Spontaneous and relative categorisation.

Edwards, Darren J

Spontaneous and Relative Categorisation

Darren J. Edwards

Submitted to the University of Wales in fulfilment of the requirements for the Degree of Doctor of Philosophy

Swansea University

September 2010

Supervision provided by Dr. Emmanuel Pothos

Abstract

This thesis concerns two different aspects of categorization, the first is an investigation with a novel paradigm, which relates to relative vs. absolute (supervised) categorization, and the second, an investigation of the simplicity model (Pothos & Chater, 2002). The first investigation was motivated from the relative judgment model (Stewart et al., 2005). According to this model, classification judgments in absolute identification tasks are influenced by the relative context in which they are presented. We examine the generality of this conclusion in categorization. In the present study, we tested 320 participants in 5 experiments, in which participants had to classify new items into predefined artificial categories. In three experiments, we observed a (predominantly) relative mode of classification, and in 2 experiments we observed an absolute mode of classification. These results suggest three factors which promote a relative mode of classification; when there are fewer items per group, more training groups, and the presence of a time delay. Overall, we propose that less information about the distributional properties of a category and/or weaker memory traces for the category exemplars (induced, e.g., by smaller item numbers per category, or a time delay respectively) can encourage relative judgment. For the simplicity model, we conducted three experiments, a free sort task, a learning task and a memory task. In the free sort task, we asked 169 participants to spontaneously categorize nine sets of items. A category structure was assumed to be more intuitive if a large number of participants consistently produced the same classification. Our results provide a rich empirical framework for examining the simplicity model of unsupervised categorization (Pothos & Chater, 2002).

Declaration and Statements

Declaration

This work has not been previously accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ....Darren Edwards..................................

Date ....29/9/2010.................................

Statement 1

This thesis is the result of my own investigations, except where otherwise stated.

Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .... Darren Edwards ....

Date ....29/9/2010..................................

Statement 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter library loan, and for the title and summary to be made available to outside organizations.

Signed .... Darren Edwards ....

Date ....29/9/2010................................

Table of Contents

Acknowledgements

In doing this work, I have benefited greatly from my association with excellent researchers, Dr. Emmanuel Pothos, my academic supervisor, and Dr. Irene Reppa. Emmanuel's continuous and enthusiastic support has helped me greatly to carefully refine my ideas, and also, gave me the patience to discover what real research is about. Irene's influence was significant in ensuring a smooth continuation of my research. My interaction with Emmanuel's group has initiated my interest in categorization. This has allowed my current research focus to develop, which is my application of this modelling work to the areas of autism and brain injury, for which I have applied to carry out further work, to the ESRC in terms of a postdoctoral Fellowship.

The research experience of Dr. Amotz Perlman, has helped me obtain a publication in JEP:HPP in the area of unitization, and the help from Prof. Phil Reed and Prof. Rodger Wood has now allowed me to develop my thesis wok into the applied areas of autism and brain injury. The association with these people have ensured that I have received an ideal environment for research, where ideas and advice were freely shared. More specifically, I would like to thank Neil Carter for Matlab, and computational help whose computer knowledge spared me long hours.

The moral support from several friends of mine, and my family, has been an important factor in the completion of this work. Acknowledgements go to Carlo, Rob, Gary, Ting, Katy, Karl, Maria, Raff and many others to whom I am grateful for consistent assistance and thoughtful advice at all stages of my Ph.D. Also, acknowledgements go out to many other friends of mine, such as those at the archery club which I attend, and to those whom I score higher on, on a weekly basis, in my favourite computer game—Call of Duty.

Thesis peer review; conference proceedings, publications and submissions.

Aspects of my thesis work have been used in the following publications and submissions. Note, for the publications (e.g., Pothos et al., 2008), where I am *not* first author, in Pothos et al., the emphasis has not been on the experimental results (which was my work) or the simplicity model analysis (which I did), but rather on an extensive application of additional models of categorization (e.g., SUSTAIN, rational model, DIVA), with which I was not involved and I do not cover in this thesis.

- **Edwards, D. J.**, Pothos, E. M., & Perlman, A. (Submitted). Relative Vs. Absolute mode of categorization. *Memory and Cognition.*
- Pothos, E. M., Perlman, A., **Edwards, D. J.**, Gureckis, T. M., Hines, P. M., & Chater, N. (2008). Modeling category intuitiveness. In *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, LEA: Mahwah, NJ.
- Pothos, E. M., Perlman, A., Bailey, T. M., Kurtz, K., **Edwards, D. J.,** & Hines, P. (submitted). *Measuring category intuitiveness in unconstrained categorization tasks.*

I have presented my thesis work in the following conferences:

Conference proceedings 2008-2010

- **Edwards, D. J.,** Pothos, E. M., Perlman, A. (2010) *Using the simplicity model in measuring the category intuitiveness of unconstrained sorting tasks.* In proceedings of the 27th Anniversary BPS Cognitive Section Conference. Cardiff.
- **Edwards, D. J.,** Pothos, E. M., Perlman, A. (2010) *Validating the simplicity model of unsupervised learning.* In proceedings of the 25th Annual conference PsyPAG. Sheffield.
- **Edwards, D. J.** (2009) *Classification Mode in Categorization: Relative vs. Absolute Judgment.* In proceedings of the 24th Annual conference PsyPAG. Cardiff.

- **Edwards, D. J.** (2008) *Relative Vs. absolute mode of categorization.* In proceedings of the 25th Anniversary BPS Cognitive Section Conference. Southampton.
- **Edwards, D. J.** (2008) *Relative Vs. absolute categorization.* In proceedings of the 23rd Annual conference PsyPAG. Manchester.

Posters

- **Edwards, D. J.,** Pothos, E. M., Perlman, A. (2009) *Classification Mode in Categorization: Relative vs. Absolute Judgment.* In proceedings of the 26th Anniversary BPS Cognitive Section Conference. Hertfordshire.

Tables and Figures

Tables

Figures

Chapter 1

Introduction

The term 'cognitive science', was coined in 1967 by Ulric Neisser, who used this to mean all processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used. Neisser (1967) refers to people as being dynamic information-processing systems whereby a description of their mental operations can be given in computational terms. The origins of ideas in cognitive psychology, such as computational theory of mind, can be traced back to Descartes (17th century) and continued with Alan Turing (1940s-1950s). This basic foundation of cognitive psychology allowed the development of more thorough attempts to understand how we process and organise information in terms of information processing, and has led to the development of complex the categorization theories, that we have today.

Today in cognitive psychology we have very complex mathematical descriptions about how information is organised in terms of spontaneous (e.g., the simplicity model, Pothos & Chater, 2002) and supervised categorization (e.g., the generalized context model, Nosofsky; 1984, 1986, 1991). These models provide us with a rich range of predictions about how information can be organised when we have no past knowledge about it (spontaneous) and also, how information is organised when there is an external agent guiding classification (supervised). We also have evidence as to how relative properties can affect classification decisions, that is properties which do not depend on the physical appearance of a stimulus that the most spontaneous and supervised modes depend on. This is through the work carried out on shared properties in analogical mapping (Gentner, 1983, 2003; Holyoak & Thagard, 1995) and also the relative judgment model (RJM; Stewart et al, 2005). From this literature, we have produced five experiments that explore a relative vs. absolute shift in categorization and three further experiments that demonstrate that people organise information in spontaneous and supervised categorization in terms explained by simplicity model (Pothos & Chater, 2002).

In Chapter 2, I explain the simplicity model (Pothos & Chater, 2002) in unsupervised categorization and compare it with other models (such as SUSTAIN; Love, Medin, & Gureckis, 2004) in the hope to further clarify its uniqueness in categorization. In Chapter 3, I explain supervised categorization (including models such as the generalized context model, Nosofsky; 1984, 1986, 1991), how this contrasts with spontaneous categorization, and how this relates to the studies I will present on relative vs. absolute judgment. In Chapter 4, I explain analogical mapping (Gentner, 1983, 2003; Holyoak & Thagard, 1995) and provide a literature review of relative judgment in categorization (Stewart et al, 2005). I also explain how it is relevant to our relative vs. absolute judgment experiments. In Chapter 5, I explain the five relative vs. absolute judgment experiments that I have carried out and reach a conclusion of what promotes a relative judgment as compared to an absolute judgment. In Chapter 6, I explain the spontaneous categorization experiments and give my conclusion as to how this relates to the simplicity principle. In Chapter 7, I explain the results from the supervised categorization studies that relate to the simplicity model. In Chapter 8, I give my general conclusions on the three simplicity model (unsupervised, learning and memory) experiments and the five relative vs. absolute judgment experiments, and how my findings advance our knowledge of categorization in general.

Chapter 2

The simplicity model in unsupervised categorization

2.1 An Introduction

Unsupervised classification deals with the problem of understanding how people organise information into categories without any prior knowledge of the items, or how they should be categorized. For example, if someone were presented with novel items, such as seeing a novel computer game or material viewed under a microscope, then the information presented might be interpreted in terms of different groups. Crucially, unsupervised categorization deals with how we group items that we have not seen before or have any idea what the items relate to.

The main technique for exploring how people organise information in an unconstrained way is through free sort tasks. In these tasks, the participant is given a collection of items and is asked, simply, to categorize these in a way that seems most intuitive. There is no feedback instructions and therefore in this case categorization is completely intuitive. This is different to supervised categorization tasks, where constraints on categorization are included. These constraints can include feedback relating to a desired structure, general knowledge, and category labels (for a detailed specification on such constraints, see exemplar theory, Chapter 3, on supervised categorization). The objective in supervised categorization (e.g., Brooks, 1987; Hintzman, 1986; Medin & Schafer, 1978), is to identify the ways in which people categorize new items into existing groups which have been already specified by the experimenter. In such a case, the experimenter attaches a label to the group, such as 'this is a group of Chomps and this is a group of Blibs', and specifies exactly which items belong to the groups. So, the main difference between these types of categorization is that one uses constraints and the other does not.

Despite this difference, unsupervised and supervised categorization are not completely different. One shared feature of supervised and unsupervised classification is that they both make their predictions of classification (mostly) on the basis on physical similarity. More specifically, in supervised classification, the classification decisions are typically made on the basis that the new item is most similar to the items within an existing group (see Nosofsky; 1984, 1986, 1991). For example, if one category consists of triangles and another consists of squares, and then a new item is introduced which appears more like a triangle than a square, then participants are more likely to categorize the new item into the category which consist of triangles, than the category that consist of squares. Likewise, in unsupervised categorization, where, for example, free sort classification tasks are used, the participant has to sort items into groups using the similarity of the individual items, such as the length and width of the items. So, the key feature that both of these types of categorization share, is that they both make their predictions on the basis of physical similarity.

Unsupervised categorization can be conducted under different experimental conditions (e.g., Zippel, 1969; Imai & Garner, 1965) where, rather than predicting spontaneous categorization, the objective is to understand what factors influence categorization performance (e.g., different instructions or stimuli). An example of this is given when investigating whether the structure of the stimuli is made up of integral or separable dimensions and how the number of dimensions used in a task affects classification performance (see for example Handel & Preusser, 1970; Smith & Baron, 1981; Wills & McLaren, 1998). One example of how performance is affected by different conditions, relates to comparing the simultaneous presentation of stimuli vs. a sequential presentation. In simultaneous presentation, the spontaneous classifications between participants are similar, but in sequential presentations, the spontaneous classifications are dependent on the particular sequence of stimulus presentation (Handel & Preusser, 1969). In another example, when the stimuli were composed of separable dimensions, classification was based on a single dimension, but in contrast to this, when integral dimensions were used, classification was based on overall similarity (Handel & Imai, 1972). In a more recent case, Regehr and Brooks (1995; see also Medin, Wattenmaker, & Hampton, 1987) suggested that a single dimension was most frequently used in classification when the constraint of asking participants to classify items into two groups was imposed, rather than having no constraints, such as in typical free sort tasks. However, in Pothos and Close (2008), it is argued that uni-dimensional sorting is not a general constraint, rather it is an artefact of the particular task employed by Ragehr and Brooks (1995). In another example, which uses a more unconstrained method, Compton and Logan (1999) used an arrangement of dots, and examined if the proximity between elements acted as a factor in determining classification results.

The previous research considered above has typically tried to identify manipulations that influence spontaneous categorization performance rather than to actually predict the classification groupings (the Compton and Logan studies are an exception to this). The simplicity model (Pothos & Chater, 2002), Rational model (Anderson, 1991) and SUSTAIN (Love, Medin, & Gureckis, 2004) are three examples of unsupervised categorization models that make predictions on how the classification groupings are made.

As both the research traditions of supervised and unsupervised categorization have complementary explanatory objectives, it is useful to identify similarities and differences between them.

2.2 Exemplar Approach in Supervised Categorization vs. Unsupervised Categorization

Supervised and unsupervised categorization have some similar and some very different aspects. See Chapter 3 for an in depth description of models of supervised categorization. In exemplar models (Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1986), the classification of new items is made based on computing the similarity of this with each training exemplar stored in memory. In an example of how the exemplar model works, if test items are more similar to items in categories 'A' compared to items in categories 'B' or 'C' then classification of the test items into category 'A' will be made.

Supervised vs. Unsupervised categorization:

There are definitional accounts of categorization (e.g., Bruner, Goodman, & Austin, 1956; Katz, 1972; Katz & Fodor, 1963, Pothos & Hahn, 2000) which suggest that categories are characterized by necessary and sufficient features. In exemplar theories (e.g., Brooks, 1987; Hintzman, 1986; Medin & Schafer, 1978; 1989, 1988a, 1988b, 1985) a set of known instances represent the concept, where the assignment of a new instance to a category is made on the basis of similarity to each member. There is also Prototype theory (e.g. Homa & Vosburgh, 1976; Homa, Sterling, & Trepel, 1981; Posner & Keele, 1968; Reed, 1972) where categorization is made on the same basis as exemplar theory except that in this case the central tendency of the group (the prototype) is used rather than each individual exemplar. Also according to general recognition theory (Ashby & Perrin, 1988), intrinsic noise properties of perception and representation explain categorization effects.

The obvious difference between unsupervised and supervised categorization models is that in supervised categorization, there is a pre-specified group for how to categorize the training items, so that in this case the learner must infer the underlying category structure. In the case of an unsupervised task, the learner has no category structure to infer and therefore has to make a classification based on what is most natural and intuitive.

In experimental terms, participants are presented with artificial labels for the training stimuli by the experimenter in supervised categorization. The group labels are learned by the participant before a classification of a new item is made. In the real world, the application of supervised categorization seems relevant in many cases. For instance a case of supervised categorization is when a child is told that a particular item is called an apple, while other items are called oranges. In order for the child to classify correctly new instances of apples and oranges the child must infer from the category structure enough about the concepts "apples" and "oranges". This is different to unsupervised categorization where in this situation we would spontaneously categorize objects without being told the category labels, and which items belong to which category.

It could be claimed that concepts are based upon supervised categorization mechanisms exclusively such as involving the use of linguistic labels. The typical assumption in unsupervised categorization is that boundaries between groups are determined only after seeing enough exemplars of items from each within group. However, children and adults generalize from a small number of examples when learning new words (e.g., Feldman, 1997; Tenenbaum & Xu, 2000). From this, the assumption can be made that there are prior constraints on which categories are plausible, and these constraints may be determined by unsupervised categorization learning. There are also strong commonalities between schemes of categorization between different cultures (e.g., Lopez, Atran, Coley, Medin, & Smith, 1997). Therefore, unsupervised categorization may help in the understanding of how supervised learning occurs.

A crucial difference in the two approaches is that in unsupervised categorization we deal with the problem of what makes a category naturally coherent. Category coherence deals with the question of what makes a category of birds or cups a coherent category but disallows non-sensible categories such as dolphins born on Tuesday.

2.3 The simplicity principle

In 1986, Pomerantz and Kubovy formulated the simplicity principle to describe how the perceptual system sought the simplest rather than the most likely (see Helmholtz, 1962 for the likelihood principle) perceptual organisations which were consistent with the sensory input given. There was much controversy as to whether the perceptual system was governed by the likelihood or simplicity principle (e.g., Pomerantz & Kubovy, 1986). However, Chater (1996) provided a mathematical account, which linked the simplicity and likelihood principles in perceptual organisation using the mathematical theory of Kolmogorov complexity (e.g., Kolmogorov, 1965). This account provided evidence that the two theories were not in competition with one another, but instead were identical (at least when accounting for perceptual organisation).

With the controversy partly alleviated, the simplicity principle has been applied to explain how the cognitive system imposes patterns on the world. As the world is highly patterned, the cognitive system has presumably evolved to successfully find these patterns. The simplicity principle achieves two criteria: (1) It is normatively justified; (2) It appears descriptively correct. Normative justification refers to the requirement of the principle to be consistent with theoretical arguments. In this case evidence for this is presented in the formulation of 'Occam's razor' (William of Ockham, 1285-1349) and also in early positive epistemology (e.g., Mach, 1883/ 1960) and remains a standard principle in modern philosophy of science (e.g. Sober, 1975). In addition, over the past thirty years the theory of simplicity 'Kolmogorov complexity' has been developed and applied in mathematics (Chaitin, 1966, Kolmogorov, 1965, Solomonoff, 1964), in statistics (Rissanen, 1987, 1989; Wallace & Freeman, 1987), and computer science (Quinlan & Rivest, 1989; Wallace & Boulton, 1968). This evidence gives the rigorous normative justification for the simplicity principle, which suggests that the simplest account for some data leads to the best theory for the data. Regarding (2), evidence for being descriptively correct refers to whether the theory explains specific evidence accurately. The simplicity principle in this case appears descriptively correct as demonstrated in the examples in: Mach (1959/1886), Gestalt psychology (Koffka, 1962/1935) and in information processing research in perception (Buffart, Leeuwenberg & Restle, 1981; Garner, 1962, 1974; Hochberg and McAllister, 1953; Leeuwenberg, 1969, 1971; Leeuwenberg & Boselie, 1988) and in the simplicity model, Pothos and Chater (2002).

For a more thorough example of how the simplicity principle is descriptively correct we can take the example from Gestalt psychology (Koffka, 1962/1935). More specifically, we can consider the Gestalt law of good continuation which states how the cognitive system completes visual patterns when part of the visual pattern is occluded. In figure 1 (a) the vertical bar is perceived as occluding the upper left and right horizontal lines, therefore the two upper left and right horizontal lines are perceived by people as a single line as in figure 1 (b), although it could have any form as in Figure 1 (c). The simplicity principle, predicts a preference for the straight line. This is because it is more simple as there would be a shorter codelength to describe a continuation of the same pattern, as compared to altering a pattern.

When referring to the codelength of information, and the simplest codelength to describe perceived information patterns, we are referring to the measurement of information as introduced by Shannon (1948). One bit of information is the smallest piece of quantifiable information, and is a single binary decision. In categorization (e.g., Pothos & Chater, 2002), codelength of categories are computed using the simplicity principle (this will be explained in more depth in this chapter).

The simplicity principle is consistent with the Gestalt law of good continuation. In the case of the lower left and right horizontal lines of Figure 1 (a), this is perceived as two separate lines. This is consistent with both the simplicity principle and the Gestalt law of good continuation. In the case of the simplicity principle, the deviation of the two lower horizontal lines allows the minimal description to account for a possible disappearing of the hidden line. Therefore, the advantage of the simplicity principle is that it can postulate that the hidden line disappears. This means that the hidden line is not perceived, but can continue. When a naïve observer is presented with Figure 1 (d) this is perceived as a cross, occluded by a circle, and illustrated in Figure 1 (e). The Gestalt law of good continuation fails to account for this and leads to an interpretation of Figure 1 (f). Simplicity principle accounts for the illustration of Figure 1 (e) as this form is simpler (it requires less codelength to describe, as the deviation requires a greater codelength of description) than the more complex (greater codelength description) of the irregular Figure 1 (f).

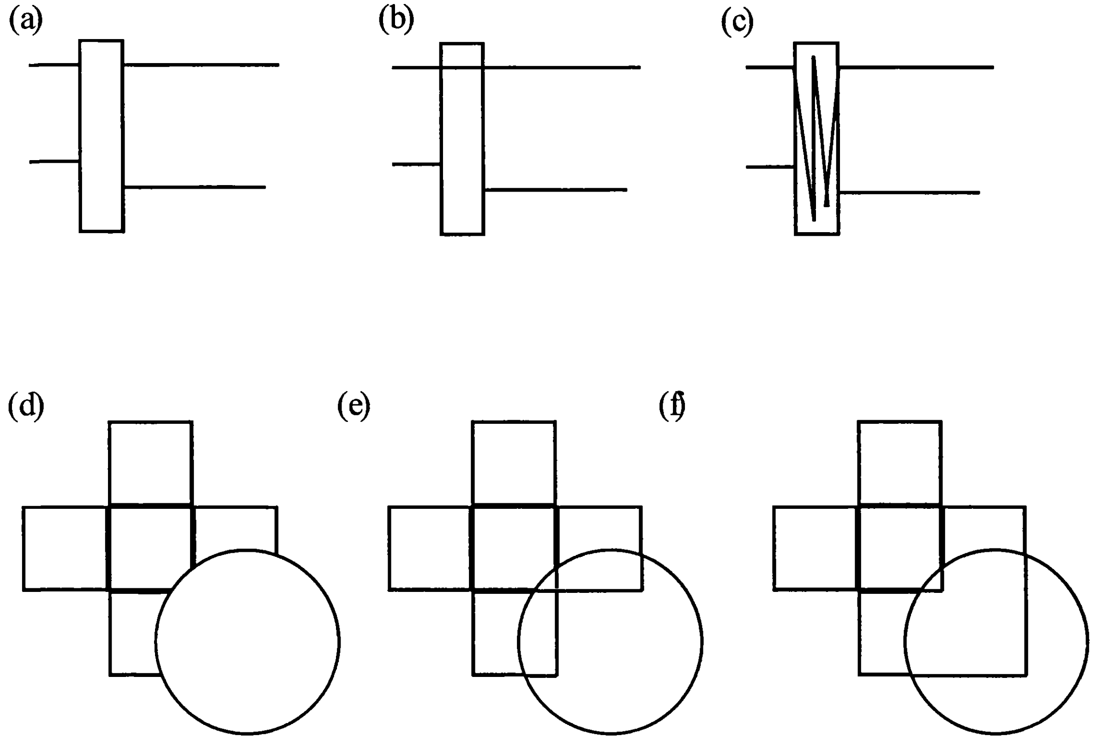


Figure 1: Simplicity in filling in occluded objects.

In addition to the evidence given regarding the preference for simpler perceptual organisations, a simple mathematical illustration in favour of simplicity can be given which supports its justification. The justification of this is given using Bayes's theorem, which states:

$$P(H \mid D) \propto P(D \mid H)P(H) \qquad (1)$$

The theorem states that the probability of a hypothesis given the data is proportional to the product of the probability of the data given the hypothesis and the prior probability of the hypothesis without the data. The H that maximizes (1) is the same as the H that minimizes (2).

$$-\log_2 P(D \mid H) - \log_2 P(H) \qquad (2)$$

Formula (2) uses Shannon's information theory for the specification of the optimal code for describing quantities such as the data, hypotheses, etc., where the optimal code minimizes the average codelength. Event x with probability $P(x)$ has the codelength $-\log_2 P(x)$. Formula (2) therefore gives the codelength for D in terms of H plus the codelength for H without D. From formulas (1) and (2) it can be seen that the most probable hypothesis is also the formula which is the simplest (i.e., is encoded with the shortest codelength). Given that both of these approaches equate to one another the general simplicity principle statement that 'when all things are equal then the simplest explanation is likely to be true can be seen as reasonable.

2.4 Measuring Simplicity

The simplicity principle predicts that the simplest possible explanation to fit the data is often the best (Chater, 1996). When using such an approach, this could lead to the prediction that a distal scene should be uniform, however the organisation must be consistent with the sensory input and this is usually non-uniform. It is important to note that the simplicity principle predicts that the cognitive system should capture the regularities in the available information to maximise descriptive power. One question is whether the consistency with the input (capturing the regularities of the information) can be traded against the simplicity of the interpretation. Again, perceptual organisation must capture the regularities in the sensory input, so the compression in the information must be compatible with the regularities in the data. If we were to ignore this point, then the simplest of explanations would be to state "anything can happen" or "group all items together as a single group" which would be completely useless as a cognitive strategy for capturing the patterns in the world. Harman (1965) suggested that the simplicity of a theory must be traded against explanatory power. However, these two factors must be stated in more specific and formalized terms in order for them to be useful as a model for category learning. The way to proceed according to Chater (1996) is to view perceptual organisation as a way of encoding information, so that the perceptual organisation which provides the simplest of encoding, is chosen. This prevents overly simple organisations that do not account for the regularities of the information to become present. This is because these encodings do not help the encoding, or the explanation of the information. Maximising explanatory power but also maximizing the simplicity in encoding are both crucially desirable in accordance with the simplicity principle (the optimal state is maximum explanatory power and minimum description). If the perceptual organisation fails to capture the regularities then it cannot provide a brief description of the data accurately, and is therefore useless. A useful example, which demonstrates this problem, is given in the Richard-Berry paradox (see Li & Vitanyi, 1997), which suggests there is a paradoxical problem when generating the following statement:

"the smallest natural number that cannot be uniquely specified in less than twenty words of English" (1)

The problem here is that out of the infinite number of numbers, the smallest number N that cannot be specified in less than twenty words can be specified with the description above (1), which contains only 16 words, and hence the paradox is clear. Kolmogorov complexity avoids this problem by specifying that the description given must construct the object. Therefore, the Kolmogorov complexity of object $K(x)$ is the length of the shortest description that generates x rather than an overly general description that does not actually generate the object directly.

The measurement of simplicity has been studied extensively in philosophy, for example by Sober (1975), who suggested that no quantitative measure of simplicity has ever been universally accepted. It has also been discussed in psychology, by Attneave (1959), who suggested that the perceptual system prefers short descriptions, and been referred to as an important goal by Atick and Redlich, (1990. It is best discussed in the context of mathematics and computer science, such as in Kolmogorov

Complexity theory, which shows that identifying simplicity with brevity provides a rigorous theory of simplicity (see Kolmogorov, 1965).

Brevity of encoding can become operational using two approaches. Shannon's (1948) information theory (Attneave, 1959; Garner, 1962) and coding theory (Simon, 1972) structural information theory is one elaboration of this (Buffart, Leeuwenberg & Restle, 1981). We now consider the quantification of brevity.

Information theory and brevity:

Brevity is quantified in terms of the number of bits required to distinguish the stimulus from an information source, which has a mutually exclusive range of alternatives.

The formula for this is given as:

$$I(A_i) = \log_2\left(\frac{1}{P(A_i)}\right). \qquad (1)$$

In this equation, each alternative A_i in an information source A has the probability of occurrence $P(A_i)$. $I(A_i)$ represents the amount of information associated with the choice of a particular alternative, A_i, and is called the surprisal (surprisal can be viewed as a measure of brevity in codelength) of A_i.

$$H(A) = \sum_j P(A_j) I(A_j). \qquad (2)$$

$H(A)$ is the entropy, and is the average surprisal of source A. It is the surprisal of each alternative, weighted by its probability of occurrence.

Information theory allows surprisal to be viewed as a measure of brevity. When choosing a sequence of alternatives according to the probabilities of the information source, these can be encoded in a binary sequence. The encoding gives each A_i an individual code word in the form of a sequence of binary digits (e.g. 001101). Sequences of alternatives can be concatenated into a single binary code. In accordance to the idea of brevity, the binary string that describes the alternatives is minimized as much as possible. The product of the sequence length and the average length of the code words within the sequence gives the length of the sequence code. One important implication here is that the average code of words should be minimized. If the binary string of length l_i gives the codelength for alternative A_i, then the specification for the average code word length for source A is given by:

$$\sum_j P(A_J)l_j \ . \qquad (3)$$

There are however some limitations of information theory in some contexts, for example when applied to individual perceptual stimuli. An example from Leeuwenberg and Boselie (1988) involves a stimulus consisting of three letters 'aaabbbbbgg'. If we assume that there is an equal chance ($\frac{1}{3}$) of choosing one of these letters (a, b or g) then the information associated with this specification for example 'a' is $\log_2 (1/\frac{1}{3}) = \log_2 (3)$ bits of information. To specify the entire 10 letter sequence is 10 $\log_2 (3)$ bits because in this case the probabilities of each item being chosen are the same for each letter. In a different situation, where for example 'b' is chosen with probability ½ and a and 'g' with probabilities ¼ then 'b's can be specified with $\log_2 (1/\frac{1}{2}) = \log_2 (2) = 1$ bit whilst the 'a's and 'g's can be specified with $(1/\frac{1}{4}) = \log_2 (4) = 2$ bits which total 15 bits of information for the entire sequence. Having more variation in the set, such as including the entire alphabet would lead, to more information required to specify it. Information theory measures the information in the stimulus relative to the probabilities of the other stimuli. This is useful in experimental settings where the range of possibilities is limited (e.g. Garner, 1962). However, in natural perception, the range of possibilities to define the stimuli can be greater, and therefore this scheme does not provide a useful measure of brevity in encoding stimuli (Garner 1962). Another problem with information theory is that it only states the number of bits required to specify the stimuli and not the best (most meaningful) code. It is both the nature and length of the code that is useful in understanding perceptual organisation (Garner, 1974). A meaningful encoding tells us something about the actual features of the stimuli whilst a meaningless encoding randomly ascribes the code without consideration of the features.

Coding theory and brevity:

Because of the problems with information theory, i.e., the fact that sometimes a meaningless code is ascribed to the sequences, a different approach has been sought to measure brevity, which allows featural detail to be encoded in the stimuli. The encoding of the organisations within the stimuli (i.e., the featural detail) is what Simon (1972) calls pattern languages. The shortest description of the expressed pattern language is the preferred organisation. It is constrained by the number of symbols in the description (e.g., Simon, 1972) and the number of parameters (e.g., Leeuwenberg, 1969). An example of bad code would be aaabbbbbgg which requires 10 parameters, whilst an example of good code would be 3(a)5(b)2(g) which requires just 6 parameters and hence economy is achieved. The problems with short description length is that (a) a new description language needs to be created for each perceptual stimulus, and (b) the prediction of the theory depend on the description language chosen, however Simon (1972) noted that description languages are highly correlated in their description lengths. Kolmogorov complexity generalizes coding theory and addresses these issues.

From the simplicity principle that suggests that simple explanations that fit the data are often the best, a formulation of a much more specific simplicity model (Pothos & Chater, 2002) was proposed about how people spontaneously categorize stimuli in their environment.

2.5 The Simplicity Model of Unsupervised Categorization

Unsupervised categorization and category coherence

The simplicity model is designed to capture category coherence (see Figure 2 for an illustration of category coherence, i.e., greater intuitiveness) in a stimulus set (it is useful in free sort tasks of unsupervised categorization), and it assumes that there are no constraints on how the stimuli should be classified. Several theories have been suggested which explain what constitutes category coherence. One theory is that some categories are grouped together through a common function they share (such as corkscrews having the function of opening bottles) rather than appearance such as size, colour etc. (Barsalou, 1985). In contrast to this, other explanations suggest that categories contain items that are judged to be similar to each other (Rosch, 1975; Wittgenstein, 1957; see also Goodman, 1972, and Quine, 1977), the simplicity model (Pothos & Chater, 2002) is one example of a model that uses similarity in classification.

Murphy and Medin (1985; see also Gentner & Brem, 1999; Lakoff, 1987; Medin & Wattenmaker, 1997) proposed the dominant theory of category coherence, according to which a concept is an element of people's naive theories about the world. This means that category coherence is not based on any specific piece of information, but rather on meaning in our general life. For example, regarding the concept of water, coherence is not based on its chemical structure, but rather on its meaning in our general life. For example, general knowledge could include information that tap water comes from reservoirs; it is wet and can soak our clothes etc. Gelman and Wellman (1991), provide support for this idea by demonstrating that young children generalize on the basis of theoretical knowledge rather than physical similarity. An example of this is in the case of categorizing a worm, a person and, a toy monkey; the worm and the person were deemed more similar because both share biological properties.

The work carried out by Murphy and Medin (1985; also Medin & Wattenmaker, 1997) provides compelling arguments for why a model of conceptual coherence cannot be based on similarity alone. The Simplicity model (Pothos & Chater, 2002) uses similarity information in its account of unsupervised categorization, but in principle could be extended to include background general information relating to particular classifications. The formalization of general knowledge has been shown to be very difficult (Dreyfus & Dreyfus, 1986; Heit, 1997; Heit & Bot, 1999; McDermott, 1987; Oaksford & Chater, 1991, 1998; Pickering & Chater, 1995). In the case of the present experimental work, stimuli that are novel and abstract are used as this avoids the problem of formalising general knowledge.

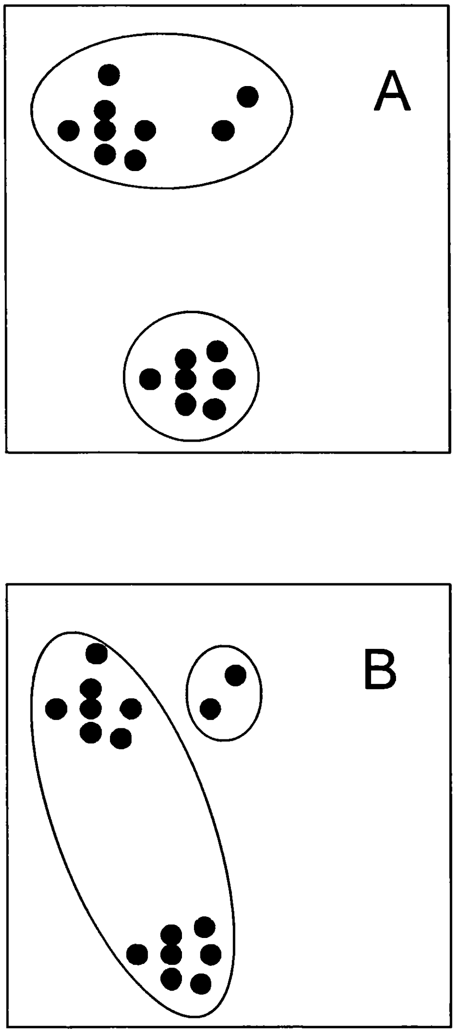


Figure 2: A simple arrangement of points in a Euclidean space. Classifications A should be more intuitive (indicating greater category coherence) for naïve observers than classification B, since it involves more cohesive clusters.

Another aspect of categorization is that of Basic level categories which identifies information according to a hierarchy where classification of new items must fit the definition of a category at its 'basic level'.

Basic level categories and unsupervised categorization:

Basic level categories deal with the explanation of a 'basic' level for categorization, which is a general (basic) category label in a hierarchy of categories which can be more specific higher up the hierarchy (Rosch & Mervis, 1975). One hierarchy could include; 'Scottish highland terrier, a terrier, a dog, an animal, a living thing' etc. The default (or basic) level of categorization, for example, for a dog called 'Fido', would be that it is a dog rather than an animal or living thing. There is a wide body of evidence supporting this argument. One example of this is where basic level categories lead to more rapid picture naming, in comparison to the superordinate or subordinate categories (Rosch, Mervis, Gray, Johnson, & Boyles-Braem, 1976). There is also evidence that suggests basic level categorization is used in naming and other category related behaviour in children (Mervis & Crisafi, 1982; Horton & Markman, 1980).

The relation between unsupervised categorization and basic level categorization can be seen if one assumes that the basic level of categorization is the category level that is most coherent, and explaining category coherence is the ultimate goal of unsupervised categorization tasks. Basic level categories have been modelled computationally (e.g., Corter & Gluck, 1992; Gluck & Corter, 1985; Gosselin & Schyns, 1997). However, basic level and unsupervised categorization do have different predictive scopes. In basic level categorization the predictive objective is to identify the basic level category from a hierarchy of three or four category levels. There is no attempt to predict the exact way in which items are partitioned within the basic level. In unsupervised categorization the aim is to identify the preferred classification (the classification which has the minimal descriptive length, if one adopts the simplicity model, see later) amongst all possible classifications for a particular data set. Another important difference is that basic level categorization is based on featural representations (e.g., a dog has several known features such as a tail, a snout, paws, etc) of objects but in unsupervised classification the items are novel and therefore cannot be typically expressed in terms of features. In unsupervised categorization, features such as short vs. long or differences in shades of colour can be used but this does not include the complex background information that is found in basic level of categorization. Because of this, it is difficult or impossible to identify features. The advantage of the simplicity model is that it can be used to compute preferred classifications on the basis of features or independent of them whereas models for the basic level categorization are restricted to feature based categorization.

From perception to unsupervised categorization:

When confronted with an unfamiliar scene, the information can often be organised into different kinds of groups. This can be viewed as a process of perceptual organisation, whereby sometimes we identify groups in the sensory input. It is also a process of unsupervised categorization. In order to form a mathematical model of unsupervised categorization theoretical insights from perceptual organisation can be considered, as we have done above. The two processes, 'perceptual organisation' and 'unsupervised categorization', can be considered related in the sense that the perceived structure of a set of objects can lead to the (unsupervised) categorization of these items into groups.

The application of the simplicity principle to unsupervised categorization is made on the assumption that perception is based upon physical similarities (Pomerantz, 1981). Therefore, groupings made in unsupervised categorization should maximize within group similarity and minimize between group similarities. Using this assumption, we can view categorization as imposing default constraints on the similarity relations between a set of to-be-categorised stimuli.

The simplicity model of unsupervised classification:

The first step in considering how the simplicity principle can be applied to grouping items into categories is to specify the data and hypotheses (a hypothesis corresponds to a possible grouping of the items). An assumption is made that the information about the similarity structure of the items corresponds to the data. The codelength required to specify the similarity structure (from standard information theory) from the objects in terms of a particular grouping is the sum of the:

codelength to specify similarity in terms of the grouping + the codelength to specify the groupings. (1)

There is a unique codelength for each possible grouping. According to the simplicity principle (e.g., see Rissanen, 1978) there is a preference for the grouping with the shortest (most compressed) codelength. The specification of the simplicity model is made in such a way that the similarity structure with the most reduction in codelength is chosen.

The form of the data:

In categorization research, there have been many kinds of representation assumptions. It is assumed that items can be embedded in a multi-dimensional space in spatial models of representation (e.g., Nosofsky, 1985; Shepard, 1980, 1987), and that similarities are negatively monotonically related to distances in such a space. An adherence to metric axioms is implied by such spatial models of representation but in some situations similarity information violates the metric axioms (e.g., Bowdle & Gentner, 1997; Tversky, 1977, see Nosofsky, 1991). A representation of objects in terms of features is an alternative to this. In this case, similarity is a function of the degree to which features are shared between items, as the items correspond to bundles of features (Tversky, 1977). The problem with features is that in unsupervised classification the use of novel objects with no prior knowledge is common and is often the case that we may not be able to express such objects in terms of features.

It is the perception of the similarity of objects that is important in the simplicity model. Similarity information can be best described in terms of internal spaces, abstract similarity relations or features depending on circumstances. The formulation of the simplicity model is designed in a way to be compatible with different types of representation assumptions. In this way, the difficult and largely irrelevant problem of psychological representation (see Goldstone, 1993; Goodman, 1972; Hampton, 1999; Quine, 1977) is avoided. Such computational principles, which are independent of a representational assumption, have been usefully employed in other areas of cognitive science (e.g., Anderson, 1990; Marr, 1982; van der Helm & Leeuwenberg, 1996, 1999).

The form of information the simplicity model assumes is illustrated in the following: Consider four objects A, B, C and D. A specification of similarity information would be needed to be specified, such that, for example, the similarity of 'A, B' is greater or less than similarity 'C, D'. The formulation of the model is made in a way that similarities are never equal, and obey minimality, such that 'A, A' = 0, and symmetry, so that 'A, B' = similarity 'B, A', but they can violate transitivity. Any combination of the metric axioms can be assumed in the simplicity model and if metric axioms can be assumed for similarity relations then these can be specified with less information. Similarity is used in the implementation of the model whether symmetry and minimality is assumed or violated but there is no reason to suggest that minimality should be violated (Tversky, 1977, discusses some of the considerations that underlie some of the metric axioms). The version of the simplicity model employed in this work was implemented without the assumption of transitivity. The assumption of transitivity does not affect the computation of the codelength in the simplicity model. Transitivity is always obeyed unless the similarity information is collected with a task in which trials have the form: 'is similarity between A, B less or greater than similarity C, D?' Assuming transitivity (and all of the other metric axioms as well) is equivalent to assuming extra constraints, such as 'A, B' > 'C, D' when given 'A, B' > 'B, C' and 'B, C' > 'C, D'. The number of groups and elements in each group determine the number of extra constraints due to transitivity. The extra constraints due to transitivity will be the same where the classifications compared have similar groups, and numbers of elements in each group, so that the assumption of transitivity does not influence the optimal classification.

From information theory it can be assumed that when deciding between two pairs A, B, and C, D, it is a binary choice to compute whether similarity (A, B) is smaller or greater than (C, D), and this is associated with one bit of information to compute. Where we have r items we require $s(s-1)/2$ bits to specify the data directly where there are $s = r(r-1)/2$ similarities between pairs of r items and $s = r(r-1)/2$ comparisons between the similarities for pairs of items.

Fig. 1 shows items represented in a Euclidean space, where distance corresponds to dissimilarity. Here, there are (5 × 4)/2 = 10 distances between these points, which can be expressed as 45 inequalities ((10×9)/2 = 45) such as:

d(a,c) < d(b,c);d(a,c) < d(a,b);d(a,c) < d(d,e);d(a,c) < d(b,d);

d(a,c) < d(b,e);d(a,c) < d(c,d);d(a,c) < d(a,d);d(a,c) < d(c,e);

d(a,c) < d(a,e)...

The regularity in the specifications of the inequalities (the redundancy), means that there may be a shorter description which captures the structure of the data. The simplicity model is one attempt to model the regularities in this structure, creating the largest saving in codelength.

Clustering by simplicity step 1: Coding group:

When computing the codelength required for specifying how r items are allocated into a set of n categories the allocation of all items into all possible classifications needs to be considered. This is given by $\sum_{v=0}^{n}(-1)^{v}((n-v)^{r}/(n-v)!v!)$ (this is Stirling's number, e.g., Graham, Knuth, & Patashnik, 1994; Feller, 1970). Using standard information theory we can assume that $\log_2 (D)$ gives us the codelength required to identify one out of D possibilities (with the assumption that each one is equally probable). Therefore, $\log_2$ $\sum_{v=0}^{n}(-1)^{v}((n-v)^{r}/(n-v)!v!)$ gives a codelength, which specifies the allocation of r

items into groups. This, however, represents only a minor contribution to the overall computation.

In general, certain category structures are more likely to be chosen than others. For example, a category structure that consist of clusters equal to the number of items divided by two, is more likely than for example a classification where each item is in its own cluster or if all the items are clustered into a single group. The computations made in the simplicity model are based on the probability of different category structures which are consistent with the simplicity approach. Pothos and Chater (2002) suggest that in future work the model could identify constraints regarding the likelihood of different category structures, and this could be in the form of a non-uniform, prior probability distribution over category structures. They also suggest that general knowledge effects could be introduced here, as some groupings using general knowledge would be more plausible and therefore more likely. Such a case could include groupings based on biological vs. non-biological kinds, and thus could reduce the codelength. Where there is no general knowledge, i.e., in a case of novel items, the codelength for the classifications can be computed as above (see also Pomerantz, 1981).

Clustering by similarity step 2: specifying the data in terms of groups:

When encoding similarity data, the definition of a cluster (or category) is that it is a collection of objects where the within cluster similarities are greater (which should be as great as possible) than the between cluster similarities (which should be as low as possible; Rosch & Mervis, 1975). Default constraints on the similarities between items are therefore introduced by a particular grouping. If the constraints are strong (i.e., many comparisons between distances are explained by them), and generally correct (i.e., there are no corrections to the constraints) then the first term in (1) is reduced.

The description of similarity inequalities that are not specified by the grouping is needed, such as between two within cluster similarities or between two between cluster similarities. If there are t of these, then the codelength will be t bits.

If there are u constraints, where e are incorrect, the encoding of e must be between 0 and u, so the encoding of u requires a binary code of length $\log_2 (u+1)$ bits. Identifying the constraints e out of u constraints is done with standard combinatorics: ${}_uC_e = (u!/e!(u-e)!)$ ways to choose e items from set u. The total code for correcting erroneous constraints, E, is $\log_2(u+1) + \log_2({}_uC_e)$.

In order to specify the errors when there are few or very many errors, a short codelength is needed. In the case of the errors, having half the number of constraints requires the greatest codelength. Pothos and Chater (2002) suggest that the number of errors should be less than half the number of constraints, and where this is not the case then no clustering should be defined, as in this case the clustering would be of dissimilar items. This additional assumption is mild, as any reasonable algorithm for finding clusters should use similarity.

The simplicity model, a summary:

Pairwise similarity inequalities between pairs of objects are the representation of the similarity structure in the simplicity model. The number of inequalities needed to be specified is reduced with use of categories. The disadvantage of using categories is that they require a codelength to describe the particular set of categories used and correct for any errors in the constraints. Using categories usually shortens the description of the similarity structure of the items, and the greater the simplification the more intuitive the category structure is predicted to appear. The simplicity model is evaluated in the experiments assessing naive observers' unsupervised categorization performance.

2. 6 Other Unsupervised Models vs. the Simplicity Model

Some of the early research into unsupervised learning helps illustrate the distinctiveness of the simplicity model. One of these early studies of unsupervised learning comes from Fried and Holyoak (1984), where categories are described in terms of density functions. They suggest that participants can infer the actual density function from a sample of exemplars presented to them. An assumption is needed that category distributions have a particular form (an illustration of Fried and Holyoak's theory is made with normally distributed category distributions). They also suggested that an external specification must be made of the number of categories sought. The difference between this and the simplicity model is that here we have a situation where learners know a priori that the category exemplars have properties whose range conforms to a certain distribution; another difference is that this model requires advanced knowledge of the number of categories. This corresponds to a situation where, for example, a bird expert has to identify bird categories in a new domain. The simplicity model, by contrast, does not make any assumption about the parametric properties of the categories or the number of categories sought.

AutoClass

Cheeseman and Stutz (1995) provided a model for unsupervised categorization called AutoClass, which comes from the machine learning literature. The model consists of two components, the first of which is a probability distribution which specifies how items belong to different categories with different probabilities, as opposed to being assigned to any particular category. There is also a probability density function. This is for the distribution of the attributes of the objects that belong to the category, which constitutes the second component. Attributes can be distributed in several ways in AutoClass, as it can model many types of attribute distribution within categories and category distributions, and is not restricted to one type of probability distribution, which is unlike Fried and Holyoak's model. However, the range of probability density functions AutoClass can employ determines the modelling scope in the AutoClass version used. This is different from the simplicity model, as here a particular distribution for categories or category attributes is not assumed, therefore it is more similar to Fried and Holyoak's (1984) model.

There are several related Bayesian approaches to unsupervised learning. Some of these do not use a specification on the number of categories sought. Ghahramani and Beal (2000) used a factor analysis procedure within the Bayesian framework to determine the number of factors required to model the data automatically (when considering the number of instances associated with each factor as belonging to the same cluster then factor analysis is similar to the process of clustering). Components are rejected from the model under particular circumstances in order for this to be achieved. In contrast to the non-parametric method used in the simplicity model, Ghahramani and Beal's (2000) computations use a Gaussian function to model the distribution of information.

CODE

The CODE model by van Oeffelen and Vos (1982, 1983) later modified by Compton and Logan (1993, 1999), is another model in which the classification of objects is guided by parametric features. It deals with the perceptual grouping of dot patterns but presumably could be used for the classification of more complex elements. In this model, a value of strength is associated with each element in a pattern that originates from the element. Group allocation is made on the basis of when the pattern of strengths from the different elements at a location, when added, are above a certain threshold. As with AutoClass, the determined classifications are predicted on the basis of the strength spread. It has a parameter which is a fixed threshold, so a single classification was predicted for a set of objects in the original formulation of CODE. The model was later adapted so that it could produce from a set of objects nested classifications (Compton & Logan, 1993, 1999).

Ahn and Medin (1992) produced a two-state model of category construction for free sort classification. The model's primary use was to evaluate the relative compellingness of a hypothesis where overall family resemblance drove the spontaneous groupings rather than sorting via a single dimension (this issue has been considered extensively in the free sort classification literature). The prediction made by this model, was that there would be as many groups as there are featural values along a dimension (but there was no attempt to predict the most salient dimension).

Kohonen neural network

Schyns (1991) proposed an unsupervised model of classification. This model used a two module neural network to investigate the spontaneous discovery of categories, and the association of these categories with labels. The Kohonen neural network was used to reduce the high dimensionality input vectors to lower dimensionality (two dimensions) output vectors. This was used to find how categories were spontaneously discovered. The segregation of the output space into distinct regions that can be identified with categories can be made by the Kohonen neural network. The similarity structure of the input distances determines the segregation into distinct regions and therefore is a spontaneous classification, rather than being determined by an external constraint. One limitation of such a model as compared to the simplicity model is that a specification of the number of categories is needed, in advance, in order to classify the information.

The rational model

The rational model is the only unsupervised categorization model which is not explicitly based on similarity (although in practice its predictions appear to converge with those of similarity-based models; Pothos, 2007). The rational model is an incremental, Bayesian model of categorization (cf. Tenenbaum & Griffiths, 2001), which classifies a novel instance in the category which is most likely given the feature structure of the instance (Anderson, 1991). For example, I would classify a novel instance in the category of 'cats', because its particular features ('meows' , 'has fur', 'has four legs', 'can purr') are particularly likely given membership to this category. This approach is analogous to the various category utility proposals in categorization, according to which categories are useful to the extent that they can be used to predict the features of their members (and vice versa; e.g., Corter & Gluck, 1992; Gosselin & Schyns, 2001; Jones, 1983; Medin, 1983; Murphy, 1982).

We briefly describe the version of the continuous version of the model described by Anderson (1991). The probability of classification of a novel instance into category *k* depends on the product $P(k)P(F|k)$, whereby $P(k) = \frac{cn_k}{(1-c)+cn}$. In this equation, n_k is the number of stimuli assigned to category *k* so far, *n* is the total number of classified stimuli, and *c* is the coupling parameter. The probability that a new object comes from a new category is given by $P(0) = \frac{1-c}{(1-c)+cn}$. Therefore, lower values of the coupling parameter will lead to the creation of more new categories and, so, the coupling parameters determines the number of categories which will be produced in classifying a set of stimuli. Also, $P(F|k) = \prod_i f_i(x|k)$, where *i* indexes the different dimensions of variation of the stimuli and *x* indicates the different values dimension *i* can take. Note that in this (the original) version of the rational model, feature values are assumed to be independent. Each $f_i(x|k)$ term corresponds to the probability of displaying value *x* on dimension *i* in category *k*, and is given by $t_{a_i}(\mu_i, \sigma_i\sqrt{1+\frac{1}{\lambda_i}})$, which is the *t* distribution with a_i degrees of freedom; $\mu_i = \frac{\lambda_0\mu_0+n\bar{y}}{\lambda_0+n}$ and $\sigma_i^2 = \frac{a_0\sigma_0^2+(n-1)s^2+\frac{\lambda_0 n}{\lambda_0+n}(\mu_0-\bar{y})^2}{a_0+n}$. In these equations, $\lambda_i = \lambda_0 + n$, $a_i = a_0 + n$, *n* is the number of observations in category *k*, $\bar{y}$ is their mean along dimension *i*, and s^2 is their variance. Finally, $a_0 = 1 = \lambda_0$, μ_0 can be set as the halfway point of the range of all instances, and σ_0 as the square of a quarter of the range.

The primary function of the rational model is to predict the optimal classification for a set of stimuli. This optimal classification will depend on the order of presentation of the stimuli.

COBWEB

There are also differences between the simplicity model and Fisher's COBWEB system (Fisher, 1987, 1996; Fisher & Langley, 1990; Gennari, Langley, & Fisher, 1989; Gennari, 1991). Corter & Gluck's (1992) category utility is used as a measure to examine what is special about basic level categories. With the use of category utility, COBWEB can predict how items should be divided amongst clusters and how many clusters there should be. It is difficult to compare COBWEB with the simplicity model as COBWEB is used for understanding basic level categorization and the relation between this and the aspect of spontaneous categorization that the simplicity model addresses is not clear; this requires further work.

Statistics and data mining approaches have also been extensively used in the study of unsupervised categorization (e.g. Arabie, Hubert, & de Soete, 1996; Fisher, Pazzani, & Langley, 1991; Everitt, 1993; Hartigan, 1975; Krzanowski & Marriott, 1995). Hierarchical agglomerative cluster analysis is one important line of research here (e.g., Jardine & Sibson, 1971), where all items are assumed to be individual clusters in the first step of analysis. In the next step an all-inclusive category is created by combining items into a single cluster two at a time. Regardless of the algorithm used this procedure results in n - 1 groups for n items. In another approach of clustering, K-means clustering, items are grouped into K categories, which involve optimizing an explicit criterion (where K is determined by the investigator; Banfield & Basil, 1977; Duda & Hart, 1973; MacQueen, 1967). The criterion (the objective function) can be viewed as a measure of category cohesiveness. When given a set of items, the criterion selected determines the discrete (non-hierarchical) set of groups.

A statistical clustering model called CLUSTER/2 (Michalski and Stepp, 1983) uses simplicity of verbal description of the categories created as one of the determinants of classification goodness (see also Ahn & Medin, 1992; Medin, Wattenmakker, & Michalski, 1987b). When dealing with several different kinds of datasets then statistical clustering may have an advantage with this flexibility, but this is less so in cognitive modelling where the number of free parameters relative to the degrees of freedom in the data needs to be watched.

SUSTAIN

SUSTAIN is an adaptive model of category acquisition, aiming to capture both supervised and unsupervised categorization in the same framework (see also Gureckis & Love, 2003). The internal representations in the model take the form of clusters, which capture psychologically meaningful sub-groupings of items. For example, when learning about categories of birds, a single cluster in the model might represent highly similar species such as robins and blue-jays separate from highly dissimilar examples such as ostriches. SUSTAIN is initially directed towards classifications involving as few clusters as possible, and only adds complexity as needed to explain the structure of a category. Two key aspects of SUSTAIN's account are the role of similarity and surprise in directing category discovery. First, SUSTAIN favors clusters organized around perceptually or psychologically similar items. Second, new clusters are created in memory when the existing ones do a poor job of accommodating a new instance. Thus, SUSTAIN adjusts its category representations in a trial-by-trial fashion to accommodate the similarity structure of the items it has experienced.

When a to-be-categorized item is first presented to the model, it activates each existing cluster in memory, in a way based on the similarity of the item to each cluster. In addition, learned attention weights in the model can bias this activation in favor of dimensions which are more predictive for categorization. Clusters that are more activated are more likely to be selected as the "winner" for the item. If there are many highly activated clusters for a particular item, then confidence in the winning cluster is reduced—i.e., there is cluster competition (regulated by a parameter). In the unsupervised learning situations considered here, if the current input item fails to activate any existing cluster above some threshold level, then a new cluster is created for the item. This is the key mechanism of 'surprise' in SUSTAIN: new clusters are created in response to surprisingly novel stimuli that do not fit with existing knowledge structures. The threshold parameter (τ) controls what level of activation is considered 'surprising' enough, so that this parameter determines the number of clusters the model creates (analogous to the coupling parameter in the rational model; Anderson, 1991).

Given that SUSTAIN is a trial-by-trial learning model, in modeling free sorting task where multiple items are simultaneously presented, SUSTAIN's fits are derived by running the model thousands of times on different stimulus orderings in order to create a distribution of plausible classifications: more psychologically intuitive classifications are considered to be the ones more frequently generated.

The unsupervised GCM

The unsupervised GCM (Pothos & Bailey, 2009) is a straightforward modification in the application of the standard GCM (Nosofsky, 1991). The objective of the standard GCM is to predict the classification probabilities of new stimuli, relative to two or more pre-trained categories. For example, suppose that participants have been taught in a training phase to associate some stimuli in category A and some other stimuli in category B. Then, the GCM-predicted probability of a category A response given a new stimulus X is: $P(A|X) = \frac{\beta_A \eta_{XA}}{\beta_A \eta_{XA} + \beta_B \eta_{XB}}$, whereby $\eta_{XA} = \sum_{j \epsilon A} exp\left\{-c\left[\left(\sum_{k=1}^{D} w_k |y_{xk} - y_{jk}|^r\right)^{1/r}\right]^q\right\}$. The β terms are category biases, η_{XA} is the sum of similarities between X and all the A exemplars, c is a sensitivity parameter, r is a Minkowski distance metric parameter, q determines the shape of the similarity function, w_k are dimensional attention weights, and y's are item coordinates (it is assumed that stimuli are represented in a putative psychological space). The input to the GCM consists of the coordinates of a set of training stimuli, information about the assignment of the stimuli to categories, and the coordinates of a set of test stimuli. Behavioral data are typically fit by adjusting GCM parameters until the classification probability GCM predicts for a test stimuli X is as close as possible to the empirically observed one. An error term for the GCM can be computed as $\sum(O_i - P_i)^2$, whereby O_i are the observed probabilities and P_i are the probabilities predicted from the model.

In an unsupervised context, instead of classifying test stimuli relative to a set of training items, we consider the relative coherence of alternative partitions of a set of stimuli, where coherence means that the classification of each stimulus is predictable given the classification of the other items. Suppose we are interested in evaluating a classification for a set of stimuli, {1 2 3}{4 5 6 7 8 9} (the numbers '1', '2' etc. are stimulus ids). We can consider each item in turn as a test item whose classification is to be predicted, and all the other items as training items whose classification is given. GCM parameters are adjusted until the predicted classification probabilities for individual 'test' items are as close as possible to 100% for the classification of interest. For example, the O_i for classifying stimulus '1' into category {2 3} would be 100%, the O_i for classifying stimulus '2' into category {1 3} would be 100%, etc. In other words, stimuli are assigned to categories in accordance with the category structure being evaluated and GCM fits are computed on this basis. Pothos and Bailey (2009) suggested that the lower the sum of all the corresponding error terms, the more coherent and intuitive a classification is predicted to be, according to the GCM.

In examining a classification, the parameters of the unsupervised GCM are automatically set in a way that the groups in the classification are as separated as possible. For example, for two-dimensional stimuli, if clusters are specified along dimension 1, but there is no classification structure along dimension 2, optimizing the unsupervised GCM will typically produce a high attentional weight for dimension 1 and a low weight for dimension 2. In other words, parameter search in the unsupervised GCM is guided by the particular classification structure examined, *not* by the need to produce specific empirical results.

The unsupervised GCM assumes that all stimuli are presented concurrently. Moreover, at present it can only produce predictions of relative intuitiveness for particular partitions of a set of stimuli; it cannot (yet) be employed to identify the best possible classification for a set of stimuli from scratch.

DIVA

The divergent autoencoder, DIVA (Kurtz, 2007) is an account of human category learning based on the autoencoder connectionist architecture (Rogers & McClelland, 2004). The DIVA model consists of a three-layer, feedforward neural network with a bottleneck hidden layer that is trained auto-associatively using backpropagation. The model operates by recoding the input at the hidden layer and then decoding (reconstructing the original input) in terms of a channel for each category (separate sets of weights connect the hidden layer to sets of output nodes that represent the feature reconstruction for each category). In supervised learning tasks, DIVA produces a construal of the input in terms of each possible category and the relative degree of reconstructive success determines the classification response. The model learns by applying the auto- associative error to adjust the weights only along the channel corresponding to the pre-determined correct category. Psychologically, the model assumes that an example belongs to a category to the extent that it can be reconstructed by the category. A category is basically a flexible representation of the statistical properties of the exemplars. For example, one category can correspond to all items that have value 1 on feature F1, or all items for which F1 and F2 are perfectly correlated, or all items such that feature F1 has value 1 unless features F2 and F3 each have value 0.

In unsupervised learning tasks, the model has no information about which stimuli belong to which category or about the number of categories, so DIVA begins as a standard autoencoder with a single channel. DIVA performs unsupervised learning by evaluating stimuli one at a time. To simulate a spontaneous classification task with all stimuli concurrently available, DIVA is trained on blocks of all stimuli presented one at a time in a random order. DIVA evaluates each stimulus by determining the reconstructive success of all existing category channels (on the initial trial, there is only one category channel and no evaluation process). A spawning threshold is used to determine whether any of the existing categories provide a satisfactory account of the stimulus (i.e., sufficiently low sum-squared error). This threshold is analogous to the coupling parameter in the rational model or the τparameter of SUSTAIN and it effectively determines the number of categories or clusters. If none of the existing categories meet the threshold, then the network architecture is altered: a new category channel is created and seeded by conducting one training trial with the current stimulus. After the evaluation of a stimulus, one self-supervised (input = target) training trial is conducted in which the error signal is applied only to the category channel with the best reconstruction of the current stimulus. Based on this learning procedure, a clustering solution arises in the form of category channels that specialize in reconstructing sets of stimuli with similar properties. See Table 1 for a summary of the key differences and similarities of some of these models.

Table. 1. An examination of how the unsupervised models differ from each other.

	Within cat. sim.[1]	Between cat. sim.[1]	Trial-by-trial	Formal principle
Geometric	Yes	Yes	No	N/A
DIVA	N/A	N/A	Yes	N/A
Rational	N/A	N/A	Yes	Bayes
Simplicity	Yes	Yes	No	Simplicity
SUSTAIN	Yes	No	Yes	Simplicity
Un. GCM	Yes	No	No	N/A

Notes: Within cat. sim. and Between cat. sim. refer to whether the models favor classification which maximize within category similarity and/or between category similarity.

2.7 Summary

This chapter has outlined the simplicity model (Pothos & Chater, 2002) and its theoretical foundation, which is of the simplicity principle (Pomerantz & Kubovy, 1986; such as demonstrated in the simplicity model). It has outlined the similarities and differences of unsupervised and supervised categorization, and compared the simplicity model to other unsupervised models. In the next chapter, I explain another type of categorization, supervised categorization, and how this relates to absolute representation (or judgment).

Chapter 3

Supervised Categorization and Absolute Judgment

3.1 An Introduction

In categorization, there are several theories which attempt to explain how people make classifications, such as unsupervised (see Chapter 2) and supervised categorization (which this chapter explains). These theories hold their own unique perspective of how categories are formed. For example, some focus on rule formation (Nosofsky, Palmeri, & McKinley, 1994); decision boundaries (Maddox & Ashby, 1993); prototype abstraction (Posner & Keele, 1968; Reed, 1972; Smith & Minda, 1988); and exemplar storage (Medin & Schaffer). The focus of the present investigation is to explore relative vs. 'absolute judgment' (or absolute-like representation). Absolute classification is a classification based on the actual physical properties of the items used. So, a judgment based upon absolute properties would be influenced by how physically similar an item is to another item. This is similar to the way that exemplar and prototype theories suggest that classification is made. This chapter explores the exemplar and prototype models for an illustration of what is meant by absolute representation (the terms representation and judgment are used interchangeably).

3.2 Exemplar Models

Exemplar models (e.g., Medin & Schaffer, 1978; Nosofsky, 1986) assume that in categorization, a new item is categorized based on its similarity with existing exemplars (items) in memory. An alternative to this is the distributional approach (Ashby & Townsend, 1986) which suggests that classification of a new exemplar is based on the relative likelihood of belonging to each distribution. These two accounts make qualitatively different predictions. Consider the case of two categories, one of which has high variability and another which has low variability. Exemplar theory will predict that a critical exemplar which is exactly half way between the two categories will be categorized as belonging to the category with low variability. The distributional model predicts that the critical exemplar should be classified into the high-variability category.

Despite the difference in qualitative prediction for categorization, the exemplar model has been successful in accounting for results which have been used in support of other models such as prototype abstraction or rule induction. An example of this, is where prototypes are classified better than exemplars (Homa, 1984). However the exemplar theorists have shown that prototype enhancement effects are predicted well by pure exemplar models (e.g., Busemeyer, Dewey, & Medin, 1984; Hintzman, 1986; Medin & Schaffer, 1978; Nosofsky, 1988, 1991; Shin & Nosofky, 1992).

3.3 Prototype vs. Exemplar Theories of Categorisation

A major controversial issue in the categorization literature, has been whether categorization for new stimuli into existing categories occurs on the basis of comparing the similarity of the individual exemplars, within a group, with the new item (exemplar theory; e.g., Medin & Schaffer, 1978; Nosofsky, 1989; Shin & Nosofsky, 1992), or by comparing the similarity of the average summary representation, of the category with the new item (prototype theory; e.g., Reed, 1972; for general discussions see Nosofsky, 1990; Komatsu, 1992; Ashby & Alfonso-Reese, 1995). According to exemplar theory (e.g., Brooks, 1978; Hintzman, 1986; Medin, 1986; Medin & Schaffer, 1978; Nosofsky, 1989, 1988a, 1988b, 1990, 1991) a person will classify a new item as a member of a category if the new item is more similar to the items in this category as opposed to another. So, in this case, the previous exemplars within the pre-specified groups shapes the way the classification of the new items is made, because, the category structure is given by the experimenter. This is different to spontaneous categorization, which is explored in Chapter 2. Spontaneous categorization is completely unconstrained, and has no existing categories which suggest how the items should be classified (see Chapter 2 on unsupervised categorization).

In contrast to this, prototype theory suggests that when learning a category, the person abstracts a central tendency across all encountered instances of the category (e.g., Rosch & Mervis, 1975; Posner & Keele, 1968, 1970; Homa et al., 1981; Homa & Vosburgh, 1976; Reed, 1972). In some cases, certain restricted types of prototype and exemplar models are equivalent (independent cue models; Nosofsky, 1990; Ashby & Alfonso-Reese, 1995). The controversy relates to which theory best describes conceptual structure (for reviews see Murphy & Medin, 1985; Komatsu, 1992; Hahn & Chater, 1997).

In the investigation for 'relative vs. absolute categorization' (see Chapter 5), I will ask the question: under what circumstances should we expect a relative as opposed to an absolute representation in categorization? Before this can be answered, however, clear definition of the terms absolute-like and relative-like representations must be given. This is the goal of the present chapter, which relates to Chapter 4 on relative judgment. In order to explain absolute representation, some of the models on supervised categorization are described in more detail. The goal here is to give a deeper understanding into exemplar theory, and absolute judgment (representation).

3.4 The Generalized Context Model of Supervised Categorization

The Generalised Context Model (GCM; Nosofsky, 1984, 1986, 1991) has been used successfully to model exemplar (absolute) representation in categorization. This model generalizes the original version of the context model proposed by Medin and Shafer (1978), and integrates this with classic theories and ideas in the area of choice and similarity (Garner, 1974; Shepard, 1958). The model uses multidimensional scaling (MDS) in modelling similarity. Exemplars are represented in multidimensional space, and similarity is a decreasing function of their distance in space.

The GCM assumes that the categorization of a new exemplar is determined by the similarity between that new exemplar and those stored in memory. The GCM sums the similarity of a new item with the items in each category and predicts that the new item will be classified in the category for which this summed similarity is greatest. For example, a new instance will be classified as belonging to category A rather than category B, if it is more similar to the A exemplars than the B exemplars. More specifically, exemplars are represented in a multidimensional space; each exemplar is stored together with its category label. In a simple, one-dimensional case, the distance between two stimuli S_i and S_j is given as:

$$d_{ij} = |x_i - x_j| \tag{1}$$

Where x_i is the absolute magnitude of S_i, and x_j is the absolute magnitude of S_j. For an m-dimensional space, the weighted Minkowski power formula is used, so that the distance between stimuli S_i and S_j is given as:

$$d_{ij} = \left[\sum_m w_m \cdot |x_{im} - x_{jm}|^r \right]^{1/r} \tag{2}$$

In Equation (2), x_{im} denotes the value of exemplar i on psychological dimension m. The r value defines the distance metric of the psychological space. For example, the city block metric is defined with $r = 1$, and the Euclidean distance metric is defined with $r = 2$ (Garner, 1974; Shepard, 1964). Shown in Equation (2) are also the attention weight parameters w_m (Carroll & Wish, 1974), which model the degree to which a participant attends to a particular dimension. The similarity between stimuli S_i and S_j is a function of their distance. Similarity is typically a monotonically decreasing function, of distance as in the equation below:

$$\eta_{ij} = e^{-cd_{ij}^q}, \tag{3}$$

In Equation (3) η_{ij} is the similarity between S_i and S_j ; where q = 1 leads to an exponential function and q = 2 leads to a Gaussian function. The sensitivity parameter, *c*, determines how quickly the similarity between stimuli S_i and S_j is reduced with distance.

The probability of classifying stimulus S_i in category A, is proportional to the similarity between S_i and all the A exemplars, as in Equation (4); in that equation, the β_A parameters are category biases, which indicate whether there might be a prior bias to identify new items as being members of a particular category.

$$H_{iA} = \beta_A \sum_{x_j \in C_A} \eta_{ij} \; , \tag{4}$$

Finally, the actual probability of making a category A response given stimulus S_i, when there are two alternative categories (A and B), is given by Equation (5).

$$P(R_A|S_i) = \frac{H_{iA}}{H_{iA} + H_{iB}} \; , \tag{5}$$

3.5 Other supervised categorization models

COVIS (Ashby et al., 1998).

Ashby et al. (1998) asked participants to learn to classify stimuli into two bivariate normally distributed categories. Ashby et al.'s COVIS (competition between a verbal and implicit system) model suggests that there are two mental systems that compete with each other in the categorization response. It suggests, that first, there is an implicit (nonverbal) system that learns the optimal decision boundary for separating a psychological space into regions corresponding to categories. In categorization, items above the decision boundary would fall into category A, and the items below this criterion would fall into category B. There is also an explicit system that learns verbal rules. The criteria set by the verbal rule are then used in categorization, so that a new item above the criterion (e.g., 10 cm) would be categorized into category A, and an item below this criterion would be categorized into category B. Ashby et al. (1998) suggested that the fact that categorization results fitted the decision boundaries (criteria) as predicted by their model was evidence in support for their model.

ALCOVE (Kruschke, 1992; Nosofsky, Kruschke, & McKinley, 1992).

A model that is closely related to the GCM is ALCOVE (Kruschke, 1992; Nosofsky, Kruschke, & McKinley, 1992), which incorporates the principles of the GCM within a connectionist framework. The advantage of ALCOVE is that it has an explicit mechanism that can learn the attention weights on a trial by trial basis. The mechanism is error driven, and therefore can learn the weights that optimizes performance, rather than the experimenter having to set the weights manually for each stimuli set presented, in the GCM.

RULEX (Nosofsky, Palmeri & McKinley, 1994).

Results from another study which suggest a limitation of exemplar models, was made by Nosofsky, Palmeri and McKinley (1994). They advocated the alternative rule-plus-exception (RULEX) model of classification. From this model, they suggested that categorization is made by forming simple logical rules along single dimensions and then storing occasional exceptions to these rules. For example, if category A consists of features 1112, 1212, 1211, 1121, 2111 (1 could mean, 'it has a feature x', and 2 could mean, 'it does not have feature x') and category B consists of 1122, 2112, 2221, 2222 then the logical value 1 can be predicted as a determining factor for what should belong in category A and logical value 2 can be predicted for category B. So, according to the model, the individual might store value 1 on dimension 1, as a test of what belongs to category A, and value 2 on dimension 1, as a test of what belongs to category B. The exceptions stored would be 2111 for category

A and 1122 for category B. The learning process in RULEX is stochastic, and a key property is that different observers can from different rules from the same information. The vast array of different rules are the result of a probabilistic learning process described by few free parameters.

One of the advantages of the RULEX model over the GCM is that it successfully predicted a distribution-of-generalization data which the GCM failed to predict (Nosofsky, Palmeri & McKinley, 1994). However, Nosofsky and Johansen (2000) demonstrated that a modified version of the GCM was successful at accounting for this data by allowing for an individual-subject parameter variability. Also, to gain further support for the exemplar based account of the distribution-of-generalization data, the ALCOVE (Kruschke, 1992) model was applied. In the GCM version, altering particular patterns of attention weights across the five subgroups was required, but in ALCOVE, this requirement was fulfilled by the model's attention-weight learning mechanism.

ATRIUM (Erickson & Kruschke, 1998)

ATRIUM (Erickson & Kruschke, 1998) is a multiple-system categorization model that incorporates both rule and exemplar representations. Specifically, there is a rule module that learns to establish single-dimension decision boundaries, an exemplar module, that learns the association between exemplars and categories, and a module that links the two together, called the competitive gating mechanism. In general, the model uses the rule module in categorization, unless there is an exception to the rule in which case it prefers the exemplar module.

Erickson and Kruschke (1998) demonstrated that when using stimuli that vary along two dimensions, the ATRIUM model accounted for the categorization performance more accurately than the GCM. Nosofsky and Johansen (2000) suggest that this was because the stimuli involved numerical data which allowed for the precise perception of the magnitude of the items. When replicated without the numerical data there was little difference in the GCM and ATRIUM model predictions.

3.6 Exemplar theory; the GCM and how this relates to absolute judgment

In the present investigation a ‘relative (or relational; see chapter 4 on relative judgment for a full account) mode of categorization (or representation)’, is a categorization process in which items are represented in terms of some relational property (e.g., ‘small vs. large’). A relational classification is therefore based on a relational property which is independent of the particular physical properties of individual exemplars, but rather depends on the relations between sets of exemplars (in different categories). The implied converse mode of categorization, ‘absolute categorization’, involves item representations which veridically correspond to the actual physical properties of the items (e.g., ‘approximately 6 cm vs. approximately 20 cm’). It is this latter kind of categorization which the GCM has been designed to capture.

To demonstrate the specific difference in the absolute and relative representations, an account is given using the GCM which models the classifications of absolute and relative properties (is is not designed of relative representation but we include it here for illustrative purposes) in this example. So, in this example, the GCM is applied on the basis of two representational schemes for the training and test items: one in which the items are represented in an absolute way (in terms of their actual physical magnitudes; e.g., 12mm, 15mm etc.) and another in which the items are represented in a relative way (e.g., in terms of a simple coding whereby ‘smaller’ items are represented with the value 1 and ‘larger’ items are represented with the value ‘5’). For this example, all the other details of GCM fits were standard.

So, consider the following example. In the absolute version of the GCM fit, there are four items in a category called Chomps which have the heights: 32, 35, 36, 40 mm, and four items in a category called Blibs with heights: 62, 64, 66, 70 mm, and four test items, with heights: 81, 85, 121, 124 mm. It can be seen that two of the test items have ‘relatively’ smaller magnitudes and the other two relatively larger magnitudes. It therefore can be asked, ‘How do participants classify the test items in this experiment?' If they represented the training and test items in an absolute way, then it would be expected that most of the test items would be classified in the category of Blibs, since the Blibs training items were most similar to the test items. Using the GCM to predict the classifications made, a sum of squares value of 2.372 is produced when assigning all of the test items into the category of Blibs (note that smaller sums of squares indicates better classifications by the GCM).

Alternatively, in the relative version of the GCM fit, the relative value 1 could be used to represent all the (small) Chomps in training and the value 5 all the (large) Blib items in training. Likewise, the values 1 and 5 are used to represent the pair of smaller and larger test items respectively. Crucially, with this representational scheme, the items are only represented in terms of small and large, there is no more specific information about their physical (absolute) properties. As before, attempting to predict the empirical classification probabilities using the GCM and a relative representation for the training and test items, a sum of squares value of 0.181 was found for the relative classification. In other words, the GCM could predict classification probabilities better when the training and test items were represented in a relative way, as compared to when they were represented in an absolute way.

In the case of using a prototype model to represent absolute judgment, the same predictions would be made. For example, in the case of an absolute representation, the physical size of the prototypes for the 'Chomp' category and 'Blib' category would be used in the classification process. So, in the case of a pre-specified category labelled 'Chomps' which consist of heights 32, 35, 36, 40 mm, and a category labelled 'Blibs' consisting of heights 62, 64, 66, 70 mm, the prototype for the Chomp group would be 36 mm and the prototype for the Blib group would be 66 mm. In the same way as described in the GCM exemplar situation, this physical size would be used in the categorization process. So, in the case of new test items being presented corresponding to heights 81, 85, 121, 124 mm, then according to prototype theory, just as was the case for the GCM exemplar theory, the new items would be classified with the category to which they are physically most similar. Crucially, the only difference between prototype theory and the GCM exemplar, is that in the GCM each of the individual items within a group are compared for similarity with the test items, whereas in the prototype model, it is only the abstracted prototype that is compared with the test items.

This simple example demonstrates a possible use of the GCM to account for relative properties, however, the model has been designed and adapted for the use of predicting categories in absolute modes of supervised categorisation, where physical sizes of magnitudes are used.

3.7 Summary

The focus of this chapter was twofold. Firstly, a detailed description of absolute representation was given, which was illustrated with several supervised models such as the GCM. A simple example of absolute and relative representation and an example was given which used the GCM account. This evidence will be used to motivate the experimental investigations into relative vs. absolute representation in Chapter 5 and unsupervised vs. supervised categorization in Chapter 7.

Chapter 4

Relative Judgment in Categorization

4.1 An Introduction

In Chapter 3, a description was given of absolute representation (or judgments) and an example was given using the generalized context model (GCM; Nosofsky; 1984, 1986, 1991). This chapter expands the literature review of absolute vs. relative representation by examining some of the literature in categorization on the subject of relative representation. Crucially, a description of the relative judgment model (RJM; Stewart et al., 2005) in categorization and analogical mapping (Gentner, 1983, 2003; Holyoak & Thagard, 1995) is given, which motivates the definition of relative representation that we will use.

4.2 Absolute identification tasks

Miller (1956) reported that the cognitive system had difficulty in processing information once the short-term capacity limit in memory was reached. He found that this limitation occurred when using many different types of information, such as loudness of tones to the magnitude of lengths and areas. Absolute identification tasks are commonly used in classification experiments when testing memory limitations. These tasks consist of presenting several items of varying size, but can be used many other situations, such as when using sound, or brightness. In all of these situations, the participant must identify from memory, the smallest item to the largest. For example, a participant is given several stimuli of varying sizes and is asked to identify them from memory first the smallest, then the second smallest etc., until all the stimuli are accounted for. One of the problems that can result from this task is that errors in judgment can occur once the limit in short term memory is reached. To be more specific, if there are too many items (i.e., if the sequence of information exceeds the capacity limits of short-term memory), the memory trace of the exemplars can be lost, which reduces identification accuracy in this task. To compensate for this loss, representation of the items in memory can shift from absolute (based on the actual physical size) to a relative representation (where the representation of the items is relative to one another. Such relative representations (e.g., see Stewart et al., 2005), utilized the relative properties of 'bigger than' or 'smaller than' the neighbouring items, which is a process similar to analogical mapping. Briefly, there are three main observations from in these tasks: a limit in information transmission; bow effects in the accuracy of identifying the stimuli; and sequential effects. Each of these will be explained in turn.

Limitations in Information Transmission

The amount of information that can be transmitted through short-term memory can be measured with absolute identification tasks (McGill, 1954). Information transmission has an input, the presented stimuli, and an output, the classification response made. Input information travels through the short-term memory channel and arrives as the classification response output. Perfect transmission of the input to the response, would equal perfect classifications where there would be no errors. However, Miller (1956) demonstrated that the memory channel is limited to just a few bits (2.5 bits) of information and therefore, perfect transmission, once this limit is reached is not possible. However, as Miller (1956) points out, the memory channel is limited to just a few bits, and thus the information cannot travel perfectly from input to output if this channel capacity is exceeded. The 2.5 bit limit corresponds to about six equally likely alternatives. The limit leads to a loss of information and thus leads to a reduction in classification accuracy. Stewart et al. (2005) have demonstrated that such a limitation of memory leads to an alternative form of representation which is based on relative properties of the items. These relative properties are based on comparisons between the present item with preceding items, in terms of how different they are to each other. For example, the present item could be represented as 'much bigger' than the previous item. Information transmission can increase with the increase in range (e.g., the difference in size between items from smallest to largest). However, this also reaches a limit once the items are easy to discriminate (Alluisi & Sidorsky, 1958; Braida & Durlach, 1972; Eriksen & Hake, 1955a; Pollack, 1952).

Bow or Edge Effects

One of the phenomena observed in absolute identification tasks is the bow effect. This is where the classification accuracy is greater at the extremes of the item set and poorer at the midrange, and hence a bow effect is observed when plotting accuracy on a graph (e.g., Kent & Lamberts, 2005; Lacourture & Marley, 2004; Murdock, 1960; Siegel, 1972). When the range of the item set increases, the classification accuracy only slightly improves (e.g., Braida & Durlach, 1972; Gravetter & Lockhead, 1973; Hartman, 1954; Pollack, 1952). This effect is not only observed with visual stimuli such as items, it is also found with other stimuli such as when tones of sound are used (Brown et al., 2002). The bow effect increases when the number of stimuli presented increases (Alluisi & Sidorsky, 1958; Durlach & Braida, 1969; Lacouture & Marley, 1995; Pollack, 1953; Siegel, 1972). Siegal (1972) found that this effect was not due to any response bias, such as the end items being more frequently used as compared to the midrange items.

Sequential Effects

Another observed phenomenon in absolute identification tasks is sequential effects. This is where the previous item has some influence over the perception and thus classification of the present item. For example, if the preceding item was much smaller than the current item, then the perception of the current item could be that it is smaller than it actually is. There are several theories that try to explain the sequential effect. One of these theories is the assimilation theory. In this theory, the current item is perceptually assimilated in memory by the previous item so that it is more similar to it than it actually is (Garner, 1953; Holland & Lockhead, 1968; Hu, 1997; Rouder et al., 2004). Ward and Lockhead (1970) demonstrated that a response bias led to the current item being biased away from the previous item. Evidence for the assimilation of the items has not been confined to absolute identification tasks, as this has also been shown with magnitude estimation tasks (e.g., Jesteadt, Luce, & Green, 1977), in matching tasks (Stevens 1975) and in relative intensity judgment tasks (Lockhead & King, 1983). Assimilation effects have been modelled by several researchers such as by Stewart et al. (2005) in the relative judgment model (RJM).

4.3 Models that account for the effects observed in absolute identification tasks

Assimilation Models

Assimilation and contrast effects (i.e., where the current item is contrasted away from a neighbouring item) can be accounted for by assimilation models (Holland & Lockhead, 1968). For this, it is assumed that the cognitive system generates a classification response by converging the judged distance between the current and previous stimulus. Assimilation occurs when, for example, a smaller item precedes a larger item and this results in the larger item being assimilated so that it is perceived as more similar to the previous item. Thus, the present item has been assimilated so that it is perceptually smaller than it actually is, which leads to the errors in classification judgments.

Lockhead and King (1983) and Lockhead (1984), provided an assimilation model, which made two assumptions: (1) that it is the successive stimuli, which are assimilated in memory, and (2) relative comparisons are made between each new item and those stored in memory from the sequence presented. The model has accounted for contrast and assimilation, because it assumes that such relative comparisons are made. However, it did not account for the information transmission limit and bow effects. Such a limitation motivated the development of other models (e.g., Stewart et al., 2005).

Modified Thurstonian Models

Thurstonian models give an account for the bow effect. The simple Thurstonian decision model has been modified many times (e.g., by Durlach & Braida, 1969). This model assumes that the items in memory are represented in a noisy way, so that the exact magnitudes are not stored in memory, but instead some unspecific representation. It is these noisy values that are used in the classification process and this leads to the errors found in the bow effect.

The model accounts for the limit in information transmission as it assumes that the noisy values are stored instead of the exact values because of the information transmission loss from input to output. So, instead of storing the exact values of the items, the cognitive system only stores the noisy values. This accounts for why the errors in classification increase when more stimuli are presented. For example, when more items are included but the range is held constant, then the items are closer together in terms of size. As the memory representation of these are noisy, then there is a greater likelihood that these will be confused with each other which would lead to greater errors in classification. The bow effect is accounted for by the fact that as there are less neighbouring items at the extremes of the presented sets, then there is less chance of confusing these items with the neighbouring items. Less confusion would lead to greater identification accuracy.

Restricted Capacity Models

Lacouture and Marley, (e.g., 2004; Marley & Cook, 1984, 1986) accounted for the limit in information transmission and bow effects in absolute identification tasks. They suggested that the cognitive system had a limited capacity to process information and it is this that led to the errors in the classification such as the bow effect. The exemplars in this model are represented on a noisy Thurstonian scale. This model did well to account for the information transmission and bow effect but could not account for the sequence effects.

In more recent work, Lacouture and Marley (1995, 2005) developed a neural network mapping model, which includes a network of one single input unit, one single hidden unit and an output unit for each response. The storage of the exemplars in memory, were assumed to be noisy values. Response classifications in this model were made through the mapping onto the hidden unit activation, and it is assumed that for each output unit, activation is accumulated through the course of the trial. Once the accumulation reaches a threshold, the response is activated. However, the model still does not account for the sequence effects. Lacouture and Marley (2004) suggested that the model could be modified so that it would account for sequence effects, such as by suggesting that the normalizing of hidden activation units could be made so that previous items could be used instead of anchor values.

Laming's (1984, 1997) Relative Judgment Model

The relative judgment model (Laming, 1984), accounts for the limit in information transmission. This model gives a starting point for our definition of 'relative representation'. More specifically, the model assumes that the classifications made are done in such a way that, item differences are represented relative to each other. For example, the current item is represented relative to its difference with the preceding item.

It is clear, that this model uses relative representations rather than those based solely upon absolute physical properties. For example, rather than classifying the items based upon their physical (absolute) properties, such as 'item one comes after item three because item one is 6mm and item three is 4mm', the classification is made on the bases of item one is (relatively) 'bigger than' item three. So, the representation is based on relative properties. Specifically, it is the relative difference information that is used here rather than just the relative property. This is different to the relative representation of Stewart et al. (2005) RJM, where it is the relative property 'bigger or smaller than' that is used and not the relative difference information which is based upon the absolute properties. This model was good to account for the limit in information transmission but failed to account for the bow and sequence effects. Laming did suggest that the model could be adapted to account for prior expectations of the distribution and thus account for these additional phenomena.

Absolute Judgment, Exemplar Models

A thorough explanation of exemplar models is given in Chapter 3. Briefly, there are models based on similarity of absolute physical sizes. An example of how a classification would result here for three items; item one, 10mm; item two, 12mm and item three, 14mm, would be, that item one would be classified with item two, rather than item three, because it is physically more similar to this, as compared to item three. There are several models which use absolute physical similarity in classification. According to the exemplar theory (e.g., Medin & Schafer, 1978; Nosofsky, 1986), each item is stored in memory with its associated label. So, for example, when presented with a chair and also the category label, chair (i.e., the participant is told that this item belongs to a category called chairs), then the item with its label 'chair', is stored in memory. When classifying a novel item, the probability of a classification is increased when the stored items and the novel items are physically more similar. So, if a chair is presented and there are two available groups, 'chairs' and 'stools', then there is a greater likelihood that the new item will be classified into the category 'chairs'. This is because its physical properties such as length, and width are more similar to the exemplars in the chair category, as compared to those in the stool group.

In terms of the absolute identification task, Brown et al. (2002) applied the data for absolute identification tasks to the exemplar model (Generalized Context Model, Nosofsky, 1986). The exemplar model accounted for bow effects, as the end items have fewer items to get confused with, but it does not explain the gradual bowing. This however, can be accounted for if the weights in the model are changed and bias in favour of responses for stimuli that have more extreme magnitudes.

The major problem with the exemplar models is that they fail to account for sequential effects. They try to account for such effects by placing more weight to neighbouring items (e.g., Nosofsky & Palmer, 1997, Elliot & Anderson, 1995), so that these become more available to memory. However, this fails to predict the sequence effect in classification of the items (Stewart, Brown, & Charter, 2002).

4.4 Stewarts Relative Judgement Model (RJM, 2005)

Stewart et al. (2005) were motivated to develop the RJM by the assumption that that the classification process in absolute identification tasks is based upon relative and not absolute judgment.

Relative vs. absolute judgment

The RJM assumes that when making classifications there is no mechanism which stores even noisy perceptual absolute magnitude. Instead, the model is based on the idea that classification in absolute identification tasks are made on the basis of simple relative comparisons of the current item with its preceding neighbours. Stewart et al.'s (2005) RJM, uses a similar mapping model as used by Lacouture and Marley (1995, 2004), and assumes that there is noise in the process of mapping several stimuli to the correct output response. This noise, they suggest, is the limitation which leads to errors in absolute identification tasks. This is different to other accounts such as the simple Thurstonian account, in that it does not require noisy representations of the perceptual exemplars. By assuming that the limit in capacity is due to mapping rather than perceptual noise, Stewart et al. (2005) suggested that there was no requirement for any further explanation to account for the lack of improvement in performance when stimulus range is increased, which makes the RJM approach more parsimonious than competing theories. One of the problems that face most models, according to Stewart et al. (2005), is that they base their assumptions on the physical magnitudes held in long-term memory, which makes it difficult to account for sequence effects. For example, in Thurstonian models, the position of criteria in long-term memory is used. In the connectionist model, Lacouture and Marley (1991, 1995, 2004), suggest that information about the most extreme stimuli is used from long-term memory. Also, in the exemplar models, the physical magnitudes of each stimulus is kept in long-term memory, and classifications are based upon similarity of these exemplars. Although these models can be modified to account for the sequence effects, the RJM explains all three observed effects in absolute identifications tasks without any need for modification.

RJM and absolute identification tasks

In Stewart et al.'s (2005) relative judgment model (RJM) for absolute identification tasks, the model accounts for all three effects (bow effect, sequential effects, and limited capacity), which all the other models fail to do. The main assumption the RJM makes, which is directly relevant to the present investigation, is that the classification judgments are made on the basis of relative comparisons and not absolute magnitudes to one another. This leads to the focus of the present investigation. The question asked is whether there might be analogous situations in categorization experiments.

4.5 Relative Judgment in Analogical Mapping

Analogical mapping

Analogical mapping is a process of comparison to identify shared relations between two knowledge systems, such as two objects. The generated comparisons are thought to play a role in relational reasoning (Gentner, 1983, 1989; Gick & Holyoak, 1980, 1983; Holyoak & Thagard, 1995); when learning and using rules (Anderson & Lebiere, 1998; Lovett & Anderson, 2005); in the appreciation of perceptual similarities (Medin, GoldStone, & Gentner, 1993); and in the production of language, science, mathematics and art. In analogical mapping when making a comparison between several objects such as elephant, truck, mouse and ball, then shared properties are identified such as elephant and truck are both 'big' and mouse and ball are both 'small'. The shared property receives a double activation and is therefore more active in the classification procedure as compared with single activated unshared properties. The shared properties can drive classification decisions: for example, because elephant and truck are both big they should be classified together, and the same happens for mouse and ball.

The development of relational thought

There is evidence to suggest that the ability to reason using relational thought occurs through development (e.g., Gentner & Ratterman, 1991; Halford, 2005). Initially, children make inferences based on whole object similarity and then later acquire the ability to develop relational thought (e.g., Gentner, 2003; Gentner & Rattermann, 1991). For example, consider the following situation: when given two pictures, one of which is a dog chasing a cat and another is a boy chasing a girl with the cat in the background. Three year old children use featural similarity to match the cat in both pictures while five year old children use relational similarity, e.g., in both cases chasing is taking place (Richland et al., 2006). This developmental trend is known as the relational shift (Genter & Rattermann, 1991).

Connectionist models based on distribution representations (e.g., Colunga & Smith, 2005) provide a good account of whole object similarity in younger children's reasoning, but do not account for more complex later relational thought (see Holyoak & Hummel, 2000; St. John, 1992). There are accounts of older children's and adults' reasoning ability (e.g., Anderson & Libiere, 1998; Falkenhainer et al., 1989), but these do not provide accounts of where the structured representations on which they rely originate from. There are accounts for both the featural (displayed in young children) and the relational (displayed in older children) representations, but there is no account for how the relational thought develops. This lack of an account for learning structured representations from unstructured examples is often cited as the most significant limitation of structured accounts of cognition (e.g., Munakata & O'Reilly, 2003; O'Reilly & Busby, 2002; O'Reilly, Busby, & Soto, 2003). Doumas et al. (2008) offer an account for how structured relational thought is produced from relationally unstructured information (i.e., no direct instructions that allow for relational thought).

Analogical mapping modelling: DORA

Doumas et al. (2008) formed an analogical model for discovering relations (the Discovery Of Relations by Analogy; DORA). They suggested that there are three crucial factors in the development of complex learned relations. These were: firstly to identify invariants in the features presented; secondly, to isolate such property relations; and thirdly, to bind such property relations to new examples. Identifying featural invariants has been found in children as young as 6 months, who can identify features such as 'more' and 'less' in properties such as size, Clerafield and Mix (1999) and Feigenson, Carey, and Spelke (2002). Doumas et al. (2008) suggested that in the next stage, the property needs to be isolated (such as 'taller'), from the rest of the environment, so it has its own independent meaning. In the final stage, is the ability to bind these property relations (e.g., 'taller'), to new items and concepts in novel situations (see Doumas & Hummel, 2005). This takes the process from simple detection of relational properties, such as 'taller', into one which can structure new arguments, from the same relational properties, but with novel items or concepts (Doumas & Hummel, 2005; Halford et al., 1998; Hummel & Holyoak, 1997).

The main goal in the development of DORA was to demonstrate how an unstructured relational example can lead to structured relational representations. It forms four basic operations: (1) the retrieval of propositions from long-term memory (LTM); (2) analogical mapping of the propositions, from working-memory (WM), to the novel situation; (3) predication and refinement; and (4) self-supervised learning (SSL). Analogical mapping, inference, and schema induction, all use these four basic operations (see Hummel & Holyoak, 2003). For the purposes of the present investigation, the main interest in this literature is the binding of the relational concepts such as 'smaller than', which is relevant for the present experiments.

4.6 Relative Judgment in our experimentation

From the Stewart et al. (2005) study, relative judgment accounts of absolute identification tasks, and of relations between 'current' and 'preceding' items, we can lay out our basic argument for the current investigation. The argument presented, is that there are situations in which relative judgments dominate over absolute judgments and representations, as identified in absolute identification tasks. Both relative and absolute representations are supported in the literature (e.g., the Generalized Context Model; GCM; Nosofsky; 1984, 1986, 1991, for absolute identification, and, Stewart et al., 2005, Gentner, 1983, 2003; Holyoak & Thagard, 1995, for relative judgment). The present investigation investigates, specifically, the circumstances which will promote a relative vs. an absolute representation. An argument for this is given in the chapter describing the experiments (Chapter 5) for relative judgment, but here, a simple description of the term relative representation (or judgment), is given.

A general description from analogical mapping is used, and Stewart et al.'s RJM offers a useful starting point for such a definition, on the basis of properties such as 'smaller', 'bigger' than etc. For the studies in this investigation, a 'relative (or relational) mode of categorization', is a categorization process in which items are represented in terms of some relational property (e.g., 'small vs. large'). A relational classification is therefore based on a relational property, which is independent of the particular physical properties of individual exemplars, but rather depends on the relations between sets of exemplars (in different categories). The implied converse mode of categorization, 'absolute categorization', involves item representations which veridically correspond to the actual physical properties of the items (e.g., 'approximately 6 cm vs. approximately 20 cm').

So, to conclude, a relative representation is a classification based upon the relative differences of the items (e.g., bigger than, and smaller then), whilst absolute representation is based upon a classification when using the actual physical properties (e.g., item 1 is 10cm tall). In Chapter 5 these definitions are used in the current investigation of absolute and relative representational shifts of classification.

Chapter 5

Experimental results; relative vs. absolute representation in categorization

5.1 Introduction

In this chapter, we wish to demonstrate a set of experiments which explore relative vs. absolute shifting in supervised categorization. This work is motivated by Stewart et al's (2005) relative judgment model (RJM), which explores relational representation in absolute identification tasks. For this investigation we produced 5 Experiments which explore the shifting between relative and absolute judgment in supervised categorization.

5.2 Relative vs. absolute representation in supervised categorization

The problem of how naive observers represent information is clearly a fundamental one in psychology. On one extreme, there is a strong intuition that psychological representations have to be veridical descriptions of the physical/ perceptual properties of the stimuli in our environment. The bulk of the modelling work in categorization involves such representations. For example, both exemplar theory (Ashby & Maddox, 1993; Nosofsky, 1984, 1986, 1991) and prototype theory (Homa & Vosburgh, 1976; Posner & Keel, 1968; Reed, 1972) typically formulate predictions in terms of items represented in a way, which directly corresponds to their actual physical properties (cf. Shepard, 1987). However, clearly, the representational capacity of human cognition is a lot more flexible than that.

There has been an abundance of evidence for the generation of abstract features/ the representation of information in terms of relative or relational features. An influential tradition of relevant evidence comes from research on analogical reasoning. Analogical mapping is a process of comparison to identify shared relations between two knowledge systems, such as two objects. The generated comparisons are thought to play a role in relational reasoning (Gentner, 1983, 1989; Gick & Holyoak, 1980, 1983; Holyoak & Thagard, 1995); when learning and using rules (Anderson & Lebiere, 1998; Lovett & Anderson, 2005); in the appreciation of perceptual similarities (Medin, GoldStone, & Gentner, 1993); and in the production of language, science, mathematics and art. In analogical mapping when making a comparison between several objects such as elephant, truck, mouse and ball, then shared properties are identified such as elephant and truck are both 'big' and mouse and ball are both 'small'. The shared property receives a double activation and is therefore more active in the classification procedure as compared with single activated unshared properties. The shared properties can drive classification decisions: for example, because elephant and truck are both big they should be classified together, and the same happens for mouse and ball.

As suggested in Chapter 4, I will use the term 'relative-like (or relational) mode of categorization', a categorization process in which items are represented in terms of some relational, abstract property (e.g., 'small vs. large'). A relational classification is therefore based on a relational property, which is independent of the particular physical properties of individual exemplars, but rather depends on the relations between sets of exemplars (in different categories). The implied converse mode of categorization, 'absolute-like categorization', involves item representations which veridically correspond to the actual physical properties of the items (e.g., 'approximately 6 cm vs. approximately 20 cm').

A categorization researcher can ask whether there might be circumstances, which spontaneously lead to a preference for a more absolute-like, or relative-like mode of categorization. In this respect, prior research is slightly uninformative. Most studies either assume one form of representation or demonstrate that a particular form of representation is plausible (e.g., in analogical reasoning, the objective is commonly to demonstrate situations in which analogies can be employed to solve reasoning problems). It is rarely the case, however, that alternative possible representations for the same stimuli have been directly contrasted within the same paradigm.

The above discussion immediately leads to an important methodological problem: how can a researcher determine whether a particular categorization reflects absolute-like or relative-like representations? In a typical manipulation in this work, participants see items varying along a single dimension in a training phase. The training items are organized into two categories, a category of 'small' items and a category of 'large' items. In test, suppose that there are only two test items, one of which is smaller than the other, but also such that they are both larger than the items in the 'large' training category (see Figure 3). It seems straightforward to assume that a relative-like categorization would mean that the Category A exemplars are represented as 'smaller' than the Category B ones, so that the shorter of the two test stimuli will be classified as a Category A instance while the larger as a Category B one. By contrast, with an absolute-like categorization, both test instances should be classified as Category B instances, since their physical properties are more similar to those of the Category B members.

Two important qualifications underwrite the robustness of this paradigm. First, an assumed relative-like representation is *not* the same as a fuzzy absolute-like representation. If the representation of the test instances is absolute-like, but inexact in some sense, then they should still be classified in Category B, as long as the difference between Category A and Category B exemplars is large enough (see Figure 3). This can be arranged in a straightforward way, for example in an experiment where we have exemplars in training Category A approximately 30 mm, exemplars in training Category B as 60mm, whilst test exemplar sizes are 80mm and 120mm. In this case, an absolute-like representation would yield a classification, where both test items would be classified into Category B. Second, one can assume that the default response bias of participants would be to select some test instances as members of one training category and other instances as members of the other. Such a response bias clearly favours a relative-like mode of categorization in our experiments. Crucially, the conclusions we are seeking to derive in this work are not whether a particular manipulation leads to absolute-like or relative-like categorizations, but rather whether it leads to *more* absolute-like or relative-like categorization, in relation to a baseline manipulation.

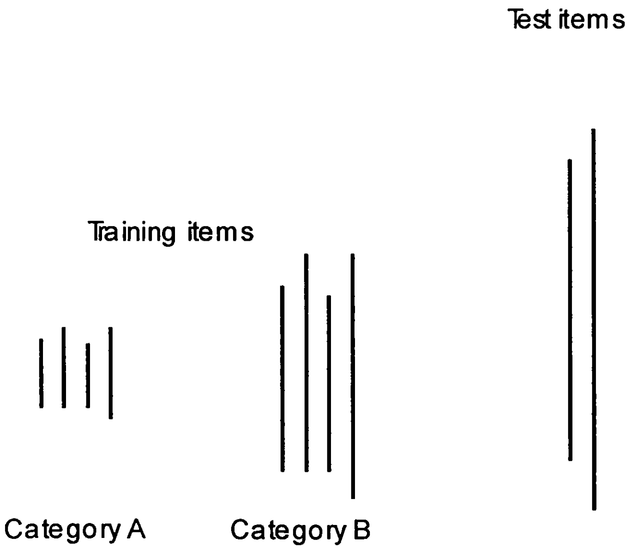


Figure 3. A schematic diagram of a typical manipulation in the present work. Each line corresponds to a stimulus. Stimuli vary along a single dimension, which is overall length.

5.3 Defining relative and absolute representation

The issue of absolute-like vs. relative-like categorizations has recently received some attention in the study of absolute identification tasks. In such tasks, a participant is presented with several stimuli of varying magnitudes along a particular dimension of physical variation, such as height. They are then asked to remember these stimuli and to place them in order from smallest to largest from memory. Using this task, Miller (1956) observed that people found it difficult to identify a particular item from a set of items that vary along a single dimensional continuum (such as length, brightness of colour or pitch of tone). Stewart et al. (2005; see also Lamings, 1984, 1997) could account for various phenomena in such tasks by assuming that the judgment for each stimulus was made relative to the previous stimuli. Also, in the work of Stewart et al. (2002; Stewart & Brown, 2004), who have demonstrated that difference information between items in sequence is used to generate the classification response, which is similar to the relative comparison that we refer to.

In our terminology, such representations would be examples of relative-like representations, in the simple sense that the representation of the stimulus does not depend on its absolute physical properties. Stewart et al. suggested that, for example, limits in memory capacity might have prevented the absolute representation of all relevant exemplars.

The work on absolute identification tasks suggests that *all* representations are relative. In categorization, it seems implausible that there are not circumstances in which the representations we employ are not absolute (e.g., see Goldstone, 1994).

So, the question becomes, under what circumstances might we expect that an absolute classification mode will be preferred? It is reasonable to suggest (e.g., in terms of a minimalist Bayesian intuition) that the absolute properties of a category (e.g., information about particular exemplars, as would be required by exemplar theory, or information about a prototype, as this would be required by prototype theory) would be inferred with more confidence if more category exemplars were studied in the training phase. In other words, suppose that the correct hypothesis about the absolute value of the prototypes of the two presented categories is such that the first prototype is P1 and the second prototype is P2.

Let's label the training exemplars of the categories as D1, D2, D3 etc. We suggest that P(P1, P2|D1, D2, D3) would be lower compared to P(P1, P2|D1, D2, D3, D4, D5, D6...). In other words, it would be possible to evaluate with more confidence a hypothesis about the absolute properties of the category prototypes, if more training exemplars are processed. Likewise, if there were four categories there would be four values to infer regarding the physical values of the prototypes, P1, P2, P3, P4. A straightforward extension of this reasoning suggests that P (P1, P2|D1, D2, D3) would in general be lower than P (P1, P2, P3, P4|D1, D2, D3). In other words, we would need more information to support or reject a more complex hypothesis (about the physical properties of category prototypes). This argument does not assume that participants represent categories with exemplars or prototypes. Rather, our claim is that a representation of a category based on (absolute) physical values of the training exemplars is more likely to be possible (and hence, we predict, preferred by the cognitive system) if there are more training exemplars per category.

The above approach leads to straightforward predictions: More training exemplars per category would lead to more absolute-like classification. More categories would lead to more confusion between the particular physical properties of each category, therefore to less absolute-like classification. Finally, other manipulations which undermine participants' confidence in the absolute physical properties of the training items, would also lead to less absolute-like classification (we employed one such manipulation, a time delay).

These predictions depend on a particular, incidental mode of category learning (cf. Milton & Wills, 2004). In our tasks, the training items were shown to participants in bundles corresponding to their intended categorizations. Thus, participants could readily perceive the training items in terms of their intended categories. In addition, participants were exposed to the training exemplars of each category relatively briefly; they did not have an incentive (and were not encouraged) to memorize the exemplars or study them thoroughly. Indeed, as they had the training exemplars available in test there would be no reason for them to do so.

Such a mode of category learning strongly contrasts with the more common supervised categorization methods. In such cases, participants are exposed over a large number of trials to the same training exemplars repeatedly, until they can perfectly reproduce their classifications. Moreover, the category structures typically learned in this way are complex: extensive training is required before learning can be achieved (by contrast, in our experiments participants were shown very simple category structures). What would be the relevant expectations under such circumstances? Clearly, the fewer the exemplars, the easier participants would find learning the required classifications, and (therefore) the more salient would participants find the taught exemplar-classification label associations. Overall, the fewer the exemplars in a supervised categorization paradigm, the more pronounced we would expect exemplar-effects to be. Indeed, this is exactly what has been found in previous work (Rouder & Ratcliff, 2004; cf. Blair & Homa, 2003).

So, with the incidental category learning paradigm approach we propose there is a prediction that fewer exemplars (typically) lead to less absolute-like classification, however, with a standard supervised categorization paradigm fewer exemplars lead to more absolute-like classification (more pronounced exemplar effects). To reiterate, the key difference between the two approaches is that in the latter case, participants have no choice but to represent the training exemplars in an absolute way. They receive typically dozens of training trials (e.g., in Rouder's study experiment 2, there were 96 trials in a block and 10 blocks in a session with each participant going through two sessions of training) in which they have to learn to associate particular exemplars with particular labels. In our case, by contrast, the training exemplars never really have to be learned (they are present throughout the experiment and the category structure is very simple). Additionally, the test exemplars could be classified by interpreting the training exemplars in two radically different ways.

In closing, our general hypothesis is that, when it is difficult to derive accurate training category representations (prototypes or exemplars) based on the physical properties of the studied objects, the cognitive system is more likely to employ relative-like (relational) representations. From this general hypothesis, as stated three more specific ones can follow: 1) We expect more relative-like categorization when there are fewer items per group; 2) We expect more relative-like categorization when there are more groups; 3) We expect more relative-like categorization when we use a time delay between the initial presentation of training stimuli and test items. For hypothesis 3 we are suggesting that the time delay will deteriorate the memory of the specific exemplar representation and therefore reduce the available information regarding the distributional properties, leading to a greater likelihood of relative-like representation.

5.4 Experiment 1

Experiment 1 provides a baseline examination of the basic experimental design. There were two training categories, a category of Chomps and the category of Blibs. There were four test items, (heights are indicated in Table 2 shows a comparison of all heights in all experiments). If participants adopted an absolute-like mode of categorization, then all four test items should be classified in the category of Blibs (that is, all the 'large' items would be classified in the same category). By contrast, if participants adopted a relative-like mode of classification, the two smaller test items should be classified with the training category of smaller items (Chomps) and the two larger test items should be classified with the category of larger training items (Blibs). Note that, as said, with a single experiment, it is impossible to gain insight into the circumstances under which absolute-like or relative-like representations are more likely to occur. For example, results in this experiment will be partly driven by a propensity to represent the stimuli in a relative-like or absolute-like way, but also by possible task demands, such as a bias to assign some of the test stimuli to all the available training categories. Experiment 1 is the baseline manipulation: We are interested in whether our additional manipulations lead to a shift in favour of relative-like or absolute-like representations. By comparing the results of subsequent experiments with those of Experiment 1, we effectively factor out such possible task demands.

Method

Participants

A total of 63 Swansea University students took part in the experiment for a small payment. Participants were tested individually and were all experimentally naïve.

Materials

12 items were created using Corel Draw (Figure 4). Each item was presented on a card and consisted of a picture of a flower grounded on a solid base. The picture of the flower was comprised of a yellow bud with eight petals, and a blue stem. Eight items, grouped into two categories, comprised the training stimuli. The group of Chomps consisted of four flowers, which were of the following heights: 32, 35, 36, 40mm. The group of Blibs consisted of four flowers with greater heights: 62, 64, 66, 70mm. There were four test items of flowers with heights; 81, 85, 121, 124 mm (see Table 2). The items in the category of Chomps and the items in the category of Blibs

were such that height differences between any successive items were computed assuming a Weber fraction of 8%. The three authors independently verified that each stimulus could readily be discriminated from all others. Note, also that the width of the flowers would increase in size as with the overall height. We only report height values, since these provide the easiest way to label the stimuli.

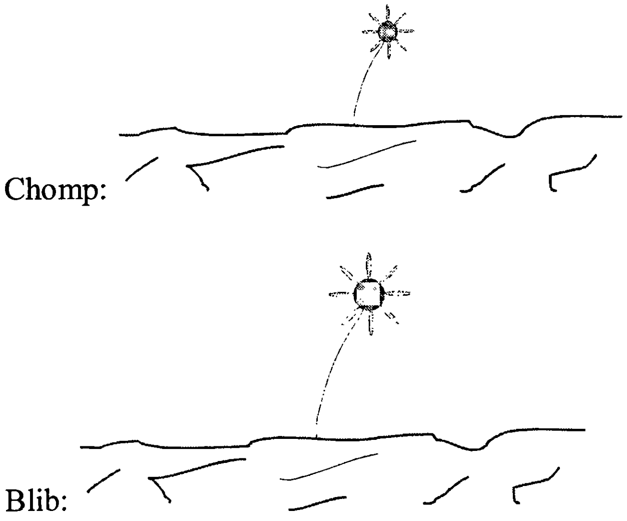


Figure 4. Two examples of the stimuli which were presented to participants. The top image is an example of an item belonging to the Chomps category (in Experiment 1) and the bottom image is an example of the Blibs category.

Procedure:

Participants were first presented with (written) instructions to the effect that they were about to see some items, which belonged to two imaginary categories (called Chomps and Blibs), and that the experimenter would tell participants which items went to which categories. At that point, participants were shown the two groups of stimuli presented on cards Chomps and Blibs (each of the items were presented on a single card). Both training groups, Blibs and Chomps, were presented together in all of our experiments (two single piles of items, Blibs and Chomps; presentation of categories was counterbalanced across participants). Participants were asked to look though every item in the two groups in their own time (this was typically less than three minutes). Subsequently, and while the training items were still available to participants, participants were presented with new (written) instructions, indicating that new items will be shown, and that each participant had to decide for each new item whether it was a Chomp or a Blib. The instructions stated, "There are no right or wrong answers! You have to classify each item as a Chomp or a Blib." After the presentation of the instructions, participants were presented with the four test items; test items were also presented simultaneously. Note that in all the experiments the training items were present when participants were asked to categorize the test items. The presentation order of items was randomized for each participant.

Results and discussion:

We define absolute-like categorization to correspond to a response pattern whereby all test items were considered Blibs and relative-like categorization whereby the two smaller test items were considered Chomps and the two larger test items Blibs. Directly analogous definitions for absolute-like and relative-like categorization were employed in the other experiments as well. Absolute- (or relative-) like categorizations will partly be influenced by whether the training stimuli are represented in an absolute or relative way. Equally, the relative proportion of absolute-like or relative-like categorization will depend on other factors, such as, for example, whether there is a respond bias to classify some test stimuli into all the available training categories. Therefore, we cannot say from the results of a single experiment whether (e.g.) absolute-like categorization constitutes evidence for absolute-like categorization and representation. This becomes possible by comparing the results of two (or more) experiments, so that we can examine whether a particular manipulation increases the tendency for (e.g.) absolute-like categorization is increased. Notwithstanding the above issues, the characterization of participant responses as absolute-like and relative-like seems like a good starting point in considering our data.

The responses of some participants were such that they did not conform to this characterization. Such participants were eliminated from the analyses here and elsewhere, since their results do not bear on the hypotheses we are interested in. Of course, if our experimental design works as intended, we would expect that relatively few participants would produce such in-between responses. In this experiment, only one participant was eliminated because he/she categorised the stimuli in a way, which did not clearly fit our definitions of absolute-like and relative-like representation. Forty eight participants adopted a relative-like categorization mode and 14 an absolute-like one.

Here and elsewhere we adopted χ^2 two-tailed tests to examine any preference for relative vs. absolute-like classification of the test items, against either what would be expected by chance, or in relation to results from other experiments. A χ^2 test against chance simply examines whether the proportion of absolute-like categorizations is equal to that of relative-like categorizations or not. We assume that a straightforward 50-50 observed-chance split is most appropriate for the chi-square analysis, as we were interested in testing against the null hypothesis that an absolute-like classification is equally likely to a relative-like classification. However, it is not the observed vs. chance analysis which is particularly interesting, as the frequency of relative vs. absolute classifications in any one experiment could be determined by demand characteristics. Rather, it is the comparison of classification performance across experiments, which supports our conclusions regarding the manipulations which promote relative vs. absolute classification.

In Experiment 1 there was a highly significant tendency for participants to prefer relative classification, against chance: $\chi^2(1) = 10.08, p < .0005$. Overall, Experiment 1 demonstrates the baseline condition and the analytical approach. It is clear that several participants (14 out of 62) did not feel obliged to assign some test instances to all the training categories. However, it is probably exactly this bias which led to the preference for relative-like classifications in Experiment 1. Accordingly, in itself, the conclusion from Experiment 1 is not interesting. In subsequent experiments, by altering the key characteristics of Experiment 1 and observing participants' performance, we will present a series of results, which support our hypothesis of when relative-like vs. absolute-like categorization is more likely to occur.

5.5 Experiment 2

In Experiment 1 we showed that relative-like categorization can be observed when each of the training categories had four test items per training group (Chomps and Blibs). In Experiment 2 we doubled the number of items per training group. We suggest that this may encourage absolute-like classification. It is possible that when more evidence is available regarding the distributional properties of the training items, concrete (absolute) information about the category exemplars (or prototypes) would be more available and so it would be such information which drives the classification of new exemplars. Conversely, relative-like classification may be encouraged by indistinct memory traces of the training exemplars. Again, having more exemplars per category is likely to strengthen the corresponding memory traces and hence promote absolute-like classification. As said, note that in all our manipulations the training items were present when the test items were categorized. However, we can minimally assume that in classifying a test item a participant has to rely on some psychological representation of the training categories.

Method and Procedure:

Fifty nine Swansea University students took part in the experiment for a small payment. Participants were tested individually and were all experimentally naïve (here and elsewhere, no participant took part in more than one of the present experiments). Materials consisted of the same two groups of flower images (Blibs and Chomps), but with eight instead of four items in each group. The heights of the members of the Chomps category were 35, 36, 40, 42, 44, 46, 47, 49 mm and the heights of the members for the Blibs category were 62, 64, 66, 70, 74, 75, 76, 77 mm (see Table 2 for a comparison of all heights in all experiments). The test items and procedure remained the same as in Experiment 1.

Results and discussion:

In this experiment, we observed 25 participants providing relative-like classification of the test items and 32 absolute-like classifications (two participants were eliminated because their responses could not be characterized as absolute-like or relative-like). Examining participants' pattern of responding against chance, as before, did not identify a preference for relative-like or absolute-like classification ($\chi^2(1) = .33$, $p =$ .573). Crucially, when comparing the results of Experiment 2 with the results of

Experiment 1, there was a highly significant interaction: $\chi^2(1) = 14.1, p < .0005$. This result indicates that in Experiment 2 participants were a lot more likely to adopt an absolute-like mode of categorization, compared to Experiment 1. Such a conclusion supports the hypothesis outlined in motivating Experiment 2.

In sum, increasing the number of exemplars per category increased the preference for representing the stimuli in an absolute-like way. As predicted by assuming that more concrete information about the training items enhances absolute-like categorization.

It is important to mention a possible range effect, which is the differences in size between the items, effecting the possible classifications (see e.g. Alluisi & Sidorsky, 1958; Braida & Durlach, 1972; Eriksen & Hake, 1955; Pollack, 1952). The range of training stimuli does change between experiments and I should have clarified in the chapter when this occurs. For example, in Experiment 1, the difference in height between the largest stimulus (the biggest Chomp) and the smallest one (the smallest Blib) is 22mm; there were four exemplars per group in Experiment 1. By contrast, in Experiment 2, in which there were eight exemplars per group, the corresponding difference is only 13 mm. Therefore, there is a decrease in range between Experiment 1 and Experiment 2. Now, the key question is whether such differences in range can (partly?) account for our results. This is extremely unlikely for the following reason: First, please note that previous research shows that it is *increases* in the range which make the stimuli more discriminable and would therefore, presumably, encourage absolute representation. However, our results are inconsistent with such an expectation. In Experiment 2 we observed more absolute representation compared to Experiment 1, however, the range of stimuli in Experiment 2 was smaller than in Experiment 1. A similar situation occurs in comparing Experiments 3 and 4. Again, when comparing Experiment 3 with Experiment 4, the range size difference for Chomp and Blib is 2 mm, Blib and Zlog, 6, and Zlog and Glab, 8; compared with, for Experiment 4, where the differences are 5, 6 and 6 respectively. As can be seen, an increase in range, can only encourage an absolute representation, therefore we can discount range effects as a possible reason for our experimental effect, when there is no such increase in range.

It is also important to mention, that in Experiment 1 the largest Blib was 70mm and the smallest test item 81mm, which is an 11 mm distance between the items. In Experiment 2 the largest Blib was 77mm and the smallest test item 81mm, which is only a 4 mm distance between the two items. In order to justify that the items being more similar, did not lead to the absolute judgment in experiment 2 we need to note two things. First, in all cases all three authors verified independent that all stimuli were discriminable from each other. Also note that participants saw the stimuli concurrently so that even relatively small differences between stimuli could be readily detected. Second, in Experiment 1 the largest Blib was 70mm and the smallest test item 81mm, but in Experiment 2 (as the Reviewer notes) the largest Blib was 77mm and the smallest test item 81mm. However, more importantly, in Experiment 4 we include two test items which are identical to two items in the training groups, one in the Zlog group and the other in the Glab (48 and 62 mm). In this Experiment there were significantly more relative vs. absolute classifications as compared with chance levels. This indicates that even when the some of the training items are not disciminable with the test items, we still can find relative judgments.

In Experiment 3 we provide a manipulation, which is intended to weaken the concreteness of information for the training items and so (if our reasoning is correct) promote relative-like categorization.

5.6 Experiment 3

In Experiment 3 we doubled the number of training groups, from two to four (the four training categories were called Chomps, Blibs, Zlogs, and Glabs). The Chomps had heights 30, 29, 36, 30 mm, the Blibs 38, 39, 40, 41mm, the Zlogs; 47, 48, 49, 50mm and the Glabs 58, 61, 62, 63mm (see Table 2). It can therefore be seen that the heights of the members of the four categories conformed to the simple ordering, smallest, small, large, and largest. The heights of the four test items were 48, 62, 113, 183mm. Accordingly, one test item was the same size as one Zlog and another test item was the same size as one Glab; the other two test items were larger than all the training items. The fact that two test items were identical to two training items might plausibly enhance absolute-like classification. However, if our suggestion that the concreteness of information for the training exemplars enhances absolute-like classification is correct, then the converse prediction is made: with more categories, one would expect that participants would be more confused about the exact physical attributes of the members of each category, so that relative-like classification would be favoured. Note that Lacouture et al (1998) demonstrated that increasing the number of possible responses reduces the frequency of correct responses. The manipulation in this experiment relies on a similar assumption, namely that with more categories, increased confusability between stimulus-category label associations would lead to more relative-like categorization. However, in Lacouture et al.'s work there was a single normative response. That is, all responses were either correct or wrong and participants had to represent the stimuli in an absolute way for correct responses to be possible. By contrast, in our experiments participants were not constrained to represent the stimuli in a specific way. They could present them in a presumably more absolute or a relative way. Our manipulation in Experiment 3 exactly follows the logic of Lacouture et al. By providing more response categories, we assumed that participants would find it more difficult to adopt absolute representations. For example, this difficulty might relate to higher confusability between stimulus-category label associations. Note that we were not interested in the difficulty with which multiple categories can be learned and, so, during the test phase, participants could observe all the training stimuli correctly arranged into their respective categories. All these points are not made in the ms.

Method:

A total of 79 Swansea University students took part in the experiment for a small payment. Materials consisted of the same flower images but with four instead of two groups (labelled Chomps, Blibs, Zlogs and Glabs) and with four items for each group. The heights of all the stimuli are given above. The procedure for this experiment was the same as for Experiment 1, except that participants were required to categorise each of the four test items into any of the four groups (whereas in Experiment 1 there were only two groups). The instructions stated that any possible classification of each test items was possible.

Procedure:

The procedure was the same as in other experiments, but for the fact that participants were presented with four training groups. Additionally, in this experiment participants were told more clearly that any possible classification was possible. Specifically, instructions prior to the test phase were augmented with the following text: "Any possible classification is allowed. The test items could be classified between all four possible categories, between three of the possible categories, between two of the possible categories or just one of the categories."

Results and discussion:

In this case, relative-like categorization corresponded to grouping the smallest test item with the training group with the smallest members (Chomps), the second smallest test item with the training group with the second smallest members (Blibs) etc. An absolute-like classification corresponded to assigning test item 48 in the category of Zlogs (since there was a Zlog of the same size) assigning test item 62 to the category of Glabs (since, likewise, there was a Glab of the same size) and, finally, assigning the two remaining test items (113 and 183) to the category with the largest members (the category of Glabs). Fifty nine participants adopted a relative-like categorization mode and 9 an absolute-like categorization mode. Responses from 11 participants were removed from the data as these did not fit the definitions of relative-like or absolute-like categorization, and so do not bear on the hypothesis tested.

We used a χ^2 test to investigate whether relative-like or absolute-like categorization was preferred (against chance) in classifying the test items into the categories of Chomps, Blibs, Zlogs and Glabs. There was a very strong tendency to form relative-like categorizations in this experiment: $\chi^2(1) = 21.26, p < .0005$. We next examined whether in Experiment 3 there were more relative-like categorizations compared to Experiment 1. There was an interaction in this direction, but it did not prove to be significant; $\chi^2(1) = 1.95, p = .163$.

5.7 Experiment 4

In Experiment 3 we doubled the number of training groups and so observed a preference for relative-like classification. This preference should be reduced, by doubling the number of members of each group, if our hypothesis regarding relative-like vs. absolute-like classification is correct.

Method and procedure:

A total of 60 Swansea University students took part in the experiment. Materials consisted of the same flower images with four groups (labelled Chomps, Blibs, Zlogs and Glabs) but with eight instead of four items for each group. These were the same training groups as in Experiment 3 (Chomps, Blibs, Zlogs, and Glabs), but each category in this case had eight members instead of four. The eight Chomps were flowers with heights 27, 28, 28, 31, 30, 29, 24, 21 mm, the eight Blibs had heights 38, 39, 40, 41, 39, 40, 41, 36 mm, the Zlogs had heights 47, 48, 49, 50, 49, 50, 51, 52mm, and the Glabs had heights 58, 61, 62, 63, 61, 60, 62, 63mm. The four test items were the same as in Experiment 3 and had heights 48, 62, 113, 183mm (see Table 2). The procedure for this experiment was the same as for Experiment 3. There was some repetition of the exemplar sizes, to prevent as much as possible a reduction in the range between groups. There is no reason to assume that such a repetition would not affect the classifications made. Some exemplars were repeated, to prevent, as much as possible differences in stimulus range regarding the members of different groups. However, our results strongly indicate that range effects (and repetitions) do not affect the conclusions we wish to draw regarding shifts towards absolute or relative representation between experiments.

Results and discussion:

In this experiment, we observed 36 participants as categorizing according to relative-like categorization and 14 according to absolute-like categorization. The responses of 10 participants were removed, as these did not fit the definitions of absolute-like or relative-like categorization. With 2 x 2 training groups and two test items, there are 2^2x2^2x2^2x2^2 = 256 possible classifications (since each item can be assigned to either group), but with 4 x 4 x 4 x 4 training groups, and four test items there are four billion possible classifications. Such a massive increase would undoubtedly lead to a greater number of classifications which we cannot characterize as absolute-like or relative-like and, hence, which do not bear on our research questions. Also, these rejected classifications did not appear to confirm to a consistent pattern of responding. We first examined whether there was any evidence for a preference of absolute-like vs. relative-like categorization against chance. A χ^2 test showed that participants preferred a relative-like mode of classification: $\chi^2(1) = 5.086, p = .024$. The crucial comparison regarding our hypothesis corresponds to whether there were a greater proportion of absolute-like categorizations in Experiment 4, compared to Experiment 3. This was indeed the case: there were significantly more absolute-like categorizations in Experiment 4 compared to Experiment 3, as predicted: $\chi^2(1) = 4.0$, $p = .045$. This result is consistent with the findings in Experiment 2, in which we also observed a shift towards absolute-like categorization when increasing the number of items per group. We also compared Experiment 4 with Experiment 2 to examine whether increasing the number of groups would have led to an increase in the number of relational categorizations (note that the number of items in each group was the same in Experiment 4 and Experiment 2). For this comparison, we found $\chi^2(1) = 8.61, p = .003$ which further shows that having more groups does increase the number of relational categorizations.

Experiments 1 to 4 examined our hypothesis in terms of manipulating the number of training categories and the size of each category. In all cases, our results were consistent with a general hypothesis regarding preference for absolute-like vs. relative-like classification, according to which when it is possible to derive more concrete information about a category, then absolute-like classification should be favoured. In Experiment 5 we attempt an alternative test of this hypothesis.

5.8 Experiment 5

Experiment 5 is based on Experiment 2 (two training categories, with eight items per group). The difference between the two experiments is that instead of asking participants to classify the test items immediately after presentation of the training items, we asked them to return one week later and then make their classifications decisions. According to our hypothesis, the time delay should deteriorate the memory traces for the training items, and thus increase the proportion of relative-like categorizations. In other words, if participants are unable to remember the exact physical characteristics of the stimuli, they might be more likely to attempt to classify the test stimuli on the basis of relational features, such as 'small vs. large'.

Method and Procedure:

A total of 59 Swansea University students took part in the experiment for a small payment. The materials were identical to those in Experiment 2. Briefly, Chomps had heights 35, 36, 40, 42, 44, 46, 47, 49 mm, Blibs had heights 62, 64, 66, 70, 74, 75, 76, 77 mm, and the test items had heights 81, 85, 121, and 124 (see Table 2). The procedure was likewise identical to that of Experiment 2, but for the fact that participants were asked to make their classifications a week after they had studied the training items. Moreover, in Experiment 5, classification of the test items took place without having the training items available.

Results and discussion:

In this experiment, we observed 47 participants classifying the test items in a relative-like one, and five classifying the test items in an absolute-like way. Results from seven participants were removed, as their classifications could not be characterized as absolute-like or relative-like. In Experiment 5 there was a strong tendency (against chance) to form relative-like categorizations $\chi^2 = (1)\, 20.3, p < .0005$. We next compared the results of Experiment 5 with the results of Experiment 2 to find that in the former the proportion of relative-like categorizations was much higher, as predicted: $\chi^2 = (1)\, 26.3, p < .0005$. This is consistent with a hypothesis such that a time delay causes decay in memory and therefore weakens the memory for the absolute-like (physical) properties of the training exemplars, which encourages a relative-like classification.

Table 2: Sizes in for the length of each item in each group.

nt 1		Experiment 2			Experiment 3					Experiment 4			
Blib group	Test items	Chomp group	Blib group	Test group	Chomp group	Blib group	Zlog group	Glab group	Test group	Chomp group	Blib group	Zlog group	Glab group
62	81	35	62	81	30	*38*	*47*	*58*	*48*	27	38	47	58
64	85	36	64	85	29	*39*	*48*	*61*	*62*	28	39	48	61
66	121	40	66	121	36	*40*	*49*	*62*	*113*	28	40	49	62
70	124	42	70	124	30	*41*	*50*	*63*	*183*	31	41	50	63
		44	74							30	39	49	61
		46	75							29	40	50	60
		47	76							24	41	51	62
		49	77							21	36	52	63

Table 3: Summary of the results or relative vs. absolute experiments.

Experiment	Classification frequency	Predominant classification	Items per group	Number of training groups	Time delay	Significance (all χ^2 tests with 1 df).
Exp1 vs. chance	48 R 14 A	Relative	4	2	none	<.0005
Exp 2 vs. chance	25 R 32 A	Non significant	8	2	none	.573
Exp 2 vs. 1	48 R 14 A vs. 25 R 32 A	Absolute	4 vs. 8	2 vs. 2	none	<.0005
Exp 3 vs. chance	59 R 9 A	Relative	4	4	none	<.0005
Exp 3 vs. 1	59 R 9 A vs. 48 R 14 A	Non significant	4 vs. 4	4 vs. 2	none	.163
Exp 4 vs. chance	36 R 14 A	Relative	8	4	none	.024
Exp 4 vs. 2	36 R 14 A vs. 25 R 32 A	Relative	8 vs. 8	4 vs. 2	none	.003
Exp 4 vs. 3	59 R 9 A vs. 36 R 14 A	Absolute	4 vs. 8	4 vs. 4	none	.045
Exp 5 vs. chance	47 R 5 A	Relative	8	2	1 week	.000
Exp 5 vs. 2	47 R 5 A vs. 25 R 32 A	Relative	8 vs. 8	2 vs. 2	1 week vs. none	.000

5.9 General Discussion

In five experiments, we explored the conditions, which might promote a relative-like vs. absolute-like mode of classifying novel instances relative-like to incidental taught categories. Experiment 1 was the baseline manipulation, in which more relative-like-like categorizations were observed compared to absolute-like ones. The predominance of relative-like categorizations could be due to a greater bias for relative-like categorization as such, but it might be also due to participants wanting to assign some test instances to all the available categories (that is, a task demand). Conclusions regarding absolute-like or relative-like categorization are possible by comparing the results of at least two experiments. In Experiment 2, we found a shift towards absolute-like categorization (compared to Experiment 1) when the number of items per group was doubled (relative to Experiment 1) from four to eight. In Experiment 3 we added two new groups with four items per group and found that participants categorized new stimuli using relative-like categorization. In Experiment 4, we used the same four categories as in Experiment 3, but doubled the number of items per category from four to eight. As was the case with Experiment 2, when comparing the results of Experiments 4 and 3 there was a shift for absolute-like classification. In Experiment 5, we used the same stimuli as in Experiment 2, but instead asked participants to classify the test items one week after the training items were presented. This manipulation led to a shift towards relative-like categorization, when comparing with Experiment 2. The results of the five experiments are summarized in Table 3.

The results can be summarized in the following way. First, smaller categories promote a relative-like mode of categorization (Experiments 1 and 3). Plausibly, when there are few exemplars per category, the cognitive system cannot confidently infer a category representation in terms of concrete information, so that a relative-like representation is adopted (by 'concrete information' we mean information which directly corresponds to physical attributes of the stimuli). Second, when the number of items per category was increased, we observed a shift towards absolute-like categorization. Presumably, and consistently with the previous assertion, more exemplars per category imply that there is more information on the basis of which the cognitive system can represent a category in a concrete (absolute) way. Third, increasing the number of category groups enhances a relative-like mode of categorization. In this case, we suggest that the cognitive system finds it more difficult to concurrently keep track of the distributional properties of several categories (and so represents them in a relative-like way). Accordingly, when there is a requirement to learn more categories, it adopts a simpler, relative-like way of representing category information. Finally, when we introduced a time delay relative-like classification was encouraged. It appears that a delay would lead to a less detailed mode of representing the information from the training phase, so that the cognitive system would abandon an absolute-like mode of representation, and instead prefer a representation in terms of (less specific) relative features, such as 'small vs. large'.

There is a consistent theme underlying all our findings. In the case of simple schematic stimuli for which there are no prior, general knowledge expectations, we assume that the default preference is for the cognitive system to derive a representation for the stimuli on the basis of their physical properties. When it comes to categories composed of such stimuli, the cognitive system appears to operate like a standard statistical engine as suggested in our Bayesian example: the more the information regarding the distributional properties for the exemplars of each category, the more likely it is that the category will be represented in way which directly relates to the physical properties of the stimuli (cf. Ashby & Maddox, 1993; Chater, 1999; Tenenbaum & Griffiths, 2001; Tenenbaum, Griffiths, & Kemp, 2006, Lee & Vanpaemel 2008). Distributional category information could be undermined in a variety of ways: by having fewer items per category, more categories (which would lead to more confusion of which item belongs to which group), and a time delay. Our experiments provide support for the importance of all these factors in relative-like vs. absolute-like categorization.

The real-life situation we are trying to model with our experiments, concerns all cases when some stimuli would be represented in a (more) relative, as opposed to absolute, way. One such situation concerns absolute identification tasks, in which it appears that judgment is relative, rather than absolute. In categorization, there are cases of objects whose representation appears to involve at least some abstract features. For example, the sun is 'large', a cheetah is 'fast', and a Christmas dinner is 'plentiful'. In such and similar cases, it is uncommon to provide a more specific (absolute) impression of the corresponding characteristics. These are exactly the kind of situations we are trying to model with our experiments, that is, situations when absolute characterizations appear inappropriate or inconvenient in some sense and so people resort to more relative representations.

This research provides evidence that the cognitive system can spontaneously represent the same items in an absolute-like or relative-like way, depending on the characteristics of the categorization problem. At face value, such a conclusion is consistent with the assumptions in Stewart et al.'s (2005) RJM for the absolute-like identification task. In that model, responses on a current stimulus are assumed to be a function of the previous stimulus, rather than an exclusive function of the physical properties of the current stimulus. In other words, responding is relative-like rather than absolute-like. Moreover, the underlying motivation for assuming relative-like responding is similar to ours: difficulty in accurately representing the physical information for all the stimuli. However, it is not clear whether there is a role for absolute representations in Stewart et al.'s RJM.

This work allows an examination of the implications of the RJM for categorization. Are there circumstances when a relative-like representation mode might be preferred to an absolute one (the latter is assumed to be the default)? The extensive research tradition on analogical reasoning has, of course, made extensive use of relational features to understand analogical reasoning. However, this work does not provide any prescription of whether absolute-like or relative-like representations are more likely to be adopted in a categorization task. With five experiments, we aimed to provide some boundary conditions on this issue. Our results show that the cognitive system appears to adopt a fairly principled and adaptive way of preferring relative-like vs. absolute-like representations.

Our findings impact most directly on categorization theory. In brief, there are two classes of models, models of supervised categorization (such as prototype or exemplar theory; Nosofsky, 1984, Homa & Vosburgh, 1976) and models of unsupervised categorization (e.g., the rational model of Anderson, 2001, or the simplicity model of Pothos & Chater, 2002). Supervised categorization models typically operate on a default representation of the stimuli, but have the ability to transform this representation typically through attentional parameters (attentional parameters effectively select out a subspace of the default representational space). Unsupervised categorization models can sometimes predict the dimension(s) participants should spontaneously prefer when categorizing a set of stimuli (Pothos & Close, 2008; Pothos & Bailey, in press).

Could our results be explained within such modelling frameworks? The principles, which guide dimensional selection in supervised categorization models, have to do with identifying the representation, which makes the required categorization easiest to learn (e.g., Shepard et al., 1961; Smith & Minda, 2000). One could conceivably propose that the representation of the stimuli in our experiments is made of both absolute and relative properties. Then, the categorization task in test makes one set of properties more useful than the other. There are two problems with this approach. First, in our experiments it appears that the emphasis on relative properties (e.g., in Experiment 1 vs. 2), has to do with the processing of both the training items and the test items. By contrast, supervised categorization models set their parameters only during the processing of the training stimuli. Second, there is an infinite number of possible relative properties. How could, as modellers, we decide a priori which are the appropriate relative properties to use in an experimental situation? Similar considerations apply in the case of models of unsupervised categorization (noting, in any case, that dimensional selection for such models is less well developed when compared with supervised categorization models).

In sum, our results show situations in which classification of test stimuli appears to have a profound influence on the representation of categories acquired in previous (training) phases. Such flexibility in category representation is difficult to reconcile with current categorization models and represents an exciting avenue for their further development.

5.10 Summary

This study explored the shifting between relative-like and absolute-like representations in categorization. While there is considerable evidence that categorization processes can involve information about both the particular physical properties of studied instances and abstract properties, there has been little work on the factors which lead to one kind of representation as opposed to the other. We tested 320 participants in 5 experiments, in which participants had to classify new items into predefined artificial categories. In three experiments, we observed a (predominantly) relative-like mode of classification, and in 2 experiments we observed a shift towards an absolute-like mode of classification. These results suggest three factors, which promote a relative-like mode of classification; when there are fewer items per group, more training groups, and the presence of a time delay. Overall, we propose that less information about the distributional properties of a category and/or weaker memory traces for the category exemplars (induced, e.g., by smaller item numbers per category, or a time delay respectively) can encourage relative-like categorization.

Chapter 6

Experimental results: unsupervised categorization; testing the simplicity model

6.1 Introduction

This chapter describes the experimental results of our unsupervised categorization investigation (this excludes the work carried out in relative vs. absolute shifting, which is given in Chapter 5, and the work carried out in unsupervised vs. supervised categorization, which is given in Chapter 7). Briefly, our goals here were twofold. Firstly, we examined the validity of the simplicity model by Pothos and Chater (2002), but with a much larger participant sample than they used. This investigation was carried out because one of the major problems with the results in Pothos and Chater (2002) was that they found high classification variability in their data sets. This led to some difficulty in identifying meaningful results with the particular sample sizes that they employed (between 10 and 29 participants). Secondly, we investigated whether introducing some general knowledge about the items would affect perception (i.e., whether they were a coherent group or not), and thus the categories made.

6. 2 Experimental investigation: validating the simplicity model and exploring background (general) information

Empirical explorations of unsupervised categorization are much more complicated than of supervised categorization. First, in supervised categorization the relevant principal dependent variable is obvious: it corresponds to classification probabilities for test instances. In other words, an experimenter will collect data about how particular novel instances are classified and he/she will then attempt to fit participant selections with models of supervised categorization. This is not the case in unsupervised categorization. As noted, some researchers have explored particular empirical issues (such as the tendency for unidimensional sorting). However, there has been uncertainty in the literature regarding a suitable dependent variable for fitting models of unsupervised categorization. Pothos and Chater (2002) suggested that one could present stimuli to participants, ask them to divide them into any number of groups in an unconstrained way, and then count either the number of distinct classifications produced or the frequency of the most popular classification (this is the procedure employed in the current study, as well; note that Medin and colleagues have employed a similar procedure in the study of cultural biases; e.g., see Atran & Medin, 2008, for a recent overview). The rationale here is that if a particular classification is psychologically more intuitive, then more participants should select it and, equally, there should be lower variability in participants' classifications (i.e., fewer distinct solutions). While conceptually such an approach seems intuitive, it quickly runs into the problem that the space of possible solutions is vast: for ten stimuli there are about 100,000 classifications (Medin & Ross, 1997). For 16 stimuli (the number of stimuli employed in the present study) there are 10.4 billion possible classifications. So, reasonably, one has to ask just how intuitive a particular classification has to be in order to be preferred by participants amongst so many alternatives. Indeed, Pothos and Chater (2002) had some difficulty identifying meaningful results with the particular sample sizes they employed (between 10 and 29).

In later work, Pothos and Chater (2005; see also Pothos & Close, 2008) employed the Rand index of classification similarity to get around the problem of response variability in unsupervised categorization. The Rand index is a measure of similarity between two classifications (Rand, 1971; see Fowlkes & Mallows, 1983, for an extension). The idea would be to specify, e.g., two target classifications, A and B (which might correspond to the predictions of two different models), and then compute the similarity of the classifications produced by participants to A and B. Higher similarity would indicate a preference for one of the target classifications, even if the actual classification had never actually been produced. This approach certainly has some merits, however, a problem is that one can never be logically certain that the target classifications appropriately characterize human performance. In other words, suppose that, as an experimenter, I choose classifications A and B for use with a Rand index analysis. It is possible that there is a third classification, C, which I have overlooked, and which is *more* similar to participants' classifications than either A or B. In such a case, the Rand index analysis may lead to slightly misleading conclusions. To conclude, while the Rand index analysis has many strengths (and may well be the most appropriate method under particular circumstances), it also has important limitations.

Another way in which researchers have attempted to study unsupervised categorization is by restricting the number of categories into which participants can divide a set of stimuli. In other words, after being shown a set of stimuli, participants might be asked to divide them into two categories, as opposed to an unlimited number of categories. Such an approach clearly leads to considerably fewer distinct classifications and, indeed, it has produced very valuable results in relation to classification strategies (e.g., Ahn & Medin, 1992; Milton & Wills, 2004). However, it has also been argued that this experimental procedure does not exactly correspond to the cognitive process of spontaneous grouping and, rather, may correspond more to problem solving (Murphy, 2004; Pothos & Close, 2008). In other words, participants are given some information (the stimuli) and a problem (find, e.g., a suitable division of the stimuli in two clusters). So, conceivably, they simply search for the simplest way to achieve a solution to the problem, without necessarily being guided by which classification is more intuitive or natural. For this reason unsupervised categorization with a fixed number of categories may be a suboptimal procedure when it comes to fitting models of unsupervised categorization.

It therefore appears that the most suitable procedure for studying the “intuitiveness” or “naturalness” of particular categories is an entirely unconstrained classification paradigm. In the present work, we overcome the problem of response variability by employing a population sample which could be considered large (169 participants) by comparison with the relevant previous studies (Compton & Logan, 1999; Pothos & Chater, 2002). As an aside, it is interesting to consider the corresponding requirements for studies of supervised categorization: In such studies as few as 10 participants can be employed per condition in order for a researcher to expect to derive meaningful results. This is because a supervised categorization task forces conformity into participants’ responses. As participants have to learn the same division of stimuli into categories, their responses to the test items likewise tend to be fairly uniform. By contrast, in unsupervised categorization there are no aspects of the procedure to prevent idiosyncratic strategies, so that the relevant dependent variables (e.g., preference for a particular classification) would be more noisy.

The study of unsupervised categorization is not uniformly more complicated than that of supervised categorization. In the latter case, one typically has to consider whether the supervised learning procedure might alter the perceived similarity of the studied stimuli. For example, Goldstone (1995) has shown that classifying stimuli in the same category can increase their similarity to each other and decrease their reported similarity for stimuli in other categories. Schyns, Goldstone, and Thibaut (1997; Goldstone, 2000) even suggested that new features may be created as a result of categorizing stimuli in particular ways. If the similarity structure of the stimuli changes, then this may complicate the fits of supervised categorization models, noting though that there is some debate as to whether such effects correspond to changes in how the stimuli are perceived or to response biases when invoking the stimuli (Goldstone, Lippa, & Shiffrin, 2001; Roberson & Davidoff, 2000). By contrast, it is unlikely that unsupervised categorization leads to changes in the similarity of the categorized stimuli (cf. Gureckis & Goldstone, 2008; Pothos & Chater, 2005).

The purpose of the above discussion was to motivate the need for more datasets in unsupervised categorization and explore possible methodologies, the most suitable of which appears a completely unsupervised categorization procedure (with a large population sample). As with previous investigators (e.g., Shepard et al., 1961), a useful dataset is one which involves several individual stimulus sets, such that each one of them reflects a different intuition about the underlying psychological process. Shepard et al. (1961) were guided in their selections by considering a range of different learning problems. Likewise, we aimed to select stimulus sets such that each one would correspond a different intuition about spontaneous categorization. With such an approach, it is clearly the case that the more the stimulus sets the greater the range of categorization intuitions which can be examined. However, there is a contrasting consideration, which is that, given that the experimental design had to be within participants, the more the stimulus sets the greater the possibility for possible experimental confounds (e.g., responses for one stimulus sets affecting those for another, fatigue, etc.). Given the amount of time it takes to carry out a spontaneous categorization task, we thought that nine stimulus sets reflected a good balance between these two considerations.

Finally, we can now revisit the question of which aspect of human cognition unconstrained unsupervised categorization in the laboratory can help us understand. The spontaneous formation of categories must be guided by a sense in which certain groupings are more intuitive than others or, in other words, the relative coherence of different groupings. Performance in such tasks must, therefore, be partly guided by the same psychological process which allows us to consider certain concepts are more intuitive (or coherent) than others. Ultimately, we wish to understand the aspects of environmental statistics which drive the development of human concepts. A related issue is that most unsupervised categorization tasks in the laboratory manipulate only similarity (in the sense that some of the presented stimuli are more similar to others). It seems clear that the sense we have that certain items must be more similar than others must be an important driving force in spontaneous grouping behavior (noting that psychological similarity may take many forms and/ or be context dependent; e.g., Goldstone, 1994). It also seems very plausible that the general knowledge we have about a set of stimuli might affect our perception of whether they form a coherent group or not (Murphy & Medin, 1984). What is the role of general knowledge in unsupervised categorization? One possibility is that it modifies the perception of similarity for some stimuli, for example, by enhancing the salience of certain dimensions and suppressing that of others (e.g., Murphy & Allopenna, 1994). However, it is possible that general knowledge may have a more complex influence on spontaneous categorization, which cannot be reduced to a manipulation of similarity (cf. Yang & Lewandowsky, 2004; Wisniewski & Medin. 1994). Current formal models of unsupervised (and supervised) categorization all depend on similarity, but proponents of such models sometimes discuss the type of extensions which could (in principle) allow incorporating general knowledge effects. Note, though, that it is still unclear whether it may even be logically possible to provide formal description of general knowledge effects (cf. Fodor, 1983; Pickering & Chater, 1995; but see Harris, Murphy, & Rehder, 2008, or Heit, 1997, for promising attempts). In sum, the most valid conclusion is probably that unconstrained unsupervised categorization tasks which manipulate just similarity provide an approximation to the cognitive process of evaluating category coherence. The previous conclusion is arguably the main difference between modeling human unsupervised categorization and statistical clustering, since, typically, the objective of the latter is to determine whether there are clusters or not, as opposed to comparing the intuitiveness of different classifications on the same or different stimulus sets (e.g., cf. Hubert & Arabie, 1985; Fraboni & Cooper, 1989; Milligan & Cooper, 1986). However, a successful model of unsupervised categorization ought to discriminate between the relative intuitiveness of different classifications.

Before we layout our experimental procedure, to summarize; we are to investigate the validity of the simplicity model with a much larger sample size, to allow for more thorough identification of the classification patterns than found in Pothos and Chater (2002). The main dependent variable in the study is the frequency of the preferred classification for each stimulus. For each stimulus, the simplicity model can provide a value for how intuitive the preferred (or optimal in some other sense) classification ought to be. Therefore, the main test for the simplicity model is the correlation between the frequency of preferred classifications and the intuitiveness predictions. Secondly, we are to investigate the possible effect that introducing some general knowledge about the presented stimuli, will have on the intuitiveness of the classifications. So, a similar analysis can be conducted for the number of times the preferred classification was produced for different stimulus sets, as a function of instructions. Below, we give a brief account of the simplicity model (see Chapter 2 for a much more in-depth account).

The simplicity model

In Chapter 2 we give a thorough description of the simplicity model, here we give a brief summary of the key mechanisms. The simplicity model of unsupervised categorized arose as a generalization of Rosch and Mervis's (1975) model for basic level categorization. Rosch and Mervis suggested that in a hierarchy of concepts there is a preferred level, so that, for example, when participants are presented with a novel object they would identify it with its basic level categorization, rather than a subordinate or superordinate one. Rosch and Mervis suggested that the basic level of categorization should be the one for which within category similarity is maximized and between category similarity minimized. The simplicity model's starting point is this intuition of Rosch and Mervis and the assumption that whatever determines preference in basic level categorization also determined preference in spontaneous categorization (cf. Gosselin & Schyns, 2001). In a sense, the simplicity model can be seen as a way to measure within/ between category similarity, and balance each term against its other in the formation of categories. More formally, the model prefers classifications that provide the greatest algorithmic simplification of the similarity structure of a set of items. It has been specified by using the principles of the minimum description length framework for formalizing the simplicity principle (Rissanen, 1987). Its original formulation is parameter-free, it is non-metric (its input is relative similarities; cf. Stewart, Brown, & Chater, 2005), and it does not require information about the number of categories sought either directly or indirectly. This last feature particularly distinguishes the simplicity model from other models of unsupervised categorization and clustering.

In the simplicity model, similarity information without categories for four stimuli A,B,C,D is specified as similarity(A,B)≤similarity(C,D) etc. Each of these inequalities requires one bit of information to be determined (since there are only two possibilities, the similarity on the left hand side is greater or less than the similarity on the right hand side). Assuming symmetry and minimality, for *n* objects there are $p = \frac{n(n-1)}{2}$ pairs so that there are $\frac{p(p-1)}{2}$ pairs of pairs. For example, the similarity structure of 10 stimuli corresponds to 990 bits of information; this is the codelength for the similarity information for a set of objects, without categories. Note that the model can easily be applied without the assumptions of symmetry and minimality. However, for the simple, schematic stimuli typically employed in unsupervised categorization tasks symmetry and minimality in similarity judgments ought to be assumed (Hines, Pothos, & Chater, 2007; Tversky, 1977).

Categories can reduce similarity codelength by using the definition that all similarities for objects within categories are greater (or equal) than all similarities for objects between categories. When there are numerous, correct such constraints, then there is more codelength reduction. However, in some cases categories may specify erroneous constraints. To correct such errors, we have to select the *e* erroneous constraints out of the total number of constraints *u* which can be achieved with a code of length $log_2(u+1) + log_2\left(\frac{u!}{e!(u-e)!}\right)$. There is also a code for specifying the particular classification employed. The length of this code is given by $log_2(Part(r,n))$, where *r* is the number of elements, *n* is the number of clusters, and $Part(r,n) = \sum_{v=0}^{n-1}(-1)^v \frac{(n-v)^r}{(n-v)!v!}$. Therefore, a classification can lead to a reduction (simplification) of the codelength for the similarity for a set of objects because of the constraints, but this advantage is moderated by the costs associated with correcting the erroneous constraints and specifying the classification. The final codelength for a set of objects with categories would be [codelength without categories] – [constraints – cost for errors - cost for specifying classification]. This final codelength corresponds to the prediction of the simplicity model: the lower its value, the more the simplification due to a classification, and the more intuitive the classification is predicted to be. Note that predictions from the simplicity model are typically expressed in terms of the ratio [codelength with categories] / [codelength without categories].

The simplicity model can produce a prediction for the optimal classification for a set of stimuli from scratch and a prediction for how intuitive particular classifications should be relative to each other. It assumes that all the available stimuli are presented at once.

6.3 Experiment 1; validating the simplicity model, and exploring background information

Participants and design

Participants were 169 students at Swansea University, who took part for a small payment. Each participant classified nine stimulus sets, one after the other. A between-participants condition related to whether the stimuli were described in a neutral way or as real-world objects (87 participants received the realistic instructions and 82 the neutral ones).

Stimuli

We created nine stimulus sets of 16 stimuli each. The stimulus sets were created so as to capture a range of intuitions about unsupervised categorization. Accordingly, we had three stimulus sets in which there were two well-separated clusters. The first such stimulus set involved two equally-sized clusters ('two clusters'), the second stimulus set involved two clusters which were not equally sized ('unequal clusters'), and the third stimulus set two spread out clusters, that is, clusters which were not very cohesive ('spread out clusters'). We then had three stimulus sets for which the most intuitive classification was deemed to be more complicated (but not necessarily less intuitive; such a notion of 'being complicated' here has entirely heuristic value and just corresponds to our prior intuitions). The first such stimulus set had three well-separated clusters ('three clusters'), the second stimulus set two clusters but with some intermediate ambiguous points ('ambiguous points), and the third stimulus set two clusters which were close to each other and which also greatly varies in size ('poor two clusters'). The final three stimulus sets were designed to that the best possible classification would be expected to be even more complicated. Accordingly, the first such stimulus set had five well-separated clusters ('five clusters'), the second reflected random variation ('random'), and the third was meant to correspond to cluster embedded in a cloud of noise ('embedded'). As noted, a greater number of stimulus sets would allow more informative model comparisons, but would complicate the data collection procedure (since the design was within participants, the higher the number of stimulus sets each participant was asked to classify, the greater the chance for response interference, problems due to fatigue etc.) All nine stimulus sets are shown in a schematic representation in Figure 5.

The stimuli were made from two continuous dimensions. We preferred this approach, as opposed to employing stimuli composed of discrete features, because, in the latter case, it is often difficult to specify complex category structures (because each feature can have a limited number of discrete levels), unless one employs several features. But, when several features are employed, then there is the issue of whether participants can create holistic representations of the stimuli (cf. Milton & Wills, 2004). The two stimulus dimensions (Figure 5) were mapped to the length of a 'body' (horizontal dimension) and the length of the ‘legs’ after the joint (vertical dimension) of schematic spider-like stimuli (Figure 6). By choosing such stimuli, both dimensions of physical variation were lengths, and so a Weber fraction in mapping the Figure 5 values to physical values could be safely assumed (8%; Morgan, 2005). For both dimensions, the actual lengths were between 40mm and 80mm.

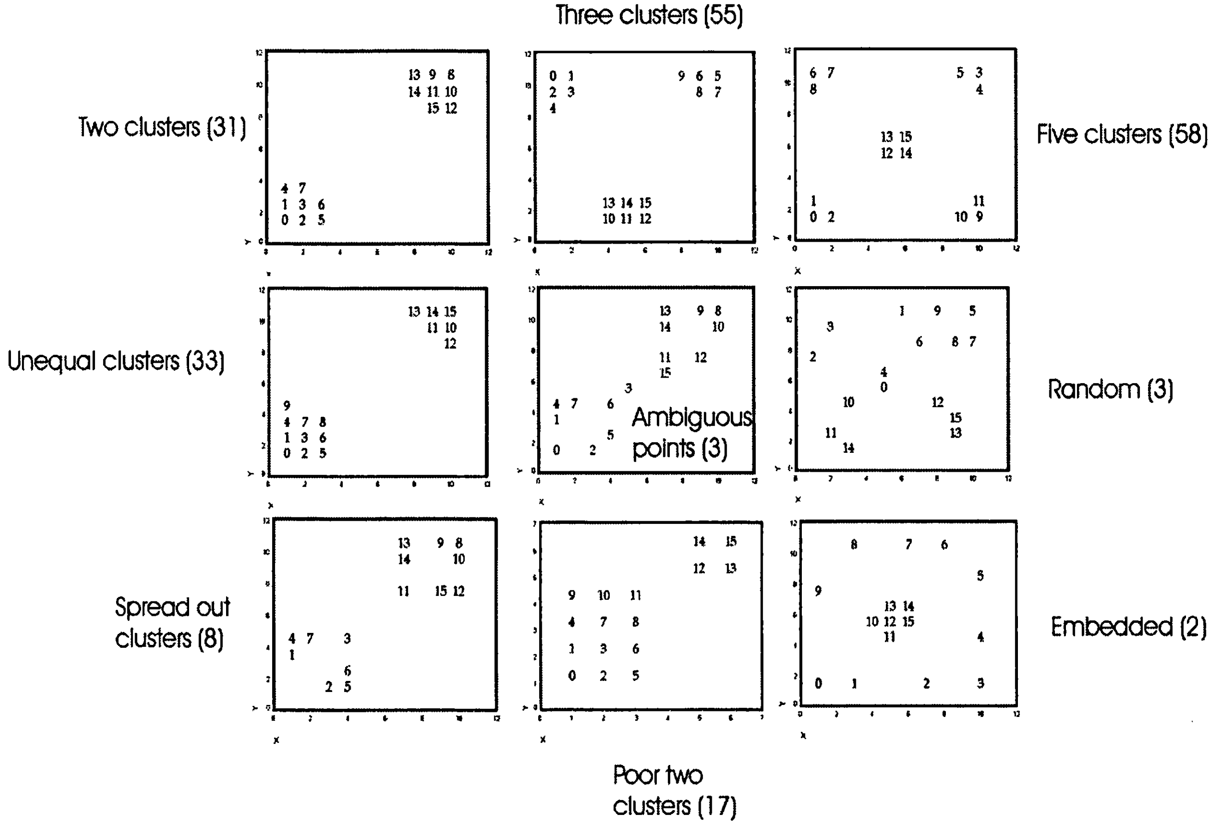


Figure 5. A schematic representation of the nine stimulus sets employed in this research. Each point in each stimulus set is indexed by a number from 0 to 15. In parentheses we show the number of times the most frequent classification was produced for different stimulus sets. A higher number indicates that the most frequency classification was more intuitive.

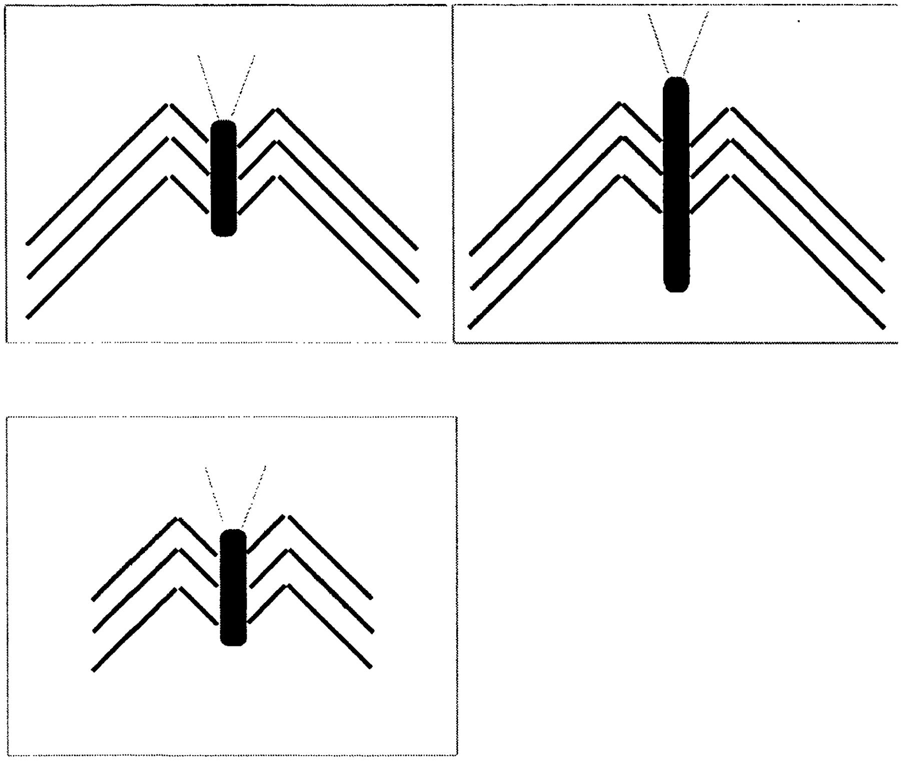
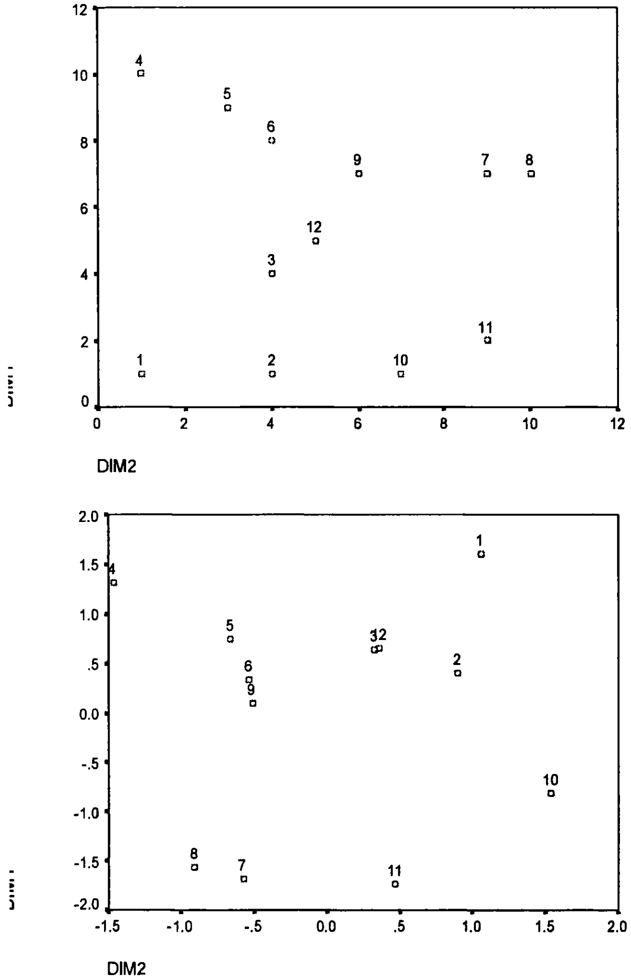


Figure 6. Some examples of the stimuli used.

A number of considerations motivated this choice of stimuli. First, as noted, we wanted stimuli for which we could assume Weber fractions for the relevant dimensions of physical variation. Most of the models of unsupervised categorization function by assuming a psychological representation of the stimuli, so that it is clearly important for the validity of model predictions to ensure that the assumed coordinate representation is as close to the underlying psychological representation as possible. (Note that even for models which can operate directly on similarity ratings, such as the simplicity model, it is typically better to employ a coordinate representation, since similarity ratings tend to be very noisy.) Second, the two dimensions of the stimuli ought to both broadly cohere together (so that it does not become an analytic process to perceive the stimuli holistically; Milton & Wills, 2004), and be perceived separately. The latter constraint arises because some models of unsupervised categorization allow spontaneous attentional selectivity. Accordingly, we wanted participants to be able, in principle at least, to allocate differential attentional salience to the two dimensions. This was verified independently by two other people that the stimuli could be perceived without effort in terms of both their dimensions concurrently and, also, that each dimension could also be attended to independently, if so desired. Finally, the stimuli had a neutral interpretation as schematic drawings (their semblance to real-life spiders was intentionally low) and also an interpretation as biological objects (with a little imagination, as spiders). This allowed an instructional manipulation (described in the Procedure).

We provided a test of whether the similarity structure of the spider-like stimuli conformed to a coordinate representation based on the length of the central bodies and the length of the legs (given the Weber fraction assumed above). We created a stimulus set of 12 stimuli which spanned all regions of the available (assumed) psychological space, as shown in Figure 7a. We then asked 30 participants (all Swansea University students, who did not take part in the categorization experiment) to provide similarity ratings for these stimuli. Specifically, each participant was shown all possible stimulus pairs in this set of 12 stimuli, excluding identities: there were 12x12 – 12 (identities) = 132 trials. Stimulus presentation and response recording were computer based. The structure of each trial was to present a fixation point for 250ms, followed by the two stimuli in a pair one after the other for 1000ms each, followed by a 1-9 Likert ratings space. Similarity results from all participants were then averaged and subjected to a multidimensional scaling (MDS; two dimensions) procedure, leading to Figure 7b; the stress associated with the MDS solution was 0.115, indicating that the spatial arrangement derived from the MDS algorithm is a good representation of the similarity information.

It can be seen that the coordinate representation (Figure 7a) and the one based on similarity ratings (Figure 7b) are very similar. We next employed the Orthosim procedure (Barrett et al., 1998) which allows the computation of various similarity indices between two sets of coordinates for the same set of items. We selected a similarity index which adopts a 'procrustes' approach (Barrett et al., 1998), according to which the coordinate configurations to be compared are first normalized and rotated/ reflected to remove any of the arbitrariness in MDS solutions. The Orthosim documentation recommends the 'double-scaled Euclidean distance' coefficient, for which 0 corresponds to complete dissimilarity, 1 to identity. This coefficient was 0.911, indicating close correspondence between the assumed coordinates and the similarity-ratings based representation. Overall, the results of this analysis support our representation assumptions.



Figures 7a, 7b. In the (a) panel we show the assumed coordinate representation of a sample of all stimuli. In the (b) panel we show the derived MDS representation for the same stimuli, from similarity ratings provided by participants. Numbers indicate stimulus ids.

Procedure

Participants received the items in each stimulus set in a pile. The two dimensions of the stimuli were described to participants and it was emphasized that they were equally important. In one instructional manipulation, the stimuli were described as 'objects' and the two dimensions as 'rectangle length in the center and thin parallel lines length on the sides'. In another instructional manipulation, a scenario was presented to participants saying how new spiders are discovered all the time around the world. Participants were then told about a recent expedition to the Amazon, during which several new spiders were identified. All these new spiders had broadly similar structure, but differed in terms of the length of their bodies and legs.

In both instructional conditions, participants were told to consider the stimuli in each set independently, that is, as if the current stimulus set was the only one they had received (this was done to avoid participants thinking that, e.g., if they used *X* groups in one stimulus set, they should also use *X* groups in another). They were asked to spread the items in front of them and classify the items in a way that seemed natural and intuitive, using as many groups as they wanted, but not more than necessary. It was stated that more similar objects should end up in the same group. Participants were told to indicate their groupings by arranging the objects in each group in separate piles.

Results

Because of the way classifications were recorded there were inevitably some transcription errors. For a stimulus set for which there were no transcription errors there were 169 classifications from participants to analyze; for the rest of the stimulus sets there were one to two classifications missing, except for the 'embedded' stimulus set for which there were 15 classifications missing. We decided not to carry out some scaling of the dependent variables (which are presented below) because the missing classifications are more *likely* to be ones which were more random and so not contribute to the frequency of the preferred classifications. Indeed, the 'embedded' stimulus set was the one for which we observed the greatest number of alternative preferred classifications (the highest frequency with which any classification was produced for this stimulus set was only two).

With 16 stimuli there are approximately 10.4 billion possible classifications. The vastness of this search space informs the complexity of the classification problem, a problem which cognitively appears trivial. Indeed, across the nine stimulus sets there were over 1100 unique classifications (many of which appear to reflect random individual variation in classification strategy; cf. Pothos & Chater, 2002).

We are interested in deriving from this data an empirical measure of classification intuitiveness. That is, under what circumstances can we say that a particular classification is psychologically more intuitive than another? The most obvious choice for a dependent variable is to measure the frequency of the preferred classification in each stimulus set. If in stimulus set A, the preferred classification is produced by 10/169 participants and in stimulus set B the preferred classification by 50/169 participants, then we can trivially conclude that the preferred classification in B is more obvious/ intuitive to naïve observers. A related dependent variable is agreement between participants for how a particular stimulus set should be classified. In other words, if in stimulus set A participants produce 40 distinct classifications and in stimulus set B participants produce 10 distinct classifications, then clearly in the latter case participants agree more on how the A stimuli should be classified, and so the corresponding classification structure must be more intuitive. A question is whether the two dependent variables of 'frequency of preferred' and 'distinct classifications' are independent or not. In principle they might be, for example, if there are more than one obvious ways to classify a set of stimuli. However, this was not the case in our stimuli: the correlation between 'frequency of preferred' and 'distinct classifications' was .97, p<.0005. Henceforth, we shall only consider the frequency of the preferred classification as the dependent variable.

Another question is how *informative* the variable of frequency of the preferred classification for different stimulus sets is. For example, suppose that the frequency of the preferred classification in stimulus set A is 40/169. This would be very informative if there were no other high frequency classifications for stimulus set A and, clearly, less informative if there were other classifications which were produced with a frequency of, e.g., 39, 38, 37 etc. In the latter case, one would be forced to conclude that there is nothing particularly special about the highest frequency classification, in light of the fact that there would also be several other very high frequency contenders. An interesting empirical finding of this research is that this latter scenario was *not* true. In other words, for the stimulus sets for which participants showed a preference for any classification, this classification was overwhelmingly preferred—there were no alternative classifications which competed with the most frequent one. That a particular classification can dominate so much in the well-structured stimulus sets (that is, stimulus sets for which there was an obvious classification) was a surprising finding, given the otherwise very high performance variability. Table 4 shows the most frequent classification for each of the nine stimulus sets.

Finally, we briefly consider the issue of the instructional manipulation. The simplest examination of the effect of this manipulation would be to consider the distinct classifications produced by participants receiving the neutral and the realistic instructions. We can then ask whether there is any difference in the pattern of responding for the two sets of instructions. Table 5 shows this was not the case. Correlating the number of distinct classifications for the nine stimulus sets with realistic and neutral instructions we obtained r=.91, p=.001. In other words, there was the same degree of classification variability for a particular stimulus set, regardless of instructions. A similar analysis can be conducted for the number of times the preferred classification was produced for different stimulus sets, as a function of instructions (Table 6; *r*=.94, *p*<.0005). Perhaps this is not a surprising finding. Both sets of instructions emphasized groupings according to similarity, described the two dimensions of physical variation, and provided similar instructions to participants regarding the ideal number of groups. The instructional manipulation will not be considered further.

Stimulus set	Most frequent classification
Two clusters	{0 1 2 3 4 5 6 7} {8 9 10 11 12 13 14 15}
Unequal clusters	{0 1 2 3 4 5 6 7 8 9} {10 11 12 13 14 15}
Spread out clusters	{0 1 2 3 4 5 6 7} {8 9 10 11 12 13 14 15}
Three clusters	{0 1 2 3 4} {5 6 7 8 9} {10 11 12 13 14 15}
Ambiguous points	{0 1 2 4 7} {3 5 6} {8 9 10 12} {11 13 14 15}
	{0 1 4 7} {2 3 5 6} {8 9 10 12} {11 13 14 15}
Poor two clusters	{0 1 2 3 4 5 6 7 8 9 10 11} {12 13 14 15}
Five clusters	{0 1 2} {3 4 5} {6 7 8} {9 10 11} {12 13 14 15}
Random	{0 4} {2 3} {1 5 6 7 8 9} {10 11 14} {12 13 15}
Embedded	{0 1} {2} {3 4} {5} {6 7} {8} {9} {10 11 12 13 14 15}
	{0 1} {2} {3} {4} {5} {6 7} {8} {9} {10 11 12 13 14 15}
	{0 1} {2} {3} {4} {5} {6} {7} {9} {8} {12 11 10 14 15 13}
	{0 1 2 3 10 11 12 15} {4 5 6 7 8 9 13 14}
	{0} {1} {2} {3} {4} {5} {6} {7} {8} {9} {10 11 12} {13} {14 15}

Table 4. The most frequent classifications for each of the nine stimulus sets. The category membership of each stimulus is indicated by a number id; these ids are the same as the ones in Figure 5. In cases in which more than one classification appears, this means that there were more than one classifications with the highest frequency of occurrence observed for that stimulus set.

Stimulus set	Realistic instructions	Neutral instructions
Two clusters	60	63
Unequal clusters	59	54
Spread out clusters	77	74
Three clusters	50	54
Ambiguous points	84	76
Poor two clusters	74	66
Five clusters	49	34
Random	80	78
Embedded	77	72

Table 5. The table shows the number of distinct classifications produced for different stimulus sets, as a function of the two sets of instructions participants could receive.

Stimulus set	Realistic instructions	Neutral instructions
Two clusters	19	12
Unequal clusters	17	16
Spread out clusters	5	3
Three clusters	32	23
Ambiguous points	1	2
Poor two clusters	11	6
Five clusters	27	31
Random	2	1
Embedded	1	1

Table 6. The frequency of the preferred classification for different stimulus sets, as a function of the two types of instructions participants could receive.

Table 7. Summary of the empirical results of the study in unsupervised categorization.

Stimulus set	Frequency of most preferred[1]	Frequency of next most
preferred[1]	Distinct classifications produced	
Two clusters	31	5
	123	
Unequal clusters	33	7
	113	
Spread out clusters	8	3
	151	
Three clusters	55	4
	104	
Ambiguous points	3	3
	160	
Poor two clusters	17	3
	140	
Five clusters	58	8
	83	
Random	3	2
	158	
Embedded	2	2
	149	

Notes: 'Preferred' corresponds to the classification preferred by participants for the corresponding stimulus set.

Modeling

We will say that there is (a lot of) category structure in a stimulus set if there is a classification for the stimuli which is particularly intuitive. In other words, category structure is an impression of whether is *some* good classification for the stimuli. Considering the results in Table 7 and the intuitive impression of the stimuli in Figure 5 readily leads to some puzzling questions. For example, the stimulus sets 'two clusters' and 'unequal clusters' both conform to a simple two-group classification and, indeed, participants were reasonably good at identifying this classification. However, the seemingly not dissimilar two-group classification for the stimulus set 'spread out clusters' turned out to be much less intuitive. Moreover, the more complex three-group and five-group classifications for the 'three clusters' and 'five clusters' stimulus sets respectively were the star performers. They were preferred by participants with a frequency which exceeded that for the preferred classifications in all the other stimulus sets. Equally, for the 'ambiguous points' stimulus set we expected that participants would identify some category structure; after all, this is a stimulus set with a reasonably obvious two-group category structure, but with some ambiguous points in between. However, in this case participants were hardly able to consistently able to identify any classification as salient. These findings illustrate that the challenge to formal models of unsupervised categorization will be profound.

The structure of model application to this data can take two forms. First, the simplicity model receives as input the coordinates of the nine stimulus sets. The model then produces a number which would reflect the intuitiveness of the preferred classification for the stimulus set. The objective of the model would be to produce category intuitiveness predictions which match as closely as possibly the empirically determined variable of category intuitiveness (i.e., the frequency of the preferred classification for each stimulus set). In other words, there are effectively nine data points with which we try to test each model.

Second, the model receives as input the coordinates and the preferred classification(s) for each stimulus set. It then computes a value of intuitiveness for a stimulus set and a particular classification. This second approach is relevant for models which can produce an intuitiveness value for particular classifications, but they are unable to predict what should be the preferred classification for a stimulus set from scratch. Note that the modeling challenge we are presently interested in is to correctly predict differences in the intuitiveness of the preferred classification across the nine stimulus set. (The related problem of predicting the preferred classification for a stimulus set from scratch is, arguably, less interesting anyway; cf. Pothos & Bailey, 2009). A related issue is that for some stimulus sets there were more than one classifications which were produced with the highest frequency. All such cases corresponded to stimulus sets with very poor category structure. Accordingly, reasonably, there is no sense in which we can assign a special status to any of these preferred classification and so we computed intuitiveness values for all of them and considered the model prediction (for the stimulus set) to correspond to the highest such value.

Simplicity model

The simplicity model was first developed to account for the spontaneous categorization results of Pothos and Chater (2002). These stimulus sets were composed of only 10 stimuli each and also the range of category structures employed was limited. Accordingly, in order to apply the simplicity model to the present stimulus certain extensions were required. It is still assumed that the primary determinant of classification goodness is codelength, so that the lower the codelength the more intuitive the particular classification will be to participants. However, it also had to be recognized that slight perturbations in the coordinates of the stimuli can lead to different predicted classifications. The effect of such perturbations, or noise, will depend on the similarity structure of the stimuli: for some stimulus sets noise does not affect the predicted classification, while for others even modest perturbations can lead to several different classifications. Accordingly, it appears that different classifications are more *stable* against noise.

Why would stability against noise be a significant consideration when modeling empirical results? Because different participants will basically perceive the available stimuli in slightly different ways. Even though the MDS analyses show that the assumed coordinate representations of the stimulus sets broadly match the psychological representations, inevitably there will be individual differences variation in stimulus perception. Therefore, regardless of how low the codelength of a particular classification is, we expect more variability in participants' responses in situations where perturbing the stimulus coordinates alters the predicted preferred classification.

The above considerations were implemented in the following way. We constructed a regression model to predict the frequency of the preferred classification for each stimulus set, on the basis of two predictors. The first predictor is the codelength of the best possible classification for a stimulus set, not the preferred classification. The reason why we applied the model in this way is that the simplicity model *has* to predict that the preferred classification ought to be the optimal one. This first predictor effectively corresponds to how the simplicity model has been originally applied (e.g., Pothos & Chater, 2002). The second predictor is a measure of the stability against noise of the category structure in a stimulus set.

To compute the second predictor, for a stimulus set, we perturbed the two coordinates of each stimulus independently, by adding a noise term of up to 10% of the range of the corresponding dimension (the value 10% was chosen because of its consistency with the Weber fraction employed in designing the stimuli). Noise could be positive or negative (i.e., the coordinate would change by at most +10% x range or -10% x range) and the new coordinates were scaled back so that the range of the new coordinates along each dimension would be the same as before (i.e., no overall stretching or shrinking of psychological space). This procedure was repeated 1000 times for each stimulus set and we simply counted the number of distinct classifications as a measure of stability against noise. For example, if the number of distinct classifications was just one, then the predicted optimal classification would be the same whether the original coordinates were employed or any of the 1000 alternative perturbed coordinates. Accordingly, in such a case we would say that the optimal classification should be extremely stable against noise.

The two predictors were combined in a linear regression model. The predictions from the simplicity model for a stimulus sets were taken to correspond to the predicted classification frequency values from the regression model—clearly, higher values correspond to more intuitive classifications. Note that the regression model was significant ($F(2,6)=10.46, p=.011$). Finally, we can ask whether the classification predicted as optimal from the simplicity model is the same as the classification preferred by participants. This was the case for all stimulus sets for which there was high classification structure (i.e., 'two clusters', 'unequal clusters', 'spread out clusters', 'three clusters', 'poor two clusters', and 'five clusters'). In the case of the stimulus sets 'random' and 'embedded' there were small differences between the optimal classification predicted by the simplicity model and the empirically preferred one (the codelength associated with the former was only very slightly lower than the codelength associated with the latter). However, for the stimulus set 'ambiguous points' there was a large difference between the simplicity prediction and empirical result. In that case, the preferred classification was produced with a frequency of three, so that one would have less confidence that this is indeed the classification most obvious to participants, as opposed to one which emerged as most popular simply by chance.

6.4 Summary

For this investigation, we were interested in examining the validity of the simplicity model, and whether introducing general knowledge would affect the spontaneous classifications made. An unsupervised categorization task was employed to examine observer agreement concerning the categorization of nine different stimulus sets. The stimulus sets were designed to capture different intuitions about classification structure. The main empirical index of category structure was the number of times the most frequent classifications was produced, for different stimulus sets. With 169 participants, and a within participants design, with some stimulus sets the most frequent classification was produced over 50 times and with others not more than two or three times. For some stimulus sets, there was good correspondence between model predictions and participant performance, but our results also revealed weaknesses in the simplicity model. Also, introducing general knowledge did not affect the way in which the classifications were made.

Chapter 7

Experimental results: unsupervised categorization vs. supervised categorization

7.1 Introduction

This Chapter provides the experimental results of our supervised vs. unsupervised investigation (this excludes the work carried out in relative vs. absolute shifting, which is given in Chapter 5 and the unsupervised categorization results of Chapter 6). Briefly, our goals here were to investigate whether there was a relationship between supervised and unsupervised categorization. So, after exploring the validity of the model with the larger set in Chapter 6, I then used the same stimuli, but adopted a supervised categorization procedure (a learning task). This was to test whether the intuitiveness of the categories would affect how easy it is to learn and remember the supervised categories.

7.2 Supervised and unsupervised categorization

Chapter 2 gives a thorough account of the simplicity model in unsupervised categorization, and explains some of the concepts below such as category coherence in more depth. Chapter 3 gives an account of supervised categorization. The literature in categorisation has, to a large extent, been organised around the distinction between supervised and unsupervised categorisation. For example, most categorisation models are specifically proposed as either models of supervised categorisation (e.g., Ashby et al., 1998; Minda & Smith, 2000; Nosofsky, 1988), or of unsupervised categorisation (e.g., Anderson, 1991; Pothos & Chater, 2002). As a consequence, supervised and unsupervised categorisation processes, previously, have been studied in separate research traditions, and only few studies have attempted to explore possible convergence between the two forms of categorisation.

In unsupervised categorization, one of the most fascinating aspects of human cognition is how we develop the concepts and categories with which we understand the world. Research into unsupervised categorization concerns the processes which enable us to spontaneously recognize/ create groupings in a set of stimuli and offers the promise to help us appreciate the causal principles underlying the richness and diversity of human conceptual knowledge. In the laboratory, in unsupervised categorization experiments there are no predetermined categories. Participants are presented with a set of stimuli and are asked to divide them into categories which appear natural and/ or intuitive (the number of categories can be fixed or unconstrained). In real life, unsupervised categorization would relate to the process which allows us to spontaneously consider a set of patterns as belonging together (cf. perceptual grouping) or to category coherence, that is the 'glue' which binds together the members of a category. The notion of category coherence has intrigued psychologists, since its initial proposal by Murphy and Medin (1985). Why do we consider a category like 'chairs' as intuitive (coherent) but a category composed of 'babies, the moon, and rules' nonsensical? A simple answer might be similarity. Even though exclusive reliance on similarity has been criticized (Barsalou, 1985; Murphy & Medin, 1985), there is no doubt that this is an incredibly powerful principle in understanding human categorization.

Research in unsupervised categorization concerns a range of topics. The focus of the present work is the spontaneous preference for certain classifications, as opposed to others, and whether such a preference can be applied to supervised categorization. Ultimately, it is hoped that understanding what drives this preference will help understand the issue of category coherence. Other research issues studied in unsupervised categorization relate to the spontaneous attentional dimensional selection (e.g., Milton & Wills, 2004; Pothos & Close, 2008) and the role of general knowledge in category coherence (e.g., Yang & Lewandowsky, 2004; Wisniewski & Medin, 1994). Note, finally, that most categorization research has so far concerned supervised categorization, which involves the teaching of predetermined categories. For example, in the laboratory, an experimenter might decide that certain stimuli are in category 'A' and other stimuli in category 'B'. In the real world, a toddler might be told by her mum that this round, yellow object, with the funny smell is a 'lemon'. In supervised categorization the key research question concerns how novel instances are classified in relation to existing categories. The default assumption would be that supervised and unsupervised categorization correspond to separate cognitive processes.

Modeling work in supervised categorization has progressed at an impressive rate. For example, the computational properties of supervised categorization models have been thoroughly scrutinized. For example, in relation to the debate between exemplar and prototype theory, there have been several studies examining the computational behavior of the models and sometimes even the role of specific individual parameters (e.g., Ashby & Alfonso-Reese, 1995; Lee & Vanpaemel, 2008; Olsson, Wennerholm, & Lyxzen, 2004; Minda & Smith, 2000; Navarro, 2007; Nosofsky, 1990, 2000; Smith, 2007; Vanpaemel & Storms, 2008). With no doubt, this work has been extremely useful. Individual researchers may have their preferences regarding, e.g., exemplar vs. prototype theory, but the crucial point is that there is a wealth of computational analyses to make an informed decision. We suggest that one reason for the sophistication of formal work in supervised categorization has been the existence of 'standard' datasets, capable of discriminating between model predictions. For example, Medin and Schaffer's (1978) famous 5-4 category structure and Shepard, Hovland, and Jenkins's (1961) finding that certain classifications are easier to learn than others, have been examined in dozens of studies (e.g., Johansen & Kruschke, 2005; Nosofsky, 2000; Smith & Minda, 2000; but see Homa, Proulx, & Blair, 2008). One might argue that so much emphasis on a particular dataset may be distracting and ultimately reduce the ecological validity of the resulting models/ model revisions. However, at the same time, there is unquestionable value in the existence of modeling 'standards' against which new proposals can be evaluated.

There is an abundance of empirical data if one is interested either in unsupervised learning (e.g., Billman & Knutson, 1996; Knowlton & Squire, 1994; Reber, 1967) or spontaneous categorization with some constraints (such as the number of categories to be produced; e.g., Ashby, Queller, & Berretty, 1999). For example, several researchers have reported the spontaneous selection of a single dimension for categorization, when participants are asked to divide objects in two groups (e.g., Medin, Wattenmaker, & Hampson, 1987; Milton & Wills, 2004; Regehr & Brooks, 1995; but see, Murphy, 2004, Pothos & Close, 2008). However, there are very few datasets with an entirely unconstrained unsupervised categorization procedure, which could serve as modeling standards in the development of unsupervised models of categorization (in the way, for example, that the Shepard et al., 1961, or the Medin and Schaffer, 1978, results have guided supervised categorization models). Compton and Logan (1999), Pothos and Chater (2002) did employ an entirely unsupervised categorization procedure, however, in both these cases there were problems: Compton and Logan employed a procedure which could only loosely be considered a spontaneous grouping of stimuli into categories (they presented participants with dot diagrams and asked them to draw curves around the dots which should be grouped together) and Pothos and Chater employed a very limited number of participants; as we shall see, a key empirical problem with unsupervised categorization experiments is that there is considerable variability in participants' responses.

So, as given above, there is much evidence on both unsupervised and supervised categorization, but little effort has been made to explore the relationship between these two different forms of categorization.

7.3 Experimental results supervised vs. unsupervised categorization

The extensive computational work in supervised categorization has led to a clear understanding of the differences and similarities of different models. In fact, the majority of studies in supervised categorization have been driven by a desire to test specific differences between supervised categorization models. This has not been the case in unsupervised categorization.

Note that some limited computational comparisons have been carried out for the rational model, the simplicity model, and an unsupervised version of the GCM (Pothos & Bailey, 2009; Pothos, 2007). These analyses did not show a particular model as superior. For example, Pothos (2007) presented a systematic examination of the models against a series of artificial stimulus sets. The stimulus sets were specified to conform to obvious intuitions about category coherence (e.g., if there are two clusters, the shorter the distance between the clusters, the less coherent the resulting classification). Under such circumstances, the predictions about category intuitiveness from the simplicity model and the rational model were nearly identical. Pothos and Bailey (2009) used data from previous studies. But, in their comparison, the only truly unsupervised data came from Compton and Logan (1999), who employed a rather artificial categorization task (stimuli were dots in a diagram and participants were asked to indicate their classifications by circling around the dots) and Pothos and Chater (2002), who employed probably too few participants for robust unsupervised categorization results.

Regardless of these limitations, there is an important conclusion we can make from these two studies: models of unsupervised categorization tend to agree on what is the best way to classify a single set of stimuli (not least because, as noted, the category structures employed in studies of unsupervised categorization tend to be fairly intuitive). Where they differ is regarding their predictions for the relative intuitiveness, or naturalness, of different classifications. For example, two classifications for the same set of stimuli set can vary in category intuitiveness but, equally, two classifications for different stimulus sets can vary in intuitiveness. In Chapter 2 we demonstrated two examples of category intuitiveness, given in Figure 2 (this is given again below to illustrate this point again, and more specifically to the experimental work). In Figure 2, classification A is more intuitive than alternative classification B for the same stimulus set. It is with respect to such predictions that Pothos and Bailey (2009) identified differences between models of unsupervised categorization. Therefore, the important conclusion is that models of unsupervised categorization are best evaluated with respect to how *intuitive* they predict different classifications will appear to naïve observers, across a number of different classifications for the same stimuli and different stimuli. The question that motivates the experimental work in the present chapter, is whether such notions of 'intuitiveness' can be applied to supervised categorization.

Research in categorization has been organized on the basis of a distinction between supervised and unsupervised categorization. The former concerns learning pre-specified categories. In a laboratory setting, an experimenter may have decided that certain stimuli are in one category, while other stimuli in a different one. The objective of a participant is to learn which stimuli go to which category, usually through a process of corrective feedback (that is, a participant sees a stimulus, guesses its category membership, and receives feedback as to whether his/her guess was correct or not). In real life, arguably many linguistic categories are taught through a process of supervised categorization. For example, a child can learn that certain objects are oranges and other objects are lemons, by guessing the category membership of a relevant novel exemplar and subsequently receiving corrective feedback from an adult (cf. Demetras, Post, & Snow, 1986; Gleitman, Newport, & Gleitman, 1984). A key aspect of supervised categorization is that there are no (apparent) limits on the complexity of the classifications which can be taught (e.g., Ashby, Queller, & Berretty, 1999; McKinley & Nosofsky, 1995).

Unsupervised categorization concerns the spontaneous impression we often have that a group of stimuli belong to the same category. Such an intuition is most obvious in perceptual grouping, whereby sometimes we have an immediate impression that there are clusters (e.g., see Figure 8; cf. Compton & Logan, 1999). With respect to real concepts, Murphy and Medin (1985) advocated the idea of category coherence: that is, for most real concepts, there is a 'glue' that binds the members of a concept together. As with the perceptual grouping example of Figure 8, certain real life concepts are more coherent than others. For example, there is very little ambiguity regarding membership into the category of 'chairs'. However, many naive observers will disagree as to what should be considered (a member of the concept) 'literature'. In experimental studies of unsupervised categorization, an experimenter is typically constrained to consider naturalistic classifications, that is, classifications which will be plausibly spontaneously produced by participants (e.g., Pothos & Chater, 2002; Pothos & Close, 2008).

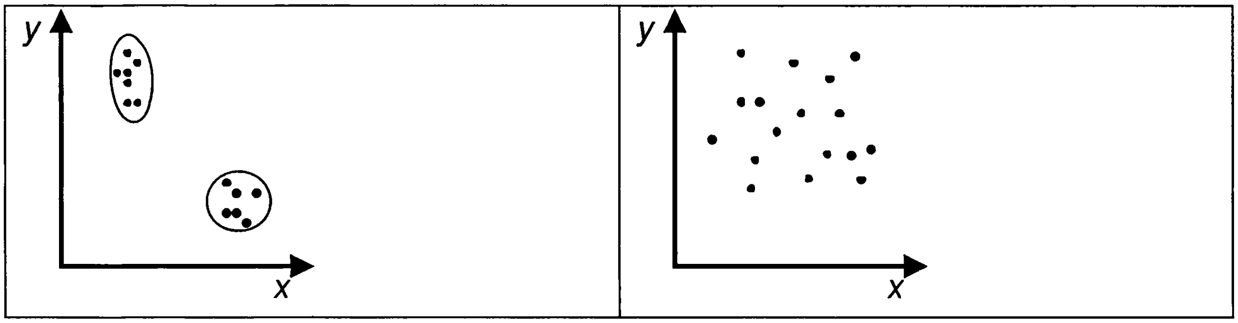


Figure 8. Assume that the diagrams correspond to some putative psychological space and that each dot corresponds to an instance in our experience. There an immediate impression that there are two clusters on the left panel, but this is not so for the right panel.

We are interested in the extent to which the distinction between supervised and unsupervised categorization is meaningful. This is an issue of central importance in the study of categorization, since, for example, it affects researchers' perception of whether there should be separate models for supervised and unsupervised categorization or not. In motivating the present experiments, we will consider relevant neuroscience, computational, and experimental work.

We can first consider whether what is known about the neuroscience of categorization can provide some clues as to whether supervised and unsupervised categorization should be considered separate cognitive processes. With respect to supervised categorization, researchers have been interested in whether one can use neuroscience methods to understand what is different about rules-based category learning and category learning based on knowledge of individual exemplars (cf. Pothos, 2005). For example, Koenig et al. (2005; see also Smith, Patalano, & Jonides, 1998) found that classifying novel instances on the basis of a rule activated the anterior cingulate cortex, parietal areas, and left inferior frontal areas, while classification on the basis of similarity to previously encountered exemplars involved anterior prefrontal areas, the posterior cingulate cortex, and bilateral temporal-parietal areas.

Participants in the rule condition of Koenig et al. were explicitly told of which rule to use (cf. Allen & Brooks, 1991). One could ask of whether there are situations when naive observers required to learn a classification might spontaneously do so in terms of a rule. Ashby and colleagues have been advocating an influential paradigm, termed COVIS (COmpetition between Verbal and Implicit Systems; Ashby et al., 1998; Zeithamova & Maddox, 2006), according to which category learning can proceed either through the development of an explicit, verbal rule (cf. Smith et al., 1998) or an exemplar similarity strategy (in the COVIS framework this is termed 'information integration'). The rule strategy is supported primarily by the prefrontal cortex, the anterior cingulate cortex, and the head of the caudate nucleus. For example, the prefrontal cortex has been widely implicated in planning, differentiating amongst conflicting goals, and identifying expectations based on actions (Banich, in press). By contrast, the information integration strategy involves the inferotemporal cortex and the tail of the caudate nucleus. This is a procedural learning system, which presumably involves the nigrostriatal dopamine pathway.

The two systems of COVIS appear to provide a reasonable framework of the range of classifications systems which might be involved in supervised category learning. We can ask whether there is any evidence that the brain areas involved in unsupervised categorization might be distinct or overlap with the areas postulated in COVIS for supervised categorization. A telling study by Op de Beeck et al. (2008) revealed that perceptual organization in the lateral occipital cortex was based on similarity (of course, in earlier visual areas, organization of information is retinotopic). Can we associate the spontaneous emergence of intuition that a set of stimuli should be categorized in a certain way, with this similarity-based organization in the later visual areas? It's unclear that we can do this, but, equally, it's unclear as to which other areas might support the spontaneous emergence of classification intuitions. Overall, the neuroscience results may tentatively indicate that separate systems support supervised and unsupervised categorization, but this conclusion is greatly undermined by the lack of neuroscience research regarding unsupervised categorization.

We can next examine the principles underlying computational models of supervised and unsupervised categorization. Influential supervised categorization models, such as exemplar theory and prototype theory (Minda & Smith, 2000; Nosofsky, 1988; see also, Van Vanpaemel & Storms, 2008), typically assume that categorization of novel exemplars is driven by their similarity to either the members or the prototype of the available categories. Similarity is typically computed as a function of distance in a putative psychological space. However, such models allow for the possibility that the process of category learning may transform the original psychological space, through the attentional weighting of different dimensions or overall stretching or compression of the space. Such transformations would take place in a way to support the process of category learning (e.g., the attentional salience of a dimension would increase if it is highly diagnostic for a required classification).

Models of unsupervised categorization also often employ a principle of similarity. For example, Pothos and Chater's (2002) simplicity model is based on the idea of Rosch and Mervis (1975) that more obvious classifications should be ones for which within category similarity is maximum and between category similarity is minimum. Other models of unsupervised categorization, such as the rational model, predict categories which maximize the posterior probability of the particular feature combination of their members, given category membership (Anderson, 1991; Sanborn, Griffiths, & Navarro, 2006; cf. Corter & Gluck, 1992). However, Pothos (2007) compared the rational model and the simplicity model and found that the predictions of these models converged across a wide range of stimulus sets. Moreover, Pothos and Close (2008) postulated a mechanism of spontaneous attentional weighting of dimensions in unsupervised categorization. According to Pothos and Close, a dimension may be spontaneously entirely ignored if it does not contribute to the intuitiveness of a classification for a set of stimuli (cf. Milton & Wills, 2004). Note, however, that the graded attentional weighting that seems to be possible in supervised categorization has not been observed in unsupervised categorization.

So, at this broad level of analysis, supervised and unsupervised categorization models appear to be based on similar principles. Love, Medin, and Gureckis (2004) were the first to try to provide a single computational framework for both supervised and unsupervised categorization, with their SUSTAIN model. However, crucially, there are separate components of SUSTAIN responsible for each type of categorization. Regarding unsupervised categorization, categories emerge for groups of items which are similar to each other. Supervised categorization is supported by a learning mechanism similar to that embodied in current versions of the exemplar theory (e.g., Nosofsky, 1988). In principle, SUSTAIN can allow the interaction between supervised and unsupervised categorization (a parameter controls the relative influence of each mechanism). Therefore, according to Love et al. (2004) there are separate computational mechanisms for supervised and unsupervised categorization.

Pothos and Bailey (2009) provided a contrasting perspective. They examined whether an influential version of exemplar theory, the Generalized Context Model (GCM; Nosofsky, 1988), could be modified to describe results in unsupervised categorization. They called their model unsupervised GCM and compared its predictions against those of two (proper) unsupervised categorization models, the simplicity model and the rational model. Overall, the comparisons of Pothos and Bailey did not reveal a model to be superior relative to the two others—the performance of the unsupervised GCM was approximately equivalent to that of the simplicity model and the rational model. Pothos and Bailey's comparisons, therefore, show that a model of supervised categorization can, with relatively little modification, be applied in the context of unsupervised categorization.

Logically, a model of supervised categorization can always be applied in unsupervised categorization, and vice versa. For example, a supervised categorization model can be used to produce an 'intuitiveness' prediction for a particular classification, by considering each instance one-by-one as a novel instance and classifying it to its respective category; this operation will result to an error term (which may be zero). Repeating this procedure for all stimuli, the sum of error terms can be used as a measure of classification intuitiveness, in the sense that when the error term is low we can say that the classification is more consistent with the model's assumptions (this is the procedure by which Pothos & Bailey, 2009, applied the GCM to unsupervised categorization data). Conversely, a model of unsupervised categorization can, in principle, be applied to predict the classification of new instances by examining how the intuitiveness of a classification is changed by assigning a novel instance to different categories.

Of course, as noted, there are possible differences between supervised and unsupervised categorization, such as the issue of attentional weighting of stimulus dimensions noted above. Moreover, the requirements of learning a particular supervised categorization may lead participants to develop complex category representations, for example, based on rules or combinations of elementary rules (Ashby et al., 1998; Kurtz, 2007). In sum, considering the computational principles relevant in supervised and unsupervised categorization provides mixed intuitions regarding a possible equivalence between supervised and unsupervised categorization.

Love (2002) carefully examined this issue. One of his hypotheses was that in supervised learning there should be no difference between linearly separable and non-linearly separable category structures (this result is supported by the data of Medin & Schwanenflugel, 1981), while in unsupervised learning linearly separable category structures appear more plausible. His results supported this hypothesis, so that Love concluded that supervised and unsupervised categorization are better understood as separate cognitive processes. However, there are some problems with this conclusion.

First, the conclusions of Medin and Schwanenflugel have been challenged, with later research indicating that in supervised categorization as well, linearly separable category structures are easier to learn than nonlinearly separable ones (Blair & Homa, 2001). Second, Love created an unsupervised categorization task using the Shepard, Hovland, and Jenkins (1961) dataset, which is a well known dataset in supervised categorization. However, importantly, he augmented the stimuli with an extra dimension of variation, which was meant to correspond to the intended classification. This manipulation effectively alters the similarity structure of the stimuli quite drastically: in all cases, it creates a very easy (and linearly separable) categorization of the stimuli into the required categories. In other words, if participants were to focus only on this additional (labels) dimension of variation, there would be no need for them to consider any of the other information about the Shepard et al. stimuli. Indeed, the intended structure of the Shepard et al. stimuli (as linearly separable categories, nonlinearly separable categories etc.) would be lost. Finally, the tasks Love employed corresponded only loosely to the more standard procedures in unsupervised categorization research. For example, participants were asked to either memorize or rate the pleasantness of the stimuli. Then, in test, pairs of stimuli were presented such that they were identical except that in one stimulus the 'classification' dimension had one value and in the other the classification dimension had the other possible value; the task was an old-new recognition task. He found that recognition accuracy was different with the memorization or pleasantness learning tasks, compared to a standard supervised categorization task. Interesting as this manipulation is, it clearly corresponds more to an incidental learning cognitive process rather than an unsupervised categorization one. The properties that emerge as more salient as a result of a memorization or irrelevant learning task (based on pleasantness) is an issue quite different from that of whether a classification is more intuitive than another.

To sum up, Love's (2002) is not a definitive test of the (lack of) equivalence between supervised and unsupervised categorization. The research reported in this paper broadly follows the design of Love's study. However, we have tried to incorporate a range of extensions which should lead to a better test of the equivalence between supervised and unsupervised categorization. Colreavy and Lewandowsky (2008) provided another comparison between supervised and unsupervised categorization, in the context of the development of learning strategies with increased exposure to a set of stimuli. In their unsupervised condition, participants could decide how to classify each stimulus into either of two available categories. In the supervised condition, participants were asked to learn twp-cluster classifications for the same stimuli. Colreavy and Lewandowsky found many similarities between the supervised and unsupervised categorization conditions, including, for example, with respect to learning rates.

The research reported in this paper broadly follows the design of Love's study. However, we have tried to incorporate a range of extensions which should lead to a better test of the equivalence between supervised and unsupervised categorization. The first extension is that the basis of the current investigation is the dataset of unsupervised categorization results, presented in Chapter 6. To our knowledge, this is currently the most extensive study of unsupervised categorization and, therefore, it provides a rich dataset against which to examine possible relations with supervised categorization. A particular advantage of this dataset is that it includes stimulus sets for which the empirically preferred classification does not always have two clusters—for some stimulus sets the preferred classification has as many as five clusters. Second, we employed exactly the same stimulus sets for unsupervised and supervised categorization. Thus, the comparison of human performance between the two types of categorization is better controlled (recall that Love, 2002, had to change the representation of the Shepard et al., 1961, stimuli, for his test of unsupervised categorization). The unsupervised categorization results are reported in detail in Chapter 6; in this work, we simply employ the main conclusions from this study, and compare them with the corresponding results from two matched supervised categorization tasks (which constitute the novel empirical work reported in this paper).

Third, there is the issue of which variables to use to characterize supervised and unsupervised categorization. The former is a straightforward issue. In this context, supervised categorization performance can be adequately characterized by the difficulty associated with learning different classifications. In this work we also employed an additional dependent variable to characterize supervised categorization, corresponding to the memory of a particular classification. The latter is a more complex issue, not least because of the enormous variability which is typically associated with unsupervised categorization experiments (Pothos & Chater, 2002). In Chapter 6, I suggested two possible dependent variables, the frequency of the preferred classification for a stimulus set and the number of distinct classifications produced by participants for a stimulus set. The logic behind both variables is the same: if for a stimulus set there is a very intuitive classification, then one would expect this classification to be produced very frequently and, equally, that there should be less disagreement in how the stimuli are classified. In fact, in Chapter 6 I reported that these two variables correlated extremely highly with each other. In this work we follow these investigators and also suggest that human performance in unsupervised categorization can be characterized by the frequency of the preferred classification in different stimulus sets.

To sum up, the purpose of this research is to provide the most straightforward possible test of the possible equivalence between supervised and unsupervised categorization. Our starting point is a large dataset on unsupervised categorization, which is reported elsewhere (Chapter 6). In this research we describe two experiments with matched supervised categorization tasks. Our overall approach follows that of Love (2002), although we have tried to improve on his specific procedure in several respects. There are two experiments that follow. Experiment 1, compares the relationship between the results of the unsupervised task with a standard supervised learning task. Experiment 2, compares the relationship between the difficulty of learning the categories (i.e. the results of Experiment 1) with the memory for category labels.

7.4 Experiment 1 unsupervised vs. supervised learning; learning condition

Participants

Participants were 180 Swansea University undergraduates, who had not taken part in any related experiments. They participated in the study for course credit or a small payment. Experimental design was between participants, so that each participant was tested with only one stimulus set (exactly 20 participants were tested with each stimulus set).

Materials

The materials employed in this study are identical to those of Chapter 6 but for the fact that in this study stimuli were presented individually on a computer screen, while in Chapter 6 unsupervised categorization study each stimulus was printed individually on a sheet of paper. We took care to ensure that the physical size of the stimuli as shown on the computer screen and as printed on the sheets of paper were the same.

We briefly summarize the stimulus details (for more information please see Chapter 6). Stimuli were created so as to broadly resemble spiders; the two relevant dimensions of variation were the length of the 'legs' (after the joints) and the length of the central body. We adopted lengths as the relevant dimensions of variations, since this makes it relatively straightforward to assume a Weber fraction (in both cases 8%; Morgan, 2005). For both dimensions, the actual lengths were between 40mm and 80mm. An example of the stimuli is shown in Figure 6 of Chapter 6. The stimuli were intentionally created to resemble some real-life creature, as a manipulation to increase the coherence of the two dimensions. It was important that the two stimulus dimensions could be perceived together without analytic effort (cf. Milton & Wills, 2004; Pothos & Close, 2008).

The key design aspect of this research concerns the range of stimulus sets employed. In Chapter 6, I employed nine different stimulus sets, each having 16 stimuli, which were meant to capture a range of intuitions regarding unsupervised categorization. For example, in one stimulus set there was a fairly salient two-cluster classification, in another a two-cluster classification whose salience was undermined by some ambiguous points, in a third a five-cluster classification etc. The considerations guiding the selection of stimulus sets are considered extensively in Chapter 6. In this work we aim to simply employ the results regarding category intuitiveness from this research (summarized in Chapter 6) and motivate the creation of matched supervised categorization tasks. The nine stimulus sets can be referred to as 'two clusters', 'unequal clusters', 'spread out clusters', 'three clusters', 'ambiguous points', 'poor two clusters', 'five clusters', 'random', and 'embedded'. These names are meant to correspond to the key aspect (in terms of prior, experimenter intuitions) of category structure in each stimulus set. All stimulus sets are shown in Figure 9.

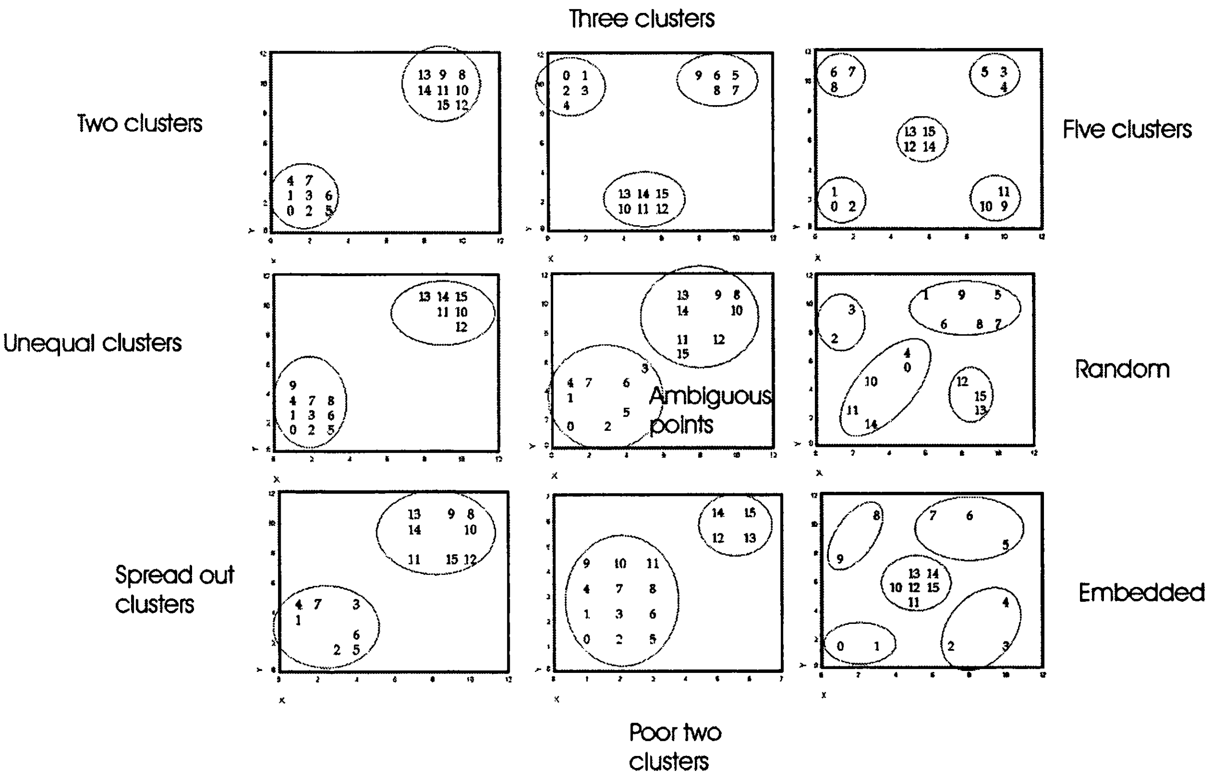


Figure 9. A schematic representation of the nine stimulus sets employed in this research. Each point in each stimulus set is indexed by a number from 0 to 15. The curves show the classifications taught to participants in each case.

Procedure

We adopted a standard supervised categorization procedure. The experiment was organized in units, such that each unit consisted of one presentation of all the stimuli with their correct category labels, and two presentations of the stimuli without the labels—in the latter case, the participant had to guess the correct label and corrective feedback was provided after each response (as is standard in experiments of supervised categorization). When participants were not required to make a response each stimulus was presented for 1000ms, when participants were required to respond, a stimulus would be shown until a response was made. The learning criterion was to go through all the stimuli in a learning unit without making any errors (the experimenter was able to determine when this happened, because a sound indicated an incorrect response). When a participant managed to do this, the experiment stopped. Otherwise, the participant would be presented again with the stimuli in a unit. A different randomized order of stimulus presentation was employed each time.

The classifications taught to participants for each stimulus set are shown in Figure 9. Note that the number of categories varies from two to five. These are the classifications predicted as most intuitive by the simplicity model (Pothos & Chater, 2002). The simplicity model has been shown to predict the classification preferred by participants in all cases in which there is a salient category structure. Moreover, for stimulus sets for which there is no salient category structure, there tends to be very high variability in participant classifications. In such cases, it *appears* that a certain classification may be preferred not because of any intrinsic structural properties but, rather, by chance. This observation provides justification to use the classifications predicted by the simplicity model in the supervised categorization task, rather than the ones preferred by participants.

It is clearly an empirical issue whether this assumption can be justified in general. Regardless, it does seem to be appropriate in the present case: For the stimulus sets shown in Figure 9, the simplicity model correctly predicted the preferred classifications in the cases of the 'two clusters', 'unequal clusters', 'spread out clusters', 'three clusters', 'poor two clusters', and 'five clusters'. In the case of the stimulus sets 'random' and 'embedded' there were small differences between the optimal classification predicted by the simplicity model and the empirically preferred one. For the stimulus set 'ambiguous points' there was a large difference between the simplicity prediction and empirical result. Importantly, in the three cases in which there was a discrepancy between the prediction of the simplicity model and the classification preferred by participants, the frequencies with which the preferred classifications were produced were just 3, 2, and 3, respectively for the 'random', 'embedded', and 'ambiguous points' stimulus sets (note that there were 169 participants in the study of unsupervised categorization in Chapter 6; therefore, a frequency for the preferred classification of 2 means that, out of 169 participants, only 2 produced this classification). To reiterate, our assumption is that when a classification is produced with a frequency as low as 2 or 3, then we are not warranted to conclude that there is something special or particularly intuitive about this classification (so that we are better off employing the predictions of a reasonably well-motivated model of unsupervised categorization, such as the simplicity model).

Results

We recorded two dependent variables, the number of learning units required to achieve criterion and the total number of errors before criterion had been achieved (note that each learning unit consisted of a presentation of all the stimuli with their labels and two presentations of the stimuli without the labels—in the second case participants had to guess the correct classification of each stimulus). There was a highly significant correlation between the two variables (r=.64, p<.0005). Accordingly, we will restrict the analyses to only one of the variables, the number of learning units required to reach criterion.

Table 8 shows how the number of units differed for the nine stimulus sets we employed. Also, it summarizes the key dependent variable from the unsupervised categorization results of Chapter 6 (this is the frequency of the preferred classification). Note, first, that there are differences between the ease of learning of different datasets: F(8,171)=35.22, p<.0005. This result confirms the expectation from Table 8, that it was much easier to learn the required classification for certain stimulus sets, compared to others.

Stimulus set	Frequency of most preferred[1]	Mean number of units[2]	Range[3]
Two clusters	31	4.10	2—10
Unequal clusters	33	4.15	2—11
Spread out clusters	8	7.40	2—17
Three clusters	55	9.30	3—21
Ambiguous two clusters	3	14.45	3—27
Poor two clusters	17	9.65	3—24
Five clusters	58	13.45	4—28
Random	3	25.40	12—33
Embedded	2	22	9—35

Table 8. A summary of the unsupervised categorization results of results from Chapter 6 and the supervise Experiment 1.

Notes: The frequency with which the preferred classification was produced. The mean number of learning criterion. The lowest and highest number of learning units required to reach criterion. The standard deviat: learning units required to reach criterion.

The critical research question concerns a possible relation between the unsupervised and supervised categorization results. From an unsupervised categorization perspective, the higher the frequency of the preferred classification, the more psychologically intuitive this classification should be. From a supervised categorization perspective, the lower the number of units required to reach the learning criterion, the easier (and hence more intuitive) the taught classification should be (cf. Pothos & Bailey, 2009). The objective in the analyses below is to examine whether these two measures of category intuitiveness, from an unsupervised and supervised categorization task, are related or not.

A simple test of a putative association between the measures of category intuitiveness from the unsupervised categorization results of Chapter 6 and the supervised categorization results from the present experiment is a correlation, for each stimulus set, between the frequency of the preferred classification and the number of learning units required to reach criterion. This correlation was low and not significant, although in the right direction (r=-.47, p=.20). However, this test does not take into account the fact that there is a differential role for the number of category labels in the supervised and unsupervised categorization procedure. Specifically, an increased number of category labels is likely to affect executive function and working memory resources, both of which would disrupt a process of supervised learning (Maddox et al., 2004). Indeed, there was a correlation between number of units required to achieve criterion and number of category labels (r=.72, p=.03). By contrast, there is no evidence that a spontaneous classification involving more clusters will be more (or less) demanding than one with fewer clusters. We therefore first regressed the number of learning units on category labels and recorded the unstandardized residuals—these residuals provide us with an estimate of the variance in the number of learning units which cannot be accounted for by differences in the number of labels. The regression was significant, as expected, showing (as before) that the number of labels participants had to keep track of affected learning difficulty (F(1,7)=7.37, p=.03). Subsequently, we correlated the residuals with the frequency of the preferred classification. The correlation was now highly significant and in the right direction: r=-.811, p=.008.

Discussion

The literature in categorization has, to a large extent, been organized around the distinction between supervised and unsupervised categorization. For example, most categorization models are specifically proposed as either models of supervised categorization (e.g., Ashby et al., 1998; Minda & Smith, 2000; Nosofsky, 1988) or models of unsupervised categorization (e.g., Anderson, 1991; Pothos & Chater, 2002). There is no doubt that the distinction between supervised and unsupervised categorization is a highly intuitive one. However, the present empirical results have failed to provide support it.

In brief, Experiment 1 was a standard supervised categorization learning paradigm. We asked different participants to learn a particular classification for nine different stimulus sets. A natural dependent variable in this context is the difficulty with which different classifications are learned (cf. Shepard et al., 1961). Certain classifications were easier to learn than others. Are these the same classifications which are spontaneously produced more frequently by participants? We utilized the unsupervised categorization results of Chapter 6 for the same stimulus sets. Factoring out the variance due to the number of category labels, we found that classifications which were easier to learn were indeed the ones more likely to be produced spontaneously. Our results therefore show that the aspects of category structure which make a classification easy to learn are the same as the ones which make a classification 'stand out' in a spontaneous categorization setting (cf. Colreavy & Lewandowsky, 2008).

In Experiment 1 we considered one possible hypothesis of how we can decide whether a categorization taught to participants is intuitive or not: if a categorization is easier to learn, then it should be more intuitive. There is an alternative perspective: we can ask whether a particular association between category labels and stimuli is more resistant to forgetting. If a classification for a set of stimuli is better remembered several days after it has been taught, then we should conclude that this classification is more intuitive. Accordingly, we can examine whether category intuitiveness in terms of remembering a taught classification is associated with category intuitiveness in terms of preference in a spontaneous categorization task. Experiment 2 addresses this issue.

7.5 Experiment 2 unsupervised vs. supervised learning; memory condition

Participants

Participants were 195 Swansea University undergraduates, who had not taken part in Experiment 1 or any other related experiments. They participated in the study for course credit or a small payment. Experimental design was between participants. Participants were divided between the nine stimulus sets as shown in Table 9.

Materials and Procedure

The materials were identical to those employed in Experiment 1. Experiment 2 consisted of two parts. First, there was a part in which participants had to learn the given classification. This part proceeded in a way analogous to that of Experiment 1, although some modifications were introduced. The learning part was organized in units consisting of a presentation of each stimulus with the correct label, followed by five presentations of all the stimuli without the labels—in these presentations, as before, participants had to guess the correct answer and received corrective feedback. Moreover, in the trials when participants did see the correct label, the stimulus and label appeared on the screen until the participant pressed the key with the corresponding label. In this way, we hoped to reinforce the stimulus—label associations. The learning criterion was, in a way analogous to what we had before, responding to all the stimuli once without making any errors. Unlike Experiment 1, a learning unit could be cut short when participants achieved the learning criterion. After the learning criterion had been achieved, participants saw all the stimuli three more times, in a way that each stimulus with its correct label appeared on the screen, and participants had to press the key with the corresponding label before proceeding to the next stimulus. This 'fixed exposure' manipulation was added to ensure that participants would experience the same number of label—stimulus associations, after they had learned the correct classification.

Participants were invited to come again to the laboratory seven days later (a deviance of one day was tolerated). To encourage participants to do so, they would not receive any compensation until they came for the second time. Nearly all participants did attend both experimental sessions. The second experimental session was identical to the learning unit described above (five presentations of all stimuli), but without the presentation of the correct stimulus—category label associations at the beginning. In other words, this was a standard recall test for the correct label for each stimulus.

Results

We first consider the dependent variables which are analogous to those in Experiment 1, the number of blocks required to achieve the learning criterion and the errors made before criterion could be achieved (note that a learning block in Experiment 2 corresponds to one presentation of the 16 stimuli, so that it differs from the learning unit defined in Experiment 1). Table 9 shows these results. As before, there was a highly significant correlation between number of blocks and errors (*r*=.92, *p*<.0005). It is also interesting to check that the supervised learning results in Experiment 2 were equivalent to those in Experiment 1, which turned out to be the case (*r*=.87, *p*=.002). This result is reassuring, since there were only superficial differences between the training procedure in Experiment 1 and that of Experiment 2.

Stimulus set	Participants	Mean number of blocks[1]	Range[2]	Standard deviat
Two clusters	25	1.36	1—3	0.64
Unequal clusters	27	2.04	1—8	1.58
Spread out clusters	32	2.22	1—11	1.93
Three clusters	13	9.23	2—37	9.33
Ambiguous two clusters	21	3.57	1—18	3.98
Poor two clusters	18	6.39	1—17	4.25
Five clusters	19	10.42	3—31	7.42
Random	20	18.15	3—47	10.99
Embedded	20	24.95	6—60	15.05

Table 9. The supervised categorization results obtained in Experiment 2.

Notes: The mean number of learning blocks required to reach the learning criterion. The lowest and highe reach criterion. The standard deviation associated. The number of errors in reproducing the category label

In Experiment 2 there was a unique dependent variable, the number of memory errors in recalling the category label—stimulus associations a week after training (Table 9). The memory variable correlated highly with the number of blocks required to reach criterion (r=.97, p<.0005). This result illustrates that classifications which were easiest to learn were indeed the easiest to remember a week later as well. Moreover, the frequency of preferred classifications correlated highly with the memory variable, once the variance due to category labels had been eliminated as in Experiment 1 (r=-.73, p=.026).

Discussion

The memory for a particular classification is a dependent variable which has not featured prominently in categorization research. However, it is an important empirical variable, since it informs our insight of what kinds of classifications might be more resistant to forgetting. Presumably, as categorization researchers, we would like to conclude that classifications which are remembered better are ones which are cognitively 'special', in some sense. A classification which is easy to learn is not necessarily the same as a classification which is resistant to forgetting. For example, categories which are closer to each other may be more prone to interference from forgetting, even if they are straightforward to learn in the first place. Equally, learning a categorization sometimes appears to involve particular transformations of the psychological space for the corresponding stimuli (e.g., Nosofsky, 1988). There has been no research as to how long-lived such transformations are. For example, a particular classification may be easy to learn after a fairly radical transformation of psychological space (e.g., involving the projection of all stimuli along a single dimension). However, if this transformation is short-lived, then one would expect that memory for the corresponding classification to likewise decay quickly.

Despite the above considerations, the present results showed that the memory for a particular taught classification correlated highly with the ease of learning the classification in the first place and, moreover, with the likelihood that the classification would be spontaneously produced in an unsupervised setting. This provides compelling demonstration that a convergence in the theoretical accounts for supervised and unsupervised categorization may be desirable, at least in some cases.

General discussion

We have examined two measures of supervised categorization, with nine different stimulus sets, and related the results to spontaneous preference for the taught classifications in an unsupervised categorization task. Each of the different categorization tasks can be seen as providing a different measure of category intuitiveness. For example, a standard supervised categorization task (Experiment 1) can discriminate between classifications which are easy to learn and ones which are more difficult to learn. Clearly, we can suggest that the former are more intuitive compared to the latter (cf. Kurtz, 2007; Shepard et al., 1961). The supervised categorization task augmented with a recall task (Experiment 2) allowed us to identify the classifications which are more resistant to memory decay and forgetting. Classifications which are better remembered must be more obvious and intuitive. Finally, the unsupervised categorization procedure that I employed in Chapter 6 provides a measure of spontaneous preference for a categorization. More intuitive categorizations would be the ones that are spontaneously produced more frequently.

All the three measures of category intuitiveness related closely to each other, consistently with the findings of Colreavy and Lewandowsky (2008). This conclusion suggests that whatever it is that makes a classification more obvious in an unsupervised task, also makes the classification is easier to learn in a supervised task. If such a conclusion proves to be general, it would have important implications for the development of models of categorization. Currently, nearly all categorization models are specifically proposed either as models of supervised categorization (e.g., Minda & Smith, 2000; Nosofsky, 1988) or models of unsupervised categorization (e.g., Anderson, 1991; Pothos & Chater, 2002). Some researchers have sought to modify models of supervised categorization so that they can function as models of unsupervised categorization (e.g., Pothos & Bailey, 2002; cf. Kurtz, 2007). Also, there have been attempts to integrate a supervised model and an unsupervised one within the same formalism (e.g., Love et al., 2004). However, few models have been proposed from the outset as purporting to account for both supervised and unsupervised categorization with exactly the same computational principles.

How general are the conclusions in this paper? A key point is that the taught classifications were all ones which were very likely to be produced spontaneously. Supervised learning can allow a naïve observer to learn classifications which would never be produced spontaneously (e.g., McKinley & Nosofsky, 1995; Maddox et al., 2004). For such very complex classifications, it seems meaningless to talk about a putative equivalence between supervised and unsupervised categorization.

A related issue is this: models of supervised categorization typically employ mechanisms which appear to go beyond those relevant in unsupervised categorization. For example, in supervised categorization researchers have advocated a process of fine tuning of the attentional salience of each stimulus dimension, non-linear compression/ stretching of the entire psychological space, response parameters which affect whether a categorization decision is more probabilistic or deterministic, and separate learning systems to distinguish between classifications which can be learned with a simple rule vs. ones which require a more passive, information integration procedure (Ashby et al., 1998; Minda & Smith, 2001 Nosofsky, 1988; Vanpaemel & Storms, 2008). It seems extremely unlikely that all these mechanisms have analogues in unsupervised categorization. Indeed, supervised and unsupervised categorization appear to share only a handful of computational principles. Similarity is one such principle, since most models of both supervised and unsupervised categorization embody some function of similarity. Attentional weighting of stimulus dimensions may be another common principle, noting, however, that only 'crude' attentional selection has been observed in unsupervised categorization (that is, a stimulus dimension may be spontaneously ignored if it does not appear to add to the overall intuitiveness of a classification; Pothos & Close, 2008). Conversely, in unsupervised categorization it has been suggested that general knowledge plays an important part (e.g., Murphy & Medin, 1985); in supervised categorization, general knowledge effects appear to be restricted to enhancing the attentional salience of certain stimulus features (e.g., Murphy & Allopenna, 1994). Note, however, that the effect of general knowledge in categorization has been incredibly difficult to formalize and so, in the absence of formal models, it is difficult to appreciate exactly how much of an effect it has on categorization (cf. Pickering & Chater, 1995; but see Harris, Murphy, & Rehder, 2008; Heit, 1997).

The upshot of the above discussion is that a putative equivalence between supervised and unsupervised categorization must only hold for classifications which 'naturalistic' in the first place (i.e., classifications which are likely to be produced naturally). A possible hypothesis forthcoming from this research is that the features of supervised categorization models which do not appear relevant in unsupervised categorization are relevant only when learning more complex classifications, that is, ones which are very unlikely to be produced naturally. Whether the learning of complex classifications is supported by the same cognitive process as that of simple classifications is very much an open issue. It is possible that learning of complex classifications should be better understood in the context of learning models in general, rather than as a cognitive process of concept formation. An alternative possibility is that categorization models should rightly incorporate the ability to learn both simple and complex concepts, so that only their features corresponding to the former ability can be extended to support unsupervised categorization processes as well. Such possibilities suggest exciting new avenues for further research.

In sum, we showed that when it comes to naturalistic classifications, supervised and unsupervised categorization processes converge. This finding raises several interesting possibilities regarding the way supervised and unsupervised models can be developed, in a way that a corresponding convergence of the relevant computational principles can be achieved.

7.6 Summary

Supervised and unsupervised categorization have been studied in separate research traditions. Only a handful of studies have attempted to explore a possible convergence between the two. This Chapter provided a research investigation which built on these studies, by comparing the unsupervised categorization results from Chapter 6 with the results from two procedures of supervised categorization. In two experiments, we tested 375 participants with nine different stimulus sets, and examined the relation between ease of learning of a classification, memory for a classification, and spontaneous preference for a classification. After taking into account the possible confounding role of the number of category labels in supervised learning, we found the three variables to be closely associated with each other. Our results provide encouragement for researchers seeking unified theoretical explanations for supervised and unsupervised categorization.

Chapter 8

Conclusions

8.1 Summary of findings

This thesis set out to explore three separate phenomena in categorization. Firstly, it tested the validity of the simplicity model (Pothos & Chater, 2002). Secondly, it investigated the relationship between unsupervised and supervised categorization. Thirdly, it explored the circumstances which would cause a relative and absolute shift in representation. The experimental traditions explored in this thesis related broadly to categorizing with learning, (supervised), without learning (unsupervised), and when learning is impaired through interference (interference in our case was implemented by increasing group size and group numbers in the relative experimental work).

8.2 Summary of the relative vs. absolute experimental work.

In Chapter 5, I explored the conditions which cause traditional 'absolute representation' of supervised categorization to be abandoned. This work was based upon the work into absolute judgment experiments and, specifically, the theory implemented by Stewart et al. (2005), regarding the use of the RJM, to explain some of the sequential effects. In these effects, judgments about the serial position of the 'current' item in a sequence is thought to be determined by the neighboring items, in terms of relational properties, such as 'bigger than and 'smaller than'. I applied this theory, in a general way, to my current investigation into relative and absolute representations, as little related work has been done in the area of categorization and in the area concerning relative, absolute representational shifts.

I generally found that reducing exemplar numbers, increasing the categories available and by introducing a time delay between presentation and response, was sufficient to induce changes in the representation. More specifically, I found that the representation of 'absolute' decision making shifted towards a more 'relative' form, where relational properties such as 'bigger than', 'smaller than' became important. I related this finding to work in the area of analogical mapping and the relative judgment model.

8.3 Summary of the work carried out in unsupervised categorization.

In Chapter 6, evidence was sought to test the accuracy of the simplicity model to predict unsupervised categorization results. This was an extension to the work carried out by Pothos and Chater (2002) in using more complex stimulus sets and using a much larger participant sample. The basic form of the model predicts that unsupervised categorization can be predicted with the 'simplicity principle', and that this principle can be formalized more specifically in categorization through the simplicity model (Pothos & Chater, 2002). The simplicity principle, suggests that, generally, given some data, the simplest hypothesis that leads to the best description (or explanation) of this data, is most likely to be the correct one. This has been formalized in the simplicity model of unsupervised categorization, which suggest that the categories are chosen on the basis that lead to greatest reduction in codelength Broadly, this means that the categories that lead to the shorted codelength are those which remove (take advantage of) redundant information.

The results did demonstrate that some aspects of the simplicity model were able to predict accurately the unsupervised categorization results, however, for some conditions it did not do as well. These results have motivated additional research (not covered in my thesis), and modeling work to accommodate these new findings.

8.4 Summary of the work carried out in supervised categorization

In Chapter 7, I examined the relationship between the unsupervised categorization results and the potential intuitiveness of supervised categorization. This is an area in categorization that has received little attention. Supervised categorization differs from unsupervised categorization, because it is categorization where the group structure and labels are indicated by the experimenter. For example, the experimenter gives corrective feedback to the participant, which indicates that item 1 belongs to a group called 'Chomps' and item two belongs to a group called 'Blibs'. This differs significantly from the work carried out in unsupervised categorization, which is based on free sort tasks, where the participants are not given any feedback, and can sort the items into categories on any basis they choose. As indicated by the simplicity model, unsupervised categorization decisions are made on the basis of information reduction, which loosely corresponds to the simplicity principle. The motivation for associating results in supervised categorization, was to identify whether the 'intuitive' categories (low codelength) were also those which could be learned and remembered more easily. This work thus was intended to find some common theme between supervised and unsupervised learning, through the simplicity principle.

The findings from this work were promising, as I found a general relationship between intuitiveness of categories predicted by the simplicity model of unsupervised categorization, and the ease with which participants could learn the supervised categories and also remember these categories a week later. This work provides promising evidence that the simplicity principle can be applied outside of unsupervised categorization, in the area of supervised categorization.

8.5 Some broader theoretical thoughts.

On the whole, these quite separate categorization research traditions have been related through the common theme of information reduction. Occam's razor (see Chapter 2), seems to be prevalent through this categorization work, and has not been explicitly explored across so many domains of categorization work. Occam's razor is a general theoretical attempt to formalize the notion that simple explanations are generally preferred. In unsupervised categorization this has formally been implemented in the form of simplicity principle, which was the basis for the simplicity model. However, to date, no such attempts have explored how this general theory of information reduction can be applied more generally throughout categorizations work. The work carried out on the relationship between unsupervised and supervised categorization in Chapter 7 formally investigated the application of such a theory to supervised categorization, with promising results. It is more difficult to relate the simplicity principle with the work on relative vs. absolute categorization formally (i.e., through a specific mathematical framework). However, in principle, this more formal approach could be applied, where in some situations relational properties could be considered 'simpler' by the cognitive system. A simple example of this is where the category 'Chomps' is smaller than the category 'Blibs', this requires just one bit of information to compute whereas in absolute mode of representation, each absolute representation would have to be considered according to exemplar theory, and thus would require more computation.

8.6 Future work and directions; potential applications of the measured used in this thesis work

8.6.1 Theoretical

This work provides a great deal of evidence which can lead to more investigations at both the theoretical and applied levels. For example, at the theoretical level, more work could be done to bridge the gap between the simplicity principle and relational representation. One plausible question is to what extent can we apply the simplicity principle, formally, outside of unsupervised categorization. Of course, the work here looked at the relation between supervised and unsupervised categorization, and, also, we have some speculative results that indicate that representation can change to 'simpler forms' (i.e., relative representations) when the information content is overly complex, but it is yet unclear if these ideas can be formalised more specifically.

Another area of direct theoretical work, could be to first refine the simplicity model, to fit the data more accurately, and then examine the ability of the model to account for changes in stimulus presentation and number of dimensions. For example, can the simplicity model predict the unsupervised categorization, accurately, using stimuli that is comprised of three of four dimensions? Also, would attention weights (as used in the GCM) need to be introduced to account for such more complex stimuli. At this stage, this is unclear.

There is yet another area that could be explored, which deals with modelling background information. For example, the simplicity model could, potentially be extended to deal with general knowledge effects through the simplicity principle. However, a formal account of such an approach would be clearly very difficult (Dreyfus & Dreyfus, 1986; Heit, 1997; Heit & Bot, 1999; McDermott, 1987; Oaksford & Chater, 1991, 1998; Pickering & Chater, 1995). The closest attempt to date at solving this problem directly is from Heit (1999), who used an exemplar account for addressing the knowledge selection problem. Heit's Baywatch model involves a supervised process where the experimenter provides the background information to the program, which allows its expert systems to select sub descriptions for the categories given. In one example, the category 'buildings', could be subdivided into 'unique buildings' by the expert system, such as identifying 'churches' as different to 'schools'. This involves a process where new information is integrated with old and observed category members have a greater effect when they are consistent with background information (see Heit, 1994). So, the simplicity model, could in principle be applied in a similar way.

8.6.2 Applied

In one potential application, the simplicity model could be applied directly to the area of autism, where there is much debate over the mechanisms behind over-selectivity. For instance, several suggestions attempt to explain why individuals with ASD have problems in discrimination learning with complex cues. These include attention deficits (Dube et al., 1999; Lovaas et al., 1971), encoding problems (Boucher & Warrington, 1976; Reed & Gibson, 2005), and post-processing or retrieval problems (e.g., Leader et al., 2009). Unsupervised categorization could be employed in this area to determine whether over-selectivity is caused by deficits in attention as there is already evidence from supervised categorization experiments showing learning deficits (Klinger & Dawson 1995; 2001; Bott et al., 2006).

Similarly, as the autistic population have shown to have limited 'absolute representation', demonstrated by their reduced supervised categorization performance, this could lead to more 'relative mode' representations when using the relative vs. absolute experimental paradigm. So, there is a lot more room for extending this work to other populations, especially in relation to cognitive deficits, such as an autistic population.

Likewise, this could also be applied to the area of traumatic brain injury, where over-selectivity has been shown by Wayland and Taplin (1985), and potentially many other clinical areas. This would therefore have important implications in terms of interventions that could be used. This could lead to further research into how to ameliorate dysfunctional over-selectivity in category learning. Therefore, there is a clear impact on potential 'users' of this research in a directly practical and applied way. I plan to use the experimental data to investigate further what interventions are most appropriate for the specific attention- or learning-based deficits, as it might be the case that certain individuals with ASD would need specially catered interventions, based on their specific attention or learning deficit needs.

8.7 Closing comments

The present investigation has explored themes in the categorization area. This includes work in unsupervised, supervised categorization, as well as relational shifting (within supervised categorization). I have demonstrated that the simplicity principle is useful as a general means of describing and predicting categorization. However, it is clear that much additional work needs to be carried out, with additional dimensions, and modelling work in the area of background knowledge. There is also potentially a lot of applied work that can be considered in the areas of autism, and clinical populations.

References

Ahn, W. & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science*, *16*, 81-121.

Allen, S. W., Brooks, L. R. (1991). Specializing the Operation of an Explicit Rule. *Journal of Experimental Psychology: General, 120*, 3-19.

Alluisi, E. A., & Sidorsky, R. C. (1958). The empirical validity of equal discriminability scaling. *Journal of Experimental Psychology, 55,* 86–95.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1991). The Adaptive Nature of Human Categorization. *Psychological Review, 98*, 409-429.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Arabie, P., Hubert, L.,&de Soete, G. (Eds.) (1996). *Clustering and Classification.* River Edge, NJ:World Scientific.

Ashby, G. F. & Alfonso-Reese, A. L. (1995). Categorization as Probability Density Estimation. *Journal of Mathematical Psychology*, *39*, 216-233.

Ashby, G. F., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A Neuropsychological Theory of Multiple Systems in Category Learning. *Psychological Review, 105*, 442-481.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, *37*, 372-400.

Ashby, G. F. & Perrin, N. A. (1988). Towards a Unified Theory of Similarity and Recognition, *Psychological Review*, *95*, 124-150.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, *61*, 1178-1199.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154-179.

Atick, J. J., & Redlich, A. N. (1990). Towards a theory of early visual processing. *Neural Computation. 2*, 308-320.

Atran, S. & Medin, D.L. (2008). T*he Native Mind and the Cultural Construction of Nature.* Boston, MA.: MIT Press.

Attneave, F. (1959). *Applications of information theory to psychology.* New York: Holt, Rinehart & Winston.

Banfield, C. F., & Bassill, S. (1977). A transfer algorithm for non-hierarchical classification. *Applied Statistics*, *26*, 206–210.

Banich, M. T. (in press). Executive Function: The search for an integrated account. *Current Directions in Psychological Science.*

Barrett, P. T., Petrides, K. V., Eysenck, S. B. G., & Eysenck, H. J. (1998). The Eysenck Personality Questionnaire: An examination of the factorial similarity of P, E, N, and L across 34 countries. *Personality and Individual Differences*, *25*, *5*, 805-819.

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

Billman, D. & Knutson, J. (1996). Unsupervised concept learning and value systematicity: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 458-475.

Blair, M. & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition, 29*, 1153-1164.

Blair, M. & Homa, D. (2003). As easy to memorize as they are to classify: The 5-4 categories and the category advantage. *Memory & Cognition, 31*, 1293-1301.

Bott, L., Brock, J., Brockdorff, N., Boucher, J., Lamberts, K. (2006). Perceptual similarity in autism. *The Quarterly Journal of Experimental Psychology, 59*, *7*, 1237–1254.

Bowdle, B. F., & Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology*, *34*, 244–286.

Braida, L. D., & Durlach, N. I. (1972). Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America, 51,* 483–502.

Brooks, L. R. (1987). Decentralized control of categorization: The role of prior processing episodes. In U. Neisser (Ed.), *Concepts in conceptual development: Ecological and intellectual factors in categorization* (pp. 141–174). Cambridge, MA: Cambridge University Press.

Brown, G. D. A., Neath, I., & Chater, N. (2002). *A ratio model of scale-invariant memory and identification.* Manuscript submitted for publication.

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

Buffart, H. F. J. M., Leeuwenberg, E. L. J., & Restle, F. (1981). Coding theory of visual pattern recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 241–274.

Carroll, J. D., & Wish, M. (1974). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 57-105). San Francisco: W. H. Freeman.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science, 10,* 408–411.

Chater, N. (1996). Reconciling Simplicity and Likelihood Principles in Perceptual Organization. *Psychological Review*, *103*, 566-591.

Chater, N. (1999). The Search for Simplicity: A Fundamental Cognitive Principle? *Quarterly Journal of Experimental Psychology*, *52A*, 273-302.

Colreavy, E., & Lewandowsky, S. (2008). Strategy development and learning differences in supervised and unsupervised categorization. *Memory & Cognition, 36*, 762-775.

Colunga, E., Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review, 112,* 347–382.

Compton, B. J. & Logan, G. D. (1999). Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation. *Perception Psychophysics, 61*, 1320-1335.

Corter, J. E. & Gluck, M. A. (1992). Explaining Basic Categories: Feature Predictability and Information. *Psychological Bulletin, 2*, 291-303.

Cheeseman, P., & Stutz, J. (1995). *Bayesian classification (AutoClass): Theory and results.* In M. F. Usama, P. S.

Compton, B. J., & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping. *Perception & Psychophysics*, *53*, 403–421.

Compton, B. J., & Logan, G. D. (1999). Judgments of perceptual groups: Reliability and sensitivity to stimulus transformation. *Perception & Psychophysics*, *61*, 1320–1335.

Corter, J. E., & Gluck, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, *2*, 291–303.

Demetras, M. J., Post, K. N., & Snow, C. E. (1986). Feedback to first language learners: the role of repetitions and clarification questions. *Journal of Child Language, 13*, 275-292.

Descartes, R; Lafleur, L. J. (translation) (1960). *Discourse on Method and Meditations*. New York: The Liberal Arts Press.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* New York: The Free Press.

Dube, W.V., Lombard,K.M., Farren,K.M., Flusser,D.S., Balsamo, L.M., & Fowler,T.R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behavior Bulletin, 17,* 8-14.

Duda, R. O., & Hart, P. E. (1973). *Pattern recognition and scene analysis.* New York: Wiley.

Doumas, L. A. A., & Hummel, J. E. (2005). Modeling human mental representations: What works, what doesn't, and why. In K. J. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 73–94). New York: Cambridge University Press.

Doumas, A. A., Hummel, J. E., Sandhofer, C. M. (2008). *Psychological Review. 115,* 1, 1–43

Durlach, N. I., & Braida, L. D. (1969). Intensity perception. I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America, 46,* 372–383.

Elliott, S. W., & Anderson, J. R. (1995). Effect of memory decay on predictions from changing categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 815–836.

Eriksen, C. W., & Hake, H. W. (1955a). Absolute judgments as a function of stimulus range and the number of stimulus and response categories. *Journal of Experimental Psychology, 49,* 323–332.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107-140.

Everitt, B. (1993). *Cluster analysis* (3rd ed.). London: Heinmann.

Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure mapping engine: Algorithm and examples. *Artificial Intelligence, 41,* 1–63.

Feigenson, L., Carey, S., & Spelke, E. S. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology, 44,* 33–66.

Feldman, J. (1997). The structure of perceptual categories. *Journal of Mathematical Psychology, 41,* 145–170.

Feller, W. (1970). *An introduction to probability theory and its applications.* New York: Wiley.

Fisher, D. (1987). Knowledge acquisition *via* incremental conceptual clustering. *Machine Learning, 2,* 139–172.

Fisher, D. (1996). Iterative optimization and simplification of hierarchical clusterings. *Journal of Artificial Intelligence Research*, *4*, 147–179.

Fisher, D., & Langley, P. (1990). The structure and formation of natural categories. In B. Gordon (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 241–284). San Diego, CA: Academic Press.

Fisher, D., Pazzani, M., & Langley, P. (1991). *Concept formation: Knowledge and experience in unsupervised learning*. San Mateo, CA: Morgan Kaufmann.

Fodor, J. A. (1983*). The modularity of mind.* Cambridge, MA: The MIT press.

Fowlkes, E. B. & Mallows, C. L. (1983). A method for comparing two hierarchical clusterings, (with Comments and Rejoinder). *Journal of the American Statistical Association*, *78*, 553-584.

Fraboni, M. & Cooper, D. (1989). Six clustering algorithms applied to the WAIS-R: The problem of dissimilar cluster analysis. *Journal of Clinical Psychology*, *45*, 932-935.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 234–257.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology, 46,* 373–380.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: LEA.

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

Gennari, J. H. (1991). Concept formation and attention. *Proceedings of the thirteenth annual conference of the cognitive science society* (pp. 724–728). Hillsdale, NJ: Erlbaum.

Gennari, J., Langley, P., & Fisher, D. (1989). Models of incremental concept formation. *Artificial Intelligence*, *40*, 11–62.

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

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

Gentner, D. (2003). Why we're so smart. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in the mind: Advances in the study of language and thought* (pp. 95-235). Cambridge, MA: MIT Press.

Gentner, D., & Brem, S. K. (1999). Is snow really like a shovel? Distinguishing similarity from thematic relatedness. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the twenty-first annual conference of the cognitive science society* (pp. 179–184). Mahwah, NJ: Erlbaum.

Gentner, D., & Rattermann, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelation in development* (pp. 225-277). London: Cambridge University Press.

Ghahramani, Z., & Beal, M. (2000). Variational inference for Bayesian mixture of factor analysers. In S. A. Solla, T. K. Leen, & K. R. Muller (Eds.), *Advances in neural information processing systems* (Vol. 12, pp. 449–455). Cambridge, MA: MIT Press.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1-38.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 2*, 43-79.

Gluck, M. A., & Corter, J. E. (1985). Information, uncertainty, and the utility of categories. *Proceedings of the seventh annual conference of the cognitive science society* (pp. 283–287). Hillsdale, NJ: Erlbaum.

Goldstone, R. L. (1993). The role of similarity in categorization: Providing a groundwork. *Cognition*, *52*, 125–157.

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

Goldstone, R. L. (1995). Effects of categorization on color perception. *Psychological Science*, *6*, 298-304.

Goldstone, R. L. (2000). Unitization during category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 86-112.

Goldstone, R. L., Lippa, Y., Shiffrin, R. M. (2001). Altering object representations thought category learning. *Cognition*, *78*, 27-43.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman, *Problems and projects* (pp. 437–447). Indianapolis: Bobbs-Merrill.

Gosselin, F., & Schyns, P. G. (1997). Debunking the basic level. *Proceedings of the 19th meeting of the cognitive science society* (pp. 277–282). Hillsdale, NJ: Erlbaum.

Gosselin, F. & Schyns, P. G. (2001). Why do we SLIP to the basic-level? Computational constraints and their implementation. *Psychological Review, 108*, 735-758.

Graham, R. L., Knuth, D. E., & Patashnik, O. (1994). *Concrete mathematics: A foundation for computer science*. Wokingham: Addison-Wesley.

Gravetter, F., & Lockhead, G. R. (1973). Criterial range as a frame of reference for stimulus judgment. *Psychological Review, 80,* 203–216.

Gureckis, T.M. and Love, B.C. (2003). Towards a Unified Account of Supervised and Unsupervised Learning. *Journal of Experimental and Theoretical Artificial Intelligence, 15*, 1-24.

Gureckis, T. M. & Goldstone, R. L. (2008). The effect of the internal structure of categories on perception. *In Proceedings of the 30th Annual Meeting of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.

Hahn, U., & Chater, N. (1997). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts, and categories* (pp. 43–92). Hove, UK: Psychology Press/MIT Press.

Halford, G. S. (2005). Development of thinking. In K.J. Holyoak & R. G. Morrison (Eds.*),* *The Cambridge handbook of thinking and reasoning* (pp. 529-558). New York: Cambridge University Press.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21,* 803–864.

Hampton, J. A. (1999). The role of similarity in natural categorization. In M. Ramscar,U. Hahn, E. Cambouropoulos,& H. Pain (Eds.), *Similarity and categorization.* Oxford: Oxford University Press.

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics, 12*, 108–116.

Handel, S., & Preusser, D. (1969). The effects of sequential presentation and spatial arrangements on the free classification of multidimensional stimuli. *Perception & Psychophysics, 6*, 69–72.

Handel, S., & Preusser, D. (1970). The free classification of hierarchically and categorically related stimuli. *Journal of Verbal Learning and Verbal Behavior, 9*, 222–231.

Harris, H. D., Murphy, G. L., & Rehder, B. (2008). Prior knowledge and exemplar frequency. *Memory & Cognition*, *36*, 1335-1350.

Hartman, E. B. (1954). The influence of practice and pitch-distance between tones on the absolute identification of pitch. *American Journal of Psychology, 67,* 1-14.

Harmon, G. (1965). The inference to the best explanation. *Philosophical Review. 74*, 88-95.

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley.

Heit, E. (1994). Models of the effects of prior knowledge on category learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, *20*, 1264-1282.

Heit, E. (1997). Knowledge and Concept Learning. In K. Lamberts & D. Shanks (Eds.), *Knowledge, Concepts*, and Categories (pp. 7-41). London: Psychology Press.

Heit, E., & Bott, L. (1999). Knowledge selection in category learning. In D. L. Medin (Ed.), *Psychology of learning and motivation.* San Diego, CA: Academic Press.

Hines, P., Pothos, E. M., & Chater, N. (2007). A non-parametric approach to simplicity clustering. *Applied Artificial Intelligence*, *21*, 729-752.

Hintzman, D. L. (1986). Schema-abstraction in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural goodness. *Journal of Experimental Psychology*, *46*, 361–364.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics, 3,* 409–414.

Holyoak, K. J., & Hummel, J. E. (2000). The proper treatment of symbols in a connectionist architecture. In E. Dietrich & A. Markman (Eds.), *Cognitive dynamics: Conceptual change in humans and machines* (pp. 229–263). Mahwah, NJ: Erlbaum.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Homa, D., Proulx, M. J., & Blair, M. (2008). The modulating influence of category size on the classification of exception patterns. *The Quarterly Journal of Experimental Psychology*, *61*, 425-443.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory*, *7*, 418–439.

Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2*, 322-330.

Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, *51*, 708–719.

Hu, G. (1997). Why is it difficult to learn absolute judgment tasks? *Perceptual and Motor Skills, 84,* 323–335.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review, 104,* 4 27–466.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–263.

Imai, S.,&Garner,W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, *69*, 596–608.

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects of the judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 92–104.

Jones, G. V. (1983). Identifying basic categories. Psychological Bulletin, *94*, 423-428.

Johansen, M. K. & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1433-1458.

Katz, J. (1972). *Semantic theory*. New York: Harper & Row.

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

Kent, C., & Lamberts, L. (2005). An exemplar account of the bow and set-size effects in absolute identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 289–305.

Klinger, L. G., & Dawson, G. (2001). Prototype formation in autism. *Development and Psychopathology*, *13*, 111–124.

Koenig, P., Smith, E. E., Glosser, G., DeVita, C., Moore, P., McMillan, C., Gee, J., & Grossman, M. (2005). The neural basis for novel semantic categorization. *NeuroImage, 24*, 369-383.

Koffka, K. (1965) *Principles of Gestalt psychology*. New York: Harcourt, Brace & World. (Original work published 1935).

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information and Transmission, 1, 1*, 1-7.

Kurtz, K. J. (2007). The divergent autoencoder (DIVA) model of category learning. *Psychonomic Bulletin & Review, 14*, 560-576.

Knowlton, B. J. & Squire, L. R. (1994). The Information Acquired During Artificial Grammar Learning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20*, 79-91.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99*, 22-44.

Krzanowski,W. J., & Marriott, F. H. C. (1995). *Multivariate analysis, Part 2: Classification, covariance structures and repeated measurements*. Arnold: London.

Lacouture, Y., & Marley, A. A. J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science, 3,*401–433.

Lacouture, Y., & Marley, A. A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383–395.

Lacouture, Y., & Marley, A. A. J. (2004). Choice and response time processes in the identification and categorization of unidimensional stimuli. *Perception & Psychophysics, 66,* 1206–1226.

Lacouture, Y., Li, S. C., & Marley, A. A. J. (1998). The roles of stimulus and response set size in the identification and categorisation of unidimensional stimuli. *Australian Journal of Psychology, 50,* 165–174.

Laming, D. R. J. (1984). The relativity of "absolute" judgements. *British Journal of Mathematical and Statistical Psychology, 37,* 152–183.

Laming, D. R. J. (1997). *The measurement of sensation.* London: Oxford University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind.* Chicago: University of Chicago Press.

Lee, M. D., & Vanpaemel, W. (2008). Exemplars, prototypes, similarities and rules in category representation: An example of hierarchical Bayesian analysis. *Cognitive Science 32*, 1403-1424.

Leeuwenberg, E. (1969). Quantitative specification of information in sequential patterns. *Psychological Review, 76*, 216–220.

Leeuwenberg, E. (1971). A perceptual coding language for perceptual and auditory patterns. *American Journal of Psychology, 84*, 307–349.

Leeuwenberg, E., & Boselie, F. (1988). Against the likelihood principle in visual form perception. *Psychological Review, 95*, 485-491.

Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications* (2nd ed.). New York: Springer.

Lockhead, G. R. (1984). Sequential predictors of choice in psychophysical tasks. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 27–47). Hillsdale, NJ: Erlbaum.

Lockhead, G. R., & King, M. C. (1983). A memory model of sequential effects in scaling tasks. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 461–473.

López, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folk biological taxonomies and inductions. *Cognitive Psychology*, *32*, 251–295.

Lovaas, O.I.., & Schreibman, L., (1971) Stimulus overselectivity of autistic children in a two stimulus situation. *Behaviour Research and Therapy. 4, 9, 305-310.*

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review, 9*, 829-835.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309-332.

Lovett, M. C., & Anderson, J. R. (2005). Thinking as a production system. In K. J. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 401-429). New York: Cambridge University Press.

Mach, E. (1959/1906). *The analysis of sensations and the`relation of the physical to the psychical*. New York: Dover Publications.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the fifth berkeley symposium in mathematical statistics and probability* (pp. 281–297). Berkeley: University of California Press.

Maddox, W. T., Filoteo, J. V., Hejl, K. D., Ing, A. D. (2004) Category Number Impacts Rule-Based but not Information-Integration Category Learning: Further Evidence for Dissociable Category Learning Systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 227-235.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of classification. *Perception & Psychophysics*, *53*, 49-70.

Marley, A. A. J., & Cook, V. T. (1984). A fixed rehearsal capacity interpretation of limits on absolute identification performance. *British Journal of Mathematical and Statistical Psychology, 37,* 136–151.

Marley, A. A. J., & Cook, V. T. (1986). A limited capacity rehearsal model for psychological judgments applied to magnitude estimation. *Journal of Mathematical Psychology, 30,* 339–390.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

McDermott, D. (1987). A critique of pure reason. *Computational Intelligence, 3,* 151–160.

McGill, W. J. (1954). Multivariate information transmission. *Psychometrika, 19,* 97–116.

McKinley, S. C. & Nosofsky, R. M. (1995). Investigations of exemplar and decision bound models in large, ill-defined category structures. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 128-148.

Medin, D. L. (1983). *Structural principles of categorization.* In B. Shepp & T. Tighe (Eds.), Interaction: Perception, development and cognition (pp. 203-230). Hillsdale, NJ: Erlbaum.

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

Medin, D. L., & Ross, B. H. (1997). *Cognitive psychology* (2nd ed.). Fort Worth: Harcourt Brace.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207–238.

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

Medin, D. L., Wattenmaker, W. D., & Michalski, R. S. (1987). Constraints and preferences in inductive learning: An experimental study of human and machine performance. *Cognitive Science, 11*, 299–339.

Mervis, C. B., & Crisafi, M. A. (1982). Order of acquisition of subordinate-, basic-, and superordinate-level categories. *Child Development, 53*, 258–266.

Michalski, R., & Stepp, R. E. (1983). Automated construction of classifications: Conceptual clustering versus numerical taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*, 396–410.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review, 63,* 81–97.

Milligan, G. L. & Cooper, M. C. (1986). A Study of the compatibility of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research, 21*, 441-458.

Milton, F. & Wills, A. J. (2004). The influence of stimulus properties on category construction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 407-415.

Munakata, Y., & O'Reilly, R. C. (2003). Developmental and computational neuroscience approaches to cognition: The case of generalization. *Cognitive Studies, 10,* 76–92.

Minda, J. P., & Smith, J. D. (2000). Prototypes in category learning: The effects of category size, category structure, and stimulus complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 775-799.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review, 67,* 16–31.

Morgan, M. J. (2005). The visual computation of 2-D area by human observers. *Vision Research, 45*, 2564-2570.

Murphy, G. L. (1982). Cue validity and levels of categorization. *Psychological Bulletin*, *91*, 174-177.

Murphy, G. L. (2004). The big book of concepts. MIT Press: Cambridge, USA.

Murphy, G. L. & Allopenna, P. D. (1994). The locus of knowledge effects in concept learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 904-919.

Murphy, G. L. & Medin, D. L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review, 92*, 289-316.

Neisser, U. (1967). *Cognitive psychology.* Appleton-Century-Crofts New York

Navarro, D. J. (2007). Similarity, distance, and categorization: a discussion of Smith's (2006). warning about "colliding parameters". *Psychonomic Bulletin & Review, 14*, 823-833.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *10*, 104-114.

Nosofsky, R. M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Journal of Experimental Psychology: Perception and Psychophysics*, *38*, 415–432.

Nosofsky, R. M. (1986). Attention, similarity, and the identification categorization relationship. *Journal of Experimental Psychology: General,* *115,* 39–57.

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

Nosofsky, R. M. (1988b). Similarity, frequency, and category representation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 54–65.

Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34*, 393-418.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance, 17*, 3-27.

Nosofsky, R. M. (2000).Exemplar representation without generalization? Comment on Smith and Minda's (2000) Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1735-1743.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 18*, 211-233.

Nosofsky, R. M., & Johansen, M. K. (2000) Exemplar-based accounts of "multiple-system" phenomena in perceptual categorization. *Psychonomic Bulletin & Review 7, 3*, 375-402

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus exception model of classification learning. *Psychological Review, 101*, 53-79.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language, 6*, 1–38.

Oaksford, M., & Chater, N. (1998). *Rationality in an uncertain world.* Hove, UK: Psychology Press.

Olsson, H., Wennerholm, P., & Lyxzen, U. (2004). Exemplars, prototypes, and the flexibility of classification models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 936-941.

Op de Beeck, H., Torfs, K., & Wagemans, J. (2008). Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. *The Journal of Neuroscience, 28*, 10111-10123.

O'Reilly, R. C., & Busby, R. S. (2002). Generalizable relational binding from coarse-coded distributed representations. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems (NIPS)* (Vol. 14, pp. 75–82). Cambridge, MA: MIT Press.

Pickering, M., & Chater, N. (1995). Why cognitive science is not formalized folk psychology. *Minds and Machines*, *5*, 309–337.

Pollack, I. (1952). The information of elementary auditory displays. *Journal of the Acoustical Society of America, 24,* 745–749.

Pollack, I. (1953). The information of elementary auditory displays: II. *Journal of the Acoustical Society of America, 25,* 765–769.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 141–180). Hillsdale, NJ: Erlbaum.

Pomerantz, J. R.,& Kubovy, M. (1986). Theoretical approaches to perceptual organization: Simplicity and likelihood principles. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, volume II: Cognitive processes and performance* (pp. 1–45). New York: Wiley.

Posner, M. I.,& Keele, S.W. (1968a). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*, 304–308.

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

Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral & Brain Sciences*, *28*, 1-49.

Pothos, E. M. (2007). Occam and Bayes in predicting category intuitiveness. *Artificial Intelligence Review, 28*, 257-274.

Pothos, E. M. & Bailey, T. M. (2009). Predicting category intuitiveness with the rational model, the simplicity model, and the Generalized Context Model. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Pothos, E. M. & Chater, N. (2002). A Simplicity Principle in Unsupervised Human Categorization. *Cognitive Science, 26*, 303-343.

Pothos, E. M. & Chater, N. (2005). Unsupervised categorization and category learning. *Quarterly Journal of Experimental Psychology*, *58A*, 733-752.

Pothos, E. M. & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition, 107*, 581-602.

Pothos, E. M., & Hahn, U. (2000). So concepts aren't definitions, but do they have necessary or sufficient features? *British Journal of Psychology*, *91*, 439–450.

Quine,W. V. O. (1977). Natural kinds. In S. P. Schwartz (Ed.), *Naming, necessity, and natural kinds* (pp. 155–175). Ithaca, NY: Cornell University Press.

Quinlan, R. J., & Rivest, R. L. (1989). Inferring decision trees using the Minimum Description Length Principle. *Information and Computation*, *80*, 227–248.

Rand,W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, *66*, 846–850.

Reber, A. S. (1967). Implicit Learning of Artificial Grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*, 855-863.

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

Reed, P., & Gibson. E. (2005). The effects of concurrent task load on stimulus overselectivity. *Journal of Autism and Developmental Disorders*, *35*, 601-614.

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 347–363.

Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology, 94,* 249–273.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, *14*, 465–471.

Rissanen, J. (1987). Stochastic complexity. *Journal of the Royal Statistical Society, Series B*, *49*, 223–239.

Rissanen, J. (1989). *Stochastic complexity and statistical inquiry*. Singapore: World Scientific.

Roberson, D. & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, *28*, 977-986.

Rogers, T. T. and McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.

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

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

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyles-Brian, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Rouder J.N., Ratcliff R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General. 133*, 63– 82.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. (2006). *A more rational model of categorization.* In R. Sun & N. Miyake (Eds.), Proceedings of the 28th Annual Conference of the Cognitive Science Society.

Schyns, P. G. (1991). A modular neural network model of concept acquisition. *Cognitive Science*, *15*, 461–508.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. (1997). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Shannon, C. E. (1948). The mathematical theory of communication. *Bell System Technical Journal. 27*, 379-423, 623-656.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54-87.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, *210*, 390–398.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, *75*, 517.

Siegel, W. (1972). Memory effects in the method of absolute judgment. *Journal of Experimental Psychology, 94,* 121–131.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review. 79*, 369-382.

Smith, D. J., & Baron, J. (1981). Individual differences in the classification of stimuli by dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 1132–1145.

Smith, J. D. (2007). When parameters collide: A warning about categorization models. *Psychonomic Bulletin & Review*, *13*, 743-751.

Smith, J. D. & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 3-27.

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

Sober, E. (1975) *Simplicity*, Oxford University Press.

Stevens, S. S. (1975). *Psychophysics.* New York: Wiley.

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 3–11.

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in categorizing tones varying in frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 416–430.

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*, 881-911.

St. John, M. F. (1992). The story gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science, 16,* 271–302.

Tenenbaum, J. B.,& Xu, F. (2000).Word learning as Bayesian inference. *Proceedings of the 22nd annual conference of the cognitive science society* (pp. 517–522). Hillsdale, NJ: Erlbaum.

Tenenbaum, J. & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629-641

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*, 309-318.

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

Turing, A.M. (1950). Computing machinery and intelligence. *Mind*, *59*, 433-460.

van der Helm, P. A.,& Leeuwenberg, L. J. (1996). Goodness of visual regularities: Anon-transformational approach. *Psychological Review*, *103*, 429–456.

van der Helm, P. A., & Leeuwenberg, L. J. (1999). A better approach to goodness: Reply to Wagemans (1999). *Psychological Review*, *106*, 622–630.

van Oeffelen, M. P., & Vos, P. G. (1982). Configurational effects on the enumeration of dots: Counting by groups.*Memory & Cognition*, *10*, 396–404.

van Oeffelen, M. P., & Vos, P. G. (1983). An algorithm for pattern description on the level of relative proximity. *Pattern Recognition*, *16*, 341–348.

Vanpaemel, W. & Storms, G. (2008). In search of abstraction: the varying abstraction model of categorization. *Psychonomic Bulletin & Review, 15*, 732-749.

von Helmholtz, H. (1910/1962). *Treatise on physiological optics*. In J. P. Southall (Ed.) (Vol. 3). New York: Dover.

Wallace, C. S., & Boulton, D. M. (1968). An information measure for classification. *Computing Journal, 11*, 185–195.

Wallace, C. S.,& Freeman, P. R. (1987). Estimation and inference by compact coding. *Journal of the Royal Statistical Society, Series B*, *49*, 240–251.

Ward, L. M., & Lockhead, G. R. (1970). Sequential effect and memory in category judgment. *Journal of Experimental Psychology, 84,* 27–34.

Wayland, S., & Taplin, J. E. (1985). Feature-processing deficits following brain injury. Overselectivity in recognition memory for compound stimuli. Brain and Cognition, *4*,338–355.

Wills, A. J., & McLaren, I. P. L. (1998). Perceptual learning and free classification. *The Quarterly Journal of Experimental Psychology*, *51B*, 235–270.

William of Ockham, 1967-88. *Opera philosophica et theologica*. Gedeon Gál, *et al*., ed. 17 vols. St. Bonaventure, N. Y.: The Franciscan Institute.

Wisniewski, E. J., & Medin, D. L., (1994). On the interaction of theory and data in concept learning. *Cognitive Science 18,* 221-281.

Wittgenstein, L. (1957). *Philosophical investigations* (3rd ed.). Oxford, UK: Blackwell.

Yang, L. & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1045-1064.

Zeithamova, D. & Maddox, W. T. (2006). Dual-task interference in perceptual category learning. *Memory & Cognition, 34*, 387-398.

Zippel, B. (1969). Unrestricted classification behavior and learning of imposed classifications in closed, exhaustive stimulus sets. *Journal of Experimental Psychology*, *82*, 493–498.