency: The sequence of items on which the membership decision task was based.
In bold: Chi-square probability (df = 1) for levels 3 and 4, alpha level of significance < .05.

Normative category-types

As can be seen in Table 8, the normative-based items showed a different pattern of results to the idiosyncratic items. All three normative-based category-types showed highly significant agreement amongst judges about which category-levels an item should belong in, when all the five levels were taken into consideration in the analysis. However, to draw conclusions about the category boundary, only Levels 3 and 4 needed to be analyzed. On these two levels, it was found that agreement amongst judges about category boundaries was less marked, but still significant for natural superordinates ($p < .015$) and property ($p < 0.049$) category-types. For ad hoc category-types, high disagreement was present ($p > .05$). Such results lead to the conclusion that ad hoc category types based upon normative word-stimuli are ill-defined; but superordinate and property membership boundaries are well-defined.

Finally, Results Summary One below summarizes the combined results of the first and second analyses dealing with the unitary theory's prediction three (as set out in Tables 6, 7, 8).

Results Summary One: Internal Membership Structure
Combined results of analyses, Assumption Three

Category-Types:	Predictor*	Boundaries**	Implications***
Idiosyncratic:			
Natural Superord	Significant	Ill-defined	Single
Properties	Significant	Ill-defined	Single
Ad Hocs	Non-signif.	Ill-defined	Fuzzy core
Normative:			
Natural Superord	Significant	Well-defined	Dual
Properties	Significant	Well-defined	Dual
Ad Hocs	Non-signif.	Ill-defined	Fuzzy core

* Goodness-of-Example as a predictor of category membership structure
** Definedness of boundaries between categories in a category-type
*** Implications for category representation, from previous two columns (Predictor, Boundaries).

Where the idiosyncratic stimuli were concerned, the results for the superordinate and property category-types supported the predictions made by the unitary theory. It was shown that the membership of an item in a category can be predicted by its degree-of-typicality; and that categories have fuzzy, ill-defined category boundaries. Consequently, it would be logical to conclude that these types, when they consist of idiosyncratic items, should have a single representation - which contains both information about the members' degrees-of-typicality and the criteria for category membership.

Where the normative items were concerned, however, the superordinate and property category-types did not reflect the claims made by the unitary theory. Where membership of the item in such category-types can be predicted by typicality, the category-boundaries were shown to be well-defined, with participant judges agreeing significantly about an item's membership or non-membership. This combination of results does not agree with the unitary theory's assumptions, but it does fit the results found in other studies, that is, typicality structures combined with well-defined category boundaries (Osherson & Smith, 1981; Armstrong, Gleitman & Gleitman, 1983).

For both idiosyncratic and normative stimuli, the ad hoc category-types failed to support the unitary theory's prediction that membership structure could be predicted by typicality of the items. The presence of ill-defined boundaries, however, suggests a membership concept which is fuzzy and imprecise. Such a combination of results - gradience is not based upon typicality, and yet fuzzy boundaries are present - implies a single representation which does not involve similarity, with the fuzzy boundaries perhaps being best modeled by a dimensional schema. The results of the preceding analyses imply a dual representation for some of the category-types, and this also is tested in the next analysis by multiple regression.

4.11 RESULTS: REPRESENTATION OF CATEGORIES

Prediction Four: Assumption four of the unitary theory is that the concept of a category is mentally represented as a prototype. Such a claim can only be tested *in*directly by looking at whether a significant number of participants use more than one predictor of representation. It is predicted that participants will represent more than typicality information in their concepts of the three category-types.

Various models of semantic memory have used measures of production frequency (shortened to "prodfreq" in the tables below); and generation order (shortened to "genord") to measure the representation of category members in memory. Simultaneous multiple regressions using frequency-of-instantiation, goodness-of-example, and ranks as predictor variables, were carried out in the three category-types. Order of generation was used as the criterion measure for the idiosyncratic items-based data; and production frequency was the criterion measure for the normative items-based data. A separate multiple regression analysis was carried out for each individual subject. To avoid needless detail, Tables 9 to 14, Appendix F, show only those predictor variables with a significant T-statistic. If a subject's number is not listed in a Table, then no predictor variables were found to be significant in his/her regression analysis for that category-type. Summaries of the results set out in Tables 9 to 14 of the regression analyses are set out in the text below for convenience.

The first question asks what information constitutes the representation of each of the category-types. That is, which task is the best predictor of mental representation in each individual? If more than one significant predictor is found in a participant, then the second question must necessarily be whether this is occurring in a significant number of participants for that category-type. If so, that would be a strong indication of a dual representation of the same concept.

Squared Semipartial Correlations from individual analyses of participants' data with order-of-generation (or production frequency in the case of normative-based tasks) as the criterion measure, are presented in Tables 9 to 14 in Appendix F. Inclusion of predictor variables in a Table depended upon a significant T-statistic being found for that variable in each individual's multiple regression analysis. The Poisson approximation to the binomial distribution with small p-value (significance level = .05, number subjects = 13, then m = 0.65) was used to evaluate the number of subjects using a *sole predictor* task (frequency-of-instantiation; *or* goodness-of-example; *or* ranks). Also, the binomial distribution was used to evaluate the number of subjects who were using more-than-one predictor task, as this was taken to be an indication of use of a dual representation for the category.

Table 9.1 Superordinate Category - Types, Idiosyncratic items
Number of participants whose individual regression analysis found a significant independent variable predicting order-of-generation (criterion measure for category representation)

Sole predictors of Genord	No. participants	(*p-value*):
Frequency of Instantiation	8 people	0.000*
Goodness of Example	2 people	0.156
Ranks	0 people	1.000
More than one predictor	2 people	0.156

Table 9.1 is a summary of the information contained in Table 9, Appendix F. As can be seen in Table 9.1, the one significant predictor of superordinate categories based on idiosyncratic items (using the order-of-generation measure as the criterion variable) is frequency-of-instantiation ($p < .001$). Two participants were found to be using more than one significant predictor of the category representation. This number of participants is not statistically significant according to the binomial distribution ($p = 0.156$), so it must be concluded that overall most people use a single (rather than dual) representation of the natural superordinate category-type.

These results are confusing, given the results gained from the analysis run on assumption three concerning internal membership structure. That analysis showed that goodness-of-example (or typicality) *did* underlie membership structure in idiosyncratic superordinates (see Table 6). Representation, however, is found to be best predicted by frequency-of-instantiation (see Table 9). On the other hand, the results here also show participants using single representations, and this *does* provide converging evidence that idiosyncratic superordinates had ill-defined category boundaries (see Table 7).

In these present analyses, only 2 people used more than one predictor, implying the presence of a single representation for this category-type. What needs to be explained is the result presented in Table 9.1, which indicates that it is frequency-of-instantiation (that is, experiential information) which constitutes the representation (although Table 6 indicates that goodness-of-example underlies structure).

Table 10.1 Superordinate Category - Types, Normative items
Number of participants whose individual regression analysis found a significant independent variable predicting production frequency (criterion measure for category representation)

Sole predictors of ProdFreq	No.Participants	(*p-value*):
Frequency of Instantiation	0 people	1.00
Goodness of Example	7 people	0.000*
Ranks	0 people	1.00
More than one predictor	6 people	0.001*

Table 10.1 is a summary of the information contained in Table 10, Appendix F. The main predictor task of superordinate categories based on normative items is goodness-of-example ($p < .001$). This was a strongly significant result, with all participants using goodness-of-example as a significant predictor of production frequency (the normative measure of mental representation). As can be seen, more than one predictor was found to be significant in six participants. This is a significant number of people ($p < .001$), and also relevant might be the fact that five of these people used a combination of goodness-of-example and frequency-of-instantiation. This implies that natural superordinate categories based upon normative stimuli can be mentally represented in a dual fashion. The results of Table 10 in Appendix F provide converging evidence that goodness-of-example constrains representation, and thus support the results presented in Table 6. Furthermore, the presence of a dual representation (6 people using more than one predictor task) provides converging evidence for the findings presented in Table 8, for well-defined boundaries in these normative types.

Table 11.1 Property Category - Types, Idiosyncratic items
Number of participants whose individual regression analysis found a significant independent variable predicting order of generation (criterion measure for category representation)

Sole predictors Genord	No. Participants	(*p-value*):
Frequency of Instantiation	2 people	0.156
Goodness of Example	1 people	0.503
Ranks	5 people	0.001*
More than one predictor	2 people	0.156

Table 11.1 is a summary of the information contained in Table 11, Appendix F. Table 11.1 shows that the main predictor task of property categories based on idiosyncratic items was found to be Ranks ($p=0.001$). With property categories, the participants' data were much less clear-cut than for superordinate categories. Table 11 in Appendix F shows that Participant 13's analysis resulted in all three measures being significant; whilst all the predictor variables in the Participants 2, 3 and 7 analyses failed to reach significance levels, and so are not included in Table 11. Only two people were found to have more than one kind of information representing their criterion variable (order of generation), and this number of participants does not reach the significance level. It is to be concluded from the results that most people hold a single representation of property categories based upon idiosyncratic stimuli.

The results in Table 11 only partially support the implications of the results gained for membership structure (assumption four's analyses), which are set out in Table 6. Table 11.1 shows that goodness-of-example did not predict category representation, although typicality was shown to underlie membership in Table 6. Instead, Table 11.1 shows that Ranks provides the main information content for the majority of participants in idiosyncratic property types. On the other hand, the results here do provide converging empirical support for the results set out in Table 7, which showed that idiosyncratic property types have ill-defined boundaries. Table 11.1 also shows that the majority of participants used a single significant predictor (Ranks) for the order of generation criterion variable, implying their use of a single representation.

Table 12.1: <u>Property Category-Types, Normative items</u>
Number of participants whose individual regression analysis found a significant independent variable predicting frequency of production (criterion measure for category representation)

<u>Sole predictors of ProdFreq</u>	<u>No. Participants</u>	(*p-value*):
Frequency of instantiation	0 people	1.00
Goodness of example	5 people	0.001*
Ranks	3 people	0.034*
More than one predictor	3 people	0.034*

Table 12.1 is a summary of the information contained in Table 12 in Appendix F. It can be seen in Table 12.1 that the main predictor task of property categories based on normative items is goodness-of-example ($p < .001$), with a secondary sole predictor variable Ranks reaching significance level ($p=0.034$). Goodness-of-example, which only one participant used as a sole predictor for property categories based on idiosyncratic items, was used here by five people when the category-type was based on normative items. Analyses of data from Participants 6 and 8 yielded no significant predictors. The number of participants using more than one kind of information in their representation of the category-type was three, and this is a significant number, leading to the conclusion that normative property types can be represented in a dual fashion, at least by some participants.

With reference to the membership structure analyses carried out to test assumption four, these results provide unreserved converging evidence.

Goodness-of-example is shown as the main predictor of representation here also, as it was shown to underlie membership structure in Table 6. Furthermore, the presence of well-defined category boundaries in normative property types (see Table 8) was supported by this analysis, in that the majority of participants used more than one predictor for the representational criterion of production frequency, implying a dual representation of the category.

Table 13.1: <u>Ad Hoc Category-Types, Idiosyncratic items</u>
Number of participants whose individual regression analysis found a significant independent variable predicting order of generation (criterion measure for category representation)

<u>Sole predictors of Genord</u>	<u>No.Participants</u>	(*p-value*):
Frequency of Instantiation	1 people	0.503
Goodness of Example	5 people	0.000*
Ranks	0 people	1.00
More than one predictor	5 people	0.001*

Table 13.1 is a summary of the information contained in Table 131 in Appendix F. From Table 13.1, the main predictor task of ad hoc categories (idiosyncratic items) is shown to be goodness-of-example ($p < .001$). Participants 1 and 9 failed to give a significant result for any of the three predictors, while for Participants 12 and 13, all three predictors were significant. More than one kind of information was used by five people, reaching significance level. Hence, it is concluded that idiosyncratic ad hoc category-types can be represented in more than one way.

Table 14.1: <u>Ad Hoc Category-Types, Normative items</u>
Number of participants whose individual regression analysis found a significant independent variable predicting production frequency (criterion measure for category representation)

<u>Sole predictors of ProdFreq</u>	<u>No.Participants</u>	(*p-value*):
Frequency of Instantiation	3 people	0.034*
Goodness of Example	5 people	0.001*
Ranks	0 people	1.00
More than one predictor	5 people	0.001*

Table 14.1 is a summary of the information contained in Table 14 in Appendix F. Table 14.1 shows that the main predictor of mental representation of normative ad hoc category-types was goodness-of-example ($p < .001$), with a significant secondary predictor being present in frequency-of-instantiation ($p < .034$). A significant number of participants were found to use more than one kind of information in their mental representation of the category, and this leads to the conclusion that normative ad hoc types can be represented in a dual fashion.

For ad hoc types, the results on representation contained in Tables 13.1 and 14.1 do not provide converging evidence for those on membership structure set out in Tables 6, 7 and 8. The analyses testing assumption three implied that ad hocs had a single representation for both idiosyncratic and normative stimuli (see Results Summary One). They also suggested that goodness-of-example did not underlie ad hoc membership structure (see Table 6). Yet the results here, as set out in Tables 13.1 and 14.1, indicate that most participants were using a dual representation (more than one predictor task), and goodness-of-example was found to be a significant predictor of representation. Overall, the results from Experiment 2 on ad hoc types were confusing, and did not point to any firm conclusion. The main predictors of category representation are summarized in Results Summary Two.

Results Summary Two Significant predictors of category representation, with implications (in italicized parentheses) for the single/dual issue.

Category-Type	Idiosyncratic *	Normative **
(i) Superordinate	Frequency-instantiation *(single)*	Goodness-of-example Frequency-instantiation *(dual)*
(ii) Property	Ranks *(single)*	Goodness-of-example and Ranks *(dual)*
(iii) Ad Hoc	Goodness-of-example *(dual)*	Goodness-of-example Frequency-instantiation *(dual)*

* Idiosyncratic stimuli were based upon the dependent variable, order-of-generation, as the criterion for category representation.

** Normative stimuli were based upon the dependent variable, production frequency, as the criterion for category representation.

In general, Results Summary Two shows that the implications for a single or dual representation of superordinate and property types match those found in the analysis testing prediction three, and set out in Results Summary One. To summarise the main points, where idiosyncratic stimuli were used the results of these regression analyses on individual subjects' data indicate that the representative information contained in each category-type can be measured solely by one of the three tasks: either Ranks, frequency-of-instantiation judgments, or goodness-of-example judgments. This is not true of all participants, but for superordinate and property category-types, the majority of participants did use only one of these kinds of information in their mental representation of those categories.

For the normative-based version of those category-types, the results show that, across participants, more than one predictor variable was found to be significant, suggesting that a significant number of people was using more than one kind of information in their mental representations of the category.

4.12 DISCUSSION: INTERPRETATION OF RESULTS

4.12.1 Categorization Process

The effects of normative and idiosyncratic word stimuli upon performance in a membership decision task were compared, but no significant differences were found in either response times or accuracy. If Lakoff's (1987) claim that emotively-based materials are richer in information is accepted, then the idiosyncratic items should have taken longer to process and been more accurate. This was the slight trend in the results, but not to an extent that was significant.

The lack of significant differences between membership decision tasks based upon idiosyncratic and upon normative stimuli could be due to participants inability to distinguish between subjective, autobiographical knowledge (idiosyncratic stimulus items), and cultural, normative knowledge (normative stimulus items). Instead, a more likely explanation for the fact that they are using both kinds of knowledge equally well, might be that a subjective coding of experience is as important as the use of category norms during categorization. However, such an interpretation needs supportive converging evidence, since the analysis lacked statistical power, so that the null hypothesis is not retained unreservedly, thus running the risk of a type II error.

This support was found when comparing the results of the next two analyses (see Tables 4 and 5). These two analyses involved the two independent variables of category-type by level of high, middle, or low representativeness of items within the category. A diminished sensitivity in the data from normative items was found, in that the data analysis of these items did not detect any significant difference between category-types. In contrast, the idiosyncratic based data analysis yielded significant results between all three levels of item representativeness *and* between category-types. Thus, the least ambiguous point to be derived from these two analyses is that idiosyncratic items do differ from normative items in a qualitative way, if not quantitatively (as shown by the lack of significant differences in the previous comparison of stimulus-types). Also, Freyd's (1983) claims for a *subjective* organization of category knowledge, and Lakoff's (1987) claims for an *experientially* based organization receive some support from the more finely-grained results yielded by data from the membership decision task based upon idiosyncratic stimulus words.

The main aim of the two analyses presented in Tables 4 and 5 was to discover if different category-types would elicit different categorization processes: serial, additive processes which take place in two stages, or computational, parallel processes which are complete in one stage. Both analyses found a significant difference between levels-of-representativeness, which were borne out by the post-hoc comparison tests. In the normative based analysis, the significant differences were found to lie between the high and the middle/low levels of production frequency. The lack of a significant main effect for the second factor, that of category-type, rules out a serial or two-stage model of processing as a possible interpretation of the normative data analysis (Hampton, 1984). Rather, it would indicate a linear increase in the response time curve (rather than a flat curve), and so must be interpreted as a model of limited processing capacity (Townsend, 1990). It is concluded that the different times between the high and the middle/low levels of gradience indicate differing levels of accessibility in participants' retrieval of the instance-category membership relationship.

The idiosyncratic based analysis yielded a greater number of significant differences, and hence more inferences were possible. The results of the post hoc tests indicated significant differences between the representativeness of items at all three levels of generation order: first, middle, last. Also, a main effect was found for the three category-types, with the natural superordinate types being significantly different from the other two types. The lack of interaction between order of generation levels and category-type suggests that the response times for the various items at the different levels of generation order were not influenced by the type of category to which they belonged. Townsend (1990) claims such a lack of interaction between two significant main effects indicates *additivity* because the factors are affecting reaction time in a separately additive fashion. It is strong evidence against the unitary view's assumption that represented structures are a direct reflection of the environment's category structures (Smith, 1978).

The results support a serial, two-stage feature comparison model for categorization (Hampton, 1984; Smith, Shoben & Rips, 1974), since no significant interaction was found between the two independent variables, the hallmark of a two stage process being in operation (Townsend, 1990). In Smith's (1978) description of a serial processing model, the two stages of processing are independent in that different kinds of information are being used: the first stage uses general encyclopaedic knowledge comprising both characteristic and defining features; and the second stage uses dictionary knowledge of defining features alone. In the first stage, features are compared in a rapid parallel process of comparison. The more features there are, the faster the membership is confirmed or contradicted. The second stage is only initiated if there are insufficient features (in other words, information) upon which to base a decision. The second stage resolves the indecision by using the presence (or absence) of defining features only, meaning that this second stage is a classical process. The swift parallel comparison of characteristic and defining features which occurs during the first stage would be sufficient for the items at the first order of generation; and the slower, more diagnostic analysis of the defining features said to occur in the second stage would be initiated only for the items at the second and third orders of generation.

Finally, taken together, these data analyses (see Tables 3, 4 and 5) provide strong evidence against the validity of the unitary approach to categorization, since this view would interpret the information-processing data as directly and simply as possible. The results found in this experiment provide support against the view of categorization processes being an automatic direct retrieval (or one-stage computation) of instance-category relationships, which are already encoded in memory. On the contrary, they imply that subjects make "indirect" or "inferential" membership decisions, first accessing knowledge about the category, then making decisions about membership based upon that knowledge (Smith, 1978).

4.12.2 Internal membership structure

The clearest results came from the idiosyncratic-based natural superordinate and property types, and supported prediction three, as shown in the combination of results in Tables 6 and 7. In both natural superordinates and property types, typicality could predict membership structure (see Table 6). Furthermore, both types were shown to have ill-defined category boundaries (subjects disagreed about levels 3 and 4 in Table 7) The unitary view provides a fairly straightforward interpretation of this combination (fuzzy structure and ill-defined boundaries). Its assumptions about structure are upheld in the superordinate types, but not in the property types, which it predicts as showing a combination of an ungraded membership structure and well-defined boundaries.

Where the *normative* stimulus words are concerned, the evidence points to a dual representation of the concept. The combination of results (fuzzy but well-defined) does not agree with what a unitary view would predict: if the category structure is fuzzy, then its criteria will be *ill*-defined. The unitary approach cannot provide an account for such results. On the other hand, this particular combination may be interpreted as evidence for a dual representation of the concept. Tables 6 and 8 show a contradictory state of affairs exists. The natural superordinate and property types were both fuzzy (typicality was a significant predictor of membership structure in Table 6); and *well*-defined (subjects agreed significantly about category boundaries in Table 8). In the property types, subject agreement did not quite reach significance level (p = .052), but it is likely that the high number of missing cases for that specific type influenced the results.

These empirical results raise the possibility that the same category might have different representations. The conceptual representation of a category 's criteria (its definedness) does not agree with the typicality structure of the items (fuzzy), so that a split exists between representation of the concept's intension

(Table 8), and the category extension (Table 6). For example, a category-type with well-defined criteria and boundary should be found to have a membership structure *unpredicted* by typicality. This did not always happen, and was most evident in the superordinate types, where typicality predicted membership structure. The results suggest that people may have one representation for information about items internal to the category (for example, their typicality information); and another representation for membership criteria in the category (such as what differentiates one concept from another).

Concerning the general question asked by Experiment 2, "can different concepts be represented differently?", it would seem at first that this must be the case, especially for the combination of results for the tasks based on idiosyncratic stimuli. One possibility would be that they differ in the number of their representations: a single representation might be sufficient for one category-type, but a dual representation might be necessary for another category-type. An alternative answer might be that, rather than different concepts being represented differently, category behaviour may best be captured by the same concept being represented in more than one way. If so, then the structures derived from goodness-of-example rating tasks *under*determine mental representations, limiting the possibilities of other potential representations. The next analysis should indicate which of these answers is the viable one.

4.12.3 Representational models

<u>Natural superordinate and property category-types (idiosyncratic stimuli)</u>

The results from the data analyses on representation of natural superordinates and property types, based on *idiosyncratic* word stimuli, suggested a single representation by most participants, with category-types differing as to the nature of the information represented. Natural superordinates were constituted of information about frequency of instantiation (see Table 9 in Appendix F). These results support Barsalou's claim (1985) that frequency of instantiation and experiential information does have a role to play in category representation. The property types had the same structure as superordinates (fuzzy) but differed from them in the kind of represented information about the category (see Table 6, and Table 11 in Appendix F). The main predictor task for these property types was Ranks (i.e., membership criteria), and it is argued that criteria would be determined by the name of the category, for example, the label *Red Things* necessitates that the objects be wholly or partially coloured *red.*

If people represent category knowledge in an idiosyncratic organization, yet understand normative structures also, the idiosyncratic based structure might be more detailed with regard to information than the normative structure, allowing a more distinctive account of an individual's use of categories to be made (Freyd, 1983; Barsalou, 1987). For example, it would seem from the results that categories showed a more basic membership structure when idiosyncratic stimuli were used.

The experiment showed that the membership structure of categories (well-defined or ill-defined) can differ according to whether the stimuli used in the task were idiosyncratic or normative, rather than differ according to category-type. It was found that a natural superordinate based upon idiosyncratic word-stimuli could be differentiated from one based upon normative word-stimuli, by the addition of goodness-of-example as a significant predictor of concept representation. The same pattern occurred across the normative and idiosyncratic property types. Conceptual representation of the idiosyncratic property types was predicted by the Ranks variable, whilst the representation of the normative property types was predicted by the Ranks variable *and* the goodness-of-example predictor. This pattern of an added predictor in the form of goodness-of-example (typicality) would suggest that normative item-stimuli are less basic to the category information represented. Perhaps they elicit a typicality structure based upon cultural norms, whilst the idiosyncratic item-stimuli provide the subjective experiential information. This possibile interpretation is further supported by the membership structure of all three types changing from fuzzy and ill-defined (in the idiosyncratic stimuli) to a precise and well-defined structure for these normative-based category types.

One interpretation of this pattern is suggested by Barsalou (1987; 1989), who found that one of the factors which influenced representation of graded structure was the point-of-view taken by the categorizer. In Experiment 2, then, subjects might be said to be taking two points-of-view: their own, idiosyncratic view formed by their own personal experience; or that point-of-view which they have learnt at second hand from the society in which they live. There might not be a great range of difference between the two points of view, but rather they might differ on a continuum of fuzziness, with the idiosyncratic items producing a fuzzy structure whilst the normative items produce a more clear-cut structure. As argued in chapter 1, people need both kinds of similarity: a sense (or metric) of subjective similarity; and of normative similarity consisting of social or cultural knowledge. Each metric interacts towards production of overall conceptual stability.

Another interpretation of this pattern is suggested by Quine's (1969) theory of an innate similarity metric, and by an eventual distinction being made between "subjective" and "objective" similarity. Originally, the overall physical appearance of an object (or its frequency of occurrence) is used as a basis for a holistic comparison. With time and adult development, however, this initial similarity metric evolves from a purely subjective, appearance-based response into a theory-based similarity metric. Quine believed that children used the first, but eventually replaced it with the second metric, which no longer used physical appearance as a basis for grouping together of objects into a category. Instead, it consisted of theories as to the meaning of the objects' physical appearance. It would seem from these results that people have retained the first "subjective" similarity metric, at least as part of the final "objective" or "normative-based" similarity.

This "subjective" similarity metric might best be accounted for by Brooks' (1978) instance category model, which describes a parallel process of holistic comparison of instances. Brooks (1987) instance account views some categories as "natural", with the observations of the world represented by a distribution of instances acting as individuals. Brooks (1978) processing-dependent account posits similarity organized in the form of schemata, and dependent upon context frequency. Brooks (1987) and Jacoby and Brooks (1984) have put forward a theory of nonanalytic cognition because they believe that "undue faith" has been placed in " the role of centralized, abstracted models as representations of everyday human knowledge" (Brooks, 1987, page 141). According to the theory of nonanalytic cognition, similarity is assessed, not by the analysis of an object's features and comparison of them with a prototype's independent feature list, but on the basis of prior processing episodes in the person's life. The weighting of a feature's similarity depends upon prior processing episodes and the *context, task,* and the *item* in which they occurred. This would agree with the research (e.g., Gati & Tversky, 1984; Tversky & Gati, 1978) which ascribes more flexibility to perceived similarity than is allowed in feature comparison models, such as family resemblance.

Representations based upon normative stimulus words

The results from the data analyses on representation of all three normative category-types support the dual representation theory (see Tables 10, 12 and 14 in Appendix F). They show that a significant number of participants were capable of using more than one kind of information when generating items for the production frequency task (used as the criterion of category representation). Taken together, these results would suggest that normative category-types are represented by more than one kind of information, and a significant number of participants are using more than one kind also. The fact that different predictor variables were found in each of the three category-types (as listed in Results Summary Two) suggests that the three types can be distinguished according to the nature of the information they represent. For the natural superordinate and property types at least, this does not mean that different category models are required. Possible category-models for category-types are described below.

Natural superordinate and property category-types (Normative stimuli)

Goodness-of-example or typicality was found to be the only significant predictor of representation in the superordinate types based upon normative stimuli (see Table 10.1), However, a significant number of participants (five) were found to be using frequency-of-instantiation as a secondary source of information as well, so that a category-model for natural superordinates would need to include two kinds of information: both typicality *and* individual frequency of experience.

The dual representations implied by the results would best be explained by the "core and identification" model (Miller & Johnson-Laird, 1976; Johnson-Laird, 1983; 1990). In this model, the conceptual core is well-defined and consists of a membership rule, whilst typicality performs identification functions such as how easily an object is distinguished or recognized as an individual member in relation to other members.

Property category-types were shown to be the same as superordinate category-types in structure (both were well-defined), but not in the kinds of information represented (compare Tables 10 and 12 in Appendix F). The main predictor tasks for property types were found to be Ranks and goodness-of-example, rather than frequency-of-instantiation and typicality as in natural superordinates. The "core and identification" model is applicable here also, and although the two category-types would be similar as regards the identification function (goodness-of-example), they do differ as to the information contained in their conceptual cores.

Ad hoc category-types, both normative and idiosyncratic stimuli

Ad hoc types were found to have roughly the same structure and representation, irrespective of whether the measures were based upon idiosyncratic or normative stimuli, so they will be dealt with together (see Tables 6, 13.1 and 14.1). The only point on which they differed was that the normative ad hocs had two significant predictor variables (frequency-of-instantiation and goodness-of-example); whilst the idiosyncratic ad hocs had one significant predictor (goodness-of-example). In Table 6, the ad hoc categories were shown to have ill-defined boundaries and a membership structure which could not be predicted by goodness-of-example or typicality. Results presented in Tables 13 and 14, Appendix F imply that ad hocs were represented in a dual fashion. That is, they could be predicted in a significant number of subjects by more than one kind of information.

This combination suggests a category-model involving a fuzzy core (not a well-defined one), and an object's relations with other members expressed by its goodness-of-example as a member (a significant predictor of representation, but not of membership structure). As can be seen, these results are ambiguous in that whilst goodness-of-example was not found to significantly predict membership structure, it does do so where representation is concerned. A tentative conclusion, then, is that membership criteria for ad hoc types is not based upon degrees-of-typicality, but that this information is contained in the representation of the category extension. Since the membership criterion is fuzzy (see Tables 7 and 8) in that category boundaries are ill-defined, the category-model which might best fit the ad hoc types is a schema constituted of typicality and experiential information, constrained by an imprecise and fuzzy core (Rumelhart, 1980).

In conclusion, the normative superordinate and property types might be described in terms of a Core and Identification model, whilst the idiosyncratic superordinateand property types can be described by a holistic, analogical model of categorisation based on experiential information, such as described in Jacoby and Brooks' (1984) theory of nonanalytic cognition. The two are not necessarily mutually exclusive, since it is likely that the normative stimuli-based results are reflecting Quine's (1969) "objective similarity" metric, with people representing their categories in both an idiosyncratic and a normative organization.

4.13 IMPLICATIONS OF RESULTS

Briefly, the main consequence of the unitary view is that similarity is taken to be the underlying basis for both the internal structure of a category (the category's extension), and for its external relations with other categories (the category's conceptual intension). The prototype view of graded structure defines it as including items both internal *and* external to the category, that is, information about membership status and typicality of the items in the one representation of the category. Whilst internal typicality gradients are obviously ubiquitious to all category-types, using them to explain the concept of what constitutes criterial membership in the category results in an impoverished account of category representation.

The more recent dual representation theories try to reconcile the structures of a category's extension and a concept's intension by combining the two in dual representations, and a number of "hybrid" models have been developed and described in the literature (Komatsu, 1992). B. Landau (1982), in an early empirical study into the psychological reality of dual meaning representations, captured the essence of the debate when she gave her article the title "Will the real grandmother please stand up?" She found that people can quite happily use two distinct kinds of information when deciding whether someone is a grandmother or not: symptomatic (prototypical) descriptions of grandmotherhood, and diagnostic criteria (necessary and sufficient conditions) for being a grandmother.

Landau (1982) presented pictures to her participants, who included both adults and children, asked them to select exemplars of a grandmother (or some other kin-terms) from them, then justify their choice. The *physical symptoms* (or indicators) of grandmothers included bifocals, gray hair and old age, whilst the *criterion* for being a grandmother was to have a grandchild beside her. Landau's pictures included both symptom and criterial information, so she was really testing which justification her subjects might give for the selection of a picture, rather than the categorization *per se*. She found that, whilst adults tended to use the criterial information more often as a justification, all subjects used both kinds of category-information. Her conclusion was that different task requirements will bias persons to use either one kind of information or the other. As she pointed out, whilst the use of symptomatic information might provide *likely* or *probable* indicators of an item's representativeness as a member, there are times when certainty becomes very important, even a matter of life or death, and criterial information is needed and used to make the categorization decision (Landau, 1982; Wittgenstein, 1953).

Wittgenstein (1953) also believed that these two kinds of meaning are represented, and are used for different purposes. In certain situations, only a tentative categorization is required, so symptomatic indicators are used to estimate a likely categorization for an object. In more serious situations, such as medical diagnosis, criterial information is used to decide absolutely about the object's category, that is, illness. Definitional criteria need only be used when a diagnostic decision is necessary, or when payoffs are high, as when diagnosing an illness correctly; or categorizing a plant as a toadstool rather than a mushroom (Barsalou, 1989; Medin, 1989).

In summary, dual representation theories posit one representation for the category's definitional core (i.e., its intension) and one representation for the category's similarity structure (i.e., its extension). In such concepts, typicality variations may reflect how well an instance fits some kind of heuristic recognition procedure for set members, whereas true category membership (in, for example, *Tiger, Lemon, Gold*) may rely on less accessible criteria, such as unseen essential features (Putnam, 1975a, 1977). This is a theoretical issue which cannot be directly tested from Experiment 2, but testing for informational cores would indicate the nature of information contained in the conceptual core.

The unitary approach, restricted as it is to similarity and typicality information, cannot capture people's use of diagnostic, criterial information. The principle of parsimony (or cognitive economy) argues against dual representation models, but single representations can be a two-edged sword, in that category models become too simplistic and cannot account for some of the generic knowledge used when making membership decisions. Whilst natural categories such as *Vegetables* or *Felines* might find a single representation sufficient, research on other category-types has raised questions about the adequacy of physical similarity as a basis for membership decisions.

Questions for Experiment 3

Experiment 2 showed that the unitary model of category behaviour could not provide an adequate account of people's representation, structure and processing of categories. The variables predicting category representation included measures of prototypicality (goodness-of-example judgments), of accessibility to the conceptual criterion (rankings derived from the response times in the membership decision task), and of judgments of direct experience with the items *in the category-context* (frequency-of-instantiation).

The results from performance with these predictor variables were ambiguous, and left some questions unanswered. The regression analysis produced high values in accountability for variance (r-squared) by some of the predictor variables. However, their high accountability for variance in the criterion variable was not always borne out, when their corresponding semipartial correlations with the criterion variable were calculated More often than not, the semipartial correlation of a predictor variable with the criterion was shown to be quite small (see Tables 9 to 14 in Appendix F).

Two possible interpretations of these results can be made. Firstly, it is possible that the experimenter's choice of stimulus category-labels is to blame for the low semipartial correlations. For example, in the stimuli for the superordinate category-types, both artifact and animal category-labels were included. Some metaphysical philosophers claim that artifact and animal concepts are inherently different, because artifacts are nominal concepts, whose meaning is a product of society, whilst animals are natural concepts whose meaning is derived from their innate essence (Putnam 1975a; 1975b). The divided nature of the item stimuli (at least those in the superordinate types) might have lessened their predictability, especially by the goodness-of-example task. The question of a possible difference in conceptual representation of artifacts as compared with that of animals will be investigated in Experiment 3.

A second possible interpretation for the low predictability values of the typicality judgment task might simply be that the unitary model is an inadequate account of category behaviour. The low semipartial correlations might be explained by the exclusion of a factor which was not allowed for in the experimental design, and not included as a predictor of representation. One such variable, as shown by recent research, is used by subjects during categorization: general background knowledge about the category. In other words, the membership, typicality and perceived frequency of instantiating the category might have all been influenced by an underlying factor: the subject's theory-based knowledge about the category.

For example, although the goodness-of-example was found to be the significant predictor of superordinate types (at least in the normative item task), its semi-partial correlation values were very low. Furthermore, it was found to be a significant predictor in the property types and ad hoc types, categories where similarity and typicality of the member-items are irrelevant. It is doubtful that the goodness-of-example task was measuring physical similarity of appearance of the members in these category-types, in which case what was happening? It is possible that, during task performance, participants were using background knowledge of the category to guide them in their judgments of typicality, direct experience, and membership. If so, that might also explain the high level of confounding amongst the three predictor variables.

This is the main question left partially unanswered by Experiment 2: why do we have the categories we do, and not others? The experiment has shown that people's category representations can be based on both typicality (goodness of example) *and* direct experience (frequency of instantiation). It would seem, from the experiment, that people use categories to cognitively organize their own experience of the objects and events in the world around them. The experiment did not answer why we have these particular representations, rather than others The results did suggest the use of an untested factor, general background knowledge, as guiding judgments of category items. The deciding factor in why people categorize an item as a member of some category rather than another, may lie with the theories and beliefs they hold about that object or creature. This possibility will be tested in Experiment 3, where it is hypothesized that people use theories and beliefs to determine the category identity of animals and artifacts.