

TRANSFORMATION  
AND  
REPRESENTATION  
IN  
SIMILARITY

Carl J. Hodgetts

A thesis submitted for the degree of Doctor of Philosophy

Cardiff University  
School of Psychology

UMI Number: U584498

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U584498

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.




ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

---


DECLARATION

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

Signed  (candidate) Date .....10/02/11.....


STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed  (candidate) Date .....10/02/11.....


STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references.

Signed  (candidate) Date .....10/02/11.....

STATEMENT 3

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 organisations.

Signed  (candidate) Date .....10/02/11.....

STATEMENT 4 - BAR ON ACCESS APPROVED

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans after expiry of a bar on access approved by the Graduate Development Committee.

Signed ..... (candidate) Date .....

# Table of Contents

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| <b>TABLE OF CONTENTS</b> .....                                       | <b>4</b>  |
| <b>ACKNOWLEDGEMENTS</b> .....                                        | <b>6</b>  |
| <b>THESIS SUMMARY</b> .....                                          | <b>7</b>  |
| <b>SIMILARITY</b> .....                                              | <b>8</b>  |
| <b>SIMILARITY'S 'DECLINE'</b> .....                                  | <b>10</b> |
| <b>PSYCHOLOGICAL CONSTRAINTS ON SIMILARITY</b> .....                 | <b>13</b> |
| <b>THEORIES OF SIMILARITY</b> .....                                  | <b>17</b> |
| <i>The Spatial Account</i> .....                                     | 18        |
| <i>The Contrast Model/The featural approach</i> .....                | 22        |
| The Three Violated Axioms .....                                      | 23        |
| <b>THE NEED FOR "DEEPER" REPRESENTATION</b> .....                    | <b>30</b> |
| <b>STRUCTURAL ACCOUNTS OF SIMILARITY</b> .....                       | <b>38</b> |
| <b>STRUCTURAL ALIGNMENT/STRUCTURE MAPPING</b> .....                  | <b>39</b> |
| <i>Types of commonality</i> .....                                    | 41        |
| <i>Alignable and Nonalignable differences</i> .....                  | 43        |
| <b>MODELS OF STRUCTURAL ALIGNMENT</b> .....                          | <b>43</b> |
| <i>The Structure Mapping Engine (SME)</i> .....                      | 44        |
| <i>Similarity as Interactive Activation and Mapping (SIAM)</i> ..... | 45        |
| <i>Other models</i> .....                                            | 46        |
| <b>REPRESENTATIONAL DISTORTION</b> .....                             | <b>47</b> |
| <b>THE CURRENT INVESTIGATION</b> .....                               | <b>50</b> |
| <i>Coding Scheme and Predictions</i> .....                           | 55        |
| <b>EXPERIMENT 1</b> .....                                            | <b>58</b> |
| <i>Method</i> .....                                                  | 59        |
| Participants.....                                                    | 59        |
| Materials.....                                                       | 59        |
| Procedure.....                                                       | 62        |
| <i>Results</i> .....                                                 | 63        |
| Models of structural alignment.....                                  | 64        |
| <i>Discussion</i> .....                                              | 67        |
| <b>EXPERIMENT 2</b> .....                                            | <b>67</b> |
| <i>Method</i> .....                                                  | 69        |
| Participants.....                                                    | 69        |
| Materials and procedure .....                                        | 70        |
| Procedure.....                                                       | 70        |
| <i>Results</i> .....                                                 | 71        |
| <i>Comparing experiments</i> .....                                   | 73        |
| <i>Interaction items</i> .....                                       | 74        |
| <i>Models of structural alignment</i> .....                          | 76        |
| <i>Discussion</i> .....                                              | 77        |
| <b>GENERAL DISCUSSION</b> .....                                      | <b>78</b> |
| <i>Summary</i> .....                                                 | 83        |
| <b>THE TIME COURSE OF SIMILARITY</b> .....                           | <b>84</b> |
| <b>EXPERIMENT 3</b> .....                                            | <b>88</b> |
| <i>Method</i> .....                                                  | 90        |
| Participants.....                                                    | 90        |
| Materials and procedure .....                                        | 90        |
| <i>Results</i> .....                                                 | 91        |
| Transformations .....                                                | 92        |

|                                                                 |            |
|-----------------------------------------------------------------|------------|
| Feature matches.....                                            | 94         |
| MIP/MOP weighting.....                                          | 96         |
| Discussion.....                                                 | 98         |
| <b>EXPERIMENT 4.....</b>                                        | <b>99</b>  |
| Method.....                                                     | 101        |
| Participants.....                                               | 101        |
| Materials and procedure.....                                    | 101        |
| Results.....                                                    | 102        |
| MIP/MOP weighting.....                                          | 109        |
| Discussion.....                                                 | 112        |
| <b>GENERAL DISCUSSION.....</b>                                  | <b>113</b> |
| Summary.....                                                    | 126        |
| <b>TRANSFORMATION AND ASYMMETRY.....</b>                        | <b>127</b> |
| <b>MEASURING ASYMMETRY.....</b>                                 | <b>132</b> |
| <b>ASYMMETRIES IN THE TRANSFORMATIONAL APPROACH.....</b>        | <b>133</b> |
| <b>EXPERIMENT 5.....</b>                                        | <b>134</b> |
| Method.....                                                     | 139        |
| Participants.....                                               | 139        |
| Materials and procedure.....                                    | 140        |
| Results.....                                                    | 140        |
| Discussion.....                                                 | 142        |
| <b>EXPERIMENT 6.....</b>                                        | <b>142</b> |
| Method.....                                                     | 142        |
| Participants.....                                               | 142        |
| Materials.....                                                  | 143        |
| Results.....                                                    | 144        |
| Self-similarity and complexity.....                             | 145        |
| Discussion.....                                                 | 146        |
| <b>GENERAL DISCUSSION.....</b>                                  | <b>147</b> |
| Summary.....                                                    | 152        |
| <b>TRANSFORMATIONS IN SPONTANEOUS CATEGORISATION.....</b>       | <b>153</b> |
| <b>SIMILARITY'S DECLINE IN CATEGORISATION.....</b>              | <b>155</b> |
| <b>EVIDENCE FOR SIMILARITY IN CATEGORISATION.....</b>           | <b>157</b> |
| <b>A (NOT SO) NEW APPROACH TO UNDERSTANDING SIMILARITY.....</b> | <b>159</b> |
| <b>EXPERIMENT 7.....</b>                                        | <b>161</b> |
| Method.....                                                     | 163        |
| Participants.....                                               | 163        |
| Materials.....                                                  | 163        |
| Procedure.....                                                  | 165        |
| Results.....                                                    | 166        |
| Transformations.....                                            | 166        |
| The spatial account.....                                        | 167        |
| <b>GENERAL DISCUSSION.....</b>                                  | <b>170</b> |
| <b>CONCLUSIONS.....</b>                                         | <b>175</b> |
| <b>COMPARING REPRESENTATION SCHEMES WITHIN A DOMAIN.....</b>    | <b>176</b> |
| <b>COMPARING SIMILARITY MEASURES.....</b>                       | <b>181</b> |
| <b>THE TRANSFORMATION 'FRAMEWORK'.....</b>                      | <b>182</b> |
| <b>FUTURE CHALLENGES AND DIRECTIONS.....</b>                    | <b>184</b> |
| <b>REFERENCES.....</b>                                          | <b>193</b> |
| <b>APPENDICES.....</b>                                          | <b>222</b> |
| <b>APPENDIX A.1 - MODEL PREDICTIONS FOR RD.....</b>             | <b>222</b> |
| <b>APPENDIX A.2 - TABLE OF PREDICTIONS.....</b>                 | <b>223</b> |
| <b>APPENDIX A.3 - MODEL PREDICTIONS FOR SA.....</b>             | <b>224</b> |
| SME.....                                                        | 224        |
| SIAM.....                                                       | 225        |

---

# Acknowledgements

I owe my first handful of thanks to my supervisor, Ulrike Hahn. Ulrike struck the balance just right - she always expected my best but never made me feel pressured or isolated; I guess she knew, more than I did, that things would be okay! As well as making research a joy, her humble approach kept me calm in times of panic. If I manage to adopt an ounce of Ulrike's attitude to research then I am confident that my future career will work out just fine.

I would also like to thank Nick Chater at UCL, who, apart from being a fantastic collaborator, supported my work throughout my PhD. I look forward to maintaining this collaboration in the future.

Completing my PhD at Cardiff was never a guarantee. My first year, that was initially an MPhil, was only made possible by the EU grant 'The Making of Human Concepts'. In relation to my final two years, I must acknowledge Rob Honey, Dylan Jones and Ulrike Hahn for ensuring that I could continue until the end.

I would also like to thank the lab group. As well as helping me with the daily toils of data collection, they were always fantastic, like-minded company. I should say a special thanks to James Close, with whom I shared an office at the very beginning and at the very end.

Of course, an enormous thank you must go to the brilliant friends I made over my PhD - you made it an unforgettable, happy time. Also, how can I possibly forget the chaps from Draw Me Stories; writing and performing music during this time was just immense!

I must also thank, especially, my girlfriend Emer, with whom I shared this whole experience.

Finally, I cannot thank my parents enough for their endless support. I hope, more than anything, to use what I have learnt and make them even more proud in the future.

---

# Thesis Summary

Similarity, being a psychological notion, involves the comparison of finite object representations. The specific nature and complexity of these representations is a matter of fierce theoretical debate; traditionally, similarity research was dominated by the spatial and featural account. In the spatial account, similarity is determined by the distance between objects in a psychological space. Alternatively, the featural account proposes that similarity is determined by matching objects' features. Despite the empirical success of these accounts, the object representations they posit are regarded too simple and specific to deal with more complex objects. Therefore, two structural accounts have been developed: structural alignment (SA) and Representational Distortion (RD).

This aim of this thesis was to further establish one particular structural account - RD - as a general framework for understanding the similarity between object representations. Specifically, RD measures similarity by the complexity of the transformation that “distorts” one representation into the other.

This RD approach is investigated in detail by testing a detailed set of transformational predictions (coding scheme) within a rich stimulus domain. These predictions are tested through experiments and modelling that utilise both a) explicit measures (ratings, forced-choice), and, for the first time, b) implicit measures (reaction time, same-different errors & spontaneous categorisation). Moreover, RD is compared empirically with both traditional and alignment models of similarity.

Overall, the results suggest that similarity can be best understood by transformational relationships in a number of contexts. The performance of RD in both explicit and implicit measures is made more compelling by the fact that rival accounts fundamentally struggle to describe the sorts of relationships that are easily captured by RD. Finally, it is emphasised that RD is actually compatible with supposedly rival approaches and can incorporate theoretically these accounts, both traditional and structural, under one general framework.

# 1

---

## Similarity

*“Upon those who step into the same rivers, different and again different waters flow.” - Heraclitus*

The environment is changing constantly in an infinite number of ways. Although a few of these changes may interest us as perceivers, many, however, will be subtle and irrelevant. In the famous adage (above), the Greek philosopher Heraclitus alludes to an essential cognitive ability, i.e., that despite constant environmental flux, we are able to successfully perceive relationships between objects in the world. Even though the environment is changing, in an objective and molecular sense, we are not representing this change psychologically. If it were unconstrained, this environmental complexity would present a clear problem for a finite cognitive system. As a result, therefore, we possess cognitive and perceptual mechanisms that 1) simplify environmental inputs (by representing only those properties that serve intelligent behaviour), and 2) reduce the novelty that arises from environmental and perceptual flux. The importance of similarity lies in the fact that it is considered to be one such mechanism.



Similarity is a perceived relationship between objects that will necessarily involve a *comparison* between two or more objects. This comparison is assumed to be characterised by a cognitive process or computation that, in some way, exploits object information to derive a final ‘similarity’ judgement. Generally speaking, similarity comparisons help identify consistencies and reduce what would otherwise be overwhelming novelty. It is this simple and yet fundamental function that has placed similarity at the core of cognition; cognitive and perceptual theories of categorisation and object recognition have assumed a central role for similarity-based processes (Graf, 2002, 2006; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984, 1986). Similarity is also considered to underlie many other cognitive processes such as memory retrieval (Chater & Brown, in press; Hintzmann, 1986), reasoning (Riesbeck & Schank, 1989) and induction (Osherson, Smith, Wilkie, Lopez & Shafiret, 1990), as well as perceptual phenomena such as object recognition and apparent motion (Graf, 2002; 2006; Oyama, Simizu & Tozawa, 1999; Shechter et al., 1988). Given this prominent role within many areas of psychological study, there is a considerable need to better understand similarity.

Although similarity seems fairly basic in relation to our everyday experiences, understanding similarity and its psychological underpinnings is far from so. Its intimate connection with a broad range of both cognitive and perceptual theories requires an understanding of a range of related concepts: mental representation, feature processing, abstractness, analogy, feature-binding, local/global processing, attention and so on. By understanding these, one can begin to determine how the cognitive system might translate object information into some measure of similarity that, in turn, can lead to intelligent behaviour – e.g., categorisation or problem solving. The theoretical debate surrounding these issues will be addressed in detail

throughout this introduction. Specifically, the aim of this chapter is to provide a backdrop to current debate about the notion of similarity within cognitive psychology by discussing well-known, fundamental criticisms of similarity and the corresponding theoretical responses to these criticisms. The remainder of the thesis will then focus on a particular psychological theory of similarity – the Transformational Approach. This approach will be investigated both through experiments and modelling, involving transformational predictions and those of rival approaches to psychological similarity.

### Similarity's 'decline'

Intuitively, similarity is a function of “the properties that two objects have in common”, that is, it is a degree to which two objects share ‘things’. What these ‘things’ are and how they come to be compared is the fundamental empirical question for similarity research. Crucially, similarity is a *psychological* notion and is not simply concerned with the properties that objects possess in the purely objective sense: to say that two objects are similar is equivalent to saying that they are *perceived* to be similar. Similarity is not an inherent relation between objects, but rather something we impose *onto* objects to guide intelligent behaviour and allow us to efficiently manage environmental complexity. This distinction, however, was not always so transparent. In fact, our contemporary view on similarity owes much to the philosopher Nelson Goodman (1972), who provided the first comprehensive attack on similarity and its supposed importance.

Goodman's (1972) criticism dissected the commonly held intuition that “two objects are similar if they have a lot of properties in common”. He challenged this statement by arguing that all objects have *something* in common. For example, a

mouse and a chair both share the property ‘weighs less than 5 tonnes’, ‘weighs less than 4 ½ tonnes’ and so on (Murphy & Medin, 1985). This logic suggests that all objects are both infinitely *similar* and infinitely *dissimilar* to all others; while a mouse has the property ‘weighs less than 100 grams’, a chair has the properties, ‘weighs more than 5 kilos’, ‘weighs less than 5.1 kilos’ and so on. However, if all objects are equally similar to all others, similarity ceases to provide an explanatory notion.

Consequently, Goodman (1972) suggests that similarity is useless until it is known in what *respect* two objects are similar, that is, concerning specific properties or features. To say that two objects are similar is meaningless unless we specify in what ‘respect’ we consider them similar, such as “oranges and footballs are similar in respect to *shape*”. It follows therefore, says Goodman, that ‘respects’ do all the work, leaving similarity a vacuous concept; a “pretender, an imposter, a quack” (Goodman, 1972, pp. 437). For example, to say that “object *x* belongs to category *Y* because it is similar to the category members of *Y* with respect to the property *square*”, is basically saying that it is categorised because it is, in fact, a *square* – similarity is a blank to be filled in (Medin, Goldstone & Gentner, 1993). In short, Goodman concluded that ‘respects’ convey the true meaning of the comparison, not similarity *per se*.

Respects, in turn, take on a number of problems of their own: if two objects are similar how do we determine the respects in the first place? There is, as Goodman suggests, no one answer to the similarity between two objects, that is, the choice of this ‘respect’ seems entirely unconstrained. Therefore, if similarity is arbitrarily flexible then generating reliable theories of similarity process and computation may be unattainable.

One source of flexibility that has been subject to not just philosophical, but also psychological attention, is the context in which a comparison takes place. Barsalou (1982, 1983) gave participants a number of comparisons with or without an explicit context. His results indicate that context can systematically change the similarity between two objects by changing the ‘respect’ in which those objects are compared. A snake and a racoon, for example, were rated similar when in the context of ‘pets’ but rated dissimilar when compared in isolation. In short, the properties that are made salient by a particular context profoundly affect the rated similarity, even if the comparison is identical in the objective sense. Tversky (1977) conducted a number of experiments investigating contextual effects in similarity judgements using a range of stimuli (schematic faces, countries). Tversky highlighted what he referred to as *diagnostic factors*, that is, the ‘classificatory significance’ of features (Tversky, 1977, pp. 342). These factors refer to a general tendency to reduce information into clusters. So, when participants are presented with a ‘base’ object and are asked to rate the similarity of a number of ‘target’ objects, the rated similarities will depend on what clusters are made salient. In one study, the base stimulus Austria was presented with three further countries, Sweden, Poland and Hungary. In this context, Austria was considered most similar to Sweden. However, when Poland was swapped with Norway for a second group of subjects, the similarity to Sweden decreased. As Poland and Hungary form a cluster in the first example, the similarity between Austria and Sweden increases. However, the introduction of Norway allows Sweden and Norway to be clustered together based on their geographical proximity, which in turn increases the similarity between Austria and Hungary.

Close, Hahn and Honey (2009), finally, demonstrated that similarity can be contextually modulated even in rats, contrary to previous assertions (Chater & Heyes,

1994). Initially, rats learnt to associate certain pairs of stimuli with an outcome (food) whilst in a particular context,  $X$  ( $A$  and  $B \rightarrow$  food,  $C$  and  $D \rightarrow$  no food). Other pairs of stimuli were also associated with these outcomes in a different context,  $Y$  ( $A$  &  $D \rightarrow$  food,  $B$  &  $C \rightarrow$  no food). When  $A$  was subsequently paired with an electric shock there was a greater fear response to  $B$  in  $X$ , when compared to  $D$  in  $X$ , thus implying the  $A$  and  $B$  are most similar. When in context  $Y$ , however, there was a greater fear response to  $D$  when compared to  $B$  - the opposite pattern. These results suggest that rats, like humans, are affected by contextual factors when determining similarity. The similarity between  $A$  and  $B$ , or  $A$  and  $D$ , for example, is not fixed, but dependent on the context in which they are considered to be similar, much like in the human studies reported above (Barsalou, 1982; Medin et al., 1993; Tversky & Gati, 1978).

Studies of expertise further highlight the flexibility of similarity. In general, it has been shown that ‘novice’ participants tend to prefer surface similarities whereas expert participants generate a high-level interpretation of the particular comparison or problem (Chi et al., 1991; Hardiman et al., 1989; Kolstad & Baillargeon, 1991; Suzuki et al., 1992). In addition, it is a well-studied phenomenon in both humans and animals that, with increased experience, objects become more dissimilar or distinctive (Gibson, 1963; Goldstone, 1998) - a finding that, once again, supports the idea that similarity is a flexible notion.

So, does this flexibility of similarity undercut its explanatory power?

### **Psychological constraints on similarity**

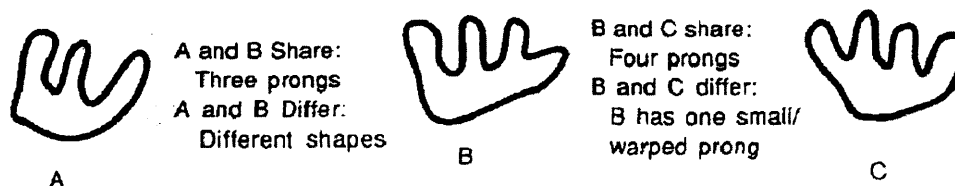
A number of authors have attempted to argue that it does not. It has been claimed that similarity, in a psychological sense, is not subject to the unconstrained

comparison process that Goodman implies (Medin et al., 1993; Goldstone, 1994a). One possibility is that perception constrains similarity.

The idea that perceptual similarities are hard-wired is quite appealing; after all it would be difficult to deny that a 400Hz and 402Hz tone sound similar (Goldstone, 1994a). Although this appears satisfactory, it is in fact victim to the inverse problem – how can the perceptual system provide a sufficiently flexible input that can account for the comparison of high-level properties or knowledge-based conceptual knowledge (see Murphy and Medin, 1985). To reduce similarity to perception is to reduce its explanatory power once more. Therefore, whilst this may eliminate properties that are non-perceptual, such as ‘taller than 1cm’, from our actual similarity comparisons, it also eliminates other types of property that are shown to influence our similarity judgements, such as relational attributes or functional similarities. As Medin et al. (1993) state, “similarity cannot be completely inflexible for the work it has to do”.

The problem converges on Goodman’s (1972) notion of ‘respects’ - is the claim that similarity is useless without ‘respects’ an accurate one? And, if respects are doing all the work, are they psychologically constrained in some way? As discussed above, many studies have shown that contextual factors can dramatically alter what ‘respects’ are relevant. However, as Medin et al. (1993) state, it is not the case that these respects vary arbitrarily – their variation seems constrained and systematic. In the context of ‘pets’, for example, a snake and racoon are considered similar, but if racoons were to become common household pets this similarity would reduce systematically as a function of this newly assimilated knowledge, that is, the change in similarity would be non-arbitrary.

Likewise, Goldstone (1994a) argues that the *comparison* itself constrains similarity. According to Goldstone, the comparison process acts as a sort of ‘relevance filter’; as objects enter into comparisons certain properties become relevant and others are discarded. By considering objects in isolation, it is impossible to determine what properties might be relevant. Medin et al. (1993) provide an elegant demonstration of this point by showing that the properties attributed to ambiguous objects depend on the comparisons in which they appear. For example, stimulus B in Figure 1 is considered to have three prongs when compared with A, and four prongs when compared with C. In such an example, where a certain property is ambiguous, or even irrelevant, the comparison itself actually disambiguates the property and increases its relevance. While certain constraints, such as similarity-as-perception (see above), may provide a more stable notion of similarity, an explanation that is based in the comparison process itself may be necessary if we are to resurrect the role of similarity in general (Goldstone, 1994a). After all, perceptual constraints can still be regarded as ‘respects’, i.e., A and B are similar in ‘respect’ to colour, frequency and so on.



*Figure 1.* An example of the stimuli used by Medin et al. (2003, pp. 262). Stimulus B is compared to either A or C. The common and distinctive properties associated with each comparison are listed.

In terms of similarity itself having no role at all, Goldstone (1994a) also argued that a notion of similarity is still essential in integrating ‘respects’ into a single judgement of similarity. Specifically, Goldstone emphasises how similarity judgements are often made over a number of dimensions or properties, (size, shape, colour etc) not a single *respect*. Therefore, similarity is necessary to integrate and weight this information in order to form a single, unitary judgement. If two objects are similar in respect to their appearance, then saying that respects do all the work is highly simplistic. ‘Visual appearance’, as a respect, is comprised of many dimensions, and so it is unlikely that all these will be equally relevant in a given context. The challenge for similarity is integrating these properties and determining the relative weight of each.

Overall, then, Goodman’s claim that respects ‘do all the work’ is, to some extent, a fair conception of the problem. However, as research has shown, these ‘respects’ do not vary in arbitrary ways, indicating that we, as intelligent animals, are able to impose constraints on what properties are relevant in a given context, as well as able to integrate multiple respects into a unitary judgement of similarity. Hence ‘respects’ do not obviate the notion of similarity itself.

The question of what we mean *psychologically* by constraining similarity remains. To address this question we must return to the fundamental distinction between the ‘objective’ and the psychological. To restate, similarity is not an objective relation between objects in the world, it is imposed onto objects in order to connect our experiences, reduce novelty and make the most of limited information. Similarity, therefore, must be flexible, but moreover, it must be psychologically plausible. Similarity comparisons are carried out by a finite brain, thus ruling out the infinite number of properties that are central to Goodman’s argument. Goodman’s



points have forced researchers to realise that similarity is a psychological notion: although objectively objects may share an infinite number of properties, subjectively, for a given cognitive agent, they most certainly do not. Human judgements of similarity are not based directly on the physical properties of objects as such, they are based on the way those objects appear to us, that is, they are based on our *mental representations* of objects (Hahn & Chater, 1997). When we represent a pair of objects, and the relationship between them, we represent only those elements and relations that are important in guiding problem solving, categorisation, and so on.

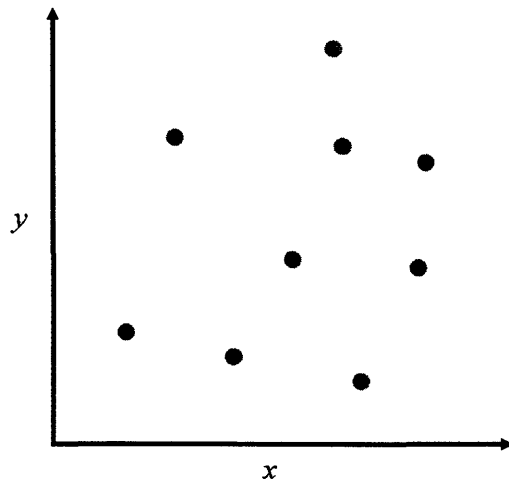
While mental representation and respects are equivalent under this view, they provide a very different focus (Hahn & Chater, 1997). Mental representation, as an explicit psychological notion, is naturally compatible with theories of cognition that assume comparisons between non-objective, non-physical identities – for example, memory representations, such as those central to models of categorisation and recognition (e.g., Nosofsky, 1984, 1986; Graf, 2006). In addition, representations can assume more than a single respect, that is, representations can be comprised of multiple dimensions without any need to speak of multiple respects. Finally, the notion of mental representation provides a clear basis for theories of similarity to make detailed predictions about object representation and how representations are compared.

### **Theories of similarity**

In light of this, it is unsurprising that the mental representation of objects has, in many ways, formed the focus of similarity theory. Although no model of similarity can be expected to provide a fully specified account of mental representation, any model must make assumptions about how certain representations are compared – and

this is not possible without assumptions about the sorts of properties that are represented in a particular context.

The empirical debate, until recently, has been dominated by two theories of similarity: the spatial account and the Contrast Model (or featural account).



*Figure 2.* A simple illustration of a multi-dimensional psychological space. Objects occupy locations based on their values on dimensions  $x$  and  $y$ . The closer two objects are, the more similar they are.

### *The Spatial Account*

The spatial account of similarity is perhaps the most widely tested model of similarity to date. The spatial account defines similarity as a distance within a multi-dimensional, psychological space (Shepard, 1957) and can be considered either a psychological theory or a convenient method for expressing and summarising similarity data (Gati & Tversky, 1982). Objects are represented as points within this space based on their values on underlying continuous dimensions. Once objects are represented within this coordinate space, similarity is determined most commonly by

the Euclidean distances between objects in this underlying spatial representation, with distance inversely related to similarity (see Figure 2). Distance in geometric models is defined as:

$$d_{ij} = [\sum |x_{im} - x_{jm}|^r]^{1/r} \quad (1)$$

where  $d_{ij}$  is the distance between objects  $i$  and  $j$ ,  $x_{im}$  is the value of object  $i$  on dimension  $m$  and the parameter  $r$  refers to the spatial metric employed. If the value is  $r = 2$  then the difference between objects is the length of the shortest line that relates them, i.e., the Euclidean distance. Alternatively, if the value is  $r = 1$  then the distance between objects is equal to their summed differences on each dimension (known as the city-block metric). Of course, proximities will vary depending on the metric used: the city-block metric, for example, will penalise differences over more than one dimension whereas the Euclidean metric will not. In addition, this parameter highlights how different dimensions are integrated in determining the similarity between object representations – emphasising similarity’s role in not only what dimensions are relevant but how dimensions or properties are combined (i.e., Goldstone, 1994a).

Crucially, varying the parameter  $r$  has been shown to have non-arbitrary effects on model accuracy in certain contexts. The suitability of each metric may depend on the stimuli used and the strategies that participants employ (Dunn, 1983; Garner, 1974; Goldstone, 1994a; Melara, Marks & Lesko, 1992; Nosofsky, 1985a; Shepard, 1987; see Shepard, 1991 for review). For example, Shepard (1987) states that the Euclidean distance metric ( $r = 2$ ) is more suitable for integral stimulus dimensions, such as the brightness and saturation of an object (see Garner, 1970; but

see also Hyman & Well, 1968) and that a city-block metric ( $r = 1$ ) is more appropriate when dimensions are separable (e.g., the length and width of a rectangle). Notably, however, determining the relative integrality or separability of a stimulus is far from straight forward in a given context, making such a distinction less practical. For example, a number of studies have shown that the integrality of two dimensions is highly flexible and subject to task-specific properties/instructions (Felfoldy & Garner, 1971; Imai & Garner, 1965; Potts, Melara & Marks, 1998), developmental factors (L. B. Smith & Evans, 1989) and individual differences (J. D. Smith & Baron, 1981; Ward, 1985).

In addition to being a model of similarity and mental representation, the spatial model also provides a technique for summarising and displaying all kinds of similarity data (ratings, confusability data, reaction time, same-different errors etc). The statistical procedure of multidimensional scaling (MDS; Shepard, 1957) will generate, as its output, a spatial representation of minimum dimensionality that preserves the proximities in the input data as best possible. The goodness of this fit is reflected in the ‘stress’ of a given solution (known as Stress-1). A high ‘stress’ value indicates that the input data has been distorted greatly in generating the output. The distances between objects in this output representation can be used to predict patterns of data in appropriate psychological models or fit related sets of similarity data (Frost & Gati, 1989; Nosofsky, 1983, 1986; Shepard, 1987).

As stated above, many psychological models have successfully adopted the notion that (dis)similarity corresponds to distance in a multidimensional space. The spatial model’s dominance is most pronounced in exemplar models of categorisation, namely the Generalised Context Model (GCM; Nosofsky, 1984, 1986) which is undoubtedly one of the most well tested cognitive models to date. The GCM – a

generalised version of Medin and Schaffer's (1978) Context Model – assumes that the categorisation of a particular stimulus is based on the summed similarity of that stimulus to all exemplars in the candidate category. In this context, similarity does not correspond directly to distance; instead it is an exponential function of distance, that is

$$s_{ij} = \exp(-c \cdot d_{ij}) \quad (2)$$

where the variable  $c$  refers to a 'general sensitivity parameter' that determines the 'spread' of similarity. Increasing the value of  $c$  will increase the influence of 'close' similarities. Hence, high  $c$  values will mean that close exemplars in similarity space will exert a larger influence on categorisation than distant exemplars. This distance metric also incorporates further free parameters,  $w_m$ , that can differentially weight stimulus dimensions (thus emphasising the requirement of similarity-based models to both integrate and weight relevant dimensions). The successful implementation of this account across a range of stimuli and tasks is compelling support for this exponential 'law of generalisation' (Nosofsky, 1984, 1986; Shepard, 1987; see also Kruschke, 1992 for a connectionist implementation of the GCM). Relatedly, however, some studies have also shown that a Gaussian similarity function is superior in many contexts (Ashby, 1992; Nosofsky, 1985a, 1985b, 1986), such as when performance is asymptotic rather than nonasymptotic, for example, learning data (Nosofsky, 1985b).

The success of similarity-based models of categorisation provides considerable support for similarity in general, at least in some contexts. Fundamentally, these conceptualisations of similarity (as exponential and Gaussian functions of distance in psychological space) have provided concrete interpretations

of a construct once considered too vague to be any use at all in cognitive psychology (Cutzu & Edelman, 1998).

Despite the empirical success of the spatial account it has received considerable theoretical criticism. The most notable critic is Amos Tversky (1977), who launched a comprehensive attack on spatial models in support of his own theory of similarity - the Contrast Model (Frost & Gati, 1989; Gati & Tversky, 1982, 1984; Tversky, 1977). In general, Tversky's criticisms focussed on the inadequacy of spatial representations and their underlying assumptions to deal with a range of similarity-based phenomena.

#### *The Contrast Model/The featural approach*

The Contrast Model, or the featural approach to similarity, states that similarity comparisons involve matching object features. The theory assumes that objects are represented as bundles of binary or nominal features. It follows that relevant features are selected through a process of extraction and compilation, both of which are determined by task and stimulus properties (Tversky, 1977). The Contrast Model is defined as follows

$$\text{SIM}(A,B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad (3)$$

A and B refer to the feature-sets of objects A and B respectively, so that the similarity between A and B is a function of three arguments: 1) features common to both A and B ( $A \cap B$ ), 2) features that belong to A and not B ( $A - B$ ), and 3) features that belong to B and not A ( $B - A$ ).  $\theta$ ,  $\alpha$ ,  $\beta$  are weight parameters that allow for a number of different similarity scales if these are weighted differentially; for example, if  $\theta = 1$  and

$\alpha$ ,  $\beta$  disappear from the equation, then only the first argument is used to determine similarity, that is,  $(A \cap B)$ . Likewise, if  $\alpha = \beta = 1$ , and  $\theta$  is not used, then only the differences between A and B are taken into account, that is,  $f(A - B) + f(B - A)$ . Discrepant values of  $\alpha$  and  $\beta$  will differentially weight the distinctive features of each compared object allowing for certain similarity relations to be born out, as will be explained in more detail in the next section.

### *The Three Violated Axioms*

Crucially, Tversky (1977) directly challenges geometric representations by demonstrating systematic violations of the metric axioms that can be readily explained by a featural account. If similarity is best represented as distance within a psychological space then it must obey the assumptions that underlie distance-based measures in general. The three axioms are as follows:

1) *Minimality*:  $d(a, b) = d(a, a) = 0$

The similarity between an object and itself is the same for all objects.

2) *Symmetry*:  $d(a, b) = d(b, a)$

The similarity between two objects is the same regardless of the direction.

3) *Triangle inequality*:  $d(a, b) + d(b, c) = d(a, c)$

The distance between two objects must be smaller, or equal to, the sum of two other distances.

For the axiom of minimality, Tversky (1977) asserts that the probability of two identical objects being identified correctly as the same is not the same for all objects.

If identification probability is regarded as measure of similarity, then this poses a threat to the spatial account. However, equating confusion probability and similarity is not unproblematic. Rather than inferring differences in self-similarity from indirect similarity measures, such as identification errors of reaction time, it seems equally plausible to assume that differences in those indirect measures reflect differences in the time course of the encoding and compiling of object features, that is, more complex objects will demand a greater level of processing relative to simple ones. While this may indicate a relationship between the complexity of feature-based representations and the processing required, it does not necessarily mean that a graded notion of self-similarity is meaningful.

Out of the three axioms, symmetry has received the most attention empirically. Before Tversky's (1977) critique, symmetry had been widely regarded as an essential property of similarity judgements; after all, it seems intuitive that the similarity between two objects should be fixed regardless of the direction. Tversky famously challenged symmetry by showing a systematic preference for the statement "North Korea is similar to China" over the opposite. As a result, he argued that the similarity between two representations is dependent, in some contexts, on the direction of the comparison, that is,  $SIM(A, B) \neq SIM(B, A)$ . Empirically, such asymmetries have been claimed to exist using both direct measures of similarity (Bowdle & Gentner, 1998; Catrambone, Beike & Niedenthal, 1996; Op de Beeck, Wagemans & Vogels, 2003; Rosch, 1978; Tversky, 1977; Tversky & Gati, 1978; White, 2008) and indirect measures (Kuhl, 1991; Rothkopf, 1957; Wish, 1967). Although asymmetry could be considered another example of similarity's boundless flexibility, the Contrast Model makes specific predictions about when and how asymmetries arise.

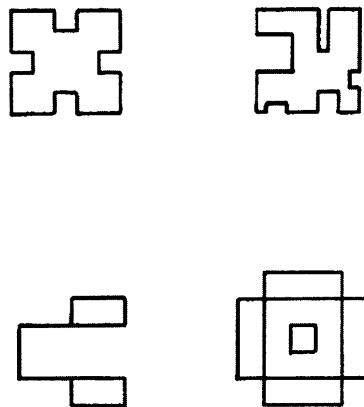


According to Tversky (1977) asymmetries can only occur in *directional* similarity statements, such as “how similar is A to B?”. Tversky argued that similarity statements are special because they, like the linguistic statements that underlie them, assume certain roles for the objects under comparison, that is, there is an ‘object’ and a ‘referent’ (or ‘base’ and ‘target’). As in syntactic structures, the roles of certain objects in the comparison are unlikely to be arbitrary. Similes and metaphors can help demonstrate this point; the two similes “a man is like a tree” and “a tree is like a man” take on very different meanings as a result of the direction assumed in each case. The former suggests that a man has roots while the latter may suggest that the tree has a life history (Tversky, 1977). In short, the role of an object in the similarity statement affects what properties are transferred to the compared objects, or how properties are weighted in the comparison. It is this phenomenon that the Contrast Model attempts to capture. Tversky’s ‘focussing hypothesis’ states that people prefer the direction where similarity is maximised. This is achieved by placing the most salient object in the base/referent role of the comparison. Given Equation 3, asymmetries are achieved by differentially weighting the distinctive features of A and B by altering the free parameters  $\alpha$  and  $\beta$ . In a non-directional similarity comparison these parameters will necessarily be equal, that is,  $\alpha = \beta$ . In a directional similarity comparison, however, the distinctive features of one object may be weighted more than the other (e.g.,  $\alpha > \beta$ ). If the compared feature sets differ in ‘salience’, that is,  $f(A) \neq f(B)$ , the above weighting will result in an asymmetric relationship. The notion of ‘salience’ refers, generally, to differences in complexity or goodness of form, with goodness of form differences (if present) being preferred to complexity differences when distinguishing between two objects (see Figure 3).

While asymmetries appear robust in some contexts, it is debatable whether they translate beyond explicit, verbal judgements (i.e., Bowdle & Gentner, 1998; Rosch, 1978; Tversky, 1977). Gleitman et al. (1994) argued that asymmetry need not be something we attribute to similarity as a ‘concept’ but instead could arise from a number of ‘linguistic-interpretative’ principles. In support of this view, Gleitman et al. empirically demonstrated asymmetries for other clearly symmetrical predicates, such as ‘equal’, ‘identical’ and ‘meets’. Crucially, Gleitman et al.’s results necessitate more research into asymmetric similarities using non-linguistic, implicit measures of similarity.

In terms of spatial models, Nosofsky (1991) argues that a notion of differential stimulus bias can account for asymmetric similarity data, as opposed to committing to the notion that similarity proper is an asymmetric relation. Whereas similarity refers to a relationship between objects, this notion of ‘bias’ is associated with individual objects. Hence, under a bias account, asymmetries arise from *processing* and not from the internal representational space or the similarity comparison specifically (Hahn & Chater, 1998). As Nosofsky states, a number of influential cognitive models have incorporated this notion of bias and have successfully fitted asymmetric confusion matrices (Holman, 1979; Krumhansl, 1978). Generally speaking, the notion of stimulus-specific bias can capture many influential cognitive phenomena, such as density (Krumhansl, 1978), frequency of exposure (Nosofsky, 1987; Polk et al., 2002), prototypicality (Rosch, 1973) and feature loss vs. feature gain (Garner & Haun, 1978). Crucially, Nosofsky emphasises that the Contrast Model, which is often portrayed at odds with the spatial account, can also be understood mathematically in terms of differential stimulus bias. This is because asymmetries require differential salience of the individual objects being compared in order to arise (see above).

The final metric axiom, the triangle inequality, is empirically the hardest to address. The assertion that one distance must be smaller than two other distances is hard to refute using ordinal or interval data (Tversky, 1977). Although such violations may exist, and have been shown empirically (Gati & Tversky, 1982), it is difficult to assess just how fundamental they are in understanding similarity in general. Although the metric axioms constrain spatial representations they do not necessarily constrain similarity, if similarity and distance are not considered equivalent. For example, spatial models are not necessarily inflexible in terms of equating similarity and distance; the dimension weight parameters can allow, like the Contrast Model, for different dimensions to take precedence in particular comparisons. Also, as similarity is considered to be an exponentially decreasing function of distance this means that the similarity between two objects is related to both psychological distance and the rate of exponential decay that is considered relevant in a particular context.



*Figure 3.* Different sources of salience and asymmetry (Tversky, 1977, pp. 335). The top pair differs in goodness of form whereas the bottom pair are said to differ in complexity.

In addition to highlighting these fundamental violations, Tversky (1977) mounted some general criticisms of spatial representations. Tversky claimed, for example, that feature sets are better than spatial representations for dealing with certain types of ‘feature-rich’ representations, such as faces, countries or personalities (a number of domains that feature in his seminal paper). This, he argues, is because the number of qualitative features provides a better representation of high-level conceptual domains when compared to a small number of quantitative dimensions (see also Kruschke, 1992). Many researchers, in turn, have supported the view that spatial models are only capable of representing continuous dimensions (Carroll, 1976; Choi, et al., 1993; Tenenbaum, 1996; see also Navarro & Lee, 2003). As a result, some spatial models have been reformulated to better deal with nominal features (Lee & Navarro, 2002; Navarro & Lee, 2003). Nosofsky (1990; see also Kruschke, 1992), however, rebutted this notion by arguing that discrete features do not necessarily present a major difficulty to spatial models, and that spatial models have been shown empirically to deal with binary and nominal features (i.e., Shepard, Hovland & Jenkins, 1961) - although this may be at the cost of model complexity (e.g., see Nosofsky, 1991). In addition, it has been suggested that conceptual stimuli could be made amenable to spatial models by being structured so as to give rise to hierarchical featural groupings or ‘pseudo-dimensions’ (Garner, 1978; Hahn & Chater, 1998).

Crucially, a general theory of similarity should be able to tolerate both discrete features and continuous dimensions within the same general representational framework – whether features or spatial models achieve this is clearly debatable.

Another prominent criticism of spatial models is their inability to account for the increase in perceived similarity between two objects when the number of common features also increases (Frost & Gati, 1989; Gati & Tversky, 1982). Fundamentally,

in an unmodified spatial model, the distance between two objects will not decrease as more common features are added, as distance is only determined by dimensional differences. Furthermore, by default, spatial models are unable to account for the total number of dimensions in a given comparison. Hahn and Chater (1997) provide a simple example to demonstrate this point: if two objects, represented by three dimensions, differ substantially on all these dimensions, then one would assume them to be quite dissimilar. However, if the two objects were represented by 10,000 dimensions and, as before, differed substantially on three of them, the objects would be very similar. The fundamental difference between these instances is that in the latter case, the objects differ on much less of the dimensions relative to the *total*. As before, spatial models can account for this effect, post hoc, by adjusting the dimension weights,  $w_m$ . Given the constraint that the weights,  $w_m$ , must equal 1, the introduction of new dimensions will necessarily reduce the relative influence of previous dimensions resulting in a decrease in the relative impact of mismatches along those dimensions. Similarity, as a result, will increase as more dimensions are introduced.

Oppositely, however, this assumption is also problematic for the Contrast Model; the assumption that adding common features increases similarity implies that similarity has no upper bound. Thus, according to the Contrast Model, no point exists at which true identity is realised. It also supposes that sameness is not a stable property and that sameness or identity can in fact vary in some way between objects – again, a rather unintuitive notion given that direct measures of similarity seldom reveal that sameness differs between objects. Once again, we must ask ourselves whether such effects are really part of similarity proper, or merely a by-product of some other process, such as the time course of stimulus encoding. As for these accounts, such criticisms may prove to be just the tip of the iceberg, as we shall see.

### The need for “deeper” representation

The spatial and featural accounts share one fundamental criticism – the representation of structure. To some degree, this criticism was first made by Murphy and Medin (1985), in support of the theory-based view of conceptual structure, or the ‘knowledge approach’ (Murphy, 2004). They argued that the then current notion of similarity was insufficient for explaining categorisation because category assignment could, in some cases, involve more complex properties such as relations and causal structures (Medin & Wattenmaker, 1987; Murphy & Medin, 1985).

More recently, this limitation has been directly addressed within the similarity literature. Generally, it has been argued that the representations posited by spatial and featural models – multidimensional spaces and feature sets - are too specific and simple for the representation and comparison of high-level properties (Hahn, Chater & Richardson, 2003). In particular, Hahn et al. (2003) stress how the representations central to traditional models are incompatible with the way we represent natural objects, such as visual scenes, linguistic structures, faces, visual textures, complex categories and geometric figures. A set of features provides only one level of description; for example, a face is made up of two eyes, a nose, and a mouth. However, this description is not sufficient; to accurately represent a face we need to know the how these features are interrelated, that is, the structure (Biederman, 1985, 1987). In other words, a face with all the necessary features, but in arbitrary locations, is not a face. Moreover, mere feature-overlap, or a spatial model, cannot easily distinguish a jumbled facial configuration from a normal configuration – despite their obvious difference. Crucially, the relations *between* features cannot be represented easily by either theory.

It is possible, as Markman & Gentner (1997) describe, to code a particular relation as a feature or dimension. For example, two objects that ‘mirror’ one another may be argued to possess the feature, or dimension, ‘symmetry’. Likewise, the difference between a square above a triangle and the opposite, a triangle above a square, could be captured by a ‘shape-above-shape’ feature. However, adding relational features or dimensions becomes increasingly complex and ‘unwieldy’ when more and more relations are relevant for a particular comparison (Markman & Gentner, 1997). Furthermore, in some contexts, both higher order relations (above, below etc) and their arguments (square, triangle) will need to be matched simultaneously in a hierarchical representation and this is simply not possible in a purely spatial or featural model. From this perspective, therefore, it is essential that new theories of similarity can tolerate structured object representations (Biederman, 1985, 1987; Chater & Vitányi, 2003; Fodor & Pylyshyn, 1988; Gentner, 1983, 1989; Markman & Gentner, 1993a).

Importantly, the criticisms of Murphy and Medin (1985) outlined above need not apply to similarity in general - only to these specific theories. As discussed throughout, both featural and spatial accounts of similarity provide accurate fits of psychological data, and in a range of contexts. The representational schemes they posit, however, lack the generality to provide explanations that match the supposed ubiquity of similarity; both features sets and multidimensional spaces seem respectively well-suited to binary and continuous properties, but are equally ill-suited to their respective opposites.

Fundamentally, the object representations they posit are too simple and, as a result, fare particularly well with artificial stimuli comprised of a few well-defined features or dimensions. Likewise, these theories of similarity may be perfectly

adequate at modelling similarity and generalisation in species with more basic representational capacities or in tasks where less complex representations are utilised, i.e., speeded same-different tasks (see Markman & Gentner, 2005). When object representations become hierarchical, that is, when both relations and surface features must be considered simultaneously, then these models are completely inadequate. The challenge, then, is to develop and test new theories of similarity that can cope with a range of arbitrary representations while still preserving the role of similarity in various prominent models of cognition, that is, in models of categorisation, induction and so on. Of course, a novel, more general approach to similarity does not necessarily make previous accounts redundant but merely widens the scope of similarity theories to account for psychological phenomena, such as similarities within knowledge-based, high-level domains.

Furthermore, it may be beneficial to develop less complex, parameter-laden approaches to similarity. While featural and spatial models have fitted a range of data, this is often at the cost of model complexity. Spatial models, in particular, are permitted considerable leeway by varying dimension weights. Comparably, the Contrast Model is able to account for asymmetries, and other effects, without having to specify anything about how objects are compared or how relevant features are selected and weighted in the first place. In fact both models assume, through the inclusion of stimulus biases, that such effects occur outside the frame of comparison, that is, they arise from the inherent properties of individual objects – not the *process* of comparison. Crucially, this enables both models to leave their representational scheme untouched in the face of seemingly conflicting evidence. However, such object-specific biases fail to account for comparisons where no discernable differences in object complexity or salience arise (see Medin et al., 1993). As we



have discussed at length, similarity *needs* to be flexible in order to account for as much as it does; stimulus biases, being stimulus-based and not comparison-based, are arguably too inflexible to account for the dynamic nature of similarity.

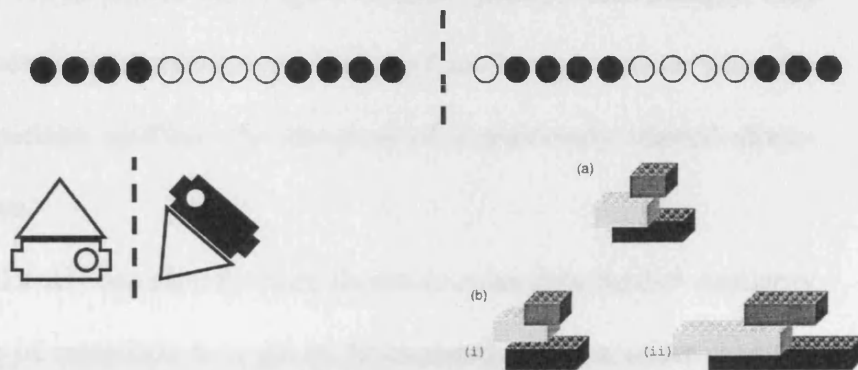
The exposition above, in many ways, presents basic criteria, or a starting point, for theories of similarity. To summarise, one may present these criteria, simply, as follows:

- 1) As similarity is a flexible construct, a theory of similarity should also be sufficiently flexible to account for similarity judgements from different domains, measures and modalities.
- 2) A theory of similarity should be able to explain basic similarity-based phenomena such as asymmetric similarities - if robust (see Gleitman et al., 1996, above).
- 3) A theory of similarity should be applicable to representations of real world objects. This necessitates, for example, a capacity to compare *and* represent structure, knowledge and relations.
- 4) A theory of similarity should be able to make some assertion about how object representations are compared and why properties are represented in the first place.
- 5) A theory of similarity should be useful in psychological models that assume a central role for similarity, i.e., categorisation, induction and so on.

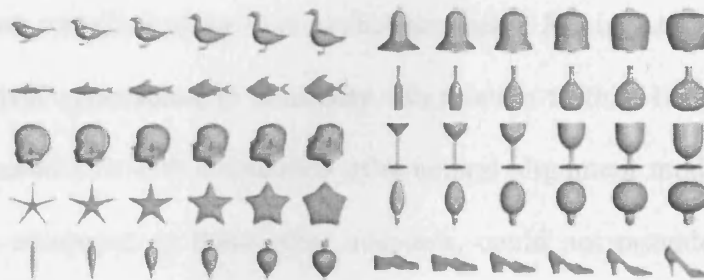
Over the course of this thesis some of these important criteria will be addressed in relation to a new theory of similarity called Representational Distortion,

or the transformational approach to similarity (RD; Hahn et al., 2003). In basic terms, RD theory states that similarity is determined by the number of transformations required to change one object representation into another. The only requirement for RD is that a plausible set of mental transformations can be derived for the stimulus set in question. In particular, the aim of this thesis is to demonstrate how this notion of ‘transformation distance’, that is both parsimonious and well-founded in the study of conceptual knowledge and perception, can provide a sufficient and broad account of similarity in a specific domain.

Hahn, Chater & Richardson (2003)



Hahn, Close & Graf (2009)



The current investigation

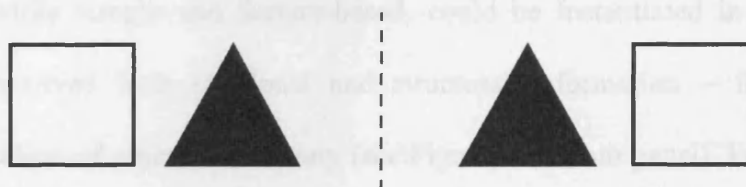


Figure 4. The range of materials used in testing RD.

The difference between this investigation and previous studies of RD is a matter of approach; in the first in-depth study of RD, Hahn et al. (2003) tested the transformational approach using a wide range of different materials, including dot sequences, unitary geometric objects and Lego brick arrangements (see Figure 4, top panel). Specifically, Hahn et al. generated a set of transformations for each of these domains, and then correlated these predictions with participants' pairwise similarity ratings (for details on specific predictions see Chapter 2). More recently, Hahn, Close and Graf (2009) demonstrated that RD could account for asymmetric similarities using naturalistic, real world stimuli (see Figure 4, centre panel). Interestingly, they showed that the rated similarity between one object and another was greater when the direction of the comparison matched the direction of a previously viewed shape-changing transformation.

Therefore, whilst RD has already been shown to accurately predict similarity ratings across a range of materials, it is yet to be explored across a wider range of tasks within a single stimulus domain. Also, as previous studies have sought mainly to establish the basic notion that transformations *can* predict similarity, RD is yet to be contrasted, in detail, with rival approaches to similarity. In relation to this, Larkey and Markman (2005) contrasted RD with a selection of structural alignment models and argued that RD, when compared to these other accounts, could not provide a sufficient account of similarity ratings. The materials they used were pairs of geometric objects that, while simple and feature-based, could be instantiated in a number of ways that involved both relational and structural information – for example, the left/right location of objects could vary (see Figure 4, bottom panel). For this reason, this domain is ideal for a full exploration of RD and its relationship with rival approaches to similarity. Furthermore, this domain is appealing because it was

not selected with transformations specifically in mind, and can therefore not be considered to favour the theory *a priori*.

Firstly, therefore, the claim that RD cannot predict similarity in this domain is re-examined by 1) generating a set of psychologically plausible transformations (that will be used throughout the thesis), and 2) performing a more detailed contrast between RD and models of structural alignment. Following a discussion of Larkey and Markman's (2005) original study and a new, more detailed empirical contrast, the idea that these theories may offer complimentary rather than competing accounts of similarity is discussed. In Chapter 3, these RD predictions are then tested for the first time within an implicit, online task – the *same-different paradigm*. As similarity in speeded tasks is considered to be based on less complex, featural similarity (Markman & Gentner, 2005), a basic feature-matching model is introduced into the model comparisons in order to look at how task constraints affect mental representations and, in turn, affect what similarity model fits the data most accurately.

In Chapter 4, the asymmetries that arise naturally from the coding scheme are examined. As shown by Hahn et al. (2009) the ease of transforming two objects is not necessarily equal in both directions, for example, transforming a tadpole into a frog is easier than the opposite. In addition, as the Contrast Model predicts asymmetries based on differential salience, the salience of individual stimuli is also investigated by measuring and comparing their 'self-similarities' (Tversky, 1977). This chapter also contains the first direct study of asymmetric similarity in humans using an implicit task; as the *same-different* task allows for sequential stimulus presentation, it is possible to present stimuli in an order that is analogous to the similar statement "how similar is A to B?", that is, the base of the comparison is first in the sequence and the target is second.

In the last empirical chapter, a spontaneous categorisation task is used to see whether the transformations posited for this domain can predict stimulus classification. Although the importance of transformations in categorisation has been implied previously, as will be discussed in Chapter 5, the RD theory of similarity has not yet been tested formally in a categorisation task. In terms of model comparison, RD is contrasted with a spatial model here because, despite its ailing reputation in similarity research, it is still the most commonly used model of similarity in categorisation research (see GCM above).

First of all, the current structural approaches to similarity will be introduced - namely RD theory and structural alignment. Once Larkey and Markman's (2005) contrast is discussed in detail, a set of concrete operations will be presented that will form the focus of the empirical chapters that follow.

# 2

---

## Structural Accounts of Similarity Transformation and Alignment<sup>1</sup>

As discussed in Chapter 1, traditional accounts of similarity are fundamentally limited when it comes to representing *structure*. To recap, these accounts define similarity over very specific and simple kinds of mental representation, that is, points in a psychological space or feature sets. Crucially, these representations are incompatible with the way in which we represent natural objects (Hahn et al., 2003). In other words, many natural objects are simply too complex to be represented *solely* by a list of features or by a number of continuous dimensions. Instead, theories of similarity must be able to compare not only the composite elements but the relations between them (Biederman, 1985, 1987; Gentner, 1983, 1989; Markman & Gentner, 1993a, 1993b). For example, as noted in Chapter 1, an accurate description of a human face must include the features it possesses and, more importantly, their interrelations. Capturing both these aspects requires *structured* representations.

---

<sup>1</sup> This chapter is based on Hodgetts, Hahn and Chater (2009).

In response to this criticism, two structural accounts of similarity have emerged: structural alignment (henceforth SA; e.g., Gentner & Markman, 1997; Goldstone, 1994b; Markman & Gentner, 1993a, 1993b) and, more recently, the transformational approach or Representational Distortion (henceforth RD; Hahn et al., 2003). Fundamentally, both of these approaches can, in theory, address the many criticisms that have been put forward against theories of similarity and similarity's role more generally (see Murphy & Medin, 1985).

Although these accounts may be related conceptually, as will be discussed later, specific models implementing the SA framework and RD can be contrasted empirically. Larkey and Markman (2005) identified an ideal testing domain for such comparisons, namely geometric object pairs of the kind used widely for the study of *feature binding* (see Bushnell & Roder, 1985; Cheries, Newman, Santos & Scholl, 2005; Kaldy & Leslie, 2003). Their initial ordinal comparisons between participants' data and the model predictions for selected items from this domain suggested that the SA model, Similarity as Interactive Mapping and Mapping (SIAM; Goldstone, 1994b) was slightly better than both RD and the Structure Mapping Engine (SME; Gentner, 1989) at matching the pattern of participants' similarity ratings.

The present chapter seeks to examine the relative performance of alignment and transformational accounts, and their relationship, in more depth - both by exploring this stimulus domain in detail and by conducting detailed quantitative comparisons.

### **Structural alignment/Structure Mapping**

Within SA, similarity comparisons are made in a manner akin to analogical reasoning (Gentner, 1983). Like analogical mapping, the SA account of similarity

focuses on methods for determining correspondences between different representations through a process of alignment (see Gentner, 1983, 1989). Crucially, this process of alignment characterises similarity not by the number of matching predicates alone but also by the nature of these compared predicates (Gentner, 1989; see Figure 5). Consequently, similarity spans a continuum where different kinds of similarity relate different kinds of mental representation. *Analogy*, for example, is thought to focus solely on aligning relations, so that water flow and heat flow may be aligned due to the corresponding roles played by attributes in each representation. *Literal similarity* will involve mappings between both relations and attributes. *Mere similarity*, however, involves only matching the object descriptions, that is, the ‘lower-order’ features of each. In essence, Gentner suggests that similarity can be understood as a continuum whereby the type of similarity that is relevant depends on the representation that is made available in a given context through task properties, expertise, and so on. Therefore, SA makes an important distinction between relations and attributes and how each influences similarity in a given context. Furthermore, SA embraces the notion that similarity is flexible by emphasising that many kinds of similarity may be relevant, as has been shown empirically.

In general, similarity in SA is driven by processes that are sensitive to both object features and the bindings between them (Gentner & Markman, 1997). Such bindings, if relevant, may require the representation of higher level relations such as causal or spatial relations. In comparing two objects, the aim should be to achieve a *structurally consistent* match between two representations. Structural consistency refers to two constraints that are central to models of SA. The first constraint, known as *one-to-one correspondence*, ensures that elements in the base representation can be aligned with, at most, one element in the target representation (Gentner, 1983, 1989).



The second, referred to as *parallel connectivity*, requires that matching relations must also have matching arguments.

SA combines structured object representations with the earlier assumptions of Tversky (1977) - who stated that similarity is a function of common and distinctive features (i.e., commonalities and differences). SA, however, extends this notion by specifying two types of commonality and two types of difference.

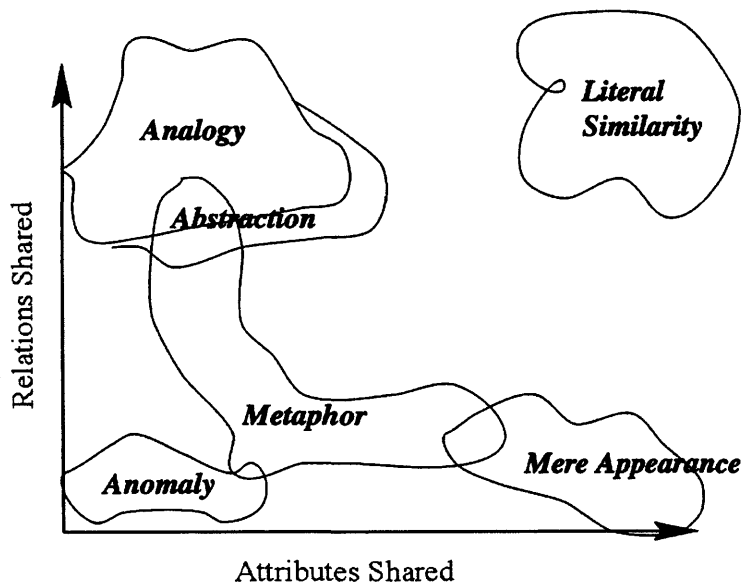
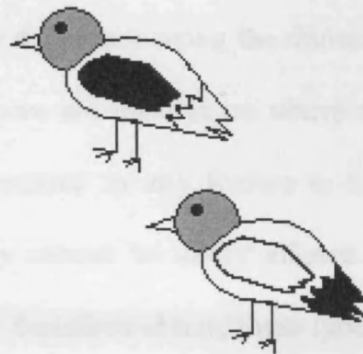


Figure 5. An illustration of “similarity space” (Gentner, 1989, pp. 207). The similarity space emphasises the notion that similarity is related to the type of information exploited when comparing two object representations.

### *Types of commonality*

Two types of commonality have been proposed: matches in place and matches out of place (Goldstone, 1994b). A match in place (MIP) is a feature match between

objects that correspond. In other words, a MIP is a match that is consistent with the global structure, that is, the structure where all objects have been placed in one-to-one correspondence. A match out of place (MOP), on the other hand, is a match between elements that do not correspond. Thusly, a MOP reflects a match outside the best global mapping (see Figure 6). Again, by distinguishing two types of commonality, SA emphasises the importance of representing the structural nature of compared object representations.



*Figure 6.* An example of MIPs and MOPs. The birds' heads reflects a match in place whereas the black of the tail compared with the black of the wing reflects a match out of place. This MOP occurs because the wing has already been aligned with the other wing on the basis of shape and so cannot be mapped onto the tail as well – doing so would violate the one-to-one mapping constraint (outlined above).

Moreover, these 'matches' have been shown to be psychologically distinct; MIPs, for example, have been shown to have a greater influence on similarity than MOPs (Goldstone, 1994b; Goldstone & Medin, 1994), demonstrating again that it is

not the mere sharing of features that increases similarity but also the degree to which features occupy *corresponding structural roles*.

### *Alignable and Nonalignable differences*

Likewise, Markman and Gentner (1993a) identified two types of differences between items. *Alignable differences* are differences that occur within the common structure (Markman & Gentner, 1993b, 1996). For example, despite the fact that a soccer ball and a rugby ball differ, their respective roles in each sport is nevertheless alignable. Alternatively, this can be expressed as a difference along a *common dimension*, that is, a difference along the dimension *balls* (Estes & Hasson, 2004). *Nonalignable differences* are differences where a feature or element in one representation does not correspond to any feature in the other representation. For example, the scrum in rugby cannot be easily aligned with any aspect of soccer. Importantly, like the matches described above, these types of difference differentially affect perceived similarity. For example, alignable differences seem to decrease similarity more than nonalignable differences (e.g., Hahn & Bailey, 2005; Markman & Gentner 1993a; Markman & Gentner, 1996; see, however, Estes & Hasson, 2004). To reemphasise, this distinction between two types of difference suggests that is not simply the case that differences decrease similarity, but that differences within the common relational structure have a greater *negative* effect on similarity than those unconnected to this common structure.

### **Models of structural alignment**

A number of models have been developed that, to some degree, assume alignments between compared representations. In general, models of SA differ in the

degree to which they obey the constraints outlined above, that is, one-to-one correspondence and parallel connectivity. Arguably, the two most prominent models of structural alignment in similarity judgements are the Structure Mapping Engine (SME) and SIAM, and the empirical tests in this Chapter will focus on these models.

### *The Structure Mapping Engine (SME)*

The structure mapping engine (Falkenhainer, Forbus & Gentner, 1989) extends Gentner's (1983) earlier structure-mapping theory of analogy. As its input, SME receives two hierarchical representations comprised of entities, predicates and functions. SME takes these representations, computes the correspondences and searches for the 'maximal structurally consistent match' (Markman & Gentner, 1993a). SME will place stimulus properties (entities, predicates and functions) into *one-to-one correspondence* and strive for structurally consistent matches that retain *parallel connectivity*. This assumption is fundamental because it implies that mappings are not based solely on the featural similarities of objects but on the role these features or entities play in the relational structure.

For an example, consider the comparison of the two following inputs: "Square above circle" and "triangle above square". In this example, the square and triangle could be aligned as they both share the role as the 'above' object. Alternatively, the square could be mapped onto the square in the second configuration. Therefore, in principle, the square in the first representation can be mapped onto either object in the second representation. However, the strict adherence to one-to-one correspondence in SME means that the square can only be mapped onto one object in the target representation. This strict constraint means that MOPs have little influence on

similarity computation (i.e., the black feature match between wing and tail in Figure 6 will not be considered).

### *Similarity as Interactive Activation and Mapping (SIAM)*

SIAM (Goldstone, 1994b) provides a connectionist implementation of the alignment process that has much in common with Holyoak and Thagard's (1989) model ACME. SIAM is a localist, connectionist model that assumes similarity to be determined through a dynamic process of interactive activation between feature, object and role correspondences. Correspondences operate reciprocally, as the degree to which objects from two scenes are placed in correspondence depends on the degree to which the composite features of those objects are also placed in correspondence. The opposite is likewise true; the strength of feature correspondences also depends on the strength of object correspondences.

SIAM's network is comprised of nodes that can form excitatory and inhibitory connections with one another depending on the consistency between different correspondences. A node in SIAM represents the hypothesis that two entities correspond across two scenes. There are three types of node: object-to-object, feature-to-feature and role-to-role (Goldstone, 1994b). Similarly to SME, there are constraints on what counts as a consistent correspondence for both features and objects. For example, correspondences where more than one feature or object maps onto another are deemed inconsistent, resulting in inhibition between inconsistent nodes. Therefore, if a feature in both object A and B correspond to a feature in object C, then the two nodes that reflect this correspondence will inhibit one another. The same system applies to object-to-object nodes.

SIAM's network activity begins by placing features in correspondence according to their basic match values. Match values represent basic perceptual similarity for both associated features and roles, and can range from 0 to 1, with 0 indicating maximal difference (or dissimilarity) and 1 indicating an identical match value. By default, SIAM's node activation commences at 0.5 in order to signify 'maximal uncertainty' about the possible alignments between objects. After featural correspondences are calculated, SIAM begins to place objects into correspondences that are consistent with the feature-to-feature mappings. Activation is then fed back to feature matches consistent with the object alignments. Similarity in SIAM is a function of the match values for each node, weighted by the current activation of that node. As a consequence, matching features increase similarity to a greater or lesser extent depending on the degree of correspondence.

SIAM differs from SME by not adhering as strictly to the one-to-one constraint; it allows for *degrees* of correspondence rather than an 'all or nothing' approach. This means that in SIAM, MOPs will have an influence on similarity (albeit a lesser influence than MIPs).

### *Other models*

These two specific models of structural alignment stem from a wider literature on analogical reasoning that contains a wealth of other models that embody the same fundamental assumption that analogies are governed by processes that seek to determine structurally consistent alignments (see e.g., ACME, Holyoak & Thagard, 1989; LISA, Hummel & Holyoak, 1997; IAM, Keane, Ledgeway & Duff, 1994). As many alignment models are primarily models of analogy, not similarity in general, many are not appropriate for the current investigation. Larkey and Markman (2005)

also considered CAB (Larkey & Love, 2003), a model of structural alignment that uses directed-acyclic graphs as its fundamental representation. CAB, however, fared poorly in predicting their similarity data. Relatedly, Taylor and Hummel (2009) attempted to model Larkey and Markman's data using LISA, a model of relational reasoning. Like CAB, this model failed to match the ordinal differences in the human similarity data (see Taylor & Hummel, pp. 238). Moreover, LISA contains 22 free parameters and is non-deterministic, thus requiring multiple runs to derive an 'average' fit for any given data point - this rules out its practical use as a measure of similarity. Consequently, the present investigation will be limited to SME and SIAM. Finally, it should be noted that the emphasis on structure-mapping is not universally shared by models of analogy (see e.g., French, 1995; Hofstadter, 1995).

### **Representational Distortion**

Representational distortion is a structural account of similarity that posits that the similarity between representations is determined by transformation distance. Transformation distance refers to the complexity required to "distort" the representation of one object into the representation of another (Chater & Hahn, 1997; Chater & Vitányi, 2003; Hahn & Chater, 1997; Hahn et al., 2003; Hahn, et al., 2009). The simpler this transformation is, the more similar the objects are deemed to be. RD asks, in concrete terms, how each representation can be mapped onto, and transformed into, the other. Of course, if there are structural relationships between the two items, this will tend to make such mappings simpler by allowing an entire chunk of a representation to be transformed as a whole, rather than piecemeal. To this extent, RD need not be viewed in opposition to structural analogy models. Nonetheless, the focus of RD is different; and specific models based on RD, such as that described below,

will be based on providing a coding language for possible transformations, so that the complexity of transformations can be operationalised in terms of the length of the code required to specify them. Moreover, perhaps not surprisingly, the detailed predictions of such a model will not typically agree with the predictions of specific instantiations of structural alignment models of similarity, such as SME or SIAM, as we shall see. The general relationship between MIPS and MOPS, however, will still often hold, in that MIPS will typically be associated with simpler transformations when compared to MOPs.

As its foundation, RD can call upon a particular branch of mathematics known as Kolmogorov complexity theory (Li & Vitányi, 1997). According to Kolmogorov complexity theory, the complexity of a transformation is the shortest computer program that is required to perform it. The intuition of RD is that a simple transformation between representations can be expressed using a short “program” or code, whereas more complex transformations require longer programs.

To bring this idea closer to psychological similarity, the RD between two object mental representations, A and B, is determined by the length of the shortest code or program which can distort A into B. If the coding language is a general-purpose computer programming language, this distortion can be mathematically expressed as the Kolmogorov complexity,  $K$ , of B given A,  $K(B|A)$ . This general notion of coding complexity or ‘information distance’ has a number of useful applications in perception, generalisation, language and mathematics. For example, Chater (1996) applied this notion of complexity to perceptual organisation, arguing that the perceptual system, when faced with a particular sensory input, chooses the simplest (and most likely) perceptual organisation, as determined by the K-complexity. In general, complexity-based coding schemes have a long history in



concretely testing the simplicity of visual inputs in perceptual research (Hochberg & McAlister, 1953; Restle, 1970; Simon, 1972). Likewise, Falk and Konold (1997) suggest that psychological judgements of randomness are best explained by the difficulty to encode a given input, that is, the coding complexity. This, in turn, may explain why participants create sequences with more alternations than would be expected by chance. Also, in relation to similarity, Chater and Vitányi (2003) demonstrate that Shepard's (1987) 'Universal Law of Generalisation' can be derived using transformation distance.

To make this type of account psychologically concrete requires specifying a particular coding language for transformations, in which concrete code lengths can be associated with specific transformations. The complexity of a transformation, relative to a coding language, is determined by the length of shortest code that carries out that transformation; and transformational complexity is assumed to be inversely related to similarity. A simple mathematical example provides a basic demonstration: 1 2 3 4 5 6 and 2 3 4 5 6 7 are similar in that one simple instruction can transform one into the other, that is, *add 1 to each number*. By contrast, 1 2 3 4 5 6 and 3 5 7 9 11 13 are less similar, as two instructions are needed to transform one into the other, *multiply by 2 and add 1*.

Prior empirical research has indicated that this view is promising. In particular, Hahn et al. (2003) found, in three experiments, that transformation distance predicted patterns of similarity ratings. In their first experiment, participants were presented with sequences of filled or unfilled circles. The basic transformations were reversal, deletion, insertion, mirror imaging and phase shift (based on the work of Imai, 1977). Transformation distance was operationalised by the number of these basic transformations required to distort the representation of the test item into the target

item. There was a high correlation between transformations and perceived similarity. Moreover, a featural description of the data did not predict perceived similarity as accurately. Hahn et al.'s second experiment used pairs of geometric stimuli that were related by a range of perceptually natural transformations. Again, transformational complexity (measured simply by the number of transformations) was correlated strongly with perceived similarity.

In Hahn et al.'s (2003) final experiment, participants had to make judgements of similarity based primarily on relational information rooted in the spatial relationships between arrangements of Lego blocks. As featural and spatial accounts are inappropriate for dealing with structured representations, this experiment directly addresses a critical limitation of earlier theories of similarity and their representation schemes. The results closely matched the relationship between transformation distance and similarity that was observed in their first two experiments, despite the fact that both the materials and the relevant transformations differed greatly. These findings provided further evidence against a purely featural or spatial view of similarity, in that neither account can tolerate structured representations, as discussed above. To this end, they provided support for structural theories of similarity, such as RD and, potentially, also SA.

### **The current investigation**

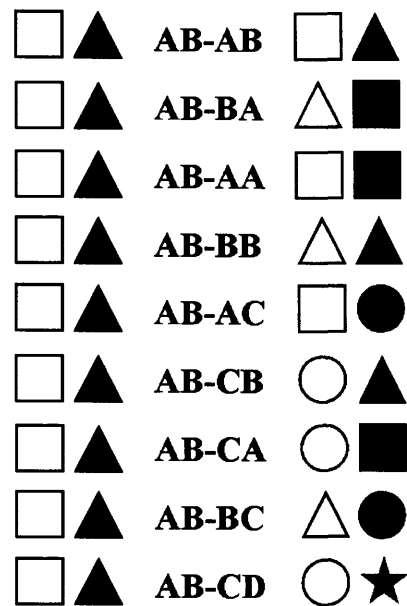
As stated previously, within the current chapter and throughout this thesis, transformational relationships will be explored using a stimulus domain that has been previously identified as an ideal testing ground for studying structural models of similarity (Larkey & Markman, 2005). As noted, this domain had been used to explore feature binding in adult humans, infants, and non-human primates (Bushnell

& Roder, 1985; Cheries, Newman, Santos & Scholl, 2005; Kaldy & Leslie, 2003). In this domain, participants compare assemblies of simple, geometric objects. In a given comparison, participants contrast two pairs of geometric shapes that vary in both shape and colour (see Figures 4 & 7). These stimuli allow for both colour and shape to vary simultaneously or for one of these dimensions to vary independently. Crucially, the former method of variation allows an investigation of 1) how separable properties or dimensions are psychologically bound into individual objects, and 2) what conditions facilitate the binding of object features. Therefore, this recommends the domain not only for its original purpose of investigating feature binding, but also for testing structure-based theories of similarity.

Stimuli will be described using a description scheme previously developed by Larkey and Markman (2005). On this scheme, each unique letter corresponds to a unique shape or colour; in other words a square might be represented by the letter A, a circle by the letter B, and a triangle by the letter C. “AB/AC”, then refers to, say, a square and a circle, that together make up the first object pair, and a square and a triangle, that together make up the second pair. The comparison at the top of Figure 7, for example, can be represented by the code AB/AB on both shape and colour. What these letters refer to specifically is arbitrary within this domain; the nature of these stimuli ensures also that the degree of featural similarity between non-matching shapes (e.g. the similarity between a square and a triangle) is of little importance when making judgements, that is, only the structural relationships between pairs are important - or the ‘what’ and ‘where’ aspects.

To further reduce the influence of any unwanted featural similarities in the data, the assignment of actual shapes to structural descriptions is randomised; that is,

on a given trial ‘A’ might correspond to a square, whereas on other trials it might be instantiated by a circle or triangle.



*Figure 7.* The nine feature combinations used by Larkey and Markman (2005, p. 1068) for the dimension of shape (the second dimension, colour, was fixed to AB/AB).

Larkey and Markman (2005) used a number of these pairs to conduct what is the only explicit comparison between SA and RD to date. Specifically, they used nine unique feature combinations for each dimension (shape and colour; see Figure 7). The data that Larkey and Markman reported were the ordinal relationships between perceived similarity for the nine combinations on one dimension when the other dimension (colour or texture) was characterised either by the simple relationship AB/AB or AB/BA (resulting in 18 stimulus combinations overall). These ordinal relationships were then compared with the predictions made by each account (see Figure 8). As can be seen in Figure 8, the leftmost column on each panel shows the

ordinal patterns in the behavioural data where each horizontal line represents a significant difference between combinations in the behavioural data; the remaining columns show the predictions ascribed to RD and the SA models. The vertical order in each table signifies descending similarity and stimuli within a box are predicted to be equally similar.

|                                                             | Colour and shape | RD      | SME     | SIAM    |
|-------------------------------------------------------------|------------------|---------|---------|---------|
| Most Similar<br> <br> <br> <br> <br> <br> <br>Least Similar | AB - AB          | AB - AB | AB - AB | AB - AB |
|                                                             | AB - BA          | AB - BA | AB - AA | AB - BA |
|                                                             | AB - AA          | AB - AA | AB - AC | AB - AA |
|                                                             |                  | AB - AC | AB - BA | AB - AC |
|                                                             | AB - CA          | AB - CA | AB - CA | AB - CA |
|                                                             | AB - CD          | AB - CD | AB - CD | AB - CD |

|  | Colour and shape | RD      | SME     | SIAM    |
|--|------------------|---------|---------|---------|
|  | AB - BA          | AB - AB | AB - BA | AB - BA |
|  | AB - AB          | AB - BA | AB - AA | AB - AB |
|  | AB - AA          | AB - AA | AB - CA | AB - AA |
|  |                  | AB - CA | AB - AB | AB - CA |
|  | AB - AC          | AB - AC | AB - AC | AB - AC |
|  | AB - CD          | AB - CD | AB - CD | AB - CD |

*Figure 8.* Results table after Larkey and Markman (2005), p. 1070; in the original table, results for CAB were also reported, but the model fared poorly and has been omitted for ease of reference. The left-hand table shows the relationships between methods of changing one dimension when the other dimension displays the “null” relationship of no change, AB/AB. The right-hand table reports the same methods on one dimension with the other dimension showing the relation AB/BA.

Inspection of Larkey and Markman’s (2005) results shows that for the first case, in which one dimension was fixed to AB/AB (left panel, Figure 8), the predictions for both RD and SIAM matched the descending similarity reported by participants but those of SME failed to do so (in particular this model ranked the similarity of AB/BA considerably lower than observed). However, both RD and

SIAM made more fine-grained differentiations than observed in the behavioural data, and SME provided a better account in this regard. Alternatively, when that dimension was fixed to AB/BA, (right panel Figure 8), only SIAM fits the pattern of descending similarity shown in the actual data set. The predictions ascribed to RD did not.

From this, Larkey and Markman (2005) concluded that SIAM provides a superior account of structural similarity. However, there are several reasons why one might want to probe the model relationships more thoroughly. First, purely ordinal comparisons provide a fairly weak test. For one, partitioning participant data into patterns that do or do not differ significantly in terms of their rated similarity depends in part on sample size. Moreover, purely ordinal comparisons do not take into account how far off a model is. Second, the stimulus set consisted of a fairly small and specific selection of the whole domain. Of course, it is quite possible that different models may fail on different kinds of items within this domain. Hence, testing a more representative sample could be informative with regards to the specific strengths and weaknesses of each model, and could alter radically their comparative performance.

Finally, it was not clear how RD's predictions were derived. Larkey and Markman (2005) seem to have assumed that each physically unique relationship between pairs corresponded to a single, unique transformation. However, as stressed in Chapter 1, similarity is a relationship between mental representations of objects, not objects themselves. Also, the coding theoretic basis of RD demands the specification of a coding scheme for the domain in question and an appropriate set of code-based transformations. However, neither a specific coding language, nor a set of psychologically plausible transformations represented in that language was specified.

That RD requires the specification of psychological transformations in order to be put to test, of course, does not mark it out from other accounts of similarity (see

also Hahn et al., 2003). SA models likewise need specification of psychologically relevant features, relations and their degree of match in order to be applied at all. Given that this stimulus domain provides such rich and yet constrained materials in which to test theories of similarity based on structured representations, a simple and perceptually natural coding scheme is presented here, from which a rich set of predictions for this entire domain can be derived.

### *Coding Scheme and Predictions*

The challenge is to generate a natural coding scheme that makes intuitive predictions about the sorts of transformations that would be relevant over mental representations. It was assumed that the relationship between the pairs could be intuitively captured by three distinct operations. Each operation takes a base pair and modifies it, as follows:

- 1) Create feature – taking the base pair we apply this operation to create a new feature that is unique to the target pair. For example, if the base pair consists of two black shapes, and the target pair consists of a grey shape and a black shape, we have to create the feature “grey,” as this is not present in the base pair.
- 2) Apply feature – this operation takes an object or entity that is currently available and applies it to *one or both* of the objects in the target pair. The ‘both’ element emphasises that applying a shape that is readily available (on a representational level) to both shapes in the target can be equivalently difficult to applying it to one. Thus, once the “grey” feature has been created (using operation 1), it may be applied to one, or both, of the black objects.
- 3) Swap – this swaps features between a pair of objects *or* swaps the object in its entirety (i.e. on both dimensions). Thus, swap can reverse the shapes in the pair of objects, leaving colour invariant; it can reverse the colour, leaving shape invariant; or it can simply swap the left-right positions of the entire object.

Figure 9 shows an example of these operations being applied to the comparison AB/BC. Using these operations, specific transformation distances can be associated with all possible stimulus combinations and subsequently compared with similarity judgements (see Figure 10 later below).

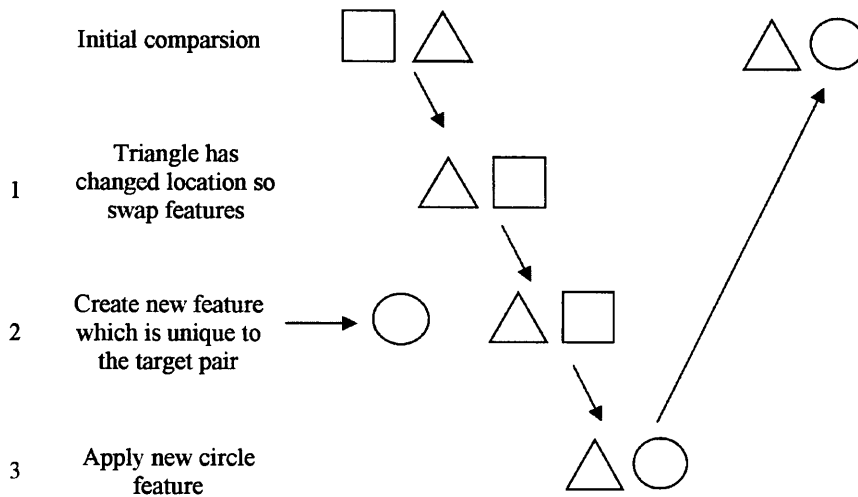


Figure 9. An example of the transformations being applied to the combination AB/BC. This comparison has a transformation distance of 3.

Importantly, this scheme must provide a natural description of how these objects are compared and represented. Firstly, the motivation for the posited operations seems consistent with previous research on feature-binding; Cheries et al. (2006) investigated whether rhesus monkeys have the capacity to bind features for stimuli very similar to those used in the current investigation. They described their materials, which consisted of two simple geometric objects varying in shape and colour, in the following way: “the monkeys were then presented with test displays in



which the objects: (i) remained identical; (ii) changed their colours (e.g., two green shapes both turning yellow); (iii) changed their shapes (e.g., two rings turning both turning into triangles); (iv) swap location; or (v) swapped features” (Cherries et al., 2006, p. 1061). After habituating to the base pair, looking times were measured during the presentation of the test pair. The monkeys in this paradigm dishabituated for both swapped cases, that is, for condition (iv) and condition (v). This suggests that the rhesus monkeys were not only sensitive to features swapping in general, but also to the notion that changes across a single dimension violate original object identity. For the present context, it provides support for the naturalness of the posited swap transformation<sup>2</sup>.

Regarding the remaining two transformations, ‘create’ and ‘apply’ operations are posited (as opposed to a single transformation that simply ‘adds’ or ‘changes’ objects) to reflect the fact that specifying a new object in its entirety requires more information than merely copying an object that is already represented. Two otherwise identical squares will require less code than a square and a circle, because the generation of the second square is parasitic on the specification of the first. Separating apply from create allows one to respect that difference (see Appendix A.1).

---

<sup>2</sup> The salience of the swap transformation was also confirmed in a pilot study. When small sample of participants ( $n = 5$ ) was asked how they could change one pair of shapes into another, they often opted to carry out a swap when both a single feature changed location and when an object changed location in its entirety. The only notable difference in participants’ reports was that they would only report the specific feature (i.e., the shape or colour) when the feature was manipulated independently of the other feature. One could argue that the assignment of linguistic labels to transformations is not evidence that the transformations themselves are actually represented during similarity computation. However, the relationship between linguistic and visual representations of visual space has received some attention empirically and may suggest that spatial language, and the spatial representations they describe, share some common underlying structure (Hayward & Tarr, 1995; Pinker & Bloom, 1990). Hayward and Tarr (1995) suggest that there must be some level of visual representation that can be accessed by our linguistic system and their results support this idea, particularly for prototypical relations such as ‘above’, ‘below’, ‘left’ and ‘right’. Crucially, it is these relations that are manipulated during the swap transformation i.e., ‘left of’ becomes ‘right of’ and vice versa. In light of this, it seems intuitive to allow a ‘swap’ transformation which can be applied to single features or whole objects - where features interact.

With this coding scheme in hand, a comprehensive test of RD is provided that expands on Hahn et al.'s (2003) initial findings while also providing more detailed comparisons with SA models, hence expanding on Larkey and Markman's (2005) original investigation. Specifically, two experiments that employ a wide, representative range of stimuli, and two different dependent measures of similarity (forced choice and direct ratings) will provide a basis for qualitative model evaluation.

### **Experiment 1**

The first experiment tested the coding scheme against a stimulus set that varied across a single dimension, namely 'shape'. For this test all unique feature combinations that are possible within the domain were tested. This expands considerably on the specific combinations tested in previous work. As discussed above, Larkey and Markman's (2005) method of testing multiple shape assignments was used again here in order to reduce unwanted effects of similarity between non-matching shapes. As a dependent measure of perceived similarity, participants were given a simple, two-alternative forced-choice task. Notably, the forced-choice task has been widely used for this purpose (e.g., Bailey & Hahn, 2005; Goldstone, Medin & Gentner, 1991; Hahn & Bailey, 2005; Tversky, 1977). Given all pairwise judgements for a set of items, this method allowed for an underlying ratio scale of similarity to be recovered from the individually ordinal judgements (see Luce, 1959). In addition, this task is conceptually easier for participants when compared to direct ratings of similarity, hence offering, in effect, a less noise-prone way of achieving the same level of measurement (forced choice tasks also avoid some of the other well-known problems of direct ratings such as compression of scale etc).

Transformation distance was operationalised by the number of transformations specified by the coding scheme. It must be noted that counting the number of transformations is an approximation of transformational complexity. Although this may not provide the most accurate account of the data in all contexts, this method does provide a useful, and simple, starting point for operationalising transformation distance (see Chapter 6 for more discussion). For example, it may be that certain transformations differ in complexity, and thus should be weighted accordingly. In such a case, a mere counting method may not provide the best account of the data when compared to a weighted model of RD. Crucially, in the current context there seems little reason to assume a priori that certain operations will wield a greater effect on perceived similarity and, as a result, all operations are given equal weight.

### *Method*

#### *Participants*

A total of 42 of Cardiff University undergraduates completed the experiment (mean age = 19.6, range = 18 to 25). Participants were either allocated course credit or paid £3 for their participation. Each participant was tested individually. As a result of a failure to follow basic instructions, the data from two participants were later omitted.

#### *Materials*

Trials were presented on a 19" LCD screen. Each display contained two separate comparisons, each comprised of two pairs. One of these comparisons occupied the upper part of the display and the other occupied the lower. Each component object within a pair was approximately 2.5 cm wide x 2.5 cm tall. Each

set was presented horizontally with the first pair in the set on the left and the second pair on the right. The distance between object pairs was 7.5 cm with a within-pair object distance of 0.5 cm. The two independent sets were separated by a vertical distance of 15 cm on the screen.












|    | Shape set                                                                                                                                                               | Object Code | Relevant transformations       | No. of transformations |
|----|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|--------------------------------|------------------------|
| 1  |       | AA/AA       | N/A                            | 0                      |
| 2  |       | AB/AB       | N/A                            | 0                      |
| 3  |       | AB/BA       | Swap                           | 1                      |
| 4  |       | AA/AB       | Create + Apply                 | 2                      |
| 5  |       | AA/BA       | Create + Apply                 | 2                      |
| 6  |     | AB/AA       | Create + Apply                 | 2                      |
| 7  |   | AB/BB       | Create + Apply                 | 2                      |
| 8  |   | AA/BB       | Create + Apply                 | 2                      |
| 9  |   | AB/AC       | Create + Apply                 | 2                      |
| 10 |   | AB/CB       | Create + Apply                 | 2                      |
| 11 |   | AB/BC       | Swap + Create + Apply          | 3                      |
| 12 |   | AB/CA       | Swap + Create + Apply          | 3                      |
| 13 |   | AA/BC       | Create + Apply; Create + Apply | 4                      |
| 14 |   | AB/CC       | Create + Apply; Create + Apply | 4                      |

Figure 10. Stimuli used in Experiment 1. Depicted in column 1 is one of six possible shape assignments (see *Materials*, Experiment 1), all of which instantiate the logical code structure of column 2. Columns 3 and 4 contain the specific transformations and their respective code lengths.

For a single dimension, pairs of objects could be combined by fourteen different predetermined methods resulting in fourteen unique comparisons in this experiment (see object codes, Figure 10). As outlined above, the specific instantiations are represented using letters: For example, a combination represented by the letters ‘AB/AC’ will have shapes A and B as one object pair and shapes A and C as the other: The letter A refers to the type of the first shape, reading from left to right; the letter B refers to the second unique shape (i.e., the next shape that was not identical to shape A); the letter C refers to the third unique shape (if any), and so on.

In an attempt to control for spurious featural similarities between non-matching objects, six unique shape assignments were used; these differentially assigned the shapes triangle, circle and square to A, B and C, depending on the assignment group. For each of the specific shape assignments, the letters consequently represented a different specific object (i.e., A might refer to a square in one group; but to a triangle in another group, depending on which specific shape happened to be leftmost). As there was no predicted difference depending on the specific shapes used, responses were collapsed across assignments for analysis, resulting in one data point for each of the fourteen combinations. On a given trial a comparison (i.e., AB/BA) would be randomly selected to occupy the top of the display and this could then be randomly paired with any of the remaining object comparisons (see Figure 11). Screen position (top or bottom half) was counterbalanced within participants, by showing each set of pairs of objects twice, with opposite screen locations. In total there were 182 trials, that is,  $14 \times 13$  (13 because no comparison was paired with itself).



*Is the similarity between the pairs in the above set greater than the similarity between the pairs in the below shape set?*

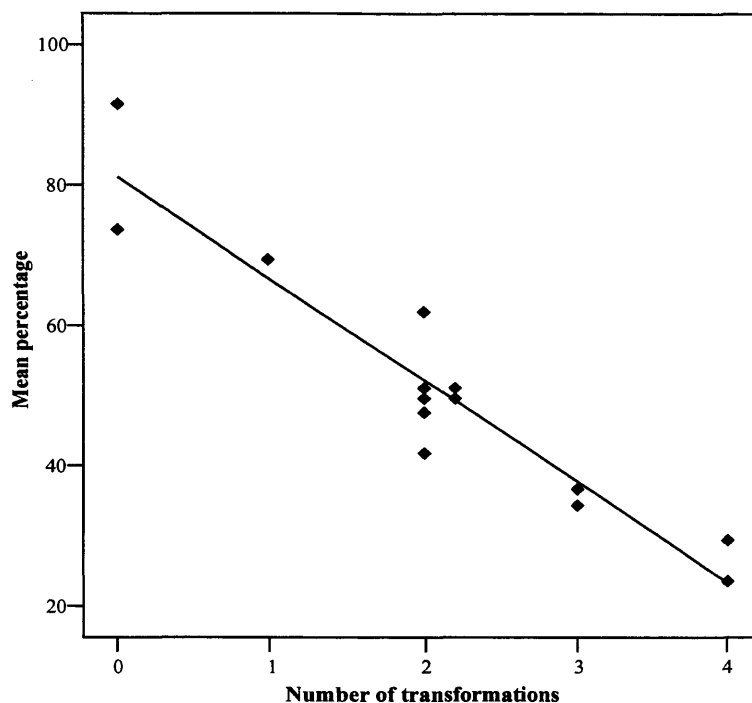


*Figure 11.* The organisation of the display in Experiment 1. A comparison occupied the upper and lower section of the display and participants were required to make two separate similarity judgements and choose the one with greatest between-pair similarity.

### *Procedure*

On each trial, participants observed two sets of stimuli (or two comparisons) where each individual comparison consisted of two pairs of shapes (see Figure 11). Their task was to judge which of the two sets of pairs had the greatest between-pair similarity. Specifically, participants were asked “*Is the similarity between the pairs in the above set greater than the similarity between the pairs in the below shape set?*”. This question was presented in the centre of the display between the two sets on each trial. Participants then made a ‘yes’/‘no’ by pressing Z on the keyboard for ‘yes’ or M on the keyboard for ‘no’. After their response, the screen was erased and two more

object sets were randomly selected for the next trial. All participants completed all 182 trials (14 x 13) and the task took approximately 20 minutes to complete.



*Figure 12.* Graph depicting the relationship between transformation distance and the mean percentage of times that each comparison was judged the most similar in a forced-choice paradigm. Each point represents the mean similarity for each object set across participants and assignments. Note: horizontal jitter is added to two points due to an overlap at two transformations on the  $x$  axis.

### Results

For each comparison (e.g., AB/AC), the percentage that it was selected as the most similar, out of all possible pairings, was the derived measure of similarity. This provides a ratio-level measure of the similarity between each pair in the stimulus set (see also Bailey & Hahn, 2005). These similarity data were correlated with the

predicted transformation distances in order to evaluate how well the posited coding language accounts for the perceived similarity between items. Inspection of Figure 12 reveals the expected negative correlation between perceived similarity and transformation distance. As in Hahn et al. (2003), this relationship between similarity and transformation distance seems linear.

The bivariate correlation between the number of transformations and perceived similarity was highly significant: Pearson's  $r$  ( $r = -0.946$ ,  $p < 0.01$ ). In addition, transformation distance, as specified by the coding scheme, accounts for 90% of the variance in perceived similarity ( $R^2 = 0.90$ ), without recourse to any free parameters.

### *Models of structural alignment*

The results for Experiment 1 provide compelling fits for RD when the stimuli vary across one dimension. However, there remains the question of how well SA models will fare, and whether they might provide equally good, or even better, explanations of the data.

For SME the number of matches in place (MIPs) was correlated with the similarity data (see Appendix A.3 for an explanation). The left graph in Figure 13 depicts the relationship between the number of matches in place (MIPS) and perceived similarity. As can be seen from the graph, there is a positive correlation between perceived similarity and MIPs, as predicted, and this correlation is statistically significant ( $r = 0.74$ ,  $p < 0.05$ ). However, MIPs account for much less of the variance when compared to RD (54% vs. 90%).

Second, how well does SIAM account for these results? Following Larkey and Markman (2005) the data were modelled using SIAM's default parameters



(Goldstone, 1994b), as these had provided excellent fits in Larkey and Markman's previous experiment. Given these parameter values, SIAM also correlates significantly with participants' similarity ratings ( $r = 0.76$ ,  $p < 0.01$ ; Pearson's  $r$ ; see Figure 13), and accounted for slightly more of the variance in perceived similarity than SME (58% vs. 54%).<sup>3</sup>

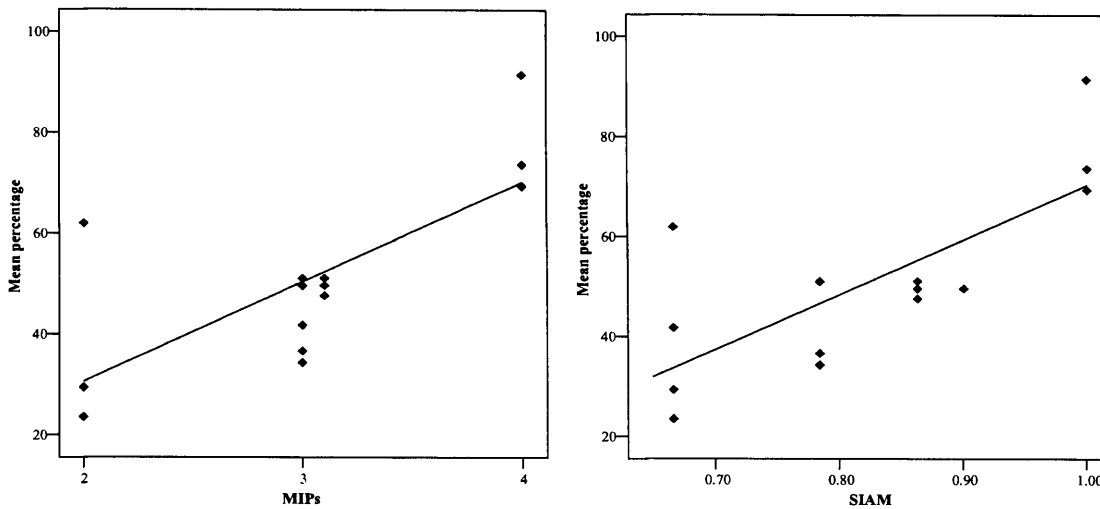


Figure 13. Left: Graph showing the relationship between similarity (as the mean percentage selected) and matches in place (MIPs). Right: Graph depicting the relationship between mean percentage and the default SIAM predictions.

As an additional model comparison, the likelihood ratio was correlated for each model pair (see Glover & Dixon, 2004). This simple statistic indicates the relative likelihood of the data given the two models. Comparing RD and SME, the

<sup>3</sup> As one reviewer of Hodgetts, Hahn and Chater (2009) noted, the combinations AA/AB, AA/BA, AB/AA and AB/BB might be considered equivalent, at least when considered in isolation. Whether they seem equivalent to the participant who sees all 14 possible combinations is, of course, an empirical question. It is, however, of interest whether the discrepancies in model performance might be driven, in part, by these 'repeats'. Consequently, the data were also analysed with these 'repeats' collapsed into a single observation. The same overall relationship between the models is maintained: for RD,  $r = -0.949$ ,  $p < 0.01$ ; R-square = 0.90; for SME,  $r = 0.74$ ,  $p < 0.01$ ; R-square = 0.54; for SIAM,  $r = 0.78$ ,  $p < 0.01$ ; R-square = 0.61.

data are 43581 times more likely to occur if RD is true than if SME is true. When compared with SIAM, the data, given RD, are 23053 times more likely. The data are 1.9 times more likely to occur under SIAM when compared to SME.

Finally, given that different instantiations of SA assign different weights to MIPs and MOPs (as in SME and SIAM), it is possible that some other account of SA, that assigns differential weights to MIPs and MOPs could yield even better fits. To assess this, a range of possible relative weights between MIPs and MOPs was considered, correlating each weighted MIP/MOP model with the similarity data. In general, when MIPs are given more weight the models account for more variance. Overall, more variance is accounted for when MOPs still have a slight influence in the model output (i.e., 0.3). As can be observed in Figure 14, no weighted combination of MIPs and MOPs accounted for as much of the variance as RD.

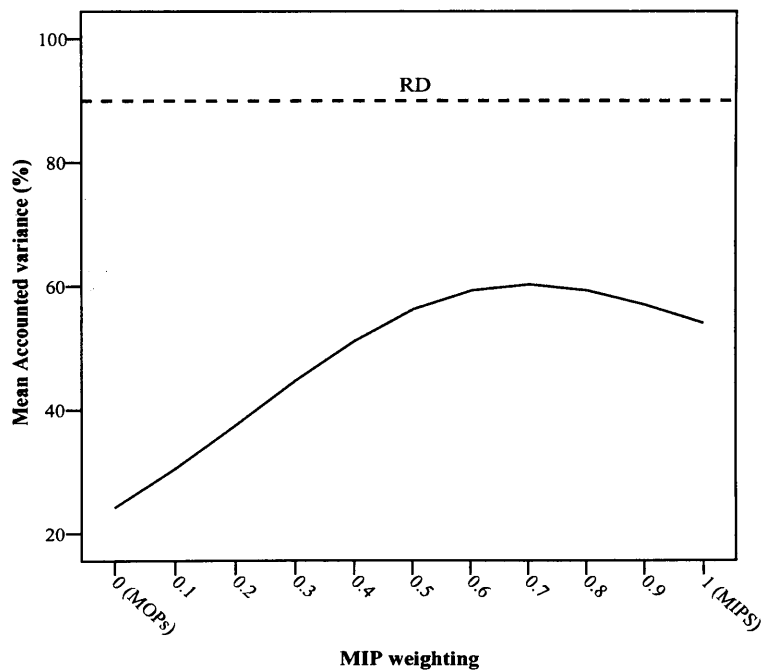


Figure 14. Graph showing the variance accounted for by MIPs/MOPs when they are weighted differentially. The dotted line represents the variance accounted for by RD.

### *Discussion*

The results from Experiment 1 provide evidence for a strong relationship between transformation distance and perceived similarity. These data provide support for RD, extending the empirical results of Hahn et al. (2003) to a new stimulus domain. Transformational relationships accounted for most of the variance in participants' responses, even though there were no free parameters involved in fitting the data, and the fits obtained were superior to those of SA models. The transformation set was next tested on materials that involved modifications over two dimensions as used by Larkey and Markman (2005).

### **Experiment 2**

For the second experiment, RD was tested using the same general stimulus domain as Experiment 1, but the stimuli varied across two featural dimensions: colour and shape. This not only provides a test with a more complex stimulus set, it also enables examination of the way these featural dimensions interact.

Larkey and Markman (2005) found that the combination AB/CA is less similar than the combination AB/AC when the relationship on the second dimension is AB/AB. However, the opposite pattern is observed when the relationship on the second dimension is AB/BA (see Figure 15). This, they argued, is problematic for RD if these relationships are assumed to reflect transformations, because the impact of a given transformation on one dimension is not independent of the transformational relationships governing the second dimension. This means that no simple scheme for weighting transformations could be invoked to explain the rank-order similarities observed in the data, because the relative weight of a particular transformation would

seem to depend also on the relationships governing the other dimension, leaving the account over-parameterised and ad hoc.

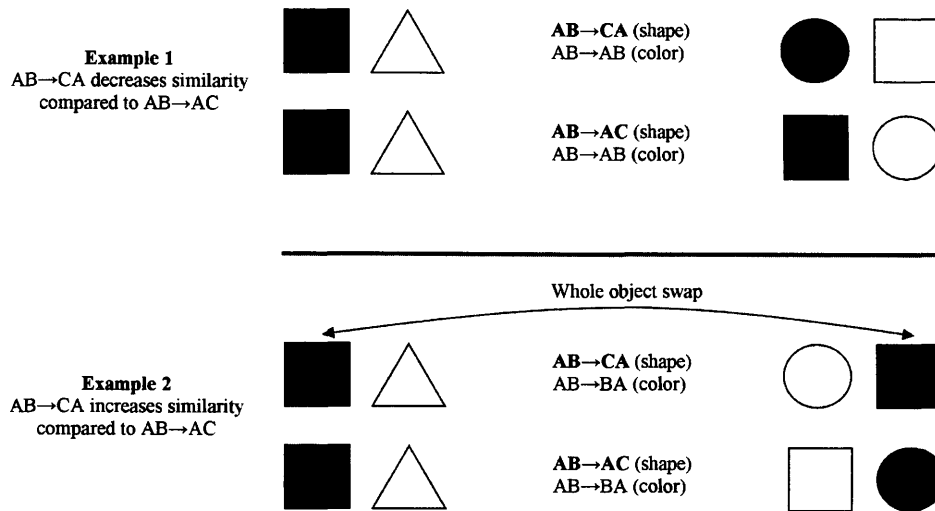


Figure 15. An example of a transformation's supposed 'non-fixed' effect on similarity, as discussed by Larkey and Markman (2005). The shape 'transformation'  $AB/CA$  reduces perceived similarity relative to  $AB/AC$  in the upper case whereas it increases perceived similarity in the lower case when paired with colour swap,  $AB/BA$ .

However, no differential weighting of transformations is required: interactions arise readily within the simple 3-operation coding scheme presented above. Figure 15 depicts the interaction discussed by Larkey and Markman (2005); on inspection of Figure 15, it is clear in the lower example that one of the shapes is naturally perceived to have swapped location. If transformations are applied to objects in their entirety, and not just to single dimensions, as is inherent in any kind of 'move-object' operation, the presence of dimensional 'interactions' is guaranteed. Consequently,

interactions do not, in and of themselves, provide evidence against a transformational account. Specifying a ‘transformation’ for each feature assumes that the representation itself separates these features, even when these dimensions are manipulated simultaneously.

Nevertheless, it remains to be seen how RD will fare qualitatively on such interactions present within this domain.

The goal was to provide a test that would give results representative of all the stimulus combinations contained within the domain. Given the 14 distinct combinations per dimension (see Figure 10), the full set of all possible two dimensional combinations would contain  $14 \times 14$ , that is, 196 items (see Appendix A.2 for the full set of possible combinations along with the corresponding transformational predictions). A forced-choice design, which pairs each item with every other item, was no longer possible for this many items (over 19,000 comparisons would be required) so a different dependent measure was adopted in Experiment 2: a rating task. However even rating 196 distinct pairs seemed unduly onerous. Consequently a random subset of items was selected; statistics obtained from such a random sample should provide estimates that are representative of the entire population.

### *Method*

#### *Participants*

A total of 33 participants took part in Experiment 2 (mean age = 19.9, range = 18 to 23). The subjects were Cardiff University students who were either allocated course credit or paid £4 for their participation. Each participant was tested individually.

### *Materials and procedure*

As in Experiment 1, participants were presented trials on a 19" LCD screen. On each trial participants were presented with a single comparison that consisted of two object pairs; the base pair was situated on the left of the display and the target pair was situated on the right. The shapes used were squares, circles and triangles. All shapes were created using the 'Autoshape' function on Microsoft Publisher. Each shape object was 2.5 cm x 2.5 cm tall. The comparisons were displayed horizontally on each trial, with a 7.5 cm distance between the pairs. The within-pair distance was 0.5 cm. The two pair set was located directly in the centre of the display.

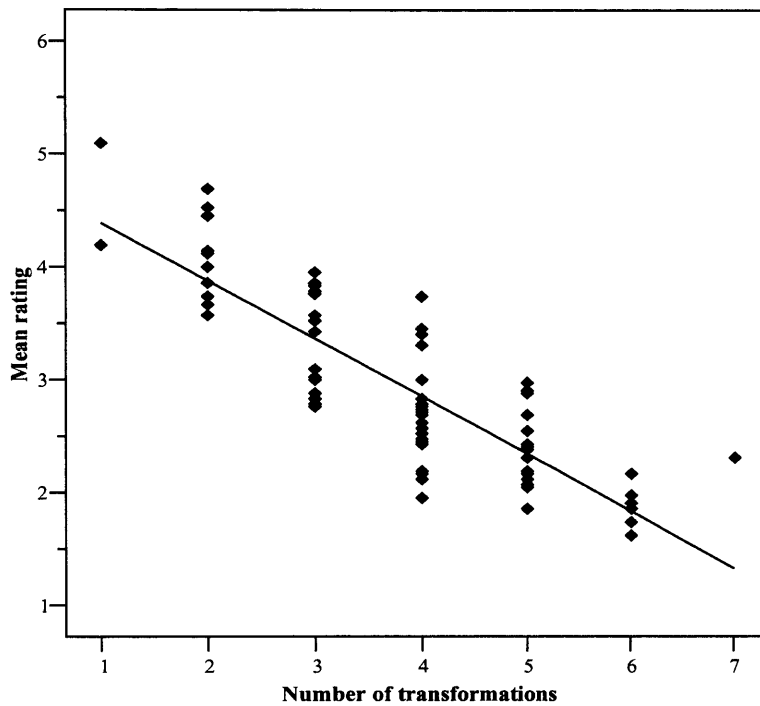
Each dimension (shape and colour) could be combined in one of fourteen predetermined methods (see Figure 10 for these methods applied to one dimension). A random subset of 81 combinations was selected from all possible combinations. This subset was presented to all participants.

As in Experiment 1, this set of 81 comparisons was instantiated in more than one specific set of shapes and colours - specifically, there were a total of three possible unique shape and colour combinations to which participants were systematically assigned. As before, there was no predicted difference between assignments, and so results across these groups were collapsed for subsequent analysis. The order in which participants saw the object pairs was randomised for each of the three assignments.

### *Procedure*

Participants were first presented instructions on the computer screen before being told to press the space bar to commence the task. On each trial, the participants were presented with a comparison consisting of a base pair and a target pair and were

required to provide a rating of similarity between the pairs. Similarity ratings were made on a Likert scale between 1 and 6 (1 = highly dissimilar; 6 = very similar). Participants made the appropriate response on the keyboard for each trial. After the response, the screen was erased and another stimulus comparison was randomly selected for the next trial. All participants rated each of the 81 trials.



*Figure 16.* Graph depicting the relationship between transformation distance and the mean similarity ratings, from 1 (very dissimilar) to 6 (very similar). Each point represents the mean similarity for an object set across participants and assignments.

### *Results*

The mean similarity ratings, averaged across participants and shape assignments, were correlated with transformation distance. Figure 16 indicates, once again, the predicted negative correlation between participants' perceived similarity

and transformation distance, thus matching the pattern exhibited in Experiment 1. In line with previous findings, the nature and fit of the scatter suggests a linear fit between transformations and similarity (see Experiment 1; Hahn et al. 2003).

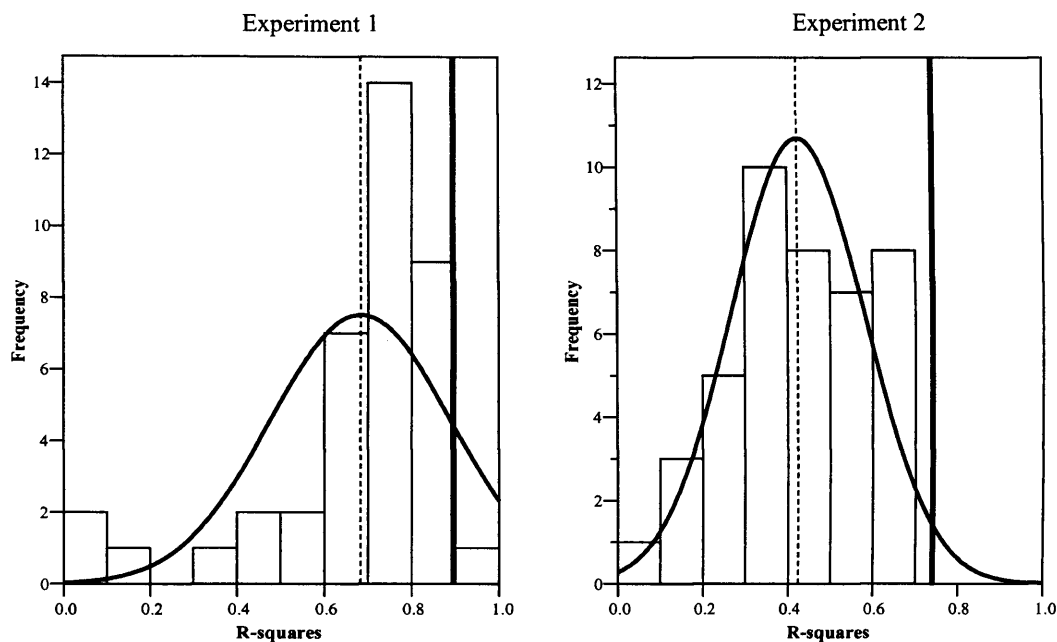


Figure 17. Left: Histogram plotting the frequency of R-square values across the whole sample for Experiment 1. Right: Histogram plotting the frequency of R-square values for Experiment 2. The dotted line signifies the mean R-square and the thick line signifies the variance accounted for by the coding scheme for RD.

As before, there was a significant bivariate correlation between transformation distance and perceived similarity ( $r = -0.86$ ,  $p < 0.01$ , Pearson's  $r$ ). Transformation distance, as determined by the three-operation coding language, accounted for 73% of the variance in perceived similarity ( $R^2 = 0.732$ ), without free parameters.



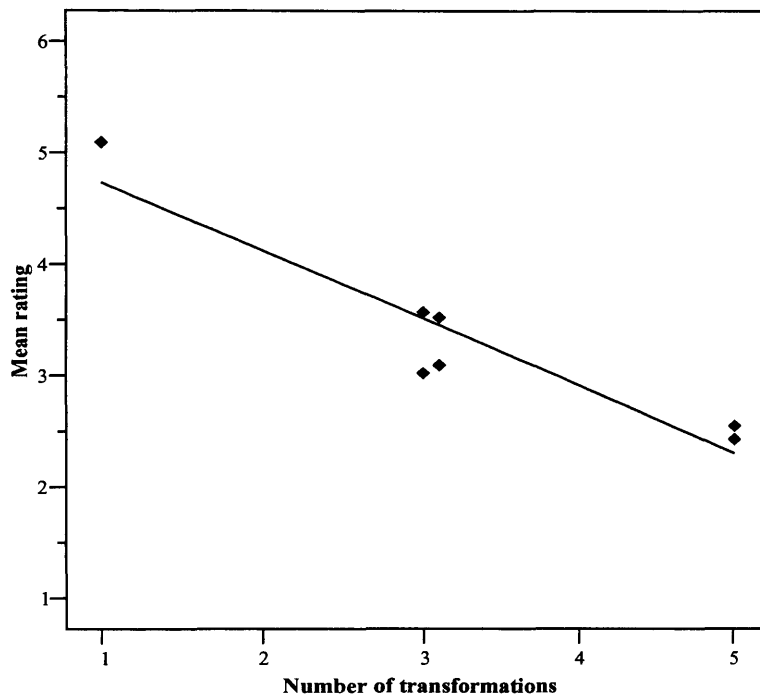
### *Comparing experiments*

Although a strong correlation was found in Experiment 2, the coding scheme accounts for less of the variance than in Experiment 1. Why is there a poorer fit of the data in Experiment 2? One possibility is that the behavioural data are simply noisier; either because of the greater complexity of the stimuli or because of the use of a rating task, rather than a two alternative forced-choice, as in Experiment 1. To investigate this possibility each participant was correlated with the rest of the sample: these correlations provide a baseline against which to evaluate model performance. Essentially, if participants themselves agree only very little, then there is only so much any given model can do.

The histogram (see Figure 17) plots R-square values for each participant, where each participant's ratings were taken as a 'predictor' of the group's average rating (calculated without that participant), that is, as if each participant was a hypothesised model. It also displays the R-square value associated with RD's predictions. The R-square value for RD in Experiment 2 (right panel, Figure 17) is higher than the mean R-square observed. In other words, RD was a more successful predictor of the average similarity rating of each item than were the majority of individual participants. Indeed RD proved a better predictor than *any* one participant. This suggests that RD is doing an excellent job at accounting for the data, especially given the levels of noise present within the data set. This is supported further by considering the same analysis carried out on the data from Experiment 1 (left panel, Figure 17), where 90% of the variance was explained. Here too, RD appears as an 'ideal' participant, in that the amount of variance accounted for by RD, relative to individuals within the sample, is very high. The main difference between the two sets of data is simply that the overall peak corresponds to a much higher R-square value in

Experiment 1 when compared to Experiment 2, that is, both the correlations between model and data and amongst individual participants are higher in Experiment 1.

In conclusion, these analyses suggest that the levels of fit displayed by the model are very good (even though the overall R-square value is reduced relative to Experiment 1), and that RD also provides a good account of a more complex set of items that display structural variations across two dimensions simultaneously.



*Figure 18.* Graph depicting the relationship between mean similarity ratings and transformation distance for interaction items only. As data points were overlapping, horizontal jitter has been added to clearly show all observations.

### *Interaction items*

In addition to assessing all comparisons, the items for which RD predicts an interaction between dimensions were extracted from the data set, as outlined above.

To reiterate, interactions arise on this coding scheme because the ‘swap’ transformation can be applied not just to a single dimension, but to whole objects. To address the unique predictions made by the coding scheme with regards to these items, all object sets that involved a whole object swap are identified (i.e. a swap on both colour and shape). These interaction items consisted of all possible combinations of AB/BA, AB/CA, and AB/BC. Figure 18 shows specifically the relationship between transformation distance and these interaction items. The expected linear negative relationship is preserved for these items and this correlation is significant ( $r = -0.93$ ,  $p < 0.01$ ). As the number of observations are small ( $n = 7$ ), an adjusted correlation coefficient is calculated to correct for this (see Howell, 1997, p. 240): this correlation is also significant:  $r_{\text{adj}} = -0.92$ ,  $p < 0.01$ .

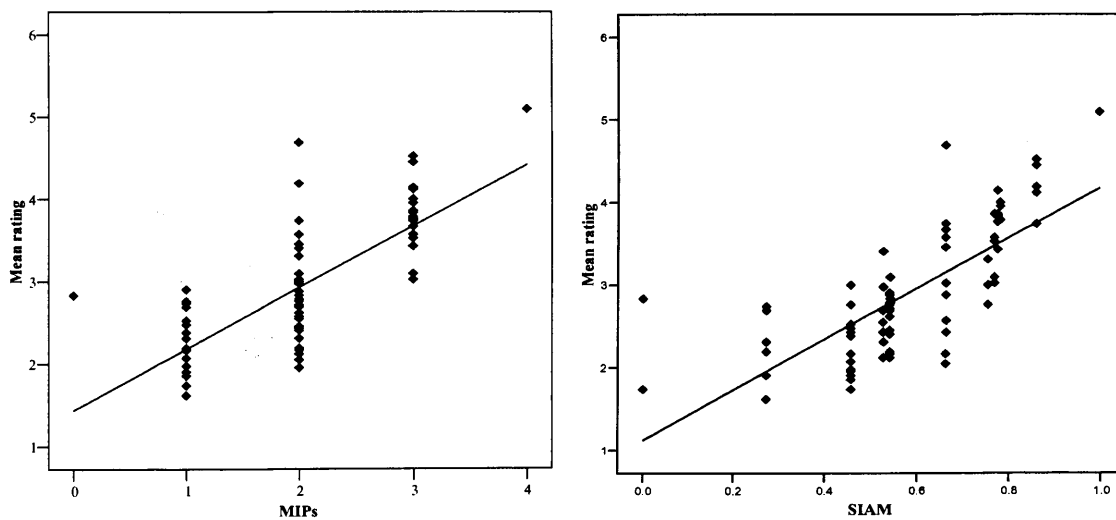


Figure 19. Left: Graph showing the relationship between mean similarity ratings and matches in place (MIPs). Right: Graph depicting the relationship between mean ratings and default SIAM predictions.

### Models of structural alignment

As in Experiment 1, SME and SIAM's respective data fits are examined against the full data set. The left panel in Figure 19 depicts the relationship between perceived similarity and the number of MIPs - as predicted by SME. The predicted positive relationship is evident here, with rated similarity increasing as the number of corresponding features also increases. This correlation is also statistically significant ( $r = 0.74$ ,  $p < 0.01$ ). However, considerably less of the variance is accounted for relative to RD (55% vs. 73%).

SIAM (Goldstone, 1994b), again using the default parameters, also correlated significantly with participants' rated similarity ( $r = 0.74$ ,  $p < 0.01$ ; Pearson's  $r$ ; see right panel, Figure 19). Fits were comparable to those obtained with SME, with an R-square value of 0.54.

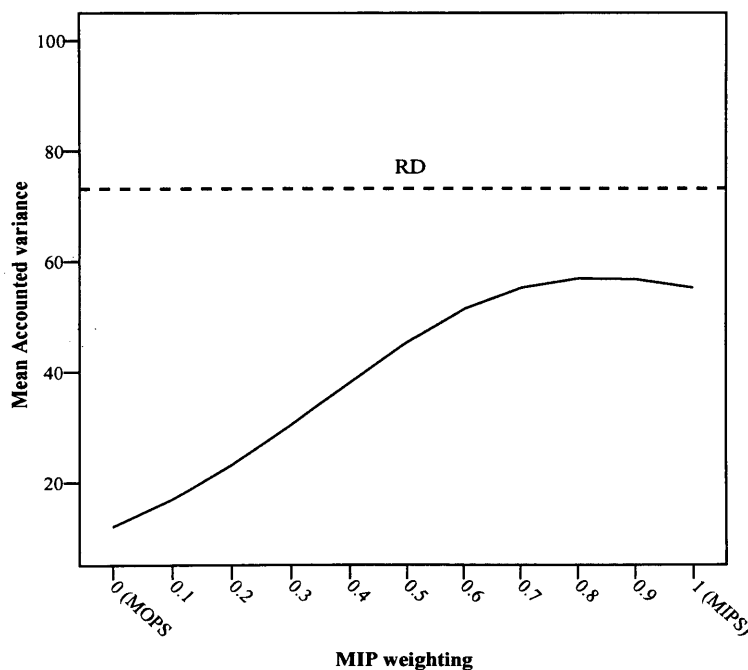


Figure 20. Graph showing the variance accounted for by MIPs/MOPs when they are weighted differentially. The dotted line represents the variance accounted for by RD.

As in Experiment 1, the likelihood ratio for each model pair is calculated (see Glover & Dixon, 2004). If RD is true then the data are 965,771,442 times more likely when compared to SME. When compared with SIAM, the data, given RD, are 2,352,121,407 times more likely. If SME is true, then the data is 2.43 times more likely when compared to SIAM.

Finally, as with Experiment 1, the performance of *any* weighted combination of MIPs and MOPs was investigated. R-square values between weighted MIP/MOP combinations and perceived similarity are plotted in Figure 20. As can be observed, the pattern is similar to that in Experiment 1. As MIPs are given greater weight in the models, relative to MOPs, both the fit improves and the overall best fit obtained – albeit with a slight influence of MOPs (i.e., 0.2). Again, RD provides fits superior to those of any weighted combination of MIPs and MOPs.

### *Discussion*

The results from Experiment 2 show a very similar pattern to those reported in Experiment 1, despite a more complex stimulus set. The results also suggest that the slightly lower fits in Experiment 2, relative to Experiment 1, are a consequence of greater noise in the data, not a specific inability of RD to deal with aspects of these more complex stimuli. These results again provide evidence for a close relationship between transformational complexity and perceived similarity. Furthermore, RD seems to provide an excellent account of the dimensional interactions within this domain. As Larkey and Markman (2005) noted, particular combinations on one dimension can decrease similarity in one case but increase similarity in another depending on the combination on the other dimension. The simple coding language

tested here, which allows the same operations to apply to both individual features and to whole objects, directly predicts, and explains, such interactions.

Finally, comparisons with models based on SA demonstrated that despite obtaining moderate data fits, models of SA could not capture the data to the extent demonstrated by RD.

## **General Discussion**

The results from two experiments support the view that similarity can be understood in terms of Representational Distortion, that is, the simpler the transformation required to distort the mental representation of one item into another, the greater the judged similarity between those items. To provide direct predictions, the RD framework requires specifying a coding language (just as, e.g., feature-based approaches require specifying a set of features in terms of which items are encoded, and so on with other models of similarity). The results demonstrate that an extremely simple coding language allows the RD approach to fit the empirical data well, with no free parameters. Moreover, because the stimuli sampled are representative of the domain (containing all possible one-dimensional variations in Experiment 1, and a representative sample of all possible two-dimensional variations in Experiment 2) this excellent performance can be taken to extend to the domain in general; it is not a consequence of selecting only particular kinds of relationships for test.

Furthermore, the use of a considerably larger and, more importantly, an evenly sampled stimulus set gives rise to very different relationships between structural models of similarity than that observed in the selective test carried out by Larkey and Markman (2005). First, RD considerably outperforms all of the SA models tested. This, in part, is undoubtedly attributable to the use of a psychologically plausible

transformation set, as opposed to Larkey and Markman's use of 'physical transformations'. However, Larkey and Markman's results concerning the relationships between SA models themselves also fail to scale to a more representative stimulus set. Whereas SIAM had considerably outperformed SME in Larkey and Markman's study, there was little to choose between the two models on the tests performed here. Whereas SIAM performed slightly better on the one-dimensional items of Experiment 1, SME slightly outperformed SIAM on the representative sample of the domain as a whole in Experiment 2. Moreover, the respective likelihood ratios indicate that these differences discriminate only weakly between the two models.

Finally, additional analyses for both experiments found that *no* model based on differentially weighted MOPS and MIPS outperforms RD on either experiment.

Although a specific version of RD (using a particular coding language) is contrasted with specific SA models here, it is important to emphasise that transformations and structural alignment are not necessarily in conflict. Determining the transformations between one object and another will typically involve specifying an alignment between the representations in question. This raises the possibility that the RD perspective on similarity might yield insights into the process by which structural alignment is achieved. In essence, because the RD framework suggests that the preferred alignment will be that which is required to provide the simplest transformation between objects, the process of alignment should involve a minimisation of transformation distance.

Such a relationship between transformations and alignment has been explored within the realm of object recognition; recognising an object can be viewed as transforming a perceptual input into one or more mentally represented prototypes or

templates where the process of alignment between perceptual input and stored template is of critical importance. One illustration of this point is that there are examples where the speed and ease of object recognition seems to be inversely related to the complexity of transformations in orientation, size and location in object recognition (Bundesen & Larsen, 1975; Cave et al., 1994; Milliken & Jolicoeur, 1992; Tarr, 1995; Tarr & Pinker, 1989; these mental transformations themselves have been extensively studied, for example, in Kosslyn, 1980; Shepard, 1984; Shepard & Cooper, 1982). As a consequence, transformation and alignment are tightly coupled, for example, in Graf's recent Transformational Framework of Recognition (Graf, 2006).

More generally, both transformation distance and alignment play a role in the visual system's determination of object identity. As we, and the objects we perceive move through space, the retinal image of an object changes. One of the most basic functions of the visual system is the maintenance of object constancy across systematic changes in retinal input such as location, rotation, size/scale changes, or changes in colour. Deciding that an object is the same object across two viewpoints, as opposed to a new and separate entity, is again a correspondence or mapping problem. That transformation distance is a factor in determining this kind of correspondence is supported by studies of apparent motion (Anstis & Mather, 1985; Bundesen, Larsen & Farrell, 1983; Farrell, 1983; He & Nakayama, 1994; Oyama et al., 1999; Shechter et al., 1988; Shepard & Judd, 1976). Apparent motion refers to the illusory motion or 'impletion' between sequentially presented stimuli that appear in different spatial locations. The experience of apparent motion is reliant on the visual system drawing correspondences between objects in different frames as a means of maintaining or preserving perceptual constancy (Oyama et al., 1999; Ullman, 1979).



Crucially, the correspondence strength between objects in two frames has been found to increase with their similarity, as determined by a number of perceptual attributes, such as colour, luminance, size, shape and orientation (Anstis & Mather, 1985; He & Nakayama, 1994; Oyama et al., 1999; Shechter et al., 1988). Moreover, a number of studies have explicitly investigated the role of visual *transformations* between objects in alternating frames. For example, the presentation of a sequentially alternating object at different orientations creates the perceptual illusion of a single object rotating through space (Farrell & Shepard, 1981; Shepard & Judd, 1976). Hence, the intimate connection between correspondence and transformation is strongly supported in this domain, by evidence that the visual system uses transformations to determine object correspondences in the service of object constancy.

All of this suggests a reciprocal relationship between mapping and transformations, whereby mappings constrain what transformations can be applied, but transformations themselves constrain what mappings are assumed.

This general relationship between mapping, transformation distance and object identity also surfaces in the present data with the interaction items. The coding scheme makes specific predictions about interactions between the two independent dimensions of these materials. A simple example is shown in Figure 16. The operations in this coding scheme need not be applied to each feature dimension separately and summed (i.e. swap (colour) + swap (shape) = 2) because the coding scheme allows operations to be applied to whole objects, as well as to individual features. Furthermore, this ability to manipulate representations of objects in their entirety seems definitional of what it means to represent a set of features as a single object in the first place.

Finally, this relates to the observation that there seems to be a bi-directional relationship between transformational relationships and object representation (Hahn et al., 2003). Given the pair of objects on the left of Figure 21, it might be natural to encode them simply as ‘a black square and a white triangle’. The relative location of the two objects within the pairs is not necessarily salient, and so might not be included in our mental representation of these objects. Once contrasted with the pair on the right, however, the relative location becomes relevant, and is effectively made salient through a simple psychological transformation, that is, a ‘swap’. Transformations, mapping, object representation, and object identity all go hand in glove.

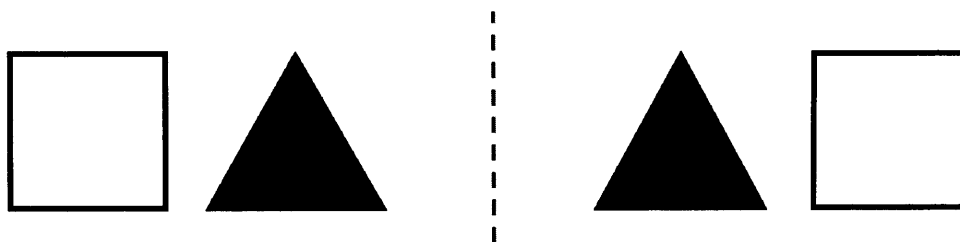


Figure 21. The stimuli used in Experiment 2.

Transformation distance in all of these contexts might be viewed as the quantity that is *minimised* when making alignments; in other words, minimisation of transformation distance might provide a functional, computational level, explanation of how alignment takes place.

Despite this close relationship between RD and alignment in general, specific instantiations of each may give rise to very different predictions. The poorer fits of the specific SA models tested suggest that the constraints they embody might be too rigid. Of course, for RD, the ability to predict perceived similarity depends on a growing

understanding of psychologically relevant transformations. Moreover, it seems inevitable that the set of relevant transformations will differ depending on the types of items being compared. The mathematical theory of Kolmogorov complexity ensures, however, that a very general theory of representation and transformation is viable, and has appealing theoretical properties (Chater & Vitányi, 2003). However, to be useful as a cognitive account, transformational predictions will need to be exhaustively validated not only across (i.e., Hahn et al., 2003) but *within* stimulus domains, particularly those which require structured representations (Hahn, et al., 2003). Given that this domain contains a number of appealing properties and is, contrary to Larkey and Markman (2005), well suited for testing transformations in an explicit similarity task, these materials are perfectly suited for such an exhaustive exploration. To this end, therefore, this domain will now be followed through both a range of tasks and a selection of theoretical comparisons. In Chapter 3, this coding language will be tested using an implicit task – the *same-different paradigm* - to see whether such transformations are relevant above and beyond direct measures of similarity.

### *Summary*

This chapter presented empirical support for Representational Distortion, a transformational theory of similarity. Two experiments provided new evidence that the similarity between objects is determined by the complexity of the transformation required to distort the representation of one object into that of another. Furthermore, the relationship between RD and other structural accounts of similarity was explored, emphasising how transformational and structural alignment processes may be complementary, rather than standing in competition.

# 3

---

## The Time Course of Similarity

Although a number of studies have, to date, supported a transformational approach, there are still significant steps to take to further establish this account. For example, it is important to determine whether there is a place for structure and transformations above and beyond direct judgements of similarity. Previously, similarity judgements have been gathered in two ways (Medin, Goldstone & Gentner, 1993): direct or *explicit* similarity assessment (e.g., subjective ratings, forced choice), and indirect or *implicit* measures, such as false positives in old-new recognition, reaction time or similarity as measured through related processes, that is, categorisation or induction (i.e., does similarity model  $x$  predict the likelihood of a particular classification). Whereas Chapter 2 provided strong evidence for RD using the first type of measure – forced choice and similarity ratings - this chapter seeks to provide further evidence for a transformational notion of similarity using an indirect measure of similarity: the same-different paradigm. So far, evidence for RD has been based wholly on explicit similarity judgements (Hahn et al., 2003; Hahn et al., 2009). Both explicit and implicit

measures have their benefits and drawbacks; in particular, ratings provide an individual's subjective and unambiguous assessment of similarity. However, it is unclear how results obtained with these ratings extend to similarity as it functions implicitly in a wide range of cognitive tasks, such as categorisation, inductive reasoning and so on. Implicit tests, by contrast, are less readily interpretable but potentially address not only specific similarity accounts but also the sorts of representations that underlie low-level stimulus comparisons in general.

Parallels between similarity research and perceptual theory provide some grounds for believing that transformations *are* relevant beyond explicit tasks. As discussed in Chapter 2, navigating our environment continuously transforms our retinal input. However, even under these circumstances we are still able to maintain object constancy, suggesting that our visual system is equipped with mechanisms and representations attuned to visual transformation (e.g., rotation, dilation). For example, in object naming/recognition tasks the speed and ease of responses have commonly been associated with the transformational distance between compared object viewpoints (e.g., Bundesen & Larsen, 1975; Bundesen et al., 1981, 1983; Graf, 2002, 2006; Lawson, 1999; Lawson, Humphreys & Jolicoeur, 2000; Lawson, Bultoff & Dumbell, 2003; Tarr & Pinker, 1989). In addition, studies of apparent motion have also suggested that, along with spatio-temporal proximity parameters, similarity and transformation distance are key factors in facilitating motion correspondence (Bunden et al, 1983; Farrell, 1983; Shepard & Judd, 1976). It is important to stress that these studies differ from studies of transformational similarity in the context of RD in that they often manipulate only a single transformation (i.e., the degree of rotation) as opposed to manipulating multiple transformations simultaneously (but see Lawson et al., 2003). Overall, however, this literature does provide a basis for the

idea that transformational relationships will be readily identified in a perceptual task, and given that similarity in RD also refers to a distance from an original object identity (particularly in a directional similarity comparison), this parallel seems both relevant and justified.

Considering the conceptual relationship between transformations in visual perception and the transformational approach to similarity, it would be desirable to manipulate structure-based transformations and their role in judged similarity more directly. To this end, a speeded same-different (or 'perceptual matching') task was used to implicitly assess similarity. Previous research has consistently confirmed that greater difference or dissimilarity gives rise to faster RTs in a speeded, same-different task (Cohen & Nosofsky, 2000; Farrel, 1985; Goldstone & Medin, 1994; Posner & Mitchell, 1967; Takane & Sergent, 1983). Likewise, similarity affects accuracy: similarity between non-identical objects will result in a response competition between *same* and *different* responses; consequently, greater similarity leads to higher error rates on different trials.

Despite the importance of similarity in the same-different paradigm, the relationship between specific similarity models and same-different performance has been relatively under-explored (but see Frost & Gati, 1989; Takane & Sergent, 1983; Tversky, 1977). Tversky (1977), for example, used the predictions of the Contrast Model to account for patterns of same-different responding between Morse Code signals of varying length (see Rothkopf, 1957). Frost and Gati (1989) compared the Contrast Model and spatial models using a same-different task and showed that adding a common feature to both objects in a comparison increased their similarity, as measured by the number of incorrect same responses. Crucially, this result was not predicted by the spatial model, as noted in Chapter 1.

Despite the benefits of measuring similarity implicitly (see above), there has been very little research into how more complex representational schemes fare in speeded tasks, such as the same-different task. The only study to specifically investigate structural models of similarity was conducted by Goldstone and Medin (1994) who used the same-different task to test the dynamic, time-course characteristics of the similarity model SIAM. SIAM, as noted earlier, belongs to the larger class of models known broadly as structural alignment models, or SA (Gentner, 1983, 1989; Markman & Gentner, 1993a, 1993b). As was described in Chapter 2, the SA framework assumes that similarity is determined via the same processes of alignment that are involved in analogical mapping (Gentner, 1989). Crucially, this can involve aligning the features, objects and relations in one scene with those in another, depending on the representations that are compared in a given context. In the case of SIAM, this function is carried out by a process of interactive activation (McClelland & Rumelhart, 1981). As was introduced in Chapter 2, there are two kinds of matches in SIAM, and SA more generally, that are fundamental in governing model behaviour: 1) matches in place (MIPs) and 2) matches out of place (MOPs). To restate, a MIP is a feature match between corresponding objects, whereas a MOP is a feature match between objects that have not been placed in correspondence (see Figure 5, Chapter 2). Optimally, SIAM will make correspondences that maximise the number of MIPs, that is, make correspondences that are globally consistent with other correspondences. Crucially, this type of alignment, given its complexity, will be associated with a particular time course, as determined by the complexity of the compared mental representations. Early on, MIPs and MOPs will be equally salient, meaning that locally consistent matches (i.e., matching attributes), as well as one-to-one matches, will influence similarity. However, over time, MIPs will grow in

salience and principally determine similarity.

Goldstone and Medin's (1994) data supported SIAM's predictions for thirteen stimuli that varied along four dimensions. Specifically, MIPs and MOPs had a near equal influence on similarity at short response deadlines. At long deadlines, however, MIPS had a much larger influence on similarity, as determined by the number of errors.

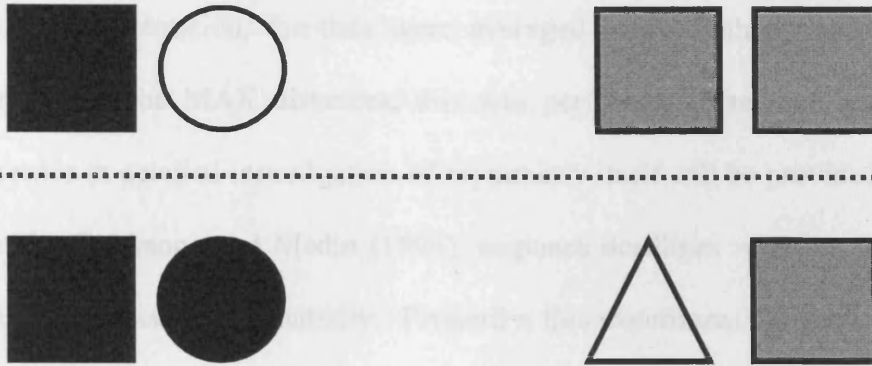
In the present study, the same-different paradigm is likewise used to provide a detailed investigation of similarity with particular focus on RD, a theory of similarity that is yet to be tested in an implicit domain. Additionally, however, RD will also be compared with both a featural model and an alignment model that relatively weights MIPs and MOPs (as in Chapter 2). Given that feature-matches provide a simpler, non-structured account of these comparisons, this SA model provides an ideal benchmark against which the importance of structure can be inferred.

### Experiment 3

For Experiment 3, the stimulus set and coding scheme from Chapter 2 were used. As in Chapter 2 (Experiment 4), a random subset of all possible two-dimensional comparisons was employed (81 comparisons), that is, those stimuli that vary in both shape and colour (see Figure 22). Likewise, this subset should provide a sufficient estimate of the entire population. To recap, the three transformations that govern similarity in this domain take the base pair and modify it as follows: 1) *Create* – taking the base pair this operation creates a new feature that is unique to the target pair; 2) *Apply* – this operation takes an object or entity that is currently available (by being present in the base or by having been created via step (1) and applies it to *one* or *both* of the objects in the target pair. 3) *Swap* – this swaps features between a pair



of objects *or* swaps the object in its entirety (i.e., on both dimensions)<sup>4</sup>.



*Figure 22.* An example of the stimuli used in Experiment 3. As in Experiment 2 objects vary in both colour and shape. Top panel: a comparison that reflects the code structure ‘AB/AA (shape) – AB/CC (colour)’. Bottom panel: comparison following the structure ‘AB/CA (shape) – AA/BC (colour)’.

Participants took part in a sequential same-different task whereby the entire random subset of 81 comparisons made up the different trials. Unlike the tasks employed in Chapter 2, the sequential same-different task has a directional component, that is, a sequential stimulus presentation is analogous to the similarity statement “how similar is A to B?”. In a non-directional similarity task/statement, such as “are A and B similar?”, the RD predictions are taken to be the MAX distance, that is, the greatest distance out of A to B and B to A (see Hahn et al., 2003; Li & Vitányi, 1997). Crucially, this measure of information distance, which was also used

<sup>4</sup> In a directional similarity comparison (i.e., ‘how similar is B to A’ as opposed to ‘A and B are similar’), the term ‘base’ refers to the referent object or A.

in the previous chapter, provides a symmetrical estimate of the distance between two objects (see also Appendix A.1)<sup>5</sup>.

As the length of transformation, for some of these stimuli, depends on the direction of presentation, the data were averaged across both directions and then correlated with the MAX distances; this was performed to remove any unwanted asymmetries (a detailed investigation of asymmetry itself will be provided in Chapter 4). Unlike Goldstone and Medin (1994), response deadlines were not imposed and stimuli were presented sequentially. Primarily, this experiment indicates whether the coding language, that provided accurate fits of similarity rating data, will successfully extend to an implicit measure of similarity, that is, response-time in a same-different task.

### *Method*

#### *Participants*

A total of 30 Cardiff University undergraduates completed the experiment (mean age = 19.2, range = 18 to 25). Participants were allocated course credit for their participation. Each participant was tested individually. As a result of a failure to follow the basic instructions, the data from two participants were later omitted.

#### *Materials and procedure*

The stimulus set used here is identical to that used in Experiment 2. These stimuli are a randomly selected subset of all possible two-dimensional comparisons – 81 comparisons. Trials were sequentially presented on a 19” LCD monitor with a refresh rate of 60 Hz (1024 × 768 pixels). There were three possible features on each

---

<sup>5</sup> A sequential paradigm is used (as opposed to a simultaneous matching task) so that the asymmetries predicted by the coding scheme can also be examined - this analysis is provided in Chapter 4.

dimension (shape = triangle, square, circle; colour = yellow, purple, green; for a detailed description see Chapter 2; Appendix A). Each shape was 2.5cm wide x 2.5cm tall and each had a strong outline (0.75pt) and bold colour. The screen location of each presented pair varied on each trial by randomly combining predetermined values on each screen axis (i.e., 10, 20, 30, 40, 50, 60, 70, 80 and 90). The stimulus duration for a given pair was 833ms (50 frames) and the inter-stimulus interval (ISI) was 16ms (1 frame). A response could be given at the onset of the second stimulus pair. All participants were seated approximately 60cm away from the screen.

Stimulus comparisons were presented in a random order. Different trials were presented in both directions (i.e., each participant saw the left pair in Figure 21 followed by the right pair, and vice versa) resulting in 162 different trials (81 x 2). The 'same trials' were generated by pairing each of the eight composite pairs (AB, AA, BA, BC and so on) with itself 16 times, resulting in 128 same trials. Participants could indicate 'same' by pressing Z on the keyboard and 'different' by pressing M. The allocation of button to response was counterbalanced. A 'different' response was given if pairs differed in any respect other than screen location, that is, on a single dimension or on both. After a response was given, the screen was erased and a new pair was randomly selected for the next trial. No response deadline was imposed but participants were urged to respond as quickly as possible.

### *Results*

For analysis, both reaction time (RT) and error responses are reported. Specifically, the RTs on correct different trials are analysed. To derive an RT for each comparison, the RT is averaged across both directions of presentation (i.e., pair A to pair B and vice versa). RTs that were three standard deviations above and below

the overall mean were removed (2% of trials were removed). 'Mean errors' in the graphs below refer to the number of false-positives per comparison (i.e., where participants incorrectly judged two different objects to be identical when they are the same) when averaged across the two response button assignments (i.e., the allocation of 'z' and 'm' to *same* and *different*). These errors did not include removed outliers.

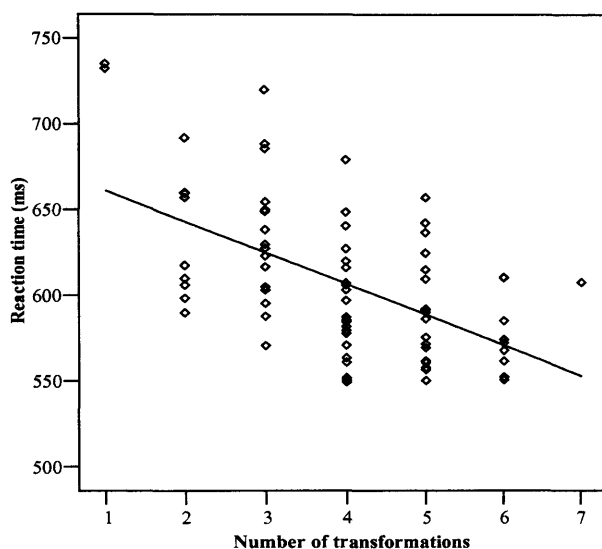


Figure 23. Graph depicting the relationship between transformation distance and reaction time ( $r = -0.55$ ).

### Transformations

As noted above, differences will be identified more quickly where items are less similar, that is, for pairs related by greater transformation distances. As seen in Figure 23, this expected negative, monotonic relationship between RT and transformation distance was indeed observed. A bivariate correlation between the number of transformations and RT for the *different* trials was found to be significant

using Pearson's  $r$  ( $r = -0.55$ ,  $p < 0.01$ ). Without free parameters, transformation distance, as specified by the coding scheme, accounted for 31% of the variance in reaction time for correct *different* responses ( $R^2 = 0.31$ ).

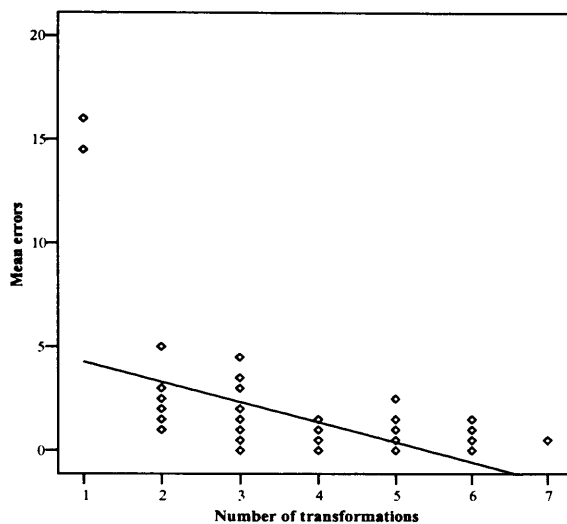


Figure 24. The relationship between error responses and transformation distance.

Error responses, like RT, should be positively related to similarity. Although response deadlines were not implemented, there were still a sufficient number of errors for each comparison across all participants to provide interpretable results (4.65% of all responses are errors). The likelihood of making an error, or falsely judging two different objects to be the same should increase with their similarity, which again implies a negative relationship with transformations: erroneous 'same' responses will be more likely to arise when the transformation distance is small, as differences will be harder to identify. As seen in Figure 24 this relationship was observed; errors decreased significantly with transformation distance ( $r = -0.52$ ,  $p < 0.01$ ), with our coding scheme accounting for 27% of the variance ( $R^2 = 0.27$ ).

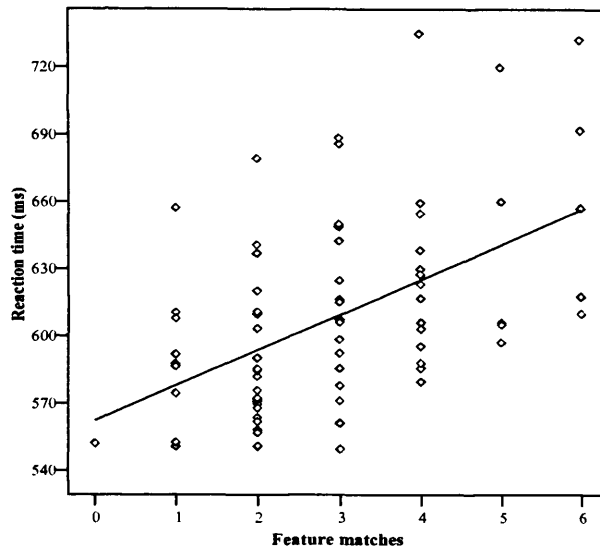


Figure 25. The relationship between the number of feature-matches and reaction time.

### Feature matches

As a model comparison the total number of features matches was correlated with the similarity data. In calculating this featural model, object features were matched regardless of where they appeared in each object and a feature could be matched with more than one feature in the other object. In the top panel of Figure 22, for example, the square in left hand pair is matched with both squares in the right hand object, resulting in two matches overall. Matching features in this way captures the notion that similarity is greater when there are two matching features in the compared object compared to the when there is a completely new additional feature, such as the triangle shown in the bottom panel of Figure 22.

Again, reaction time should increase with similarity. For the feature model, which measures similarity directly, not distance, this means a positive relationship where reaction times should increase as a function of the number of matching features. The graph in Figure 25 displays the relationship between RT and the

number of feature matches. A positive relationship is clearly evident and this relationship is born out statistically ( $r = 0.51$ ,  $p < 0.05$ ; Pearson's  $r$ ). The variance accounted for is slightly less than RD at 29% (RD for RT = 31%). As in Chapter 2, the likelihood ratio for each model contrast is calculated (see Glover & Dixon, 2004). If RD is true then the data are 3 times more likely when compared to the feature matching model.

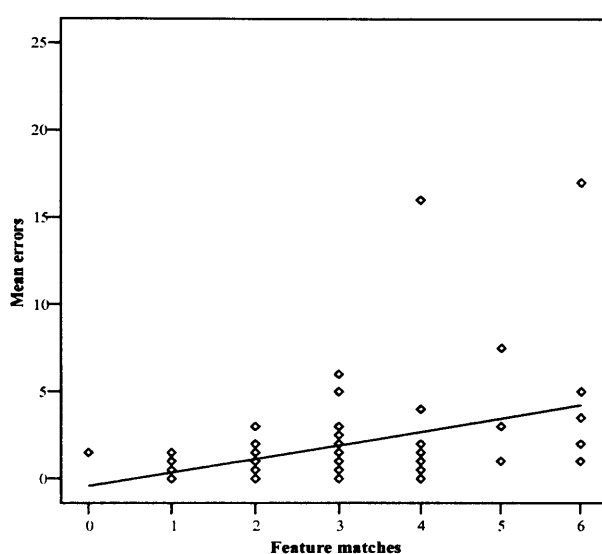
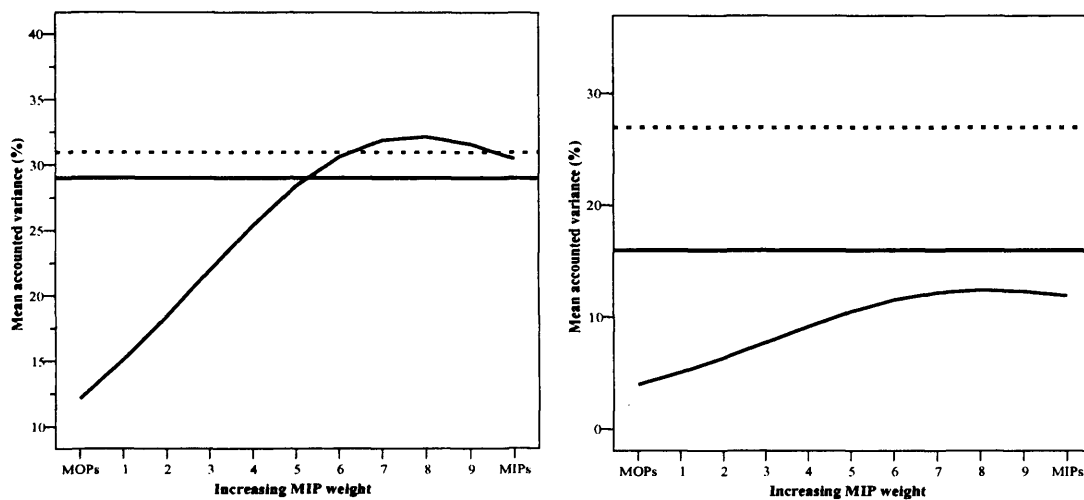


Figure 26. The relationship between the number of feature-matches and errors.

This basic feature-matching model was correlated with the error data and the predicted positive relationship was again born out (see Figure 26). Feature-matches correlated significantly with the error data using Pearson's  $r$  ( $r = 0.40$ ,  $p < 0.01$ ) and accounted for 16% of variance in error responses, providing, once again, a poorer fit relative to RD. This is confirmed by the likelihood ratio, where the data are more 294 times more likely given RD.

*MIP/MOP weighting*

As in Chapter 2, the relative influence of specific feature alignments was investigated. As stated earlier, models of SA will optimally form alignments that are maximally consistent by placing each feature into one-to-one correspondence, that is, by maximising the number of MIPs. Crucially, such alignments are considered more effortful and complex than simply matching local features, such as in the feature model implemented here (Markman & Gentner, 2005), and involves consideration not just of features but of structure.



*Figure 27.* Variance accounted for by MIPs/MOPs when they are weighted differentially for the RT (left panel) and error data (right panel). Increments on x-axis correspond to percentage weight increases in steps of 10, i.e., tick mark '3' corresponds to a weighted average that is 30% MOP/70% MIP. The horizontal dotted line and horizontal bold line represent the variance accounted for by RD and the feature matching model respectively.



As before, all weighted combinations of MIPs and MOPs are correlated against the participant data. As different models of SA place different weights on MIPs and MOPs, as demonstrated in Chapter 2, this model provides a broad approximation for many instantiations of SA. Figure 27 depicts the variance accounted for by a model that counts differentially weighted MIP and MOPs, considered over the entire range of possible weightings between the two. The marker on the  $y$ -axis signifies the variance accounted for by transformation distance alone.

As seen in the graph, differentially weighting MIPs and MOPs has a profound effect on overall fit. For RT, the optimal model fit is where MIPs are weighted at 80% and MOPs are weighted at 20% (point 8 on the  $y$ -axis). At this point, the weighted MIP/MOP model exhibits a slightly better fit than that provided by RD, and indeed the feature-matching model. Again, however, model complexity needs to be factored into the comparative evaluation; the weighted MIP/MOP model has one free parameter and this provides only a small increase in accuracy (1-2%). This is reflected in the respective AIC (Akaike Information Criterion) values, with RD having a lower AIC than the best fitting MIP/MOP model (243.5 vs. 244.9). Overall, this suggests that transformations provide the more accurate account of the data once model complexity (the number of free parameters) is taken into account.

As seen in Figure 27 (right panel) the optimal fit for errors is at point 8 on the  $x$ -axis where MOPs are weighted 20% and MIPs 80% - here it accounts for 12% of the variance. However, even at this optimal point, the model is accounting for considerably less of the data compared to RD and slightly less than the basic feature-matching model – despite having a free parameter (i.e., the relative weight of MIPs & MOPs).

*Discussion*

Experiment 3 provides the first direct support for RD in an implicit task. This instantiation of RD, based on three simple operations, achieved significant data fits when correlated with reaction time and errors. When compared to the featural model, RD achieves a marginally superior fit of the RT data (as reflected by the small likelihood ratio) and is considerably more accurate for the error data, suggesting, at least to some extent, that transformational relationships and the representations that underlie them are relevant in this sort of task. In addition, an alignment model that differentially weights MIPs and MOPs is tested in detail; crucially, MIPs and MOPs provide a more constrained featural description of these comparisons depending on the relative weight given to these matches; a MIP-only model only predicts one-to-one matches whereas a model that weights MIPs and MOPs equally will not require features to be in one-to-one correspondence and is thus less constrained. Crucially, this model performs worse than RD for both RT and errors, with RD having no free parameters. Interestingly, whilst this model performs better than a crude feature-matching model for RT, it fails to do so for the error data. This may be taken to suggest that structural, one-to-one matches are relevant, at least to some extent, but that the correspondence rules embodied in this model are ultimately not quite right (Goldstone & Medin, 1994).

The fact that the featural model performed poorer than RD (for errors in particular) provides evidence that the underlying representations in this task may be more complex than those permitted under this simpler account. In terms of the transformation set used here, it is assumed that not only separate features are perceived, through the creation and application of new and existing features, but also that structure matters: the individual features are bound together as whole objects,

hence an object, with both its colour and shape, can swap location in its entirety. By contrast, simple features matches alone will, in some cases, provide no evidence for whether something is the same or different! For example, the comparison made up of AB/BA (shape) and AB/AB (colour) has as many features in common as the identical case, AB/AB (shape & colour). Of course, one could always posit other features. In particular, one could propose location-specific features, where 'A-left' matches 'A-left' and so on, but this would then fail for the 'swap' items because the same features will then occupy different spatial locations resulting in no features matches overall. By contrast, transformations can capture the fact that in this task, with no response deadline, responses can be affected by more than just the number of matching features namely both by 'what' features are present and, perhaps more importantly, 'where' they appear in each configuration.

Of course, it is possible that these transformational relationships will be less salient if the time to respond is restricted, that is, when there is less time to encode the structure and bind relevant features into whole objects. Therefore, it might be the case that under increased time pressure, more complex transformations will add no predictive benefits over the basic featural account, or indeed the SA model. To this end, it would be informative to see if and how representations build over time for these objects.

#### **Experiment 4**

The following experiment attempted to disentangle the intimate relationship between model fits and the way representations build up over time. As in previous studies (Goldstone & Medin, 1994; Lamberts, Brockdorff & Heit, 2002), this was

achieved by presenting the same stimuli with four different response deadlines: 500ms, 600ms, '750ms', and 2000ms.

The rationale is this: manipulating the deadline, and with it the encoding and comparison process, should affect what stimulus attributes are available and hence relevant. In general, this design should determine whether there is a strong basis for the claims made in relation to the previous experiment: 1) that representations build up over time and 2) that this 'representation shift' will affect the relative efficacy of specific models over the time course.

At the moment, RD provides only a static, end-state prediction. In other words, it provides a computational level measure of similarity, not a process model. It would, of course, be possible to devise such a model. One step on this way would be to allow differential weights for individual transformations (see Hahn et al. 2003, for discussion), and such weights could change over time. This is not the path pursued here. This is because the static, end-state predictions considered so far are themselves informative when the associated model accuracy can be tracked across changes in an independent variable such as available response time. Specifically, any change in model performance over the time course will provide some insight into the emergence of certain mental representations. Consequently, the coding scheme used thus far will be examined across the time course and compared with the simple featural model. Such comparisons should provide an indicator of when more complex, structured representations arise.

As with Experiment 3, the relationship between RD, feature-matches and certain alignments will be addressed also. As in Goldstone and Medin (1994), there should be a greater influence of MOPs at shorter deadlines with that influence decreasing over the time course.

## *Method*

### *Participants*

A total of 46 Cardiff University students completed the experiment. Participants were paid £2 for their participation. Each participant was tested individually. As a result of a failure to follow the basic instructions, the data from one participant were later omitted.

### *Materials and procedure*

As in Experiment 3 trials were presented on a 19" LCD monitor (refresh rate of 60 Hz). Stimulus properties (object size, colour, shape etc) were identical to the first experiment. However, in this experiment, pairs were presented simultaneously, with one shape pair on the left of the display and another on the right. The left-right position of pairs on *different* trials was counterbalanced resulting in 162 *different* trials (81 x 2). Each individual pair (AB, BA, AA etc) was presented with itself to generate the *same* trials and these were duplicated to match the number of *different* trials resulting in 324 trials overall (162 x 2). Pairs were separated by a horizontal distance of 18.5cm. The prompt "Same or different?" was also presented at the bottom-centre of the display. Participants indicated *same* by pressing Z on the keyboard and *different* by pressing M. Here, this key assignment was reversed for left-handed participants.

Response deadlines varied between participants. From stimulus onset, deadlines could either be 500ms, 600ms, 750ms or 2000ms. As soon as the designated time elapsed the screen was erased and the next trial appeared after a post-trial interval of 500ms. There was no feedback for correct or incorrect responses.

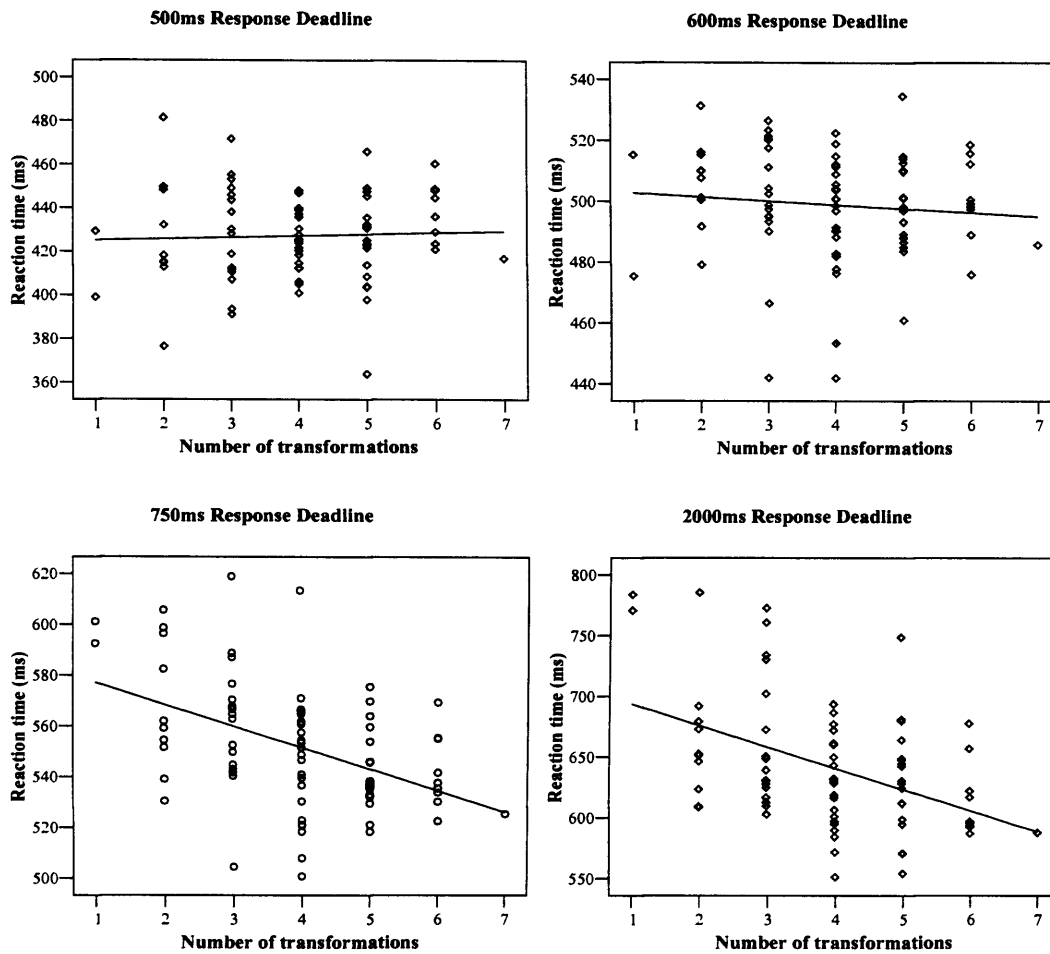


Figure 28. The relationship between transformation distance and RT for each response deadline: 500ms (top left panel), 600ms (top right panel), 750ms (bottom left panel) and 2000ms (bottom right panel). The predicted trend is absent in the top panels and emerges from 750ms onward.

### Results

Both RT and error responses were analysed separately for each response deadline. As before, RTs more than three standard deviations above the group mean and less than 200ms after stimulus presentation were removed from analysis. Only the 2000ms condition contained RTs below the 200ms cutoff and these removed

outliers only constituted 1.98% of the trial data in the 2000ms condition. The correct different trials were analysed and were averaged across screen locations (left to right and right to left<sup>6</sup>).

The graphs in Figure 28 show the relationship between the number of transformations and RT at each response deadline. Firstly, it can be seen that the strength of the relationship between RD and RT changes over time. At 500ms, there is no relationship between similarity, as determined by the number of transformations, and the speed to respond. At 600ms, the predicted relationship is still absent, although the data do begin to show a negative trend. At the 750ms response deadline, the predicted pattern of results emerges quite clearly and reflects the observed relationship shown in Experiment 1. A Pearson's  $r$  showed there to be a significant correlation between RD and RT at 750ms ( $r = -0.46$ ,  $p < 0.01$ ). Likewise, when the participants are allowed 2000ms to respond (equivalent to no deadline at all, Experiment 3), the predicted negative relationship between the number of transformations and RT is evident in the graph and the correlation is significant ( $r = -0.44$ ,  $p < 0.01$ ).

Table 1.

*Descriptive statistics for the four conditions tested.*

| <b>Response deadline</b> | <b>Mean RT</b> | <b>Errors (%)</b> | <b>RT St. Dev</b> | <b>Size of RT range</b> | <b>Distance - mean RT to deadline</b> |
|--------------------------|----------------|-------------------|-------------------|-------------------------|---------------------------------------|
| 500ms                    | 427.21         | 61.97             | 20.56             | 276                     | 72.81                                 |
| 600ms                    | 498.94         | 43.83             | 18.13             | 393                     | 101.06                                |
| 750ms                    | 551.97         | 20.42             | 24.49             | 541                     | 198.03                                |
| 2000ms                   | 641.93         | 6.98              | 51.69             | 869                     | 1358.07                               |

<sup>6</sup> Note, this averaging is only a basic experimental procedure, it is not expected, unlike Experiment 1, that the relative screen location will affect similarity. There is no theoretical basis for assuming that screen position maps cleanly onto the role of objects in directional similarity comparison.

In the data here, there seems to be little evidence for a steady increment in model accuracy over the time course between, in particular, the shortest three deadlines. Instead, the correlation between RD and similarity, as measured by RT, seems to emerge rather suddenly somewhere between 600 and 750ms.

There are three possible explanations for this observation: firstly the selected response deadlines may lack the sensitivity to detect a continuous increase in model accuracy. Alternatively, it may be that the relationship between RD and RT is not a continuous one, that is, in order to make a response based on the information assumed by the coding language, there needs to be a specific amount of time in order to represent the properties manipulated by the transformation set. Thirdly, it may be that the deadlines are so short that there is not enough time to provide responses that are sufficiently variable. The descriptive statistics in Table 1 seem to support this last interpretation. In the 500ms condition, 61% of responses are either incorrect or timeouts, resulting in less interpretable RT data to begin with. Furthermore, RTs are, on average, 72ms away from the deadline - this emphasises the difficulty of making a correct response within this time window.

Table 2.

*Summary of correlation coefficients for RD and the featural model (reaction time).*

| <b>Model\Deadline</b> | <b>500ms</b> | <b>600ms</b> | <b>750ms</b> | <b>2000ms</b> |
|-----------------------|--------------|--------------|--------------|---------------|
| RD                    | .044         | -.093        | -.46(*)      | -.44(*)       |
| Feature matches       | -.11         | -.09         | .42(*)       | .29(*)        |

\* Correlation is significant at the 0.01 level (2-tailed).



For contrast, Table 2 displays the fits of the simple feature matching model. Here, too, there was no relationship between RT and feature matches at 500ms and 600ms, but a significant correlation between the number of feature matches and reaction time at a 750ms response deadline ( $r = .42, p < 0.05$ ) and at 2000ms ( $r = .29, p < 0.05$ ). In both cases, however, the feature model accounts for less variance than RD, as can be seen in Figure 29. At 750ms the size of difference is smaller, as reflected by the likelihood ratio for this condition: the data are 6 times more likely for RD than for the feature-matching model, thus providing moderate evidence for choosing RD in this condition. At 2000ms, however, the feature model provides a much poorer fit of the data when compared to 750ms, accounting for only 8% of the variance at 2000ms. Indeed, the likelihood ratio between RD and the feature model indicates this: the data are 174 times more likely given RD in the 2000ms condition and, therefore, favor RD strongly. As the graph shows, there is no notable change in RD's fit at the longer deadlines, that is, between 750ms and 2000ms.

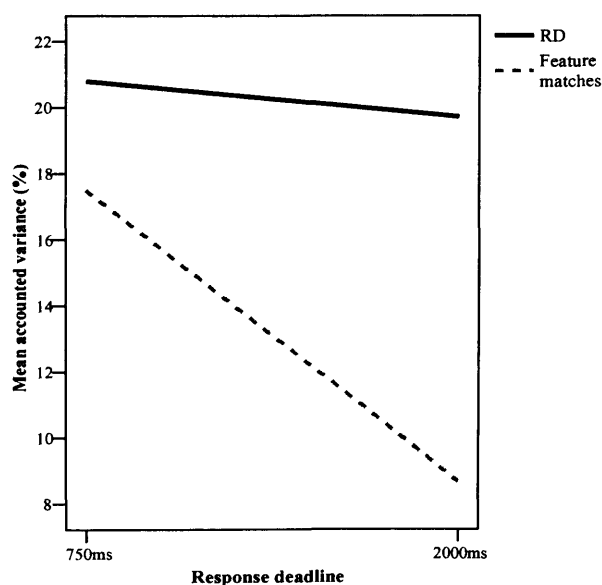


Figure 29. Graph showing the variance accounted for by RD and feature matches over the time course for RT data only.

In sum, RT fails to tell us much about either model at short response deadlines. We next consider whether the same is true of the error data. Given that it was difficult to make a response at short deadlines, not only incorrect judgments but also failures to make an appropriate response in time at all should be related to the similarity between objects.

Table 3

*Summary of Pearson's  $r$  values for each tested model across the time course (errors).*

| <b>Model\Deadline</b> | <b>500ms</b> | <b>600ms</b> | <b>750ms</b> | <b>2000ms</b> |
|-----------------------|--------------|--------------|--------------|---------------|
| RD                    | -0.38(**)    | -0.46(**)    | -0.47(**)    | -0.49(**)     |
| Feature matches       | 0.43 (**)    | 0.44 (**)    | 0.38 (**)    | 0.40 (**)     |

\* Correlation is significant at the 0.01 level (2-tailed).

For analysis, both false-positives and timeouts were classified as errors (see Goldstone & Medin, 1994). The graphs in Figure 30 show the relationship between the number of transformations and error responses at each response deadline. Unlike the RT measure, the predicted negative relationship between transformation distance and errors is evident at each response deadline. The predicted relationship is born out statistically at 500ms ( $r = -.38$ ,  $p < .01$ ), 600ms ( $r = -.46$ ,  $p < .01$ ), 750ms ( $r = -.47$ ,  $p < .01$ ) and 2000ms ( $r = -.49$ ,  $p < .01$ ).

Relative to reaction time, the predicted pattern is shown consistently across deadlines, emphasising the importance of incorporating both of these measures when assessing model performance (see Figure 30). The discrepancy across measures is most profound at the 500ms and 600ms deadlines, where neither model correlated at all with RT; the task was clearly too difficult to see sufficient variation in RTs to get

the predicted pattern of results. Like RD, the featural model performed better across the time course when correlated with errors (see Table 3) and showed significant correlations at each response deadline. The relative performance of each model (i.e., the variance accounted for) can be observed in Figure 31.

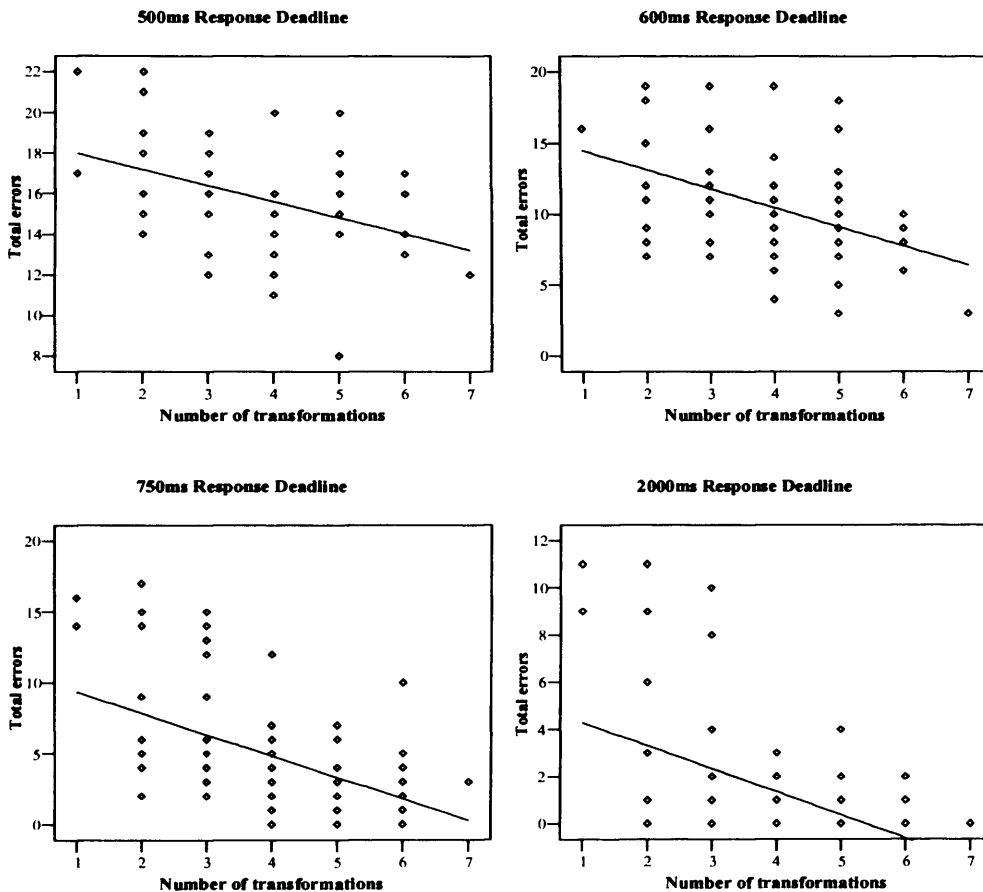


Figure 30. Graphs showing the relationship between RD and the total number of errors per item for each response deadline.

As can be seen, the feature matching model provides a slightly better fit of the data at 500ms: RD accounts 14% of the variance and the featural model accounts for

18%. Correspondingly, the data are 7 times more likely given a crude feature-matching model at 500ms, thus providing moderate support for the feature model at this point in time. At 600ms, the variance accounted for by RD shows a substantial increase whereas the feature model shows little improvement. The data are 3 times more likely given the RD at 600ms, providing slight support for RD over the feature model. As seen in Figure 31, the feature model becomes less accurate over time and performs worse relative to RD at longer deadlines (750ms & 2000ms). This improvement by RD is shown in the likelihood ratios: the data are 44 and 57 times more likely given RD in each of the respective conditions. Both by interpreting the graph in Figure 31 and by studying the likelihood ratios, it is clear that RD is by far the best model at longer deadlines. Trend-wise, RD shows a gradual increase in model accuracy over time whereas the feature model actually accounts for less of the data at longer deadlines. In addition, RD provides the best account of these data at all but the 500ms deadline – although RD provides only a moderately better fit at 600ms.

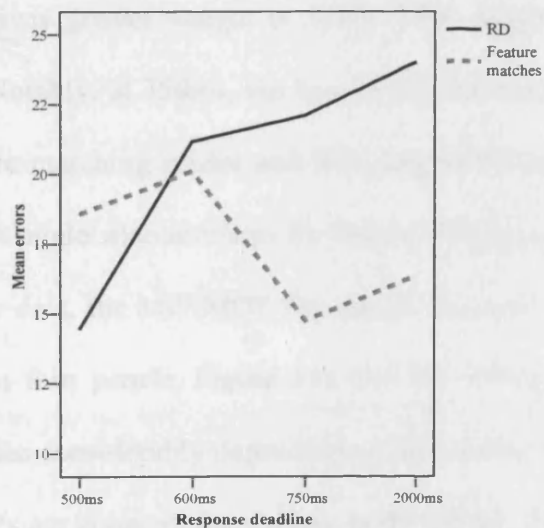


Figure 31. Graph comparing the variance accounted for ( $R^2$ ) by RD and features matches over time for the error data.

*MIP/MOP weighting*

As before, an alignment model was also tested on these data. Based on previous research participants should be more inclined to make alignments outside the best global mapping at shorter deadlines, or align features that go against the *one-to-one* constraint. Although Goldstone and Medin (1994) demonstrated that MIPs and MOPs are equally salient at short deadlines, they used much longer deadlines (1s, 1.84s & 2.86s) and more complex stimuli. Therefore, it is not completely clear what to expect in the present context. Given the much shorter deadlines used here, MOPs may have an equal influence overall, that is, both MIPs and MOPs will be equally salient throughout the range of response deadlines studied here.

The top two panels of Figure 32 depict the change in model accuracy when MIPs and MOPs are differentially weighted for the RT data (the data for earlier deadlines are not shown as no model provided accurate fits). The best-fitting SA model at 750ms is when MOPs are given greater weight in the model - i.e., MOP weighting = 0.7 (see upper left panel, Figure 32). At the 2000ms deadline, the best fitting SA model gives greater weight to MIPs (MIP weighting = 0.8; upper right panel, Figure 32). Notably, at 750ms, the best fitting SA model provides a poorer fit than both the feature matching model and RD, despite having a free parameter. At 2000ms, the best SA model also accounts for less of the variance than RD.

For the error data, the MIP/MOP fits can be observed across the whole tested time course (bottom four panels, Figure 32). For the 500ms response deadline, the model accuracy varies considerably depending on the relative weightings of MIPs and MOPs. When MOPs are given greater weight in the model, 40% to 100%, it achieves a fit that is superior to RD and the featural model and accounts for 22% of the variance at the optimal point.

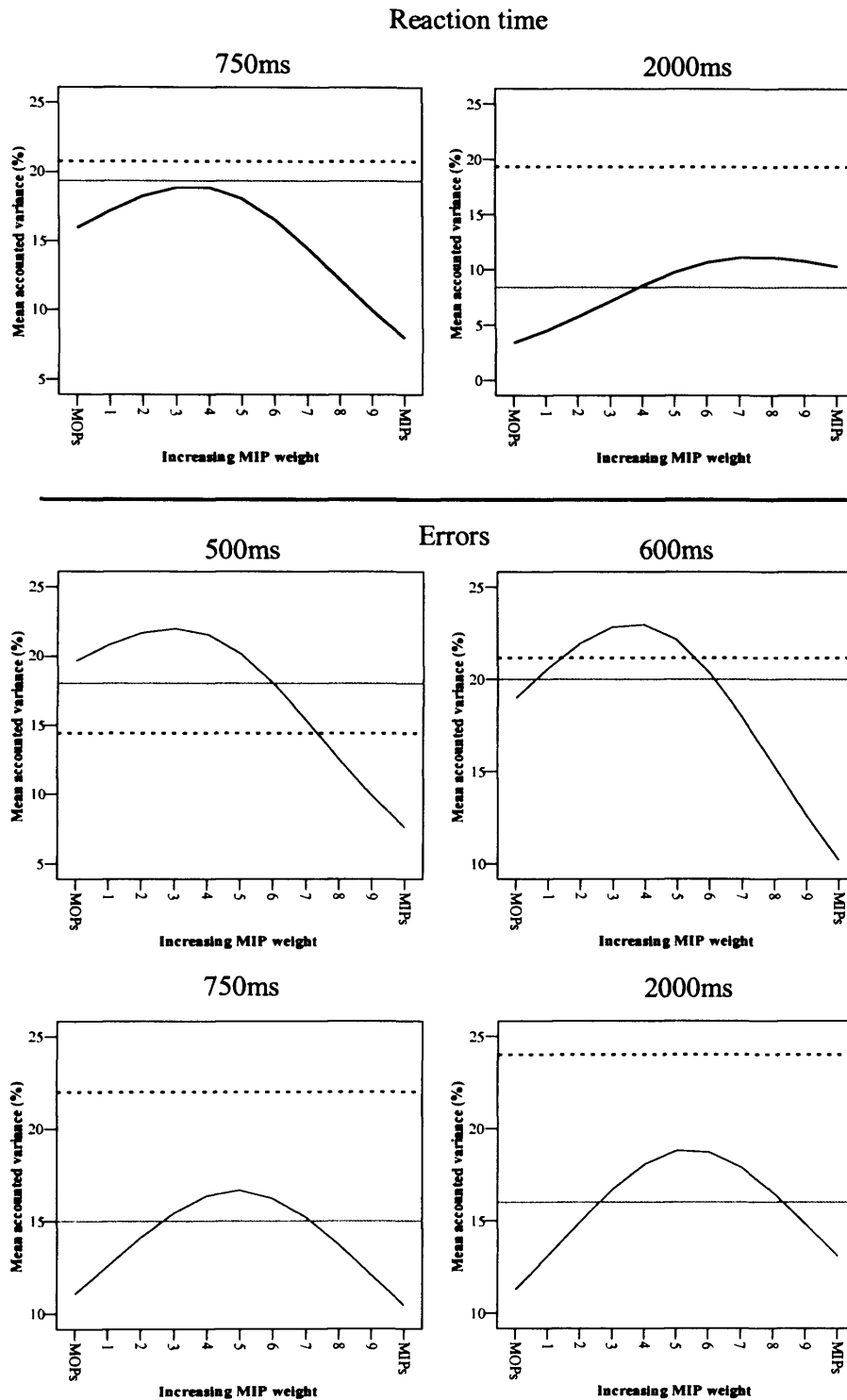


Figure 32. Variance accounted for by MIPs/MOPs when they are weighted differentially. The dotted line and bold grey line represent the variance accounted for by RD and feature matches respectively. The top two panels are the RT data and the lower panels are the error data.

Considering the free parameter, AICs are calculated to determine the cost of model complexity for the SA model (see Table 4). Indeed, once model complexity is taken into account, the SA model performs worse than both RD and the feature matching model at 500ms (see Table 4). Here, the feature model provides the best fit of the data once complexity is taken into account. At 600ms, once again, the MIP/MOP model achieves the strongest fit of the error data set when MOPs are given greater weight in the model. At this optimal point, the model slightly outperforms RD and the feature-matching model, albeit to a lesser extent (21% vs. 19%). Again, however, the SA model provides the poorest fit of the data once this free parameter is accounted for in the model evaluations. Overall, however, there is very little to tell between the models at 600ms.

Table 4

*Summary of AIC values for the 500ms and 600ms conditions. The AIC reflects the amount of information lost once complexity and accuracy are taken into account (i.e., smaller the better).  $\Delta_i$  is the distance between a model's AIC and the best fitting model and  $K$  is the number of free parameters.*

| 500ms | <b>Model</b> | <b>K</b> | <b>RSS</b> | <b>AIC</b> | <b><math>\Delta_i</math></b> |
|-------|--------------|----------|------------|------------|------------------------------|
|       | Features     | 0        | 931.37     | 63.95      | 0.000                        |
|       | RD           | 0        | 944.60     | 65.79      | 1.842                        |
|       | Best SA      | 1        | 910.72     | 69.88      | 5.932                        |
| 600ms | <b>Model</b> | <b>K</b> | <b>RSS</b> | <b>AIC</b> | <b><math>\Delta_i</math></b> |
|       | RD           | 0        | 931.37     | 85.91      | 0.000                        |
|       | Features     | 0        | 944.60     | 86.41      | 0.496                        |
|       | Best SA      | 1        | 910.72     | 87.12      | 1.211                        |

At longer deadlines (750ms & 2000ms), this free parameter does not yield superior fits of the data. Like the model comparisons between RD and features shown above, the data favor the RD account at longer deadlines. Overall, the MIP/MOP model shows poorer fits at long deadlines when compared to shorter deadlines.

### *Discussion*

In general, Experiment 4 once again provides some support for the transformational approach in an implicit, speeded task. For short deadlines, no model showed a relationship with RT, statistically or otherwise. However, at 2000ms and, to a lesser extent, 750ms, the data favor a transformational approach when compared to both a feature-matching model and an SA model that can differentially weight MIPs and MOPs. For the error data, performance could be compared across all conditions and, as expected, the relationship between models over time is more complex; RD, feature matches and the SA model achieved significant fits of the data across the time course - making errors the more valuable measure in this experiment. At 500ms, a crude feature-matching model provides the most accurate fit of the error data once the SA model's free parameter is taken into account. At 600ms, all models perform almost equivalently, with RD being the slightly preferred model once the SA model's additional complexity is considered. Overall, the data at longer deadlines (i.e., 750ms & 2000ms) favor RD and thus resemble the order of model performance observed in Experiment 3.

The SA model provides the strongest data fits for both the 500ms and 600ms deadlines when MOPs are given slight precedence in the model (i.e., 70% weight). In addition, the increasing salience of MIPs over time, as indicated by the rightward shift of the optimal fit in Figure 32, is compatible with the findings of Goldstone and



Medin (1994), despite the much shorter response deadlines used here. At longer deadlines (750ms & 2000ms), no weighted version of SA can account for as much as the data as the transformational account.

Crucially, the performance of the feature-based model at short deadlines may be indicative of the sorts of representations being exploited at this stage in time. Crucially, these tests, by comparing models that assume different underlying object representations, probe how representations, more generally, build over time for these objects. Thus, given that RD *and* the feature model provide only an end-state prediction in this context, these results can help in determining what specific stimulus properties are being encoded in this task and, more importantly, **when**.

### General Discussion

Both experiments support the notion that transformational relationships can predict the perceived similarity between object representations in an implicit task. In Experiment 3 the predictions of RD, as determined by a three-operation coding scheme, provided (to varying degrees) the best fits of both RT and error data when compared to a basic feature matching model and an alignment model in a sequential matching task. In Experiment 4, the same models were examined over the time course using a simultaneous matching task with four different response deadlines. Unlike the order of performance observed in Experiment 3, a basic feature-matching model provided a more 'likely' fit of the data than RD at 500ms. Also, at 600ms, there was no compelling evidence for selecting one model over another, with all models yielding comparable fits of the data. However, at longer deadlines, RD fitted more of the data than any other model. The superiority of the more basic feature-based models at short response deadlines, coupled with RD's superiority at long deadlines, suggests

that the encoding and accumulation of more complex object properties takes time. Before discussing how these time course changes may relate to the representation schemes of the models compared, the general implications of these findings for the study of RD will be discussed.

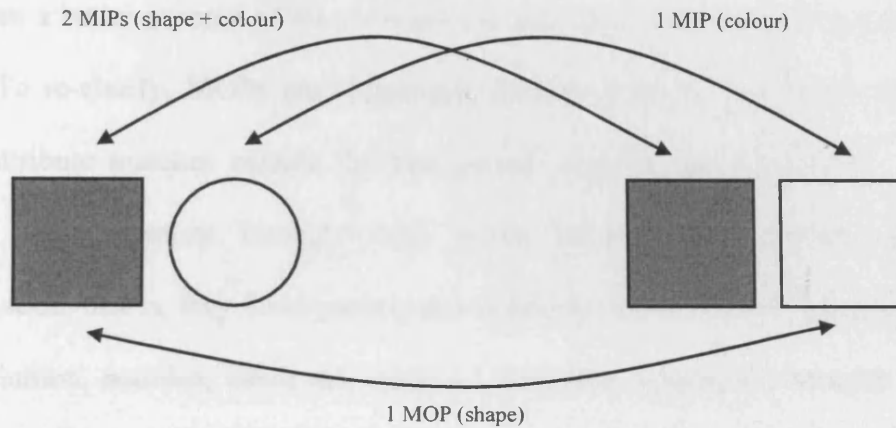
Significantly, this study is the first evidence for RD using an implicit measure of similarity. While both explicit and implicit measures have their benefits, similarity, in our everyday experience, is often derived very quickly without a conscious reference to the comparison taking place or to the properties that are implicated in that comparison (note, direct measures are not necessarily poor measures of perceived similarity - they merely represent one out of a number of possible similarity measures; for comparisons of both implicit and explicit measures, see Desmarais & Dixon, 2005). As transformations can provide accurate fits in this context, it supports the idea that transformations could be important in other areas of behaviour where implicit judgements of similarity are carried out, such as in spontaneous categorisation (see Chapter 5). As well as supporting RD generally, these results also support further the intuitiveness of the specific coding scheme used. As the coding scheme assumes the representation of feature locations and whole objects (through feature binding), these results provide further support for the representation of such information in implicit contexts (see Schoenfeld et al., 2003). In terms of similarity theories in general, these experiments also provide more insight into the role of features and alignment processes in this domain and how these may relate to the sorts of transformations that are relevant/detected.

Additionally, the reasonable fits found for the transformational account in this implicit task are compatible with the perception literature that has long assumed a role for transformations and structure in the processing of visual stimuli, namely object

recognition and apparent motion (Biederman, 1986; Graf, 2006; Shepard & Judd, 1976). More specifically, a number of behavioural studies that have employed speeded object-naming tasks have found a monotonic increase in RT with increasing transformation distance between two sequentially presented object viewpoints. This pattern has been shown for both spatial transformations, such as rotation and translation (Bundesen & Larsen, 1975; Bundesen et al., 1981, 1983), and also shape-changing transformations, such as object morphs or ‘warps’ (Graf, 2002, 2006). Graf (2002) showed that RT and errors increased with transformation distance when participants had to indicate whether the viewed objects were the same basic-level category. Correspondingly, similarity ratings also decreased with transformation distance, matching the predictions of an RD account. A later study by Panis, Vangeneugden and Wagemans (2008) also reported the same patterns between RT and transformation distance using morph stimuli across 11 natural object categories.

Overall, this literature supports the results found here, albeit indirectly. For one, the fact that a *negative* trend is observed between transformation distance and RT here signifies that transformations affect behaviour differentially in this specific version of the same-different task. Studies of object recognition, in general, demand participants to indicate whether a two viewpoints are the same object or category or not – not an identical image. Therefore, the task requires participants to search for *commonalities*, or any invariant properties, once the object has undergone some degree of transformation (i.e., spatial and/or topological transformation). When studying similarity, as opposed to recognition, participants respond *same* if and only if the compared images are identical in each and every respect. In this version of the task, then, it is object *differences* that are of interest because once a difference is identified, ‘sameness’ is falsified and the appropriate rejection response can be made

(Frost & Gati, 1989). Therefore, in the context of a same-different task, centred on strict identity, similarity has the opposite influence on RT and errors, regardless of how similarity is measured.



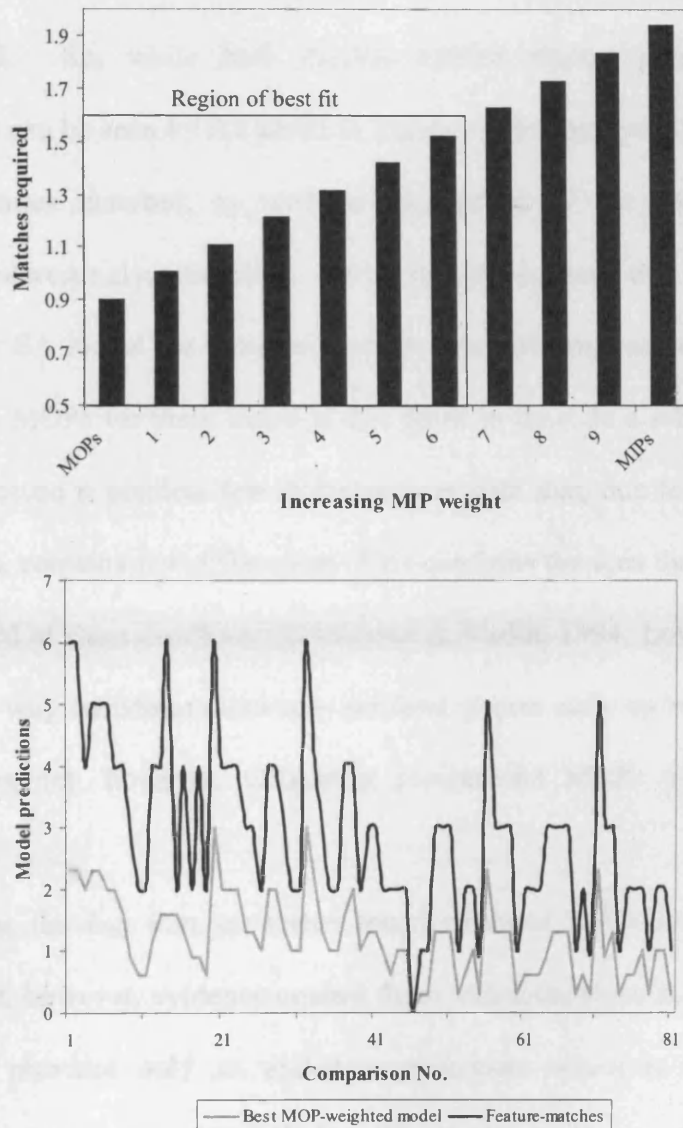
*Figure 33.* An example of a MOP in this domain. The left square in the pair 1 is placed into one-to-one correspondence with the left square in object 2 (MIP). Therefore, the leftover match between the left square in pair 1 and the right square in pair 2 is outside the best global mapping.

Experiment 4 provided the opportunity to assess the relative performance of each model over time. From these model comparisons it is possible to infer what stimulus information is represented at any one time; for example, the poorer performance of RD (in relation to other models) at 500ms within the error data suggests that all properties central to the coding language are not necessarily represented at this point in time, for example, the binding of features and the processing of structural information.

Firstly, however, it is important to address why a MOP-weighted model performed adequately at short deadlines. The data from 750ms onward are compatible with Goldstone and Medin (1994), who demonstrated that MIPs and MOPs have equal influence early on in time. These data also show that MOPs become less influential over time. However, a model which actually places *more* weight on MOPs provides a better account of the error data at very short deadlines when compared to RD. To re-clarify, MOPs are alignments between features that do not correspond (i.e., attribute matches outside the best global mapping, see Figure 33). Initially, MOPs may correlate strongly with errors because they provide misleading information, that is, they force participants to falsely respond ‘same’ because they are, by definition, *matches*, albeit non-optimal. Also, the competition between finding a feature match but not confirming object identity leads to a response slowing down - resulting in a timeout. The most important fact is that MOPs, unlike actual differences, are still *matches* and therefore invite response competition.

Furthermore, as Markman and Gentner (2005) state, “ensuring that the entire correspondence obeys the constraint of one-to-one mapping is a complex process” and is probably simply not possible at such short deadlines when the resources are not available to align many features, let alone form structure mappings. Therefore, the importance of MOPs early-on may reflect, quite simply, that fewer features have been accumulated at that point in time. For a MIP-based model to do well, all MIPs must be recognised, and given that there are, on average, two MIPs per comparison, participants will have to match at least four features for this model to perform well at short deadlines. Incidentally, there are actually fewer MOPs per comparison, thus fewer feature matches overall. In short, a model that predicts a greater influence of MOPs will just predict fewer matches to take place (see Figure 34, top panel) which

consequently will lead to better fits at short deadlines because only a few local matches are taking place at this point in time.



*Figure 34.* Graphs explaining the data fits of the MOP-weighted model early on. Top panel: the average number of matches required at each MIP/MOP weighting. Bottom panel: the difference between the MOP-weighted model and the feature model across all tested items (the x axis represents all 81 comparisons in the subset). This graph indicates that 1) the number of matches in total per comparison differs, and 2) that the spread in the predicted values is much lower for the MOP-weighted model.

The difference between the MOP-weighted model and the feature matching model indicates a related point: as can be seen in Figure 34 (bottom panel), the actual ‘spread’ of the best-fitting SA model predictions (grey line, range = 3) is much less than the featural model (black line, range = 6) - in addition to predicting fewer matches overall. So, while both models exhibit similar patterns across all comparisons (as can be seen by the peaks in Figure 34, bottom panel), the actual total number of features matched, as well as the spread of the predictions, varies considerably. Interestingly, therefore, these graphs suggest that the strong fits obtained by this SA model are incidental rather than reflecting something inherently important about MOPs for these items at this point in time. In a sense, the model is doing better because it predicts few differences in data that, due to severe response time restrictions, contains few differences. This confirms the idea that stimuli are not yet fully encoded at short deadlines (Goldstone & Medin, 1994; Lovett et al., 2009), thus explaining why transformations also perform poorer early on relative to feature models. It does not, however, ultimately recommend MOPs as a measure of similarity.

Likewise, the fact that the tested transformations fare worse at the shorter deadlines is not, however, evidence against these transformations in general. RD, as outlined here, provides only an end-state prediction, which is dependent on a particular mental representation of the objects in question. Therefore, RD’s performance necessarily depends on how representations build and change over time; mental representations are not fixed entities but are instead shaped by our perception, goals, knowledge, context and so on. Transformations should be relevant across the *whole* time course but as representations continuously change so will the relevant transformations. Quite simply, at any point in time, we compare what ever

representations happen to have built up – if these happen to be simple, the transformations will be simple also.

The posited transformation set relies on the representation of not only new and copied features, through the ‘create’ and ‘apply’ operations, but also on the representation of where these features are located within a pair – as in the ‘swap’ operation. Therefore, if short deadlines do not allow for the sufficient processing of ‘what’, and more importantly, ‘where’ information, then certain transformational relationships will not be recognised at all and non-structural, feature-based transformations will be all that is left, that is, just ‘create’ and ‘apply’-like transformations.

This can be inferred from the present data because the featural model outperforms RD early on (Figure 32). Comparing incomplete and unstructured representations will favour this model for a number of reasons; firstly, as feature models define objects as bundles of separate features, there will be no requirement for features to be bound into whole objects, unlike the coding scheme on certain comparisons. Secondly, both of these models match features regardless of spatial location, meaning that non-structured representations will pose no problems at all. Finally, as discussed above, the MOP-weighted model will be more accurate when stimuli are not fully encoded because it predicts both fewer matches to take place overall and fewer differences in the data. Therefore, if the object representations being exploited are simple and feature-based then why should the similarity model assume anything more complex?

This question does not challenge a transformational account because the adding, copying and deleting of features are, by definition, transformations, albeit non-structural, and the fact that RD still performs significantly at 500ms and 600ms



(in the error data) indicates that this coding scheme, predominantly made up of featural transformations, is still, to some extent, relevant at this stage.

So, the poorer fits at short deadlines are not evidence against transformations *per se* but solely against the representations required for the whole transformation set to be relevant - for example, structured representations. Hence, the performance of this transformation set, relative to, say, a feature-matching model, is important for understanding how mental representations build in general. Indeed, it seems that both the absolute improvement of RD over time, and the relative improvement of RD over feature-based models, confirms the intuition that structured and more complex representations emerge over time. Even the SA model used here, which does not provide the most accurate account of similarity in this task, is an analytic tool in this context for tracking the emergence of these more complex representations (see Figure 32). For example, the fact that the peak fit for the SA model shifts from being MOP-weighted at 500ms to being at the point where MIPs and MOPs are equally weighted at 2000ms (see Figure 32) suggests that more complex alignments are taking place later on. Likewise, the fact that the best SA model in Experiment 3 predicts a greater influence of one-to-one matches (i.e., is MIP-weighted) supports the general notion that similarity comparisons exploit more complex representations as more time is allowed to process objects.

In addition, if transformations that require structured representations (i.e., swaps) are relevant at all early on (see below) then this emphasises the need for similarity models that can tolerate a wider range of mental representations – particularly those that are structured.

This temporal change in mental representation (described above) has been discussed previously in relation to this task; Brockdorff and Lamberts (2000), for

example, describe representation building as an ‘information accumulation process’, whereby processing time is related to how detailed the object representation is in memory. From this process, the ‘richest’ representation can only be utilised when the essential ‘what’ and ‘where’ information has been accumulated.

Also, Markman and Gentner (2005) state that structure-based similarity comparisons are slower and more effortful than indiscriminative feature based comparisons and will emerge over a period of seconds not milliseconds. This is presumably based on the fact that the input representations, which are incomplete, are simply too limited to warrant any high-level structural alignment. Whilst this general intuition seems correct, as shown here and in other studies, the time course they assume for comparing structured representation may be far too long, at least for this domain. The end-state prediction of RD, that assumes a role for both structured representations and the binding of object features, still does well at explaining errors as early as 500ms after stimulus presentation; however, this alone does not necessitate the early emergence of structured representations given that the coding scheme, being largely comprised of featural transformations such as ‘create’ and ‘apply’, will naturally correlate with featural models to some extent at this point. Looking at these transformations in more detail, however, may help us address this.

As explained in this chapter earlier, the featural model makes some very counterintuitive predictions that are well explained by RD (if one assumes objects to be fully encoded). For example, the comparisons ‘AB/AB’ and ‘AB/BA’ have the same number of feature matches, despite two objects being identical in former and different in the latter. If you match spatial features to correct for this, such as ‘left A’ or ‘right A’, then the model then predicts no features in common at all, despite the fact that the same objects are compared in both comparisons. However, the swap

transformation can easily deal with this contrast but for a swap to give rise to this difference the relative location of features needs to be represented – if not, a featural model may do very well.

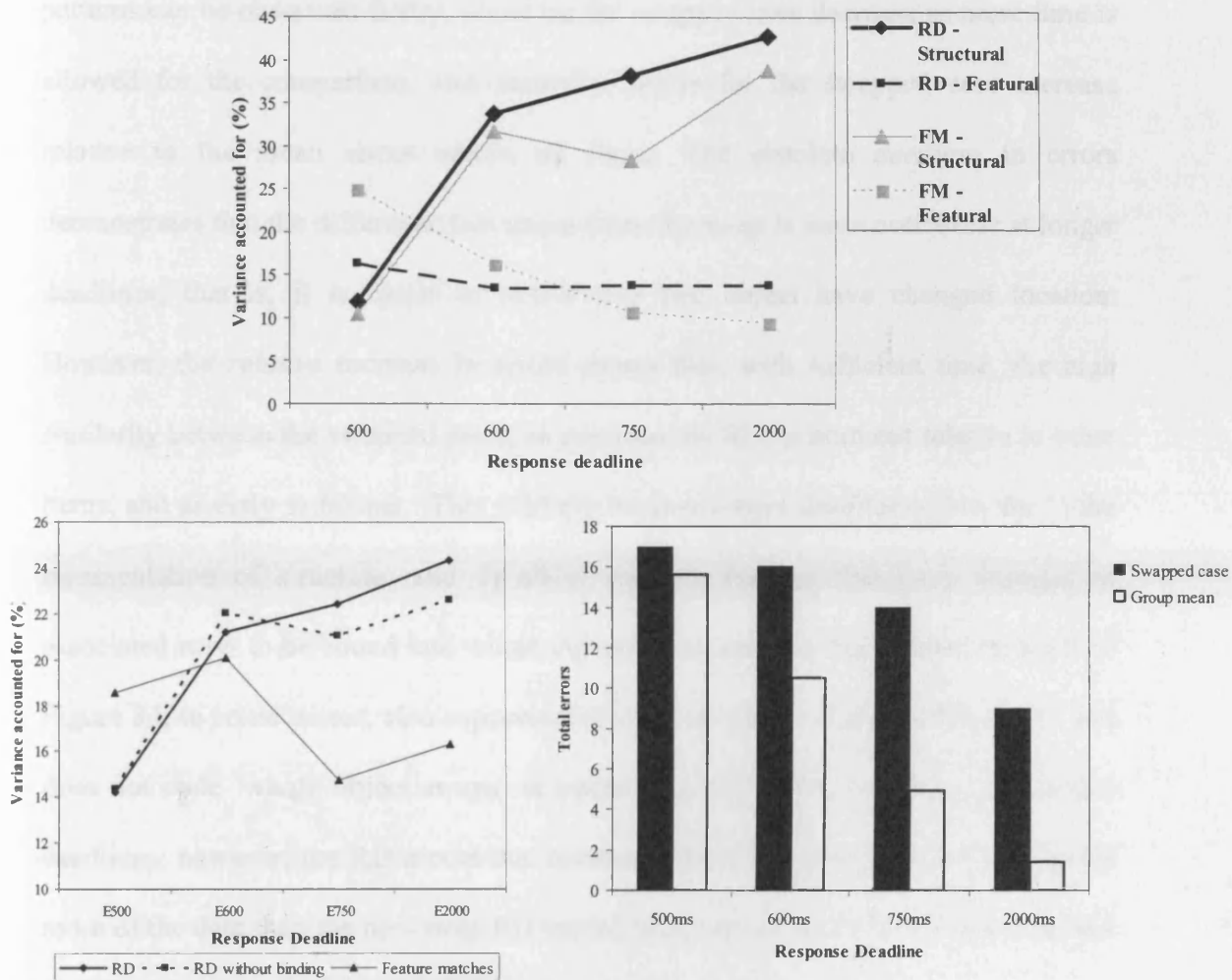


Figure 34. Graphs to that demonstrate the role of certain properties over time. Top panel: graph showing the difference between RD and the feature model at fitting comparisons with structural differences (AB/BA) and those with only featural differences (AB/AC). Bottom left panel: graph showing the difference between RD with and RD without the whole object swap transformation. Bottom right panel: graph showing the changing influence of the swap transformation over the time course.

The ‘whole object swap’ (AB/BA, colour; AB/BA, shape) also assumes, as well as the relative location of features, that features across both dimensions (shape and colour) are bound into a unitary percept. The graph in the bottom right panel of Figure 34 depicts the effect of the whole-object swap on errors over time. Two patterns can be observed: firstly, errors on the swapped item decrease as more time is allowed for the comparison, and secondly, errors for the swapped case increase relative to the mean errors across all items. The absolute *decrease* in errors demonstrates that the difference that arises from the swap is more noticeable at longer deadlines, that is, it is easier to notice that two objects have changed location. However, the *relative* increase in errors shows that, with sufficient time, the high similarity between the swapped pairs, as predicted by RD, is born out relative to other items, and as early as 600ms. This is likely because longer deadlines allow for 1) the representation of structure, and 2) allow separate features that have changed in associated ways to be bound into whole objects. The graph in the bottom left panel of Figure 34, to some extent, also supports this view; as can be seen, the RD model that does not code ‘whole object swaps’ is better than RD-proper at 600ms. At longer deadlines, however, the RD model that assumes relational bindings improves and fits more of the data than the non-swap RD model, thus supporting the claim that structure becomes more important over time in this task.

Research into the time course of feature binding supports these observations: task relevant features are selected between 120 and 180ms and the subsequent binding will take up to 300ms (Kenemans et al., 2002; Schoenfeld et al. 2003; Sharikadze et al., 2003). This time course seems compatible with the current findings as a differential sensitivity to whole object swaps is shown as early 600ms (see bottom right panel, Figure 34) – note, that the above time course needs to be applied to two

objects. To recognise that an object is the same as one previously seen in a different spatial position will firstly require the representation of a whole object – this will involve attending to the common association between colour and shape (i.e., that they both swap location) – which will then indicate the need for additional processing and facilitate the binding of these features (Markman & Gentner, 1995; Reeves, Fuller & Fine, 2005; Treisman, 1977, 1998; Treisman & Kanwisher, 1998).

Of course, it is not too surprising that binding occurs so early given the importance, and indeed naturalness, of representing objects as integrated wholes and not as bundles of separable features. A feature-based model that parses dimensions and features will necessarily fall short when object features are perceived to transform and act simultaneously as whole objects, even in implicit tasks.

Furthermore, the top panel of Figure 34 indicates a related point; the graph displays the data fit of both RD and the feature model for items with structural differences (i.e., left vs. right, such as AB/BC) and those with no structural differences (differences between features in situ, such as AB/AC). Noticeably, the feature model's performance at 500ms is being driven by its fit of these non-structured items. This suggests what has been stated previously, namely that this model is well suited to incomplete, unstructured stimuli – i.e., a feature model is quite good at predicting responses when stimulus information is still being accumulated. The most striking element of this graph, however, is that the feature model's fit of these items decreases substantially over time whereas RD shows little change. This suggests that crude feature matches alone are not enough for capturing featural differences once stimulus representations become richer and more complex – regardless of structure.

Importantly, the random subset of stimuli used here was not selected to

distinguish these accounts on the basis of structure alone - it was selected to represent this domain in its entirety. What is clear from these final observations is that RD is still accounting more data at long deadlines than the feature model, even without whole object transformations. What can not be ignored, therefore, is that these data also support the non-structural, feature-based transformations, such as 'create' and 'apply', particularly when richer, more 'complete' object representations have been formed (e.g., post 600ms).

### *Summary*

Overall, these experiments have provided the first support for a transformational account in a speeded, implicit task. Furthermore, the results of Experiment 4 show how the sorts of transformations that are relevant in a given context depend on what representations have been built. Although the stronger fits obtained by basic featural models early on suggest that representing structural information takes time, there is some evidence here (in the error data) to suggest that more complex information is relevant as early as 600ms when judging the similarity between objects.

## 4

---

# Transformation and Asymmetry

Asymmetry is arguably the most counter-intuitive phenomenon in the study of similarity and, as result, is somewhat diagnostic in distinguishing between different theoretical approaches. To recap, in the context of similarity, asymmetry refers to a difference in perceived similarity that arises from the direction of the comparison itself, that is,  $SIM(A,B) \neq SIM(B,A)$ . For example, the statement “North Korea is similar to China” is often preferred to the opposite (Tversky, 1977). As mentioned in earlier chapters, RD theory can predict asymmetries because the complexity or ease of a transformation can depend on the direction implied. Such differences in transformational complexity are predicted within the domain tested throughout and so, given the importance of asymmetries in distinguishing between different models of similarity, these asymmetries are examined in detail in the following chapter.

Indeed, asymmetries seem fairly robust; evidence of asymmetric similarity has been accrued across different stimuli (geometric shapes, rotated projections, countries,

narratives, self concepts & music), measures (confusability, ratings) and species (e.g., non-human primates; for previous evidence see Bartlett & Downing, 1988; Bowdle & Gentner, 1997; Catrambone, Beike & Niedenthal, 1996; Op de Beeck, Wagemans & Vogels, 2003; Tversky, 1977). Although evidence, on the whole, suggests that asymmetric similarities are psychologically real, this evidence is far from conclusive. As stated in Chapter 1, Gleitman et al., (1994) argued that the asymmetries observed by Tversky (1977), and necessarily others (e.g., Bowdle & Gentner, 1998), may not be related to similarity proper but to the syntactic structures in which these so called ‘symmetrical’ predicates appear. In support of this view, Gleitman et al. reproduced these asymmetries for 20 other symmetrical predicates, such as ‘equal’, ‘identical’ and ‘meets’. Fundamentally, this shows that the preference for a particular order of objects is independent of the predicate used, implying that similarity need not be reconceptualised as an asymmetric construct.

These findings may suggest that an asymmetric analysis of similarity may be unjustified based on some of the similarity experiments that have been employed so far. Gleitman et al.’s (1994) perspective strongly necessitates further research into asymmetry using non-linguistic, implicit similarity measures and also a clearer understanding of what it means to say that similarity *proper* is asymmetrical. If similarity is, in some cases, asymmetrical, then directional differences in perceived similarity should occur across a range of measures, not just within verbal tasks. In the current chapter, therefore, asymmetries will be tested using the implicit paradigm from the previous chapter - the same-different task. Furthermore, specific transformational predictions of asymmetry will be investigated, that is, comparisons where the transformation distance varies according to the direction of comparison.

Despite Gleitman et al.’s (1994) critique, asymmetries have been very



important in distinguishing between different accounts of similarity. Initially, similarity was considered to epitomise *symmetry*. For example, as described in Chapter 1, the spatial model (Shepard, 1957) embodies symmetry in its fundamental axioms (i.e., the distance between two objects in a coordinate space is the same regardless of direction). It was because of this fundamental assumption that asymmetries gained theoretical attention. To recap, Tversky (1977) specifically put forward asymmetries as evidence against spatial models, and in favour of his own featural approach, the Contrast Model. Tversky argued that judgements of similarity could not be removed from the actual statements which formed the basis of them. The statement “*A is like B*” is unique in that it has a directional component; it has a referent (or base) *B* and a subject (or target) *A*, and the allocation of objects to these respective roles is unlikely to be arbitrary (but see above). Crucially, this direction can be induced by directional similarity statements or by sequential presentation (see Chapter 3).

More specifically, Tversky (1977) noticed that participants preferred the direction where similarity was maximised, and this involved selecting the most salient or prototypical object as the base object, as opposed to the target, for example, “*North Korea is similar to Red China*” is preferred to “*North Korea is similar to Red China*”. To restate Chapter 1, Tversky’s contrast model,

$$\text{SIM}(A,B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad (3)$$

defines the similarity of representations *A* and *B* as a function of their shared features, minus those distinctive to *A*, minus again those specific to *B*. The parameters  $\alpha$ ,  $\beta$ , and  $\theta$  are weighting terms that depend on the task. In a non-directional judgement (“how

similar are  $A$  and  $B$ ?”) the distinctive feature sets of both objects are given equal weight (i.e.,  $a = \beta$ ). In this case, similarity will necessarily be symmetric. However, asymmetries will arise when objects are subject to a directional comparison, in which case the distinctive features of one object may be weighted more heavily than those of the other (i.e.,  $a > \beta$ ). This will give rise to asymmetries whenever the objects differ in salience, that is,  $f(A) \neq f(B)$ .

There is an interesting consequence of this assumption in the Contrast Model: if two objects differ in salience, then their ‘self-similarity’ will also necessarily differ, with the most salient object being more similar to itself than the least salient. In distance-based models of similarity, the similarity of an object to itself is the same, regardless of the object. In the Contrast Model, however, similarity, and with it self-similarity, has no inherent upper bound (see Chapter 1 for a discussion of this issue). In support, Tversky (1977) provides some evidence for differences in what might be construed as self-similarity in both ratings tasks (Gati & Tversky, 1982) and in the percentage of correct same responses on a same-different task (Rothkopf, 1957).

In response to Tversky’s (1977) critique of spatial models, Krumhansl (1978) put forward the distance-density model - an amended spatial model that can allow for asymmetries. Krumhansl provided a psychological interpretation of stimulus bias by assuming that the objects in denser regions of psychological space will possess greater weight. In the distance density model there is density parameter,  $d$ , associated with each object (e.g.,  $i$  and  $j$ ) and three weighting factors,  $\alpha$ ,  $\beta$  and  $\theta$ . More specifically, asymmetries will arise when both density and weighting factors differ, that is, when  $\delta(i) \neq \delta(j)$  and  $\alpha \neq \beta$ . Nosofsky (1991), then, redefined the spatial approach and argued that a spatial approach to similarity can account for asymmetries through a more general notion of a ‘differential stimulus bias’. Whereas Krumhansl provided one

example of stimulus bias, that is, density, Nosofsky's notion of bias refers to a number of possibilities that are both stimulus-based and response based, e.g., density, frequency and feature loss vs. feature gain. Crucially, differential biases are associated with individual stimuli and not with similarity per se. Nosofsky then went on to demonstrate that not only Krumhansl's (1978) model, but also the Contrast Model (at least in specific versions) is an example of the same general, stimulus bias framework. The Contrast Model, for example, does not account for asymmetry in the matching process itself but by differentially weighting model parameters.

What is common to all specific manifestations of the stimulus bias framework is that asymmetries arise as a consequence of the inherent properties of individual stimuli; they do not stem from the nature of the comparison process itself. This is in marked contrast to the transformational approach to similarity. As will be shown below, it accommodates asymmetries based on stimulus salience or complexity, but also allows asymmetries in cases where no differences in complexity or salience exist.

Hahn et al. (2009) were the first to exploit *directional* similarity judgements ('how similar is *A* to *B*?') in testing the transformational account. Transformational complexity can differ readily depending on direction: spilling water from a cup, for example, is easier than gathering the spilled water back in. Any such directional difference should give rise to attendant differences in perceived similarity, and hence asymmetric similarity between the two comparison points. Hahn et al. (2009) tested whether an inherent sense of direction could be artificially induced. To this end, they showed participants short animations of one familiar basic level object undergoing a shape-changing transformation into another. After viewing the animation, participants rated the similarity of objects drawn from the morph continuum. Directional similarity ratings for the exact same comparisons were higher when the referent object

(or base) had appeared first in the preceding animation; that is, ratings were higher when the direction of the similarity comparison corresponded with the direction of the preceding animation. In the second experiment, the start and end points of the morph sequence were visible on all trials to control for any primary or recency effects that may result in differences in endpoint salience. The same pattern in similarity ratings was shown even when these endpoints were available onscreen.

Given that the experimental manipulation involved only the direction of the preceding animation, it is hard to see how these results could be explained through differential salience or complexity of the two objects being compared. Instead, it seems that perception of the ease or naturalness of the transformation itself was being affected. In other words, it seems that the directional asymmetries that arose stemmed from the nature of the comparison process itself, not from intrinsic properties of the individual stimuli.

### **Measuring asymmetry**

Studies that have investigated asymmetric similarity directly have largely focused on using explicit measures of similarity such as preference judgements, similarity ratings or forced choice responses (e.g., Bowdle & Gentner, 1997; Deregowski & McGeorge, 1998; Tversky, 1977). It may be fruitful, given that similarities emerge rapidly in the categorisation and recognition of visual stimuli, to study asymmetries implicitly using a low-level similarity task. In addition, if asymmetry is part of similarity in general, then asymmetries should pervade through both verbal and non-verbal tasks, as long as the comparison is directional (see above). As discussed above, Gleitman et al., (1994) demonstrated asymmetries for many clearly symmetric predicates using verbal manipulations, such as ‘equal’ and

‘identical’. Crucially, this result suggests that asymmetries are related to the linguistic structures in which they appear, not necessarily to an underlying concept of asymmetric similarity. Therefore, reliable demonstrations of asymmetric similarity in implicit tasks, as well as verbal tasks, will indicate that asymmetric similarity in some contexts is more than just an artefact of syntactic structure.

To date, there is some evidence of asymmetric similarity using perceptual tasks. Rothkopf (1957) for example, found asymmetries based on the length of Morse code signals in a sequential same-different task. Specifically, there was systematic difference between the numbers of incorrect ‘same’ responses depending on whether code length went from ‘short’ to ‘long’ or the opposite. Tversky (1977), Krumhansl (1978) and Nosofsky (1991) later interpreted these findings as being compatible with their respective similarity models.

Op de Beeck et al. (2003) presented rhesus monkeys with a same-different task where stimuli varied according to their prototypicality. In line with Rosch (1975), there were fewer errors when more prototypical items were in the base position of the comparison. Unfortunately, Op de Beeck et al. did not choose to replicate this finding in humans using the same task, and instead opted for a similarity rating task.

### **Asymmetries in the transformational approach**

The transformational approach predicts asymmetries - i.e., directional differences in transformational complexity – within the stimulus domain tested throughout Chapters 2 and 3. In the current chapter these predictions will be investigated through a reanalysis of Experiment 3 and by testing *all* asymmetric predictions for this domain. Furthermore, this will be the first direct investigation of asymmetric similarity in humans using an implicit, low-level task. More specifically,

similarity will be inferred from patterns in reaction time and not from preferences ratings, similarity ratings, or forced choice responding. As the Contrast Model (i.e., a differential-salience account) provides the only testable prediction for these sorts of stimuli, the compatibility of these RD predictions with the Contrast Model is examined by looking at the relative salience of individuals pairs – i.e., the prediction that self-similarity will vary according to goodness or complexity.

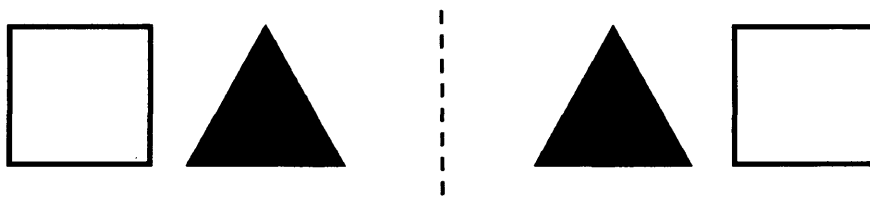


Figure 35. Stimuli from the two-dimensional stimulus set.

### Experiment 5

For Experiment 5, a reanalysis of the data reported in Experiment 3 (Chapter 3) is carried out. To recap, participants had to compare, in a same-different task, two sequentially presented pairs of shapes that varied along two dimensions (shape and colour), as illustrated in Figure 35.

In previous chapters, excellent quantitative fits were found between the transformational predictions derived from the coding language and perceived similarity for both a two-alternative forced-choice task and a direct ratings task. As the relevant test stimuli are a subset of those used in Chapters 2 and 3 these materials, and the associated transformations, will be briefly described again. On each of the two dimensions (i.e., colour and shape) there are 14 possible feature combinations. For example, the colour dimension's features in Figure 35 are combined *white, black*,

*black, white*. As colour and shape can be combined factorially, there are 196 possible comparisons in this domain. There are three general transformations, or operations, for comparing these items. These are applied to the base pair in order to modify it so as to generate the target pair. To reiterate, these operations are (in no specific order):

- 1) *Create feature* – taking the base pair we apply this operation to create a new feature that is unique to the target pair.
- 2) *Apply feature* – this operation takes an object or entity that is currently available (by being present in the base or by having been created via step (1)) and applies it to *one or both* of the objects in the target pair.
- 3) *Swap* – this swaps features between a pair of objects *or* swaps the object in its entirety (i.e. swap shape and colour features simultaneously; see Figure 9 for a possible sequence).

As stated previously, in a non-directional similarity comparison, that is “how similar are A and B?” this coding scheme takes the distance associated with the greatest complexity, the MAX-distance between the two pairs, as its overall *symmetrical* prediction. However, when a comparison is directional (i.e., when there is a base and target) the distance is simply the number of steps from whatever is portrayed as the base. Out of the 196 possible comparisons, 122 are asymmetrical; here, the transformation distance between the two pairs depends on the direction of the comparison, or, in other words, which object pair is the base of the comparison (from which the target is derived). The asymmetries themselves arise from a number

of principles embedded within the coding language outlined above; these principles are not arbitrary but are instead considered to account for how these representations are compared psychologically. For example, separating ‘create’ and ‘apply’ makes an important computational distinction between specifying a whole new and copying something that is already present in the base representation. In addition, applying two versions of the same object is as simple as applying one; this is to emphasise the fact the copying an object that is already available is ‘cheaper’ psychologically (in terms of computational complexity) than specifying a new feature in it entirety (for a detailed description see Chapter 2). An example of how this maps onto individual comparisons is depicted in Figure 36.

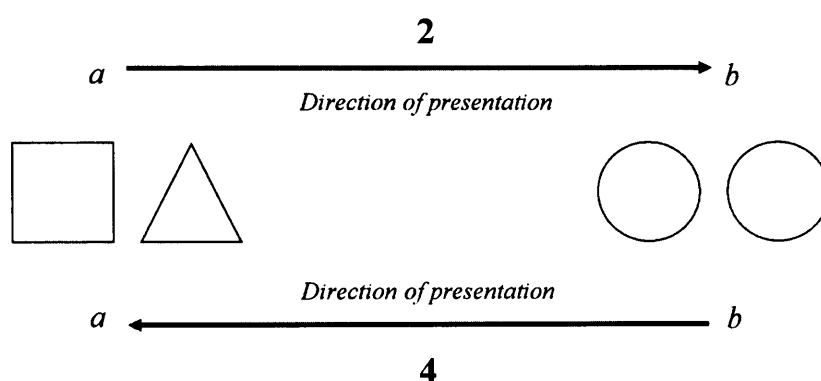


Figure 36. An example of an asymmetric relationship.

As can be observed, the transformation distance from left to right is two, whereas it is four in the opposite direction. In complexity terms, the former transformation is simpler and therefore associated with greater perceived similarity. The code is shorter, or simpler, left-to-right because ‘applying’ a feature to two shapes is as complex as applying it to one, that is:



$$\textit{create (circle)} + \textit{apply (circle(s))} = 2$$

By contrast, in the case of the longer code, both shapes must be created and applied separately;

$$\textit{create (square)} + \textit{create (triangle)} + \textit{apply (square)} + \textit{apply (triangle)} = 4$$

There are no concessions for applying two objects because the shapes differ and thus require more effort to specify.

As the differences in complexity give rise to differences in code length (that is, the specification of the instructions required to perform the transformation) these two directions will be labelled ‘long’ and ‘short’ in the following chapter, where stimuli of varying transformation distance are considered. From this, a clear prediction is made: there will be greater perceived similarity in the ‘short’ direction relative to ‘long’ for identical similarity comparisons.

It is worth noting that the transformational coding scheme introduced in Chapter 2 was not derived with asymmetries explicitly in mind, rather they are simply a consequence of the operations that were associated with specific comparisons a priori. Moreover, from simply observing these shapes, it seems difficult to assign differential salience to these pairs. Under the Contrast Model, the idea of salience is ill-defined and open to interpretation in many contexts, for example, the right hand object in Figure 36 could be argued to possess greater ‘goodness of form’ and therefore be more salient. Alternatively, the left object, by possessing a greater number of unique features could be argued to be more complex, and therefore more salient. In other words, it seems difficult, from the perspective of the Contrast Model,

to make *a priori* predictions about whether or not these stimuli should give rise to asymmetries, and, if yes, what the direction of the asymmetry should be.

To test experimentally these predictions a directional task was required. The direction of comparison can be manipulated in a number of ways; the most common method is using explicit or verbal similarity statements such as “how similar is A to B?” and comparing these to the reverse (see Hahn et al., 2009, Tversky, 1977). Less common, however, is the method of manipulating the temporal order of presentation so that the first and second object in the sequence respectively correspond to the base and target roles of the directional similarity comparison. In this instance, the stimulus currently in the visual field is compared to the memory representation of the first stimulus, making the first the ‘referent’, or ‘base’. Given that the second stimulus is currently perceived at the time of judgement, it is sensible to label this the ‘target’ or ‘subject’ of the comparison (i.e., “is the stimulus I am currently perceiving the same as what I just saw?”) rather than map this order directly onto a similarity statement (i.e., is stimulus 1 the same as stimulus 2?). This rationale, whilst justifiable, has never been explicitly stated in relation to this task. Tversky (1977), amongst others, have assumed the opposite, as this order has favoured, post hoc, the predictions of their stimulus bias accounts.

The sequential, perceptual-matching task provides an opportunity to manipulate the direction of comparison but also test these asymmetric predictions in a low-level, implicit task. In this task, participants are presented with two visual stimuli in sequence, and asked whether the second stimulus is the same (identical) or different to the first. This task provides a paradigm for measuring similarity implicitly, because, as stated in Chapter 3, response times on this task have been found to depend lawfully on the degree of similarity between compared stimuli. Specifically,

participants take longer to correctly identify two stimuli as different when they are more similar (e.g. Cohen & Nosofsky, 2000; Farell, 1985; Podgorny & Garner, 1979; Posner & Mitchell, 1967; see also Tomonaga & Matsuzawa, 1992, for matching-to-sample paradigm). Consequently, if the predictions of the transformational account are correct, participants should take longer to correctly respond ‘*different*’ when the transformation distance is short – as the differences that arise from the specific transformations will be harder to identify. When the distance, or code length, is long, differences will be rapidly perceived and a quicker ‘*different*’ response will be made.

In the previous chapter a basic perceptual-matching paradigm was used to test the predictions of the transformational approach for a subset of all possible comparisons (81 comparisons). In order to control for direction each comparison was presented in both directions and averaged for subsequent analysis. Here, however, each direction (i.e., short and long) can now be separated and compared to examine any possible asymmetries.

As the original subset was chosen at random it did not include all 122 asymmetric comparisons (48 comparisons). As a preliminary investigation, therefore, these 48 asymmetric comparisons were selected from the whole subset to see if the predicted directional differences in similarity were born out. Evidence for such patterns could then motivate a subsequent, more detailed exploration using all possible comparisons.

### *Method*

#### *Participants*

A total of 30 Cardiff University undergraduates completed the experiment (range = 18 to 25) and were all allocated course credit for their participation. The data

from two participants were later omitted due to a failure to follow basic task instructions.

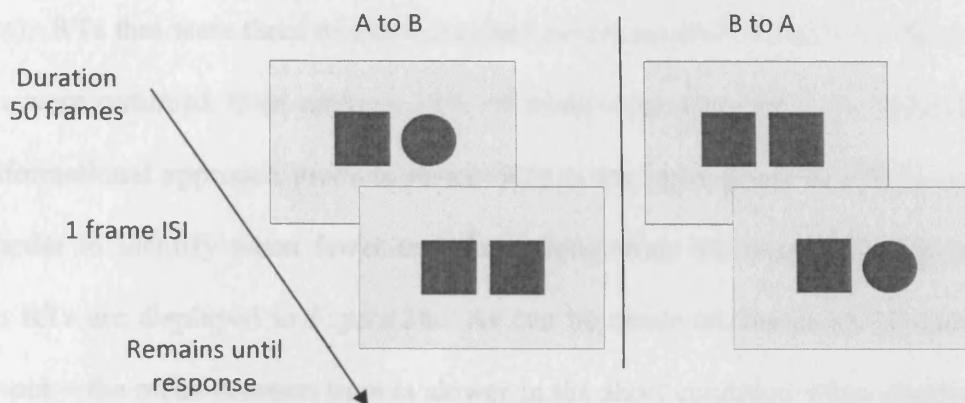


Figure 37. Testing asymmetries using a perceptual matching task. A to B, in this example, corresponds to a 'short' distance (as in Figure 36, above), whereas B to A is 'long'. instructions.

### Materials and procedure

For details regarding task parameters and specific stimulus information see Experiment 3 (Chapter 3) as this is the same experiment re-analysed. Out of the 81 tested items in Experiment 3, 48 stimuli contained a predicted asymmetry (59%). Crucially, participants were presented each comparison in both directions (see Figure 37) allowing both the symmetric (Experiment 3) and asymmetric predictions to be tested (see Figure 37).

### Results

Comparisons that contained the predicted asymmetry were selected for analysis. As each comparison was presented in two directions - *A to B* and *B to A* (see

Figure 37) - we could simply compare the reaction times for each of these directions by assigning the relevant direction to a particular distance. As there were insufficient errors in this task analysis was restricted to RT only (4.65% of all responses are errors). RTs that were three or more standard deviations above and below the overall mean were removed from analysis (2% of trials were removed). To reiterate, the transformational approach predicts slower RTs in the short group as differences will be harder to identify when fewer transformations relate the compared objects. The mean RTs are displayed in Figure 38. As can be observed, the predicted pattern is born out – the mean reaction time is slower in the short condition when compared to the long condition (607ms vs. 594ms). This difference is significant using a one-tailed within-subjects t-test ( $t(1, 36) = 2.5, p < .05$ ).

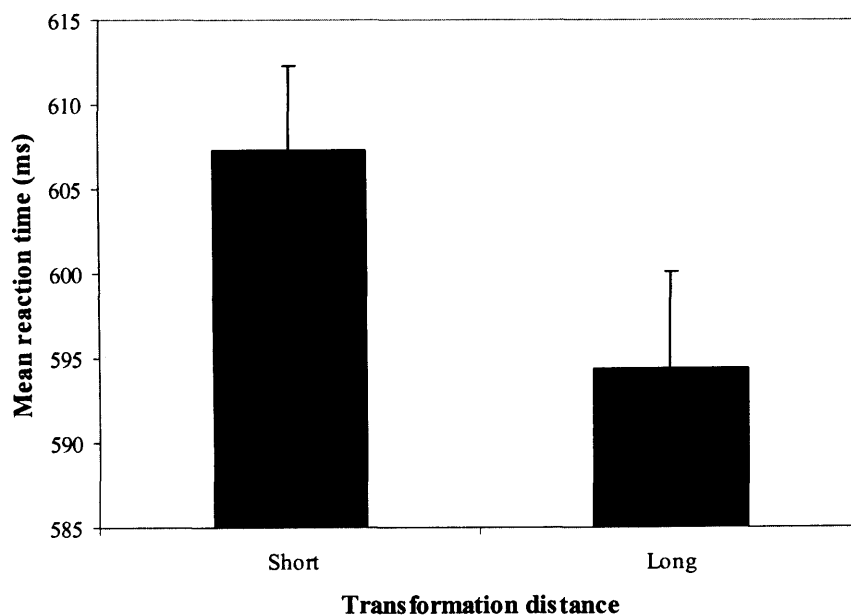


Figure 38. Graph depicting mean reaction time for each transformation distance (error bars = Standard Error).

### *Discussion*

The reanalysis of the Chapter 3 data supports the predictions of the transformation approach for this domain. Furthermore, these predicted asymmetries are born out in an implicit perceptual task, where similarities are not determined through subjective assessment, but inferred through response times when faced with rapidly presented visual stimuli. Crucially, the comparisons are identical in each condition - the only difference is the order of their presentation. Given this result, it would be useful to study these asymmetries more directly, by testing each asymmetric comparison for this stimulus domain.

### **Experiment 6**

In Experiment 6, the same stimulus domain was used. For this experiment, however, only those comparisons where an asymmetry was predicted were presented to participants (122 unique stimulus comparisons). The same stimulus and task parameters (ISI, stimulus duration) were employed. In addition to comparing directional differences in transformation distance, the predictions of the Contrast Model (Tversky, 1977) are also investigated by seeing whether the data are compatible with the notion of differential salience. Specifically, we test these predictions by looking at RTs for correct *same* trials, that is, 'self-similarity'.

### *Method*

#### *Participants*

A total of 39 psychology undergraduates participated. All were rewarded course credit for taking part. Participants were tested individually and were presented both experimental conditions.

### Materials

All task and stimulus parameters replicated those described in Experiment 5. The crucial difference was the number of asymmetric stimuli used. From the 196 comparisons available in this domain, 122 contain a directional difference in transformation complexity, or distance. These 122 comparisons constituted the different trials. As each comparison was presented in both directions, ‘short’ and ‘long’, there were a total of 244 different trials and 244 same trials across two blocks. As before, same trials were generated by taking the composite pairs (AA, AB and so on) and pairing them with themselves so to match the number of different trials. These 244 trials were divided into two blocks based on the default direction of the asymmetry (‘left to right’ and ‘right to left’), as determined by the original object codes used; so, AB/CC would be in group 1, as it is simpler left to right, whereas AA/BC would be in group 2, as it is simpler right to left. This resulted in 122 different trials in each block. Although this did not matter to the task, it made the task less exhaustive for participants.

To manipulate direction, participants were shown 61 comparisons ‘base to target’ and 61 comparisons ‘target to base’ in each block. Participants responded *same* or *different* by pressing the appropriate key (‘Z’ or ‘M’). The assignment of key to response depended on the handedness of participants – i.e., *different* was assigned to the dominant hand. Participants could respond at the onset of the second stimulus. No response deadline was imposed but participants were urged to respond as quickly as possible. The task took approximately 20 minutes to complete.

## Results

Initially, mean RTs on correct ‘different’ trials were compared for each of the tested directions (‘short’ and ‘long’). As in previous analyses, RTs more than three standard deviations above and below the overall mean were removed from analysis (2% of all trials were removed). Two participants were removed for low overall accuracy (<50% correct).

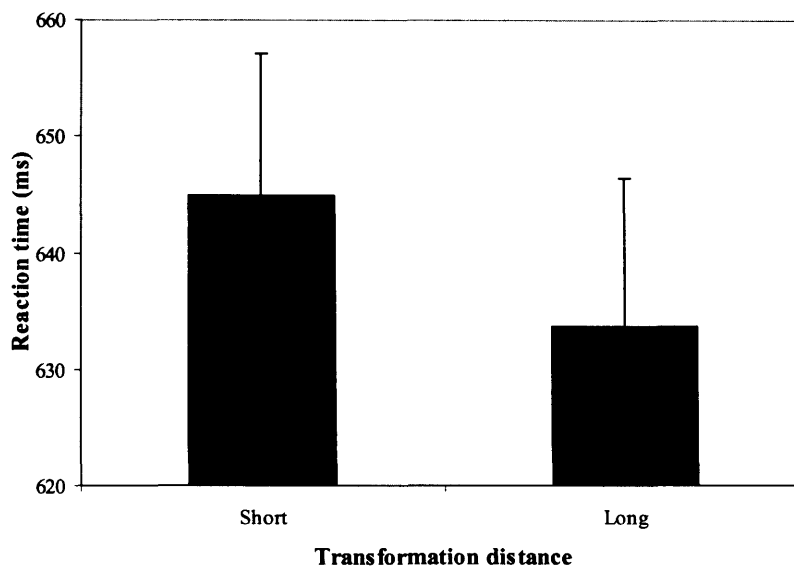


Figure 39. Graph depicting mean reaction time on correct different trials for each direction (error bars = Standard Error).

The graph in Figure 39 confirms this predicted pattern of results. The slower observed RT in the ‘short’ direction indicates greater perceived similarity (mean = 644.9ms; SE = 24.8). Correspondingly, the faster RT in direction of the long transformation indicates that these comparisons are less similar (mean = 633ms; SE = 26.8). A subsequent within-subjects t-test indicates a significant difference ( $t(1, 36) = 2.5, p < .05$ ). Moreover, 29 of the 37 participants showed the predicted difference in



RT. The specific objects compared in each case are identical; the only difference is the order of their presentation. Crucially, this order differentially affects the complexity of the transformation that manipulates the compared object representations.

### *Self-similarity and complexity*

As noted in the introduction for Experiment 5, it is unclear what the predictions of the Contrast Model are for these materials. However, it is possible to test, post hoc, whether the differences found are compatible with the model by investigating the ‘same’-trials. So, as stated above, the Contrast Model assumes that if object  $A$  is more salient than object  $B$ , that is, if  $f(A) > f(B)$ , then “ $B$  is like  $A$ ” will be preferred to the opposite. At the same time, if  $f(A) > f(B)$ , then the self-similarity of object  $A$  will be greater than that of object  $B$ , that is, the more salient, or more complex object, will always be more similar to itself than the less salient object.

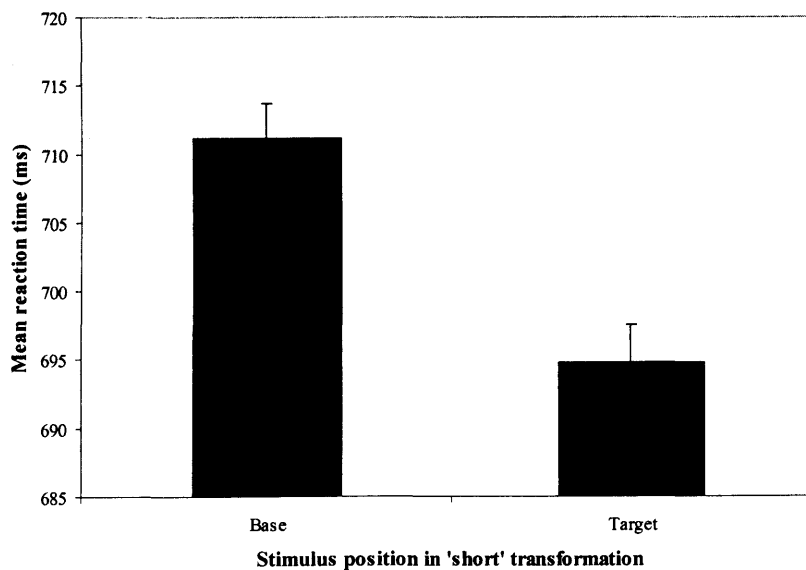


Figure 40. Graph depicting mean reaction time on correct same trials for each direction (error bars = Standard Error).

To assess ‘self-similarity’, RTs for correct same trials were compared as an index of self-similarity. If differential salience/complexity, as assumed by the Contrast Model, can explain these results, then RTs should be greater for those pairs that form the base of the short transformation, as they are more salient according to the model (i.e., pair *a* in Figure 36 will have a slower same trial RT when compared to *b*). The graph in Figure 40 shows mean same trial RTs for the base and target objects in the ‘short’ condition, that is, the condition with the shortest transformation distance. The graph shows that the base objects, on average, have slower same trial RTs when compared with the target objects in the ‘short’ condition. This relationship was also born out statistically; a within samples t-test yielded a significant difference ( $t(1, 113) = 4.5, p < 0.01$ ). The longer RTs on the same trials for the base pairs indicate that the result, while not clearly predicted by the Contrast Model, is at least compatible with the notion of differential salience.

### *Discussion*

The results of Experiment 6, once again, support the predictions of the transformational approach. Unlike Experiment 5, directional differences in perceived similarity are demonstrated for *all* asymmetric comparisons in this stimulus domain – a total of 122 comparisons. As before, these predictions are born out in a non-verbal similarity task. This result, by using a richer set of visual stimuli, gives even more strength to the claim that asymmetries are not limited to linguistic contexts and are part of similarity proper.

The observed differences in ‘self-similarity’ are compatible with Tversky’s (1977) notion that asymmetries are a result of differential salience. However, this finding, as is discussed below, is not at odds with a transformational account and is, in

fact, problematic for the Contrast model for two reasons: 1) the notion of salience is ambiguous, and 2) the direction of the asymmetry is incompatible with Tversky's original assertion about the roles of stimuli the sequential matching task.

### **General Discussion**

The results from these experiments support the idea that the unique predictions of a transformational approach can accurately predict asymmetric similarity between objects. This, in turn, supports not only the general idea that similarity can be conceptualised by transformational relationships, but also that the specific coding language, devised to reflect the representations of these objects, is making the right psychological predictions. These results extend on the work of Hahn et al. (2009), by showing transformation based asymmetries in a different domain, and with a non-verbal, implicit measure of similarity, as opposed to explicit ratings. Crucially, the asymmetries demonstrated here negate Gleitman et al.'s (1994) claim that asymmetries arise solely from linguistic structures; they can be predicted in online tasks that require rapid perceptual matching between visual stimuli.

In RD, asymmetries arise when one direction is simpler than the other in terms of transformational complexity, or code length (i.e., fewer instructions are required to transform the two objects). In the current experiment, asymmetric similarities are manifest in the longer RTs that exist in the direction of the 'short' transformation distance. Moreover, unlike the Contrast Model, the coding scheme used here makes unequivocal predictions both about the existence of asymmetries with these objects, and about their direction.

Nevertheless, in this experiment, the results are compatible with both the transformational account *and* the Contrast Model. Analysis of the same-trials showed

the preferred base objects (i.e., the base of the simplest transformation) to possess greater self-similarity. According to the Contrast Model, asymmetries arise when object A is more salient than object B; meaning the statement “B is like A” will be preferred over “A is like B”. From this, the Contrast Model predicts greater self-similarity for the salient (or preferred base) object, which is measured here by RT on ‘same’ trials. The slower responses for the base pairs of the ‘short’ transformation support the idea that they possess greater salience than the preferred target. Whilst this by no means negates a transformational explanation, it does not necessarily refute Tversky’s (1977) notion that asymmetries emerge from the differential salience of the compared objects.

Interestingly, however, Tversky’s (1977) predictions for this task suppose the opposite role of target and base (i.e., target first and base second). Not only does this order seem counterintuitive but it is also at odds with the self-similarity patterns presented here, that is, the slower RTs observed for the first object in the sequence. This relates to another issue: whilst differential salience can be inferred from ‘same’ trial response times it is not clear how this salience is defined in any given context, as noted in Chapter 1. For example, Tversky explicitly states that ‘goodness of form’ will be preferred over complexity, and this, in turn, will lead to increased salience. However, in this domain, the ‘good’ figures (where two objects are the same) tend to occupy the *target* location for the short transformation and are associated with faster RTs, not slower. Indeed, it may be that the *complex* object is more salient, but this is not stated clearly by Tversky.

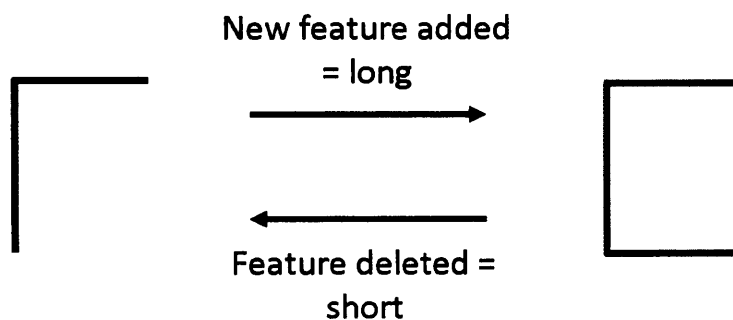
Furthermore, this notion of ‘goodness’ has been previously associated with transformations. In the case of this domain, a pair of matching shapes, that is, two squares or two circles, is ‘good’ insofar that they are preserved under a number of

transformations, such as the number of axes on which the object is mirror symmetrical (both the vertical and horizontal in this example; Palmer, 1983; Ullman, 1989). ‘Goodness’ also relates to object *simplicity* (Chater, 1996; Chater & Vitányi, 2003); the direction of the transformation that maximises similarity will move from complex to simple, not the other way round. This is because, in coding complexity terms, an object of two matching features is easier to specify than an object made up of two distinct features, as generating the second will be parasitic on the existence of the first, as in the coding scheme used here. In general, this indicates that the observed asymmetries are in fact completely compatible, unequivocally, with RD and the notion of coding complexity in general.

In this regard, the present results also complement those of Hahn et al. (2009), by demonstrating how the transformational framework applies to asymmetries that arise from differential object complexity. Although it is the complexity of the transformations relating the two objects, on this account, not the complexity of the objects themselves, there will typically be a systematic connection between the two. The fact that one of the ‘objects’ in Figure 36 contains two different shapes, whereas the other ‘object’ contains only one, has knock-on effects for the transformations that relate them; “applying” two different features costs more than applying the same feature twice. Critically, these predictions will come about naturally under a coding complexity account, without additional parameters. The same is true of other potential differences in complexity; for example, the Contrast Model also predicts that adding distinctive features to the base will increase the magnitude of the asymmetry. In this case, an attendant complexity difference in the associated transformation arises because deleting features requires a less complete specification of those features than

inserting features, leading to a shorter code overall (see also Hahn & Bailey, 2005, for evidence to this effect in the domain of word similarity).

The idea that asymmetries may arise from insertions being more ‘effortful’ than deletions has been observed previously, though not explicitly conceptualised in terms of coding complexity. Rothkopf (1957), for example, observed that going from short to long when matching Morse code signals of varying length, resulted in less confusions than the opposite. In addition, Garner and Haun (1978), using line/letter stimuli, showed that ‘many-features’ is more similar to ‘few-features’ when compared to the opposite – again, this result is suggestive of the notion that deleting is in some way simpler than inserting new object features (see Figure 41).



*Figure 41.* The type of stimuli used by Garner and Haun (1978). Principles of complexity, like those used in the transformational approach, are compatible with the direction of asymmetry found.

Interestingly, this domain contains a further possibility for testing not only the Contrast Model, but the entire notion of differential bias in spatial models of similarity. Nosofsky (1991) highlights a simple way in which the differential bias hypothesis could be falsified. Any additive similarity and bias model implies the

following transitivity condition: If  $p(i, j) = p(j, i)$ , and  $p(j, k) = p(k, j)$ , then  $p(i, k) = p(k, i)$ . Essentially, if an asymmetry exists for  $i$  and  $j$  then at least one asymmetry must exist for  $i$  and  $k$  or  $j$  and  $k$ . If one conceptualises these objects as possessing differential biases, outside the comparison, then one can see how asymmetries must be transitive in these triple scenarios. Identifying and demonstrating violations of this transitivity condition would be of enormous theoretical importance in terms of modelling asymmetries. As the coding scheme presented here allows for such ‘isolated asymmetries’, testing these potential ‘triplets’ should be a priority for future research.

Finally, the most general conceptual difference between the transformational account and both spatial and featural models is that the transformational account allows for asymmetries within a structural framework. As emphasised throughout this thesis, these traditional models assume very simple and specific representations. For the objects in the domain examined here, certain transformations, such as the swap transformation, implicitly suggest that the *left-of/right-of* relations between objects are represented and thus manipulated via the swap. Although these structure-based transformations are not exploited in all asymmetries tested here, the RD framework does allow for a level of complexity not permitted under any simple featural or spatial model.

Furthermore, there has been actually very little support for structural models of similarity predicting asymmetries in similarity judgements. The only previous study was conducted by Bowdle and Gentner (1997), who combined structure mapping theory with Grice’s (1975) pragmatic principle of informativity. Structure mapping theory states that the similarity between two objects is calculated by structurally aligning object representations (Gentner, 1983; 1989; Markman &

Gentner, 1993a). Generally, they argue that asymmetries will occur when the base is more systematic or ‘conceptually coherent’ than the target, as it then ‘lends’ structure accordingly. Like the transformational account, this model measures similarity within the frame of comparison, as object must be aligned to be compared. However, conceptualising the asymmetries found here (i.e., for geometric configurations) in terms of systematicity or informativity seems much less intuitive than it does for the sorts of domains typically used when testing structure mapping models (i.e., short narratives). Furthermore, this ‘base systematicity advantage’ requires that the compared objects are alignable in the first place, as “neither direction of the comparison will be informative if the representations are not alignable” (Bowdle & Gentner, 1998, pp. 249). Crucially, in this domain, there are asymmetries between objects that share no common features whatsoever, such as AB/CC, and thus neither object can lend structure to the other. In other words, models of structural alignment would not predict asymmetries for these items.

### *Summary*

In summary, these experiments have provided evidence that asymmetries in directional similarity comparisons can be accurately predicted by a difference in transformational complexity. Furthermore this accuracy is demonstrated in an implicit speeded task. Whilst these results are, to some extent, compatible with the Contrast Model, the transformational account embodies these asymmetries in the comparison process itself through the differing complexity of specific transformations, without any need for assuming the differential salience of individual objects. Hence, this domain recommends itself for further exploration of asymmetries as a diagnostic test for models of similarity.



# 5

---

## Transformations in Spontaneous Categorisation

As was emphasised in Chapter 1, the natural world contains an inconceivable amount of information that we, as perceivers, are required to process and ultimately simplify. Hence, the finite cognitive system needs ways of managing and simplifying this information effectively. Like similarity, our ability to spontaneously form categories is an example of this. Categorisation is the “process by which discriminably different things are classified into groups” (Nosofsky, 1986). Categorisation, or the classification of objects into ‘kinds’, has many broad and important functions: simplifying psychological inputs by placing similar objects together, generalising properties across category members, rapid object recognition, and so on. The question of what information or strategies we use to form categories is a matter of fierce debate in cognitive psychology. One method for categorising objects, and the focus of this chapter, is similarity.

To restate, similarity, considered the factotum to cognition (Larkey & Markman, 2005), is argued to hold many important roles across a range of cognitive phenomena. Out of these many roles, categorisation has been the most thoroughly studied in cognitive psychology. As a result, numerous studies and models of categorisation have regarded similarity to be central in classifying objects (Allen & Brooks, 1991; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1984; Posner & Keele, 1968; Pothos & Close, 2008; Reed, 1972). Despite assuming a central role for similarity, most models only utilise spatial models that, as discussed throughout, have undergone considerable criticism within the similarity literature. The purpose of this chapter is two-fold: 1) to examine whether transformation distance can accurately predict classification and 2) to compare RD with the model most often implemented in models of categorisation – the spatial model. Given its almost axiomatic role in cognitive models, it is important to see whether it, in fact, provides the best account of classification – even in a stimulus domain made up of simple pairs of geometric objects. Hence, model comparisons in this chapter are limited to the spatial model - not feature or alignment models. As has been discussed in previous chapters, RD can tolerate, in theory, the sorts of representations (e.g., structured representations) that cannot be embedded easily within a spatial representation (supposing that a transformation set can be postulated). In addition, as we will see later, spatial models embody other constraints that can limit its application in certain domains. Importantly, if the transformational approach is to be taken seriously as a basis for similarity computation then it must be tested as a basis for both similarity and similarity-based phenomena, such as categorisation behaviour. Also, given the potential generality of a transformational approach, such investigations can only contribute valuably to the study of similarity and categorisation in general.

Before outlining this experiment in more detail, the role of similarity in categorisation will be described.

### **Similarity's decline in categorisation**

Despite their intimate relationship, research into categorisation and similarity has rarely converged (for discussions of similarity's role in general see Close, Hahn, Hodgetts & Pothos, 2010; Goldstone, 1994a; Hahn & Chater, 1998). Whilst similarity is rarely questioned as a candidate process in categorisation, the role of similarity in general has faced considerable criticism within the similarity literature, as discussed in the first chapter. In order to place this experiment in context, certain theoretical points from Chapter 1 will be restated throughout this discussion.

If similarity is to provide an adequate explanation of categorisation then similarity must be a useful and meaningful construct in the first place. To recap, Goodman (1972) launched the first comprehensive attack on similarity's explanatory power by staunchly labelling it 'vacuous', an 'imposter'. His argument rested on the intuition that the similarity between two objects is determined by the number of properties that two objects have in common. As any two objects can share an arbitrary number of properties, Goodman argued that this explanation is useless; for example, a badger and a submarine share the property 'heavier than one pound', 'heavier than two pounds' and so on. If this holds true then all objects are somewhat similar to one another and similarity is a meaningless notion. Importantly, however, Goodman's criticism forced a number of researchers, particularly those advocating theory-based approaches, to reappraise the withstanding role of similarity in categorisation.

Murphy and Medin (1985), drawing on Goodman's critique, claimed that

similarity was too limited to explain conceptual structure. They argued that mere feature overlap was too constrained because categorisation can go beyond perceptual similarities and refer to the more complex, high level properties not included in the surface representations of objects. This led to the ‘knowledge’ or ‘theory-based approach’ to conceptual structure, which stated that categories were held together by our knowledge or theories about the world and objects (see Carey, 1985; Chi et al., 1981; Keil, 1989; Malt, 1990; Murphy & Medin, 1985; Rips, 1989). In support of this view, Rips (1989) demonstrated empirically an apparent dissociation between categorisation and similarity. In Rips’ experiment, participants were told that a bird had changed superficially into an insect (e.g. by toxic waste disaster). Crucially, participants were also told that the insect still had the ‘essence’ of a bird, that is, it could still mate with birds and produce bird offspring. Having received this information, participants would still categorise the transformed bird as a ‘bird’ despite it being perceptually similar to the insect. Likewise, Keil (1989) showed that the likelihood of children to categorise on the basis of these more complex biological properties or theories increases over childhood at the cost of perceptual similarity.

Although knowledge theories consider similarity-based categorisation to be too constrained, other evidence, as discussed in Chapter 1, has sought to expose the apparent over-flexibility of similarity. Goodman (1972) stated that similarity is useless until it is known in what *respects* two objects are similar – and once these are known it is these *respects* that govern the response. In essence, Goodman’s notion of respects is based on the fact that there is no definitive answer to the similarity between  $x$  and  $y$ , that is, perceived similarity has been shown to depend on a number of external factors, such as context (Barsalou, 1982, 1983; Close et al., 2009), expertise (Hardiman, et al., 1989; Suzuki et al., 1992) and the particular measure used

(Desmarais, et al., 1998; Podgorny & Garner, 1979; Sergent & Takane, 1987; but see also Chapters 2 and 3).

### **Evidence for similarity in categorisation**

Regardless of this criticism, similarity and similarity-based models of categorisation have continued to find substantial support. Similarity-based processes have gained considerable support in studies of *supervised categorisation* (SC). In SC, category structures are pre-defined by the experimenter and are then learnt by the participant through response feedback. Crucially, these pre-defined categories allow researchers to control what properties are exploited by participants when making category judgements.

In prototype models of categorisation, SC studies have shown that participants classify test items based on their similarity to prototypical representations, as determined by distance within a psychological space (Hampton, 1979, 1995a, 1995b; Posner & Keele, 1968; Reed, 1972).

Exemplar models of categorisation, that have used SC tasks, have also supported the role of similarity-based processes (Nosofsky, 1984, 1986, 1988). Some of the most compelling evidence for the role of similarity in categorisation comes from a number of studies demonstrating that 'exemplar similarity' affects categorisation, even when participants are given a perfectly predictive, explicit rule (Allen & Brooks, 1991; Regehr & Brooks, 1993; see also Juslin, Olsson & Olsson, 2003; Lacroix, Giguère, & Larochelle, 2005, 2003; Thibaut & Gelaes, 2006). In the seminal study by Allen and Brooks (1991) participants were required to classify a set of artificial animal stimuli. Even though participants were given a simple rule with which to classify these stimuli, similarity effects were still observed, that is, the

similarity of novel transfer items to training items influenced classification. In addition, similarity was actually orthogonal to the explicit rule meaning that similarity actually decreased task accuracy and the speed of making individual responses. Fundamentally, participants were not aware of the effect of similarity on their behaviour thus supporting the notion that similarity has a mandatory effect on categorisation, at least in some contexts.

In general, similarity-based models of categorisation have shown a bias towards spatial models of similarity. In such models, the likelihood of classifying a particular object/exemplar depends on the metric distance (dissimilarity) in this coordinate space between the object and other category members, or the prototypical representation. As stated in Chapter 1, this approach to similarity has provided the basis for a number of successful and widely tested models of categorisation (GCM, Nosofsky, 1986; for variations see ALCOVE, Kruschke, 1992; Stewart & Brown, 2005). The GCM, which is arguably one of most successful mathematical models in Cognitive Psychology, uses a general distance metric and has successfully captured supervised and unsupervised classification data (see below) for both explicit and implicit measures of categorisation (Nosofsky, 1986; Nosofsky & Palmeri, 1997; Pothos & Bailey, 2009).

Fundamentally, studies and models of SC support the notion that similarity, in some contexts, can provide a compelling account of classification behaviour. More specifically, in relation to specific models of SC, similarity seems to be a very good predictor of exemplar categorisation, particularly for simple, artificial stimuli comprised of few continuous dimensions. Therefore, Murphy and Medin's (1985) critique does not apply generally – similarity influences categorisation and, in certain stimulus domains, specific models of similarity perform excellently.

In addition to SC, similarity is an important part of many models of *unsupervised categorisation* (USC). USC differs to SC in that participants do not receive feedback for their category judgements. Although the categories, in some cases, are not defined a priori, many studies of USC implement pre-defined category structures and stimuli that will *naturally* promote certain classifications. Generally, participants are instructed to sort a set of stimuli into categories that feel natural or intuitive to them. As there is less control over the categories that people form, this form of categorisation is often considered more representative of how categorisation operates implicitly.

In terms of similarity, many models assume that spontaneous sorts reflect a 'similarity bias' whereby within-group similarity is maximised and between-group similarity is minimised (Compton & Logan, 1993; Gureckis & Love, 2002; Handel & Imai, 1972; Love, Medin & Gureckis, 2004; Milton & Wills, 2004; Pothos & Chater, 2002). Given the lack of control over the sorts, and number, of categories that people form, studies of USC will often use very simple stimuli with few dimensions. As in SC, such stimuli are amenable to models of similarity, such as spatial models, that are well suited to specific object representations where stimuli differ on a few continuous dimensions. Again, whilst some have argued that categorisation is, in some contexts, 'deeper' than perceptual similarity, the evidence within USC does suggest that similarity, and specific models of similarity, can serve as an accurate predictor for spontaneous sorting behaviour in certain stimulus domains.

### **A (not so) new approach to understanding similarity**

To restate the central point of Chapter 1, the criticisms of Goodman (1972) and Murphy and Medin (1985), whilst justified, are not psychologically important: the

representation of boundless property sets does not apply to a finite cognitive system. Crucially, similarity-based categorisation must operate within the limits of a cognitive system that attempts to simplify environmental inputs. In this regard, it is no surprise that similarity theories have been very successful in modelling categorisation in specific contexts. In similarity's defence, it has been suggested that researchers select well defined, perceptual stimuli in order to avoid the issues of psychological representation highlighted here (Pothos & Chater, 2002); this does not mean, however, that more complex representational schemes are not available, that is, schemes that can tolerate more complex object representations. As spatial models assume relatively simple object representations (i.e., locations in a coordinate space), the simple and artificial stimuli often used will not pose serious problems for such models (or individual studies). As soon as representations become structured, knowledge-based or analogically rich, these basic similarity models are limited (see Chapter 2). On the whole, the claim that categorisation is based on more than perceptual similarity is justified, but only in reference to similarity models that can not scale up easily to more complex object representations.

The onus, therefore, is not on similarity, but on categorisation, and its paradigms, to integrate new models of similarity that can cope with more complex representational properties, namely relational information and background knowledge. In short, the spatial model is being essentially 'kept alive' by models such as the GCM. In order to further strengthen similarity's role in categorisation, studies of categorisation must seek new theories of similarity that can provide a more general framework for understanding how people form concepts, both low-level and high-level.

Whilst many of these fundamental criticisms of similarity have been addressed



theoretically by structural accounts of similarity, such as the transformational account, these accounts are yet to be tested directly as a basis for categorisation. According to RD, an object is likely to be classified as a particular type if the transformation distance between that type, exemplar or prototype is short. The longer the code is, that specifies a transformation between two objects, the less similar these objects will be judged to be and the less likely they will be categorised together. Indeed, the properties manipulated by a given transformation will depend on the context in which the objects are classified, such as the expertise of the classifier or the category level at which the categorisation takes place.

In the current chapter, using the shape stimuli of previous chapters, an experiment is conducted to investigate whether the transformational approach to similarity can account for the classification of objects in an unsupervised task. Furthermore the accuracy of a transformational approach is contrasted with a spatial model – which is still prominent in models of categorisation.

### **Experiment 7**

In the present study, participants took part in an unsupervised categorisation task. Instead of presenting all stimuli from the onset, each participant was given initially a category ‘seed’, that is, a single stimulus that would act as a category reference (see Regehr & Brooks, 1995). After studying this ‘seed’, the remaining items were presented and the whole array was available. Participants were told, indirectly, to use this ‘seed’ as a basis for their judgement and to choose items that they felt ‘belonged’ with it, that is, belonged to the same arbitrary category. Hence, the likelihood of classifying items with the category seed should be related to the degree of similarity between each item and the seed. This method was selected

because engaging participants in all possible pairwise comparisons between individual items would have resulted in a number of ‘repeats’, that is, equivalent comparisons with matching predictions. Forcing participants to compare the seed with all other tasks items ensured that each comparison was unique and could be easily mapped onto the predictions of the tested similarity models. Notably, this methodology differs to the typical version where the whole stimulus array is presented simultaneously (e.g., Pothos & Close, 2008).

The same class of stimuli from previous chapters is used again here in order to provide a cohesive exploration of the transformation set and the specific domain (which is far from exhausted for testing similarity models in general; see Chapter 6). As in Experiment 1, each object in the experiment is a pair of geometric objects that can vary in shape alone (see Figure 42). The predictions for each comparison - between ‘seed’ and ‘target’ - are derived from the three-operation coding scheme that has featured throughout the thesis, that is, 1) create, 2) apply and 3) swap (see Chapter 2, for more detail).

| Category seed | Task objects | Transformation distance | Object code |
|---------------|--------------|-------------------------|-------------|
| □△            | □△           | 0                       | AB/AB       |
| □△            | △□           | 1                       | AB/BA       |
| □△            | □□           | 2                       | AB/AA       |
| □△            | △△           | 2                       | AB/BB       |
| □△            | □○           | 2                       | AB/AC       |
| □△            | ○□           | 3                       | AB/CB       |
| □△            | △○           | 3                       | AB/BC       |
| □△            | ○□           | 3                       | AB/CA       |
| □△            | ○○           | 4                       | AB/CC       |

Figure 42. All relevant comparisons between the seed object and the task objects

The category seed was fixed across participants and each 'task item' was generated from this. As the seed is the only category reference available, participants must classify in reference to it, and not in reference to other task items. Therefore, transformation distances were derived between the category seed and each task item. Figure 42 shows the transformation distances born out of the coding scheme for these items. If these transformational relationships in any way govern categorisation, and not simply the rated similarities, then a strong relationship is expected between transformation distance and the probability that a given object is placed in the same category as the 'seed'.

### *Method*

#### *Participants*

A total of 30 Cardiff University students completed the experiment. Participants were allocated course credit for taking part. Participants were tested individually and were all given the same set of objects.

#### *Materials*

The stimuli used were pairs of shapes that could vary in composition according to a number of predetermined methods (see Figure 42). As used throughout, the combination of shape features for each comparison can be represented using letters, with each unique letter referring to a unique shape feature, for example the category seed is represented by the code AB, that is, square/triangle. As in earlier chapters there were three possible features (square, triangle and circle) which, in this context, could be combined in nine unique ways given the single seed or 'base' object AB (see Figure 42). The assignment of shapes within this abstract code was fixed to

square = A, triangle = B and circle = C. As there was a single category reference (the category seed), the only relevant comparisons for modelling the data were those between task items and the category seed, that is, not all pairwise comparisons are carried out for this specific task. Using the coding language transformation distances could be calculated for each of the nine comparisons between seed object and task object.

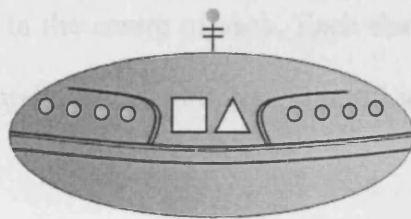


Figure 43. An example of the stimuli used.

In order to give these basic stimuli more impact as a basis for categorisation, a back-story was created that placed the shape pairs in a context that made them more engaging. The back-story was as follows:

*"A group of UFOs infiltrated Earth's atmosphere at approximately 2am. There is a danger that they will wreak havoc on mankind unless they are swiftly identified and stopped.*

*The biggest problem is that only some of these ships are hostile.*

*Your task is to help the armed forces by identifying which ships*

*you believe are hostile based on the evidence. The only*

*distinguishing marks are two symbols on the side-panel of each*

*craft. Based on these features alone, which other ships would you consider a possible danger? Rely on your intuition and select as naturally as possible. Good luck..."*

The pairs, to reflect the back-story, were presented on UFOs (see Figure 43). The back-story was presented on an A4 sheet. At the bottom of the sheet was a flap large enough to conceal the category seed. The spaceships were approximately 8cm x 5cm with the pair located in the centre of each. Each shape was 0.7cm x 0.7cm and was separated by a horizontal distance of 0.2cm. Stimuli were presented on laminated card that was 12cm x 6cm.

### *Procedure*

Participants were tested individually and were first instructed to read the back-story before inspecting the category seed. After reading the story they were asked to lift the flap at the bottom of the sheet to reveal the category seed (pair AB; see Figure 43). After inspecting the category seed they were handed the nine task items and asked to organise them as described on the instruction sheet (back-story). After deciding which spaceships they considered hostile or not, based on the shape information, they informed the experimenter that they had completed the task. The experimenter then took away the items and entered the data onto the coding sheet provided. Finally, participants were asked how they classified the items; responses were then recorded.

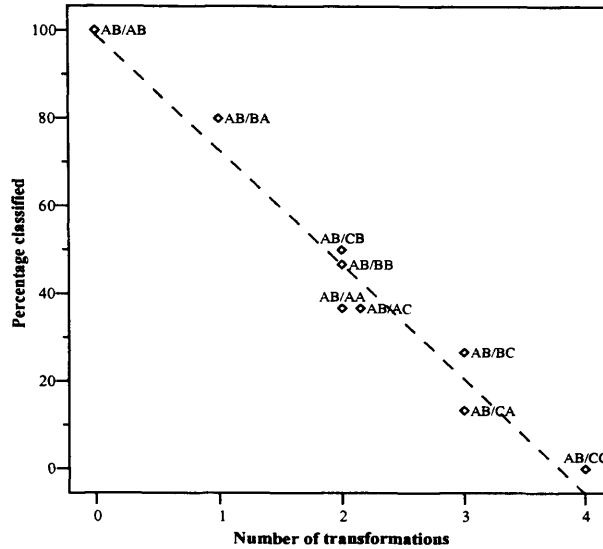


Figure 44. Graph depicting the relationship between transformation distance and classification behaviour. Note, jitter is added to data point AB/AC because of an overlap.

## Results

### Transformations

To test model accuracy, the percentage that each task item was classified with the seed was correlated with the model predictions. There are nine observations (i.e., nine task items compared with the seed) averaged across all participants. Predictions were those generated from the transformation set (see above). The graph in Figure 44 depicts the relationship between transformations and the percentage that each task item was classified with the seed.

As can be seen in the graph, there is a strong relationship between the number of transformations and probability that a given item was considered the same category as the seed overall. This relationship is also born out statistically ( $r = -0.98$ ,  $p < 0.01$ ; Pearson's  $r$ ). The predictions of the transformational approach, as determined by the

three-operation coding language, accounts for 95% of the variance indicating a very strong relationship. Also, the strength of this relationship closely matches the similarity ratings data in Chapter 2, where an almost identical set of items was used.

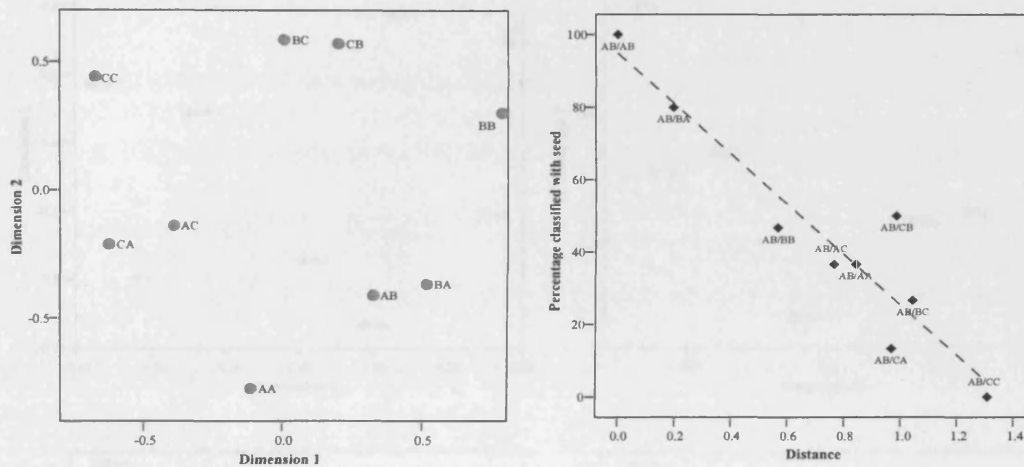


Figure 45. Left panel: The MDS representation of the tested object pairs. Right panel: The relationship between derived distances and the percentage that each object was considered the seed-object's category.

### *The spatial account*

In addition, this data was modelled using the spatial model, or MDS. A total of 10 participants completed a similarity rating task of all pairwise comparisons used in this experiment (i.e.,  $9 \times 9 = 81$ ). The MDS-generated representation is shown in Figure 45 and demonstrates what appears to be an annular configuration of the data. Although the proximities (derived from the spatial representation, left panel, Figure 45) show the expected non-positive relationship with similarity-based classification, the fit is slightly poorer than RD ( $r = -0.94$ ,  $p < 0.01$ ; Pearson's  $r$ ) with the variance

accounted for by this spatial solution at 88% (RD = 95%). A likelihood ratio for these data shows that the data are 51 times more likely to occur under RD making it quite clear that RD is the preferred model in this context.

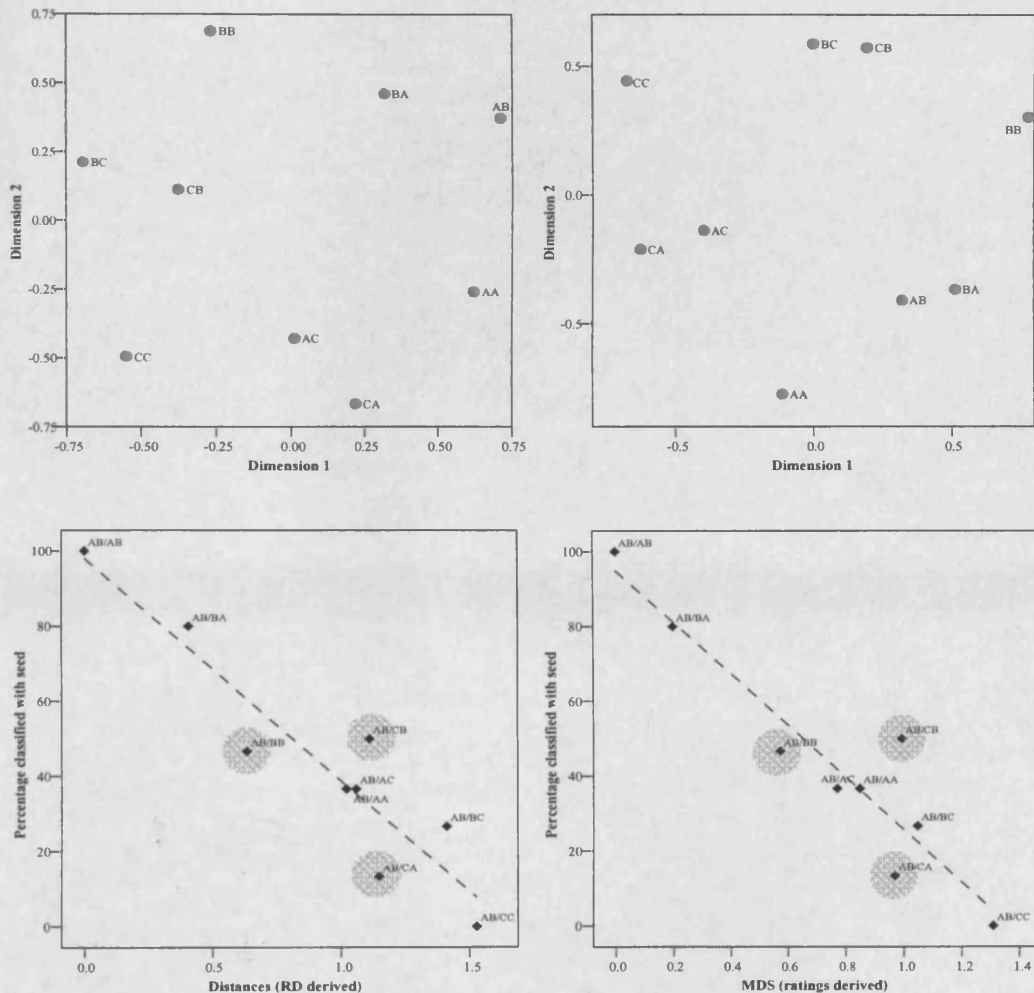


Figure 46. Top-left panel: The MDS representation of the RD predictions. Top-right panel: the MDS representation of the pairwise similarity ratings (as in Figure 45). Bottom-left panel: The relationship between MDS-RD and the participant categorisation data. Bottom-right panel: The original relationship between ratings derived MDS solution and the categorisation data.



The stress-1 value for this solution is 0.11. Although there are no definitive benchmarks for the acceptable level of stress in an MDS solution, this would largely be considered a fair degree of stress (Guttman, 1965; Kruskal, 1964). The difference in the fits reported here may suggest that RD and the spatial model are not entirely compatible for these items, that is, RD can allow similarity relationships that cannot be embedded easily into a two-dimensional spatial representation. In order to further probe this issue, an MDS solution for the RD prediction matrix was generated (MDS-RD). If RD makes predictions incompatible with a spatial model this should lead to a high stress solution when scaling these inputs.

The spatial solution for RDs predictions are shown in Figure 46 (top-left panel). The configuration is again annular and, under closer inspection, is a rotation invariant version of the original configuration shown in the top-right panel. The fit of these distances with the categorisation data is also shown in the bottom two panels of Figure 46.

Although statistically significant, the fit is less accurate than that obtained from RD's original predictions, implying that MDS distorts the RD-based similarities. This is indeed confirmed by the increased stress-1 value. The stress of the solution is 0.19, which is considered to indicate an unacceptable, or 'imprecise', level of data distortion (Guttman, 1965; Kruskal, 1964). Furthermore, there are common problematic residuals for both the RD-derived and ratings-derived solutions (with slight variation): AB/BB, AB/CA and AB/CB (highlighted in bottom panels, Figure 46). The original RD predictions fit these particular data/comparisons very well (see Figure 44) suggesting that these poor fits relate to the spatial model in general and not, for example, the sample-size from which the pairwise ratings were derived.

The graph in Figure 47 plots the correlation between the RD-derived MDS

solution (MDS-RD) and RD proper. As can be seen, the graph illustrates the degree to which the RD model predictions, which perform superiorly, are distorted by a spatial model ( $R^2 = 0.88$ ). The general difficulty of MDS to scale RD's predictions suggests that these predictions are in fact incompatible with a spatial solution.

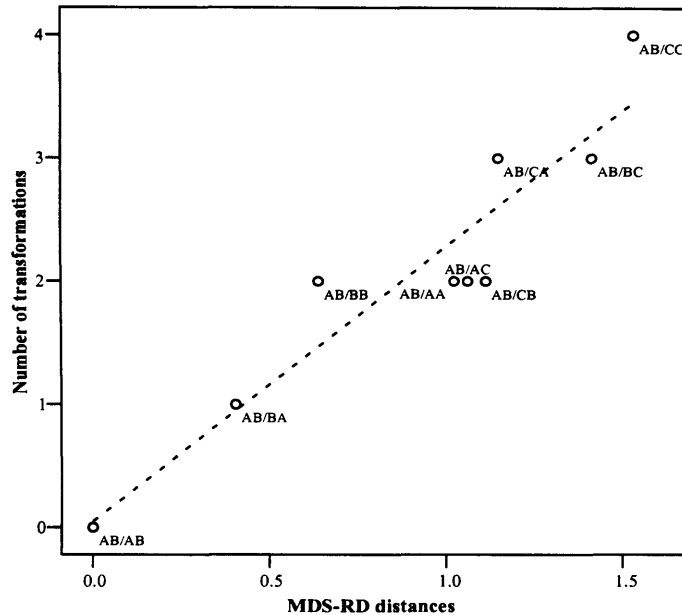


Figure 47. Graph depicting the relationship between MDS-RD and RD predictions proper.

### General discussion

This experiment provides the first support for RD as a basis for categorisation. Similarity, as determined by a combination of three concrete operations (create, apply and swap), is related strongly to the likelihood that a given object is considered the same category as a category seed. Arguably, this study provides some of the most compelling evidence for the notion that similarity is determined by the transformation distance between compared object representations. Moreover, the success of RD in

this task, relative to a spatial model, is an important step in reassessing the general role of spatial models in categorisation, particularly when the similarity relationships between objects/exemplars are more complex and thus incompatible with spatial representations. Besides, if RD is to be established a general theory of similarity, it must provide explanations in contexts where similarity is considered to be vital – that is, categorisation.

Given that similarity is not tested explicitly in this task, this experiment supports RD using an indirect measure (i.e., the probability of a certain classification). In addition, this implies that the operations within the coding scheme, and the object representations assumed to underlie them, are not only relevant in a typical rating tasks (Chapter 2, Experiment 1) but also in contexts where similarity comparisons are made spontaneously to infer category membership. Also, given that unsupervised categorisation is widely considered to reflect a more natural categorisation process, these findings support directly the idea that transformations are relevant in contexts outside basic similarity tasks and in areas where similarity is important in guiding intelligent behaviour.

To return briefly to the focus of Chapter 3, the supposed role of transformations in categorisation has been addressed, to some extent, in studies of object recognition that have shown that transformation distance is related to speed and ease of recognising that an object is a given category (Graf, 2002; but see also Panis et al., 2008). As the paradigms used are often speeded category-naming tasks, it does provide further evidence that transformational relationships are detected in a perceptual domain and are indeed born out when categorising objects spontaneously.

A number of other studies have also provided some support for a relationship between transformations and categorisation - although indirectly. In particular, several

researchers have emphasised the importance of transformational relationships (e.g., Zaki & Homa, 1999) in understanding the environment around us. Zaki and Homa (1999) argue that because objects, scenes and organisms transform in constrained and principled ways, any changes that do not obey these observed principles of growth will have profound effects on categorisation and conceptual knowledge. Specifically, transformational sequences help us bind disparate objects together making them both perceptually and psychologically more similar - an obvious example is the tadpole and the frog. In relation to this point, transformational similarity has actually been implicated previously as a basis for categorisation, although not in relation to an RD account; in the study by Rips (1989), outlined earlier, participants were presented with two stories about an animal that changed its essential properties to resemble an entirely different category of animal. Participants were told that this change occurred because of a) maturation, that is, a continuous transformation, or b) because of some catastrophe that resulted in a sudden change. In condition (a) the new animal was considered a member of the category but was also considered more similar to the original animal; the opposite was reported in condition (b). Although this experiment has been interpreted as separating similarity and categorisation, or at least supporting a notion of 'similarity chains' (Heit, 1992; Pothos, Hahn & Prat-Sala, 2009), it does, albeit indirectly, support the transformational approach to similarity.

Crucially, the current experiment highlights some of the limitations of the spatial model. While spatial representations are employed almost invariably in models of categorisation, the current finding does provide grounds for integrating new similarity approaches when studying categorisation. The criticisms that defined the so called 'decline of similarity' are relevant, but only in relation to specific theories of similarity that struggle with more complex representations and/or object relationships.

The most important aspect of this chapter is that RD can seemingly tolerate representations that spatial models find difficult (as indicated primarily by the stress values). Apart from the difference in model accuracy, the difficulty of MDS to find a spatial solution for RD's predictions is indicative of the limitations imposed by spatial representations in the first place. For example, it has been reported that spatial models have a bias towards annular configurations under certain conditions, such as for poor fits or when the true configuration is non-annular or incompatible (Goodhill, Simmen & Willshaw, 1995). As can be seen in the analysis, rotation-invariant annular representations are derived from MDS, indicative of this bias (see Figures 45 & 46). In addition, averaging data across participants, as is performed here, has been shown to hide violations of the metric axioms. Ashby, Maddox and Lee (1994) for example showed, for example, that averaging across participants eliminated violations of the triangle inequality axiom, a violation that was evident on a single subject basis. As averaging hides these violations, MDS should theoretically provide deceptively more accurate fits of the data when the data is averaged; the fact that it performs worse than RD here is a testament to RD in this domain and highlights, once again, the difficulty for the spatial model to match RD without succumbing to greater model complexity.

Tversky and Hutchinson (1986), in their discussion of featural and spatial models, highlight the spatial model's limitation on the number of 'nearest neighbours' that is, the number of items that can be most similar to another item. These 'neighbourhood' relations are problematic in contexts where numerous objects appear to share the same nearest neighbour, such as in super-ordinate categories (Tversky & Hutchinson, 1986). In RD, an object can have as many nearest neighbours as there are available transformations. Such constraints, therefore, will necessarily favour RD in this domain, as there are a number of items that transform equivalently into other

objects. The data in Figures 46 and 47 demonstrate a related point; as can be observed, a number of comparisons appear particularly problematic for a spatial account. It would appear that the neighbourhood relations in RD, that is, the distances between certain pairs, are different to those allowed in the spatial representation. For example, a number of task items are equally similar to the category seed under RD but are not permitted this equivalence on a spatial account as there is simply no way to embed these relations in a multidimensional space whilst still maintaining the similarity between the task items themselves.

Given these constraints it seems difficult to foresee how a spatial model could outperform RD in this domain, regardless of the participants used and the free parameters required to achieve it.

It is clearly not the case that MDS is a useless similarity model, simply limited in some fundamental respects. Even in the current experiment, the spatial model still provides an accurate fit of the data. What this study shows, however, is that it is not necessarily the best available similarity model, even for seemingly simple stimuli that one would consider ideal for a spatial model at first glance. Indeed, there are likely to be contexts where there is no accuracy gain for using a model like RD over a spatial model – for example, when objects vary on a few continuous dimensions. However, such a scenario would not present problems for RD because any change along a continuous dimension is still, by definition a transformation, such as those explored in the object recognition literature (e.g., Bundesen et al., 1981, 1983; Jolicouer, 1990; Lawson et al., 2003; Shepard & Cooper, 1982). It is when representations become more complex that spatial representations expose their limitations.

# 6

---

## Conclusions

Over the course of this thesis, I have provided evidence that similarity can be understood in terms of transformation distance or ‘Representational Distortion’. More specifically, I have demonstrated that transformational relationships are important in a range of contexts that involve judging the similarity between object representations.

Despite the shortcomings of featural, spatial and alignment approaches to similarity, it is still the case that these accounts have received considerable empirical attention in a variety of cognitive models that have drawn on the corresponding representation schemes. As a result, it will be important to probe further the potential role of transformations across many areas of cognition. This is no small task and is well beyond the scope of the current thesis. The experiments presented here do, however, make a significant start by taking a specific stimulus domain and 1) using both implicit and explicit measures, 2) directly comparing RD with a number of competitor models of similarity (both structural and traditional), and 3) using a range

of similarity-based phenomena (asymmetry and categorisation behaviour) to distinguish RD from rival accounts.

Crucially, a set of transformational predictions fared excellently against rival approaches to similarity in a domain that was originally put forward as evidence against an RD account. Apart from being a suitable domain for RD, the stimulus domain used, by containing both featural and relational information, provided a number of insights into the relationship between a number of similarity models – both structural and traditional.

Although many specific theoretical issues have been addressed, I consider the following to be central in understanding the role of transformations in similarity as a whole.

### **Comparing representation schemes within a domain**

As described in the opening chapter, similarity involves comparing the *mental representations* of objects. Therefore, when predicting the similarity between objects, careful consideration must be given to how objects might be represented in a particular context, or as Markman and Gentner (1997) state, “we must take representation seriously”.

Crucially, representations, unlike the objects themselves, are flexible entities that are affected by the context in which they are called upon, that is, the measure used, the goals of the perceiver, background knowledge and so on. Given this flexibility, the decision of what is psychologically relevant in the first place is a difficult one – particularly as we do not have direct access to mental representations. Therefore, depending on the domain under investigation, cognitive scientists must choose what features, relations, or indeed transformations, they consider relevant in a



given context. This, to some, may appear like too much predictive freedom. Indeed, this freedom depends to a large degree on what is (or can) be predicted by a given account. In the following section, I will reflect on the most pertinent issues concerning the representation schemes of rival approaches for the stimulus domain tested throughout this thesis.

The difficulty of deciding what properties are important or relevant in a given context is shared by all the similarity accounts tested here. As we have seen throughout, there are usually a number of ways to conceptualise object representations for a domain, and each interpretation, depending on the constraints of the model tested, will give rise to vastly different fits of the data; the challenge is deciding which, if any, are right psychologically. Crucially, all models must make a priori predictions about what representations, or properties are relevant, and these must generalise to the whole domain being tested.

The coding scheme tested throughout actually assumes very little about what information is represented explicitly; specifically it assumes only that features matter, that their spatial locations matter, and that features will become bound to each other to represent coherent objects. It is clear, from simply inspecting the objects, that only the presence/absence of features and their relative spatial locations are important for comparing pairs in this domain. If the coding scheme had to account additionally for featural degrees of similarity, that is, the similarity between a circle and a square, for example, then different and arguably more complex operations would be required. Even if this was the case, then RD predictions could still be derived as transformations could be proposed that relate individual features. This could be achieved by proposing a set of operations (e.g., insertion, creation and deletion) that could manipulate the presence or absence of specific geometric properties (i.e., lines,

corners) or, more appealingly, shape representations could be related by shape changing (morph) transformations, as have been proposed in the perception literature (Graf, 2002; Graf, Bundesen & Schneider, 2008).

When modelling similarity, selecting representations is, to a large extent, a matter of conjecture. The problem with this is that a potentially well-suited account may fare poorly if it is ascribed the wrong sorts of predictions – or indeed it may be unclear what a certain account predicts at all. To restate Chapter 2, Larkey and Markman (2005) assumed that each physically distinct set of objects in this domain was governed by a unique transformational relationship; that is, they equated physical and psychological relationships. However, this is not a natural way to conceptualise transformations nor is it, empirically, an appropriate way to capture these items. Although the location of features changes on each dimension for AB/BA (dimension one), AB/BA (dimension two), the simultaneous change of both features, gives rise to a completely unique mental operation – i.e., a whole object swap. This is in line with the fact that this domain has been used widely in the studying feature binding - a process that generates a ‘whole object’ identity (see Cheries et al., 2006).

The reduction of objects into their composite features is a typical and almost necessary aspect of feature-based models - alignment-based or otherwise – and can be problematic, as we have seen. For example, a MIP in this domain is a MIP regardless of where it appears within the target representation – the only constraint is that matches must be one-to-one (Goldstone, 1994b; Goldstone & Medin, 1994; Larkey & Markman, 2005). Indeed, it would have been reasonable to define MIPs as feature matches that occupy the same relative location in each pair (i.e., left square matches left square) and MOPs as matches between features in different locations (i.e., left square matches right square) but such an interpretation performs much worse for the

data in this thesis (see also Taylor & Hummel, 2009). Moreover, the one-to-one constraint, that is central to SA models (i.e., SME), is violated regularly regardless of the type of similarity measure employed. For example, the results of Chapters 2 and 3 indicate that MOPs are, to some extent, influential in this domain in both explicit and implicit tasks, suggesting that the strict constraints embodied in SA models are not necessarily appropriate, even when stimulus comparisons are “effortful” and carried out over many seconds (see Markman & Gentner, 2005).

Also, the sorts of matches that take place on featural accounts (including SA) mean that there is no gain for matching two features on the same *object*. This is unnatural because, as stated in Chapter 2, representing objects as integral wholes and not as lists of separable features is definitional of how we perceive ‘objects’ in the first place. The transformational approach can deal appropriately with this issue because transformations are not properties of objects themselves, they are relations *between* objects (Hahn et al., 2003), meaning that they can refer to objects differences that cannot easily be described under a purely featural, or indeed purely spatial, account. The whole object swap is, again, indicative of this point; even though object features are preserved there is nevertheless a transformation between the compared representations.

Essentially, these sorts of differences – where features are transformed in chunks rather than piece by piece – are perfectly suited to RD because transformations can manipulate objects without necessarily manipulating the number of features (but can also refer to the creation, deletion and insertion of object features). It is no surprise, therefore, that transformational relationships are often implicated in studies of object recognition where objects are perceived to change but leave the feature quantities invariant; transformational sequences, unlike crude feature matches, can

help ascertain that two images are in fact two slightly different views of the same object, even when all features are maintained (see Graf, 2002; Jolicoeur, 1990; Lawson, 1999; Tarr, 1995; Tarr & Pinker, 1990)

For traditional accounts of similarity specifically, a number of other issues have been highlighted in this thesis. In a featural model it must be determined what features are relevant, in addition to what constitutes a psychologically relevant feature in the first place. Although this may be straightforward for basic, well-defined object representations, this is less simple for complex representations. As discussed in Chapter 3, the matches counted by the basic featural model do not require features to occupy corresponding spatial locations; this is, of course, problematic for capturing the sorts of differences that occur when multiple features are transformed simultaneously. Crucially, this problem holds for the basic geometric objects used in this domain because a featural scheme that counts only spatial features (e.g., 'left A' matches 'left A') fares worse and is unable to capture the similarity between AB and the reverse, BA.

In addition, while the asymmetries predicted by RD in Chapter 4 are arguably compatible with the Contrast Model, it is not clear exactly how a difference in salience arises in the first place - and the prior work by Tversky (1977) makes questionable predictions regarding the assignment of base and target in the same-different task.

Similar problems exist for spatial models too; the results of Chapter 5 demonstrate how difficult it can be to embed certain object relationships within a spatial representation. Fundamentally, this experiment showed that spatial representations, albeit effective in many contexts, are ultimately too constrained for certain types of objects. Also, it seems that new approaches to similarity, such as RD

and, to some extent, maybe SA, can give new insights into categorisation – even for complex, knowledge-based domains (e.g., Rips, 1989; Zaki & Homa, 1999).

Importantly, it is not simply the case that the representation schemes of traditional similarity accounts are the best for capturing similarity in basic geometric stimulus domains, like the one used throughout this thesis. RD's predictions compare favourably with both featural and spatial models despite assuming very little about the representation of complex properties in this domain, such as structure – but a little, it seems, goes a long way.

### **Comparing similarity measures**

Going beyond direct ratings of Chapter 2 was important for investigating directly the role of transformations, and indeed structure, in speeded similarity judgements. By studying similarity within a single stimulus domain it was also possible to infer how certain measures – implicit or explicit - affected what representations were generated for the exact same physical objects.

The evidence presented in Chapter 2 suggests that the representations assumed by the coding scheme (see above) are relevant in direct similarity tasks. However, these direct measures, as mentioned in Chapter 3, do not necessarily tap into how similarity operates implicitly. Therefore, the results of Experiment 3 (and to some extent Experiment 4) are important for they indicate that the coding language, that fared well against ratings and forced choice data, is also relevant in an indirect, speeded task.

In Experiment 4, the relationship between transformations, mental representation and speeded task performance was investigated in more detail. Specifically, response deadlines were imposed as a means of assessing how model

performance relates to the time allowed for stimulus information to be processed and represented. Overall, the improvement of RD over time, relative to other models (featural model and SA model), indicates that representations build continuously over time as more stimulus information is processed and accumulated (see also Lamberts et al., 2003). More specifically, as structure sets apart RD from other accounts in this context, this improvement indicates that structured representations and feature binding happens gradually over time – it is not a case of structure ‘all the way’. Likewise, the relative fits of more basic similarity accounts, when tracked against RD, are important tools for measuring the emergence of more complex mental representations. Therefore, whilst the representations underlying such models are limited in general, for the reasons just described, they are, however, beneficial for our general understanding of similarity and representation.

What Experiment 4 supports is the contention that similarity tasks do not change similarity directly - they change the underlying object representations. This is an important point to consider when reconciling the systematic differences that exist between dependent measures that should intuitively address the same underlying construct (see Getty, Swets, Swets & Green, 1979; Goldstone, 1994a).

### **The Transformation ‘Framework’**

On the whole, it has been shown that RD’s predictions in this domain are problematic for featural, spatial and for particular alignment models. However, the sorts of properties that are considered central to these accounts (i.e., binary features and continuous dimensions) are still compatible with a transformational approach. As stated above, transformations need not manipulate structure to be tested; the ‘create’ and ‘apply’ operations proposed here manipulate only the presence of features. The

gain for RD, even here, is that certain principles of coding complexity can relate features in more complex ways, for example, applying two identical features is easier than applying two different features, or insertion is more complex than deletion (see Chapter 4).

Likewise, the sorts of properties that are considered amenable to spatial models can also be recast into a transformational framework. As Hahn et al. (2003) state, for example, changes along a continuous dimension can also be considered bona fide transformations, such as rotation, translation and so on. Therefore, RD can encompass these traditional accounts by providing a framework in which to relate and combine these different properties, such as spatial transformations along a continuous dimension (Bundesen et al., 1981, 1983; Jolicouer, 1990; Lawson et al., 2003; Shepard & Cooper, 1982; Shepard & Metzler, 1971), morphological/topological transformations (Graf, 2002, 2006; Hahn et al., 2009; Panis et al., 2008) and the insertion/deletion of new features (Garner & Haun, 1978; Rothkopf, 1957).

Importantly, however, RD can also allow complex transformational relationships that are outside the remit of these other approaches. In the case of object recognition, it has been shown that different degrees of plane rotation have a non-monotonic effect of recognition in an object naming task (Jolicouer, 1990; Lawson & Jolicouer, 1999). This is because there seems to be a preference for certain orientations, such as  $90^\circ$  and  $180^\circ$ , that preserve the spatial relations of the original  $0^\circ$  view. Whilst a continuous transformation may capture many intermediate representations, certain orientations could give rise to simpler, qualitatively unique transformations, such as reversal, or mirror image (in the case of  $180^\circ$ ). From this perspective, therefore, RD is not so much a new theory of similarity but a general framework of similarity in which many types of comparison, captured by a family of

similarity models, can be conceptualised simply by the distance of a transformational sequence that can vary both qualitatively and quantitatively. Theoretically, then, this can include the insertion of a square when comparing two basic configurations, or the similarity between a frog and a tadpole - the only requirement is that there must be some way, conceptually, to connect the two states by a pre-determined single, or set of transformations. As this thesis and other studies have shown (e.g., Hahn et al., 2009), RD benefits from that fact that certain principles of coding complexity can also be exploited to account for robust similarity-based phenomena, such as asymmetric similarity, and without the need to make additional assumption about the absolute complexity or differential stimulus bias. Indeed, these notions of complexity, salience and attention are embedded into the predictions in the first place, through the posited transformation set.

### **Future challenges and directions**

Despite the growing support for the transformational account, much more research is required in order to establish RD firmly as a general framework of similarity. Given its infancy, there are many possible directions for this future research. The directions that seem most beneficial to this end will be discussed in this final section.

Indeed, the domain tested within this thesis is by no means the 'litmus test' for RD and other models of similarity - to be truly useful as a cognitive scientific account of specific phenomena, RD must be tested thoroughly in a wide range of domains, particular those that can further distinguish RD from other, more constrained similarity models. However, these experiments do demonstrate the importance of exploring exhaustively a single set of materials, namely because the conclusions made



for one task do not necessarily hold for another (see Chapter 2 & 3, for example). Overall, it will certainly not harm future studies of similarity to adopt a similar approach for other stimulus domains.

Transformations, in their ability to relate arbitrary properties of varying complexity, makes them amenable to a wide range of stimulus domains, ranging from sentences, music, visual scenes, naturalistic stimuli, knowledge-based similarity and so on. The only requirement is that a set of plausible transformations and representations can be generated initially (Hahn et al., 2003). Naturally, then, it is not the case that a transformation set in one domain (i.e., the coding scheme used here) will generalise simply to others – as objects vary so will the transformations that relate them. This, of course, does not single out RD: the specific features that are matched on a featural account will change across domains, but they are nonetheless all *features*. Despite this, it would be interesting to see whether the coding scheme used here would be relevant for single objects rather than pairs, like those used by Hahn et al. (2003, Experiment 2). Further possibilities for these transformations and stimuli will also be discussed further below.

Throughout this investigation, transformation distance has been operationalised by the number of instructions required to transform one representation into another. It may be possible, as mentioned in Chapter 3 in particular, that transformations may bear different weights in certain contexts. Given that certain transformations appear to have a non-fixed influence over time (swap, Chapter 3) it may be interesting to investigate a weighted RD model in future research. For example, Cheries et al.'s (2003) study into the feature binding of rhesus macaques demonstrated that looking time recovered *more* when only one feature dimension was swapped, compared to the whole object - although this was not a significant

difference. As it is quite possible that whole object swaps are simpler than feature swaps (in particular if, from a code length perspective, objects are easier to individuate than features), it could be possible to reformulate the coding language so that swaps along one dimension are assigned a greater weight, that is, have a greater effect on similarity when they occur. However, a non-parameterised, simpler model is surely preferable unless the more complex model is really necessary - and in the context of the present studies, the simple approach works well and there seems little need to increase model complexity. Also, given that the end-state predictions of RD were still diagnostic in determining the emergence of structured representations it is not simply the case that developing process-like models is the best solution, on its own, for understanding the relationship between similarity and mental representation.

Also, if transformations do differ in individual complexity (Hahn et al., 2003) it should still hold that two operations result in a more complex transformation overall, that is, two is always more complex than one, and so on. Clearly, these issues would be interesting to explore further in future work but without further, independent motivation for a differential weighting, and ideally, some way of tying down those weights in advance, inclusion of differential weights would appear ad hoc and more in keeping with the 'leeway approaches' of featural and spatial accounts (see Chapter 1).

Furthermore, the issue of determining what transformations are psychologically relevant in a given context does bring about the question of how transformations are actually selected from the infinite pool of possible object relationships. This was highlighted by the philosopher Goldman (1986), who asks whether there is a finite set of transformations or whether the set of cognitively relevant transformations is potentially unbounded. Likewise, such a question has given rise to some researchers questioning the 'tractability' of the transformational

approach. The ‘Tractable Cognition Thesis’, or TCT (Van Rooij, 2008; Muller, van Rooij & Wareham, 2009), states that we must possess tractable algorithms for cognitive computations because our brain/mind is finite and is thus ultimately constrained. Muller, van Rooij & Wareham (2009) identify a number of parameters that could render certain instantiations of RD intractable. The following assumptions can, according to TCT, render RD tractable:

- 1) Similarity is computed on simple objects (short representations) and all intermediate transformations are equally simple.
- 2) Humans compute transformation distances only for objects that are somewhat similar.
- 3) The set of transformations in a given context is small.
- 4) The transformations that are relevant in a given context are non-inferentially provided.
- 5) The number of transformations in total, in any context, is “not too large”.

Whilst all of the above are important to consider, they do, however, provide no fundamental problems for RD. First of all, one must consider whether tractability is an issue for RD at all. Such considerations rely on what computations are necessary given the constraints imposed by external and internal factors, that is, context and our perceptual system. Assumption 1, for example, requires that ‘short’ representations be compared. Muller et al, argue that this is implausible psychologically, as we can compare many ‘objectively’ complex objects. Even for the complex ‘objects’ they mention - such as buildings, movies or faces - it is unlikely that our representations will contain all properties that make up those objects in any given comparison. It

does not follow, necessarily, that complex objects require complex representations; we need only those properties that will support efficiently what the representations are required to do (e.g., Goodman, 1972; see also Chater, 1996).

In addition, it is in a cognisor's best interest, given external constraints, to represent only a subset of an object's properties when making a similarity comparison. It is unlikely that this extraction requires much computation as some properties are given 'for free' by our perceptual system, or by previous experiences that have rendered certain properties useful or informative.

In relation to this, it has been suggested that our cognitive system, when presented with a complex input, will seek to find the simplest (or most likely) interpretation/organisation of that particular input (Chater, 1996). Essentially, to say that a movie is more complex than everyday visual inputs is missing the point; the cognitive system will try and build the simplest representation it can, given the computational constraints it possesses. Likewise, therefore, the notion of simplicity also applies to finding the simplest transformation between two object representations; for the sake of psychological plausibility, the cognitive system will select the simplest transformation, or set of transformations it can find in a particular context, even if this is not the simplest in reality and is selected from only a very small set of candidate transformations.

Also, transformation distances will only be computed for objects that are somewhat similar. According to Hahn et al. (2003), "for pairs of items for which no transformational relationship is discernable, RD predicts simply that these items are maximally dissimilar". Similarity, of course, serves a cognitive function, and thus there will be a point of dissimilarity where there is no cognitive-adaptive benefit of computing similarity at all; at this point, it would be best to delete one representation

and create the target representation from scratch. Importantly, however, this notion goes further than other models in that transformations can be specified in situations where there is no featural similarity to be exploited at all, as in the create operation for items such as AB/CC.

The requirement for sets of transformations to be small does not present any immediate challenges; in this investigation, and in previous studies, the set of possible transformations is quite small. Although many transformations could relate two objects objectively, many of them will not be psychologically plausible or perceptually salient. In the speeded tasks we report, it would not make sense to speculate about hypothetical transformations that may exist between the compared objects, as there is neither the time nor the resources to perceive or implement them. In addition, the transformations that are available in a certain context will necessarily be related to the object representations that are being compared. Goldman (1986), for example, suggested that transformations could be subject to a sort of 'preference ranking' that is activated when multiple transformational sequences are available for one comparison.

The results reported here also present new questions and new approaches. To this end, I will now describe the priorities for future research. As stated before, RD must be applied in many different contexts if it is to be established as an account of similarity. This investigation has underlined the sort of approach that is required in other stimulus domains: so, RD can accurately explain similarity ratings for a particular group of items, but do the proposed transformations scale up to more implicit measures of similarity? If not, what does this tell us about the sorts of representations being exploited in this task? Do the transformations seem an appropriate basis for categorisation? And so on. A number of domains seem

particularly well suited to a transformational approach. For example, auditory stimuli such as sounds, voices and music, seem particularly amenable to transformational relationships. Furthermore, studies of analogy have already begun to incorporate transformation distance. Leech, Mareschal and Cooper (2008), for example, provided evidence that the degree of transformation between two states (or the 'size' of a relation) is a factor in analogical problem solving - a finding that can not be easily accounted for by models of structural alignment. Further investigation into the role of transformations in analogy is important given that RD is yet to be thoroughly tested in high-level stimulus domains.

Chapter 4, on asymmetry, also outlined how stimulus bias accounts can be falsified by demonstrating the existence of 'isolated triples'. To reiterate, if there is a bias associated with object A, resulting in an asymmetric similarity with object B, then object C must be asymmetric with at least one of these objects. In other words, C cannot be symmetrically related to both A and B because they are not equally weighted. There exists, in the coding scheme tested throughout, comparisons that violate this assumption. If this could be shown empirically, in line with the predictions, then stimulus bias accounts could be falsified unequivocally.

Also, there are number of areas, and indeed dependent variables, in perceptual research that would seem well suited for further studying similarity and transformations. For example, studies that have demonstrated reliable asymmetries based on insertion vs. deletion in visual search may provide another promising avenue for studying transformations in an implicit paradigm (e.g., Treisman and Souther, 1985). Likewise, the strength of perceived apparent motion between two items could provide another measure of similarity in a perceptual domain. Interestingly, pilot demonstrations using the same stimulus domain tested here provided some evidence

for the perception of objects 'swapping' in motion for the comparison AB/BA. Formal demonstrations of this could provide further support for the plausibility of this scheme, as it suggests that our perceptual system is sensitive to these transformational relationships when forming motion correspondences.

Although similarity is a flexible construct, there are many examples that suggest that similarity has a mandatory effect on our behaviour - or possesses a 'context-independent core' (Goldstone, 1994a). Egeth (1966), for example, showed that participants were affected by similarities on an irrelevant task dimension using a same-different task – even though the stimuli were composed of separable properties such as shape and colour. Overall, this suggests that similarity is processed early on regardless of actual task demands. This has obvious parallels with the famous Stroop effect (Stroop, 1935) where participants are slow to name the colour of a written word when the word itself refers to a conflicting colour. Such an effect could be exploited empirically; for example a stroop-like task could be derived where the supposedly 'irrelevant' task dimensions are manipulated by set of pre-determined transformations. If the underlying transformational relationships on the irrelevant dimension affect response time and/or errors, this could provide evidence for the mandatory effect of both similarity and transformation.

Overall, whilst the experiments presented throughout this thesis support the notion that similarity is determined by transformational complexity, there is still much work to be done in terms of testing the unique predictions of RD in a wide range of stimulus domains – particularly those that involve more complex, structured representations.

Importantly, these experiments have highlighted how certain relationships that are easily captured by transformations (such as transforming assemblies of features)

are not necessarily compatible with other theories of similarity. Such findings, therefore, have the potential to enhance similarity's explanatory power in general; given that RD can tolerate more complex relationships between object representations it should be a priority to understand how closely related phenomena, once beyond traditional accounts of similarity, can be understood in terms of transformation distance – such areas, as discussed above, may include visual search, apparent motion and knowledge-based categorisation.

The main implication of this is that areas that once seemed distinct, or even at odds with similarity, could be unified under a general framework of transformational complexity. To this extent, RD is not so much a new 'theory' of similarity, but a convenient framework in which a wide range of arbitrary object relationships can be understood - as long as a relevant set of transformations can be specified.

Despite the truth that RD still needs to be tested thoroughly in a range of contexts, it seems that the role of transformations in understanding similarity, and indeed mental representation, is already a vital one.



---

# References

- Allen, S., & Brooks, L. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3-19.
- Anstis, S.M., & Mather, G. (1985). Effects of luminance and contrast on direction of ambiguous apparent motion. *Perception*, *14*, 167-179.
- Ashby, F. G. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 449-483). Hillsdale, NJ: Lawrence Erlbaum.
- Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*, 144-151.
- Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*(3), 339-362.
- Barsalou, L.W. (1982). Context-independent and context-dependent information in concepts. *Memory & Cognition*, *10*, 82-93.
- Barsalou, L.W. (1983). Ad hoc categories. *Memory & Cognition*, *11*, 211-227.

- Biederman, I. (1985). Human image understanding: Recent research and a theory. *Computer Vision, Graphics, and Image Processing*, 32, 29–73.
- Biederman, I. (1987). Recognition by components: A theory of human image understanding. *Psychological Review*, 94, 115-147.
- Bowdle, B. F., Gentner, D. (1997). Informativity and asymmetry in comparisons. *Cognitive Psychology*, 34 (3), 244-86.
- Brockdorff N., & Lamberts, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 77–102.
- Bundesen, C., & Larsen, A. (1975). Visual transformation of size. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 214-220.
- Bundesen, C., Larsen, A., & Farrell, J. E. (1981). Mental transformations of size and orientation. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 279-294). Hillsdale, NJ: Erlbaum.
- Bundesen, C., Larsen, A., & Farrell, J. E. (1983). Visual apparent movement: Transformations of size and orientation. *Perception*, 12, 549-558.
- Bushnell, E. W., & Roder, B. J. (1985). Recognition of color-form compounds by 4-month-old infants. *Infant Behavior and Development*, 8, 255–268.

- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: Bradford Books.
- Carroll, J. (1976). Spatial, non-spatial and hybrid models for scaling. *Psychometrika*, 41(4), 439-463.
- Catrambone, R., Bieke, D., and Niedenthal, P. (1996). Is the self-concept a habitual referent in judgements of similarity? *Psychological Science*, 7, pp. 158–163.
- Cave, K., Pinker, S., Giorgi, L., Thomas, C., Heller, L., Wolfe, J. & Lin, H. (1994). The Representation of Location in Visual Images. *Cognitive Psychology*, 26, 1-32.
- Chater, N. (1996). Reconciling simplicity and likelihood principles in perceptual organization. *Psychological Review*, 103, 566-581.
- Chater, N. & Brown, G. D. A. (in press). From Universal Laws of Cognition to Specific Cognitive Models. *Cognitive Science*.
- Chater, N., & Hahn, U. (1997). Representational distortion, similarity, and the universal law of generalization. *Proceedings of the Interdisciplinary Workshop on Similarity and Categorization, SimCat97* (pp. 31–36). Edinburgh: Department of Artificial Intelligence, University of Edinburgh.
- Chater, N. & Heyes, C. (1994). Animal concepts: Content and discontent. *Mind and Language*, 9, 209-246.

- Chater, N. & Vitányi, P. (2003). The generalized universal law of generalization. *Journal of Mathematical Psychology*, *47*, 346-369.
- Chater, N. & Vitányi, P. (2007). 'Ideal learning' of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, *51*, 135-163.
- Cherries, E. W., Newman, G. E., Santos, L. R., & Scholl, B. J. (2006). Units of visual individuation in Rhesus Macaques: Objects or unbound features? *Perception*, *35*(8), 1057 - 1071.
- Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- Close, J., Hahn, U., & Honey, R. C. (2009). Contextual modulation of stimulus generalization in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *35*(4), 509-15.
- Close, J., Hahn, U., Hodgetts, C.J., & Pothos, E.M. (2010). Rules and similarity in adult concept learning. In *The making of human concepts*, D. Mareschal, P.C. Quinn, & S.E.G. Lea (eds.). Oxford University Press.
- Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar retrieval model of speeded same-different judgments. *Journal of Experimental Psychology: Human Perception & Performance*, *26*(5), 1549-1569.

- Compton, B. J. & Logan, G. D. (1993). Evaluating a computational model of perceptual grouping by proximity. *Perception and Psychophysics*, 53, 403-421.
- Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: Psychophysics and a computational model. *Vision Research*, 38, pp. 2229–2257.
- Deregowski, J. B., & McGeorge, P. (1998). Perceived similarity of shapes is an asymmetrical relationship: a study of typical contours. *Perception*, 27(1), 35 – 46
- Desmarais, G., & Dixon, M.J. (2005). Understanding the structural determinants of object confusion in memory: An assessment of psychophysical approaches to estimating visual similarity. *Perception and Psychophysics*, 67, 980–996.
- Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, 9, 242-257.
- Egeth, H.E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1, 245-252.
- Estes, Z., & Hasson, U. (2004). The importance of being nonalignable: Structural alignment in the judgment of similarity and difference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 1082-1092.

- Falk, R., & Konold, C. (1997). Making sense of randomness. *Psychological Review*, *104*, 301-318.
- Falkenhainer, B., Forbus, K., and Gentner, D. (1989). The Structure-Mapping Engine: Algorithm and Examples. *Artificial Intelligence*, *41*, 1-63.
- Farrell, J. E. (1983). Visual transformations underlying apparent movement. *Perception and Psychophysics*, *1*, 85-92.
- Farrell, J.E. & Shepard, R.N.(1981) Shape orientation and apparent rotational motion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 477-486.
- Felfoldy, G. L., & Garner, W. R. (1971). The effects of speeded classification of implicit and explicit instructions regarding redundant dimensions. *Perception & Psychophysics*, *9*, 289-292.
- Fodor, J. and Z. W. Pylyshyn. (1988). 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, *28*, 3-71.
- French, R. (1995). *The Subtlety of Sameness*. MIT Press, Cambridge, MA.
- Frost, R., & Gati, I. (1989). Comparison of the geometric and the contrast models of similarity by presentation of visual stimuli to the left and the right visual fields. *Brain & Cognition*, *9*, 1-15.

- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, 25, 350-358.
- Garner, W. R. (1974). *The processing of information and structure*. New York: Wiley.
- Garner, W. R., & Haun, F. (1978). Letter identification as a function of type of perceptual limitation and type of attribute. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 199-209.
- Gati, I., & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology*, 16(3), 341-370.
- Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7 (2), 155-170.
- Gentner, D., 1989. The mechanisms of analogical transfer. In: Vosniadou, S., Ortony, A. (Eds.), *Similarity and Analogical Reasoning*. Cambridge University Press, Cambridge, UK, pp. 199-242.
- Gentner, D., & Markman, A. B. (1997). Structure mapping in analogy and similarity. *American Psychologist*, 52, 45-56.

- Getty, D.J., Swets, J.A., Swets, J.B., & Green, D.M. (1979). On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics*, 26, 1-19.
- Gibson E. J. (1963). Perceptual learning. *Annual Review of Psychology*, 14, 29-56.
- Gleitman, L.R., Gleitman, H, Miller, C & Ostrin, R. (1996) Similar and similar concepts. *Cognition*, 58, 321-376.
- Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple, intuitive statistic for empirical psychologists. *Psychonomic Bulletin and Review*, 11, 791-806.
- Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.
- Goldstone, R. L. (1994a). The role of similarity in categorization: Providing a groundwork. *Cognition*, 52, 125-157.
- Goldstone, R. L. (1994b). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 3-28.
- Goldstone, R. L. (1998). Perceptual Learning. *Annual Review of Psychology*, 49, 585-612.



- Goldstone, R. L., Medin, D. L., & Gentner, D. (1991). Relational similarity and the nonindependence of features in similarity judgments. *Cognitive Psychology*, 23, 222-264.
- Goodhill, G.J., Simmen, M., & Willshaw, D.J. (1995). An evaluation of the use of Multidimensional Scaling for understanding brain connectivity. *Philosophical Transactions of the Royal Society, Series B*, 348, 265-280.
- Goodman, M. (1972). *Problems and projects*. Indianapolis: Bobbs Merrill.
- Graf, M. (2002). *Form, Space and Object: Geometrical Transformations in Object Recognition and Categorization*. Berlin: Wissenschaftlicher Verlag Berlin).
- Graf, M. (2006). Coordinate Transformations in Object Recognition. *Psychological Bulletin*, 132, 920-945.
- Graf, M., Bundesen, C., & Schneider, W.X. (2008). Topological transformations and object shape. Manuscript submitted for publication.
- Gureckis, T.M., Love, B.C. (2002). Who says models can only do what you tell them? Unsupervised category learning data, fits, and predictions. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Lawrence Erlbaum: Hillsdale, NJ.

- Guttman, L. (1965). A faceted definition of intelligence. *Scripta Hierosolymitana*, 14, 166-181.
- Hahn, U. & Bailey, T.M. (2005). What makes words sound similar? *Cognition*, 97, 227-267.
- Hahn, U., & Chater, N. (1997). Concepts and similarity. In K. Lamberts & D. Shanks (Eds.), *Knowledge, concepts and categories* (pp. 43–92). Hove: Psychology Press.
- Hahn, U. & Chater, N. (1998). Understanding similarity: A joint project for psychology, case-based reasoning, and law. *Artificial Intelligence Review*, 12, 393-427.
- Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition*, 87, 1–32.
- Hahn, U., Close, J. & Graf, M. (2009) Transformation direction influences shape similarity judgements. *Psychological Science*, 20, 447-454.
- Hampton, J.A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18, 441-461.
- Hampton, J.A. (1995a). Testing the prototype theory of concepts. *Journal of Memory and Language*. 34, 686-708

- Hampton, J.A. (1995b). Similarity-based categorization: The development of prototype theory. *Psychologica Belgica*, 35, 103-125.
- Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception and Psychophysics*, 12, 108 -116
- Hardiman, P. T., Dufresne, R., & Mestre, J.P. (1989). The relation between problem categorization and problem solving among experts and novices. *Memory & Cognition*, 17, 627-638.
- Hayward, W. G., & Tarr, M. J. (1995). Spatial language and spatial representation. *Cognition*, 55, 39-84.
- He, Z. J. and Nakayama, K. (1994) Perceived surface shape determines correspondence strength in apparent motion. *Vision Research*, 34, 2125-2136.
- Heit, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, 24, 341-380.
- Hintzmann, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, 93, 411– 428.
- Hochberg, J., & McAlister, E. (1953). A quantitative approach to figure 'goodness'. *Journal of Experimental Psychology*, 46, 361–364.

- Hodgetts, C.J., Hahn, U., & Chater, N. (2009). Transformation and Alignment in Similarity. *Cognition*, *113* (1), 62-79.
- Hofstadter, D. 1995. *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York NY: Basic Books.
- Holyoak, K. J., & Thagard, P. (1989). Analogical mapping by constraint satisfaction. *Cognitive Science*, *13*, 295–355.
- Howell, D. C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.
- Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: A theory of analogical access and mapping. *Psychological Review*, *104*, 427–466.
- Hyman, R., & Well, A. (1968). Perceptual separability and spatial models. *Perception & Psychophysics*, *3*, 161-165.
- Imai, S. (1977). Pattern similarity and cognitive transformations. *Acta Psychologica*, *41*, 433-447.
- Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General*, *132*, 133-156.

- Kaldy, Z., & Leslie, A. M. (2003). Identification of objects in 9-month-old infants: integrating 'what' and 'where' information. *Developmental Science*, 6, 360-373.
- Keane, M. T., Ledgeway, T., & Duff, S. (1994). Constraints on analogical mapping: A comparison of three models. *Cognitive Science*, 18, 387-438.
- Keil, F.C. (1989). *Concepts, kinds and development*. Cambridge, MA: Bradford Books/MIT Press.
- Kenemans, J.L., Lijffijt, M., Camfferman, G. and Verbaten, M.N. (2002). Split-second sequential selective activation in human secondary visual cortex. *Journal of Cognitive Neuroscience*, 14, 1448-1461.
- Kolstad, V., & Baillargeon, R. (1991). *Appearance and knowledge-based responses to containers in infants*. Unpublished manuscript.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MIT: Harvard University Press.
- Krumhansl, C. L. (1978). Concerning the Applicability of Geometric Models to Similarity Data: The Interrelationship Between Similarity and Spatial Density. *Psychological Review*, 85, 445-463.
- Kruschke, J. K. (1992). ALCOVE: an exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.

- Kruskal, J.B. (1964a). Multidimensional scaling by optimizing Goodness of fit to a Nonmetric Hypothesis, *Psychometrika*, 29, 1-28.
- Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, 93-107.
- Lacroix, G. L., Giguère, G., & Larochelle, S. (2005). The origin of exemplar effects in rule-driven categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 272-288.
- Lamberts, K., Brockdorff, N., & Heit, E. (2002). Perceptual processes in matching and recognition of complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1176-1191.
- Larkey, L. B., & Love, B. C. (2003). CAB: Connectionist analogy builder. *Cognitive Science*, 27, 781-794.
- Larkey, L. B., & Markman, A. B. (2005). Processes of similarity judgement. *Cognitive Science*, 29, 1061-1076.
- Lawson, R. (1999). Achieving visual object constancy over plane rotation and depth rotation. *Acta Psychologica*, 102, 221-245.

- Lawson, R., Humphreys, G. W., & Jolicoeur, P. (2000). The combined effects of plane disorientation and foreshortening on picture naming: one manipulation or two? *Journal of Experimental Psychology: Human Perception and Performance*, 26, 568-581.
- Lawson, R., Bulthoff, H. H., & Dumbell, S. (2003). Interactions between view changes and shape changes in picture-picture matching. *Perception*, 32, 1465-1498.
- Lee, M.D., & Navarro, D.J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, 9(1), 43-58.
- Leech, R., Mareschal, D. & Cooper, R. (2007). Relations as transformations: Implications for analogical reasoning. *Quarterly Journal of Experimental Psychology*, 60, 897-908.
- Li, M., & Vitányi, P. (1997). *An introduction to Kolmogorov complexity and its applications*, (2nd ed.). New York: Springer-Verlag.
- Love, B.C., Medin, D.L. & Gureckis, T.M. (2004). SUSTAIN: A network model of category learning. *Psychological Review*, 111, 309-332.
- Luce, R.D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

- Malt, B. C. (1990). Features and Beliefs in the Mental Representations of Categories. *Journal of Memory and Language*, 29, 289-315.
- Markman, A. B. (1999). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Markman, A. B., & Gentner, D. (1993a). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, 32, 517-535.
- Markman, A. B., & Gentner, D. (1993b). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25, 431-467.
- Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition*, 24, 235-249.
- Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. *Psychological Science*, 8, 363-367.
- Markman, A. B., & Gentner, D. (2000). Structure mapping in the comparison process. *American Journal of Psychology*, 113, 501-538.
- Markman, A.B., & Gentner, D. (2005). Nonintentional similarity processing. In R. Hassin, J.A. Bargh, & J.S. Uleman (Eds.) *The new unconscious*. (pp. 107-137) New York: Oxford University Press.



- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, *88*, 375–407.
- Melara, R. D., Marks, L. E., & Lesko, K. E. (1992). Optional processes in similarity judgments. *Perception & Psychophysics*, *51*, 123-133.
- Milliken, B., & Jolicoeur, P. (1992). Size effects in recognition memory are determined by perceived size. *Memory & Cognition*, *20*, 83-95.
- 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.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207–238.
- Medin, D.L., Wattenmaker, W.D.. & Hampson, S.E. (1987). Family resemblance, conceptual cohesiveness and category construction. *Cognitive Psychology*, *19*, 242-279.
- Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review*, *100*, 254-278.

- Muller, M., van Rooij, I., and Wareham, T. (2009) Similarity as Tractable Transformation. In N.A. Taatgen and H. van Rijn (eds.) *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*. Cognitive Science Society; Austin, TX. 49-55.
- Murphy, G. L. (2004). *The big book of concepts*. MIT Press: Cambridge, USA.
- Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review*, 92, 289-316.
- Navarro, D. J. & Lee, M. D. (2003). Combining dimensions and features in similarity-based representations. In S. Becker, S. Thrun, and K. Obermayer (Eds.), *Advances in Neural Information Processing Systems*, 15 (pp. 67-74). Cambridge, MA: MIT Press.
- Nosofsky, R. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94-140.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104-114.
- Nosofsky, R. (1985a). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception & Psychophysics*, 38, 415-432.

- Nosofsky, R. M. (1985b). Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics*, 37(1), 89-91.
- Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
- Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13(1), 87-108.
- Nosofsky, R. M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology*, 34(4), 393-418.
- Nosofsky, R. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology*, 23, 94-140.
- Nosofsky, R.M., & Palmeri, T.J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, 104, 266-300.
- Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). Asymmetries in stimulus comparisons by monkeys and man. *Current Biology*, 13, 1803-1808.
- Osherson, D. N. (1990). Categorization. In D. N. Osherson & E. E. Smith (Eds.), *Thinking: an invitation to cognitive science*. Cambridge, MA: MIT Press.

- Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, *97*, 185–200.
- Oyama, T, Simizu, M., & Tozawa J. (1999). Effects of similarity on apparent motion and perceptual grouping. *Perception*, *28*, 739 – 748.
- Palmer, S. E. (1983). The psychology of perceptual organization: A transformational approach. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 269–339). New York: Academic Press.
- Panis, S., Vangeneugden, J., Wagemans, J. (2008). Similarity, typicality, and category-level matching of morphed outlines of everyday objects. *Perception*, *37*(12), 1822-1849.
- Pinker, S. & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences* *13* (4), 707-784.
- Podgorny, P., & Garner, W. (1979). Reaction time as a measure of inter- and intraobject visual similarity: Letters of the alphabet. *Perception & Psychophysics*, *26*, 37-52.
- Polk, T.A., Behensky, C., Gonzalez, R., and Smith, E.E. (2002). Rating the similarity of simple perceptual stimuli: Asymmetries induced by manipulating exposure frequency. *Cognition*, *82*(3), 75 - 88.

- Posner, M. I. & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–63.
- Posner, M.J., & Mitchell, R.F. (1967). Chronometric analysis of classification. *Psychological Review*, 74, 392-409.
- 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*, 35, 1062-1080.
- Pothos, E.M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science*, 26, 303-343.
- 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., & Prat-Sala, M. (2009). Similarity chains in the transformational paradigm. *European Journal of Cognitive Psychology*, 21, 1100-1120.
- Potts, B. C., Melera, R. C., & Marks, L. E. (2003). Circle size and diameter tilt: A new look at integrality and separability. *Perception & Psychophysics*, 60 (1), 101–112

- Reed, S. (1972). Pattern recognition and categorization. *Cognitive Psychology* 3, 382–407.
- Reeves, A., Fuller, H., & Fine, E. (2005). The role of attention in binding shape to color. *Vision Research*, 45, 3343-3355.
- Regehr, G., & Brooks, L.R. (1993). Perceptual manifestations of an analytic structure: the priority of holistic individuation. *Journal of Experimental Psychology: General*, 122, 92-114.
- Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, 77, 481–495.
- Riesbeck, C., & Schank, R. C (1989). *Inside Case-Based Reasoning*. NJ: Lawrence Erlbaum Associates.
- Rips, L. J. (1975). Inductive judgements about natural categories. *Journal of Verbal Learning and Verbal Behavior*, 14, 665–685.
- Rips, L. J. (1989). Similarity, typicality, and categorization. In: *Similarity and analogical reasoning*, ed. S. Vosniadou & A. Ortony, pp. 21–59. Cambridge University Press.

- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111-144). San Diego, CA: Academic Press.
- Rosch, E. (1978). Principles of categorization. In Rosch, E. and Lloyd, B., (Eds.), *Principles of categorisation, in cognition and categorisation*, (pp. 27 – 48). Erlbaum, Hillsdale, NJ.
- Rothkopf, E. Z. (1957). A Measure of Stimulus Similarity and Errors in Some Paired-Associate Learning. *Journal of Experimental Psychology*, 53, 94-101.
- Schoenfeld MA, Tempelmann C, Martinez A, Hopf JM, Sattler C, Heinze HJ, Hillyard SA (2003) Dynamics of feature binding during object-selective attention. *Proceedings of the National Academy of Science, USA*, 100, 11806-11811.
- Sharikadze M, Otto TU, Kezeli AR, Fahle M, & Herzog MH (2003). The time course of visual feature binding. *Perception*, 32, ECVF Supplement, 116.
- Shechter, S., Hochstein, S., & Hillman, P. (1988). Shape similarity and distance disparity as apparent motion correspondence cues. *Vision Research*, 28(9), 1013–1021.

- Shepard, R. N. (1957). Stimulus and response generalization: a stochastic model relating generalization to distance in psychological space. *Psychometrika*, *22*, 325–345.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.
- Shepard, R.N. (1984). Ecological constraints on internal representation: resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review*, *91*, 417-447.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.
- Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In J.R. Pomerantz & C.L. Lockhead (eds.), *The perception of structure* (pp. 53-71). Washington, DC: American Psychological Association.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Shepard, R. N., & Judd, S. A. (1976). Perceptual illusion of rotation of three-dimensional objects. *Science*, *191*, 952-954.



- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75 (13, Whole No. 517).
- Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 84, 127-190.
- Simon, H. A. (1972). Complexity and representation of patterned sequences of symbols. *Psychological Review*, 79, 369–382.
- Sloutsky, V. M., & Yarlas, A. S. (submitted). Processing of information structure: Mental representations of elements and relations.
- Smith, J. D., & Baron, J. (1981). Individual differences in the classification of stimuli by dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, 7, 1132-1145.
- Smith, L. B., & Evans, P. (1989). Similarity, identity, and dimension: Perceptual classification in children and adults. In B. E. Shepp & S. Ballesteros (Eds.), *Object perception: Structure and process* (pp. 325-356). Hillsdale, NJ: Lawrence Erlbaum.
- Stewart, N., & Brown, G. D. A. (2005). Similarity and dissimilarity as evidence in perceptual categorization. *Journal of Mathematical Psychology*, 49, 403-409.

- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.
- Suzuki, H., Ohnishi, H., & Shigemasu, K. (1992). Goal-directed processes in similarity judgment. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 343-348). Hillsdale, NJ: Erlbaum.
- Takane, Y., and Sergent, J. (1983). Multidimensional scaling models for reaction times and same different judgments. *Psychometrika*, *48*, 393-423.
- Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review*, *2*, 55-82.
- Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Science*, *21*, 233-282.
- Taylor, E. G. & Hummel, J. E. (2009). Finding *similarity* in a model of relational reasoning. *Cognitive Systems Research*, *10*, 229-239.
- Tenenbaum, J. B. (1996). Learning the structure of similarity. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo (Eds.), *Neural information processing systems*, *8*. Cambridge, MA: MIT Press.

- Thibaut, J. P., & Gelaes, S. (2006). Exemplar effects in the context of a categorization rule: Featural and holistic influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 6, 1403-1415.
- Tomonaga, M., & Matsuzawa, T. (1992). Perception of complex geometric figures in chimpanzees (*Pan troglodytes*) and humans (*Homo sapiens*): Analyses of visual similarity on the basis of choice reaction time. *Journal of Comparative Psychology*, 106, 43-52.
- Treisman, A. (1977) Focused attention in the perception and retrieval of multidimensional stimuli. *Perception and Psychophysics*, 22, 1-11.
- Treisman, A. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society, Series B*, 353, 1295-1306.
- Treisman, A. & Kanwisher, N.K. (1998). Perceiving visually-presented objects: Recognition, awareness, and modularity. *Current Opinion in Neurobiology*, 8, 218-226.
- Treisman, A., & Souther, J., 1985. Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, 114, 285-310.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327-352.

- Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 79-98). Hillsdale, NJ: Erlbaum.
- Tversky, A., & Gati, I. (1982). Similarity, separability, and the triangle inequality. *Psychological Review*, 1982, 89, 123-154.
- Tversky, A., & Hutchinson, J.W. (1986). Nearest neighbor analysis of psychological spaces. *Psychological Review*, 93 (1), 3-22.
- Ullman, S (1979). *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.
- Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, 32(3), 193-254.
- van Rooij, I. (2008). The Tractable Cognition thesis. *Cognitive Science*, 32, 939-984.
- Ward, T. B. (1985). Individual differences in processing stimulus dimensions: Relation to selective processing abilities. *Perception & Psychophysics*, 37, 471-482.
- White, J. B. (2008). Self-other similarity judgment asymmetries reverse for people to whom you want to be similar. *Journal of Experimental Social Psychology*, 44, 127-131.

Wish, M. (1967). A model for the perception of Morse Code-like signals. *Human Factors*, 1967, 9, 529-540.

Zaki, S.R. & Homa, D. (1999). Concepts and transformational knowledge. *Cognitive Psychology*, 39, 69-115.

---

# Appendices

## Appendix A.1 - Model Predictions for RD

The table in Appendix A.2 displays the transformation distances for each distinct comparison in the domain tested throughout this thesis. The numbers refer to the number of operations, from the 3-operation coding scheme, required to transform the representation of one object into that of the other as a measure of code length. The first column and the top row refer to the respective dimensions (shape or colour). As directional comparisons (see Tversky, 1977) not all comparisons are symmetrical, for example, “how similar is AB to BB?”  $\neq$  “how similar is BB to AB?” (see Chapter 4). Given that certain experiments within this thesis involve non-directional judgements (“how similar *are* pair 1 and 2?”), the simplest symmetrical measure within the Kolmogorov complexity framework was used: max-distance, which takes the larger of the two distances, in each direction, as the fixed complexity measure (Li & Vitányi, 1997). Max-distance was also the measure used in Hahn et al. (2003), and further discussion of this measure can be found there in Footnote 2.

**Appendix A.2 - Table of predictions**

| Dimension 1/2 | AA/AA | AB/AB | AB/BA | AA/AB | AA/BA | AB/AA | AB/BB | AA/BB | AB/AC | AB/CB | AB/BC | AB/CA | AA/BC | AB/CC |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| AA/AA         | 0     | 0     | 1     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 3     | 3     | 4     | 4     |
| AB/AB         | 0     | 0     | 1     | 2     | 2     | 2     | 2     | 2     | 2     | 2     | 3     | 3     | 4     | 4     |
| AB/BA         | 1     | 1     | 1     | 3     | 3     | 3     | 3     | 3     | 3     | 3     | 3     | 3     | 5     | 5     |
| AA/AB         | 2     | 2     | 3     | 4     | 4     | 3     | 3     | 4     | 4     | 4     | 5     | 5     | 6     | 5     |
| AA/BA         | 2     | 2     | 3     | 4     | 2     | 3     | 3     | 4     | 4     | 4     | 5     | 5     | 6     | 5     |
| AB/AA         | 2     | 2     | 3     | 3     | 3     | 4     | 4     | 4     | 4     | 4     | 5     | 5     | 5     | 6     |
| AB/BB         | 2     | 2     | 3     | 3     | 3     | 4     | 4     | 4     | 4     | 4     | 5     | 5     | 5     | 6     |
| AA/BB         | 2     | 2     | 3     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 5     | 5     | 6     | 6     |
| AB/AC         | 2     | 2     | 3     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 5     | 5     | 6     | 6     |
| AB/CB         | 2     | 2     | 3     | 4     | 4     | 4     | 4     | 4     | 4     | 4     | 5     | 5     | 6     | 6     |
| AB/BC         | 3     | 3     | 3     | 5     | 5     | 5     | 5     | 5     | 5     | 5     | 5     | 5     | 7     | 7     |
| AB/CA         | 3     | 3     | 3     | 5     | 5     | 5     | 5     | 5     | 5     | 5     | 5     | 5     | 7     | 7     |
| AA/BC         | 4     | 4     | 5     | 6     | 6     | 5     | 5     | 6     | 6     | 6     | 7     | 7     | 8     | 6     |
| AB/CC         | 4     | 4     | 5     | 5     | 5     | 6     | 6     | 6     | 6     | 6     | 7     | 7     | 6     | 8     |

Nb. Not included is ABCD - a dimensional structure with no within-dimension similarity effectively reduces to a comparison involving similarities on one dimension only, that is, to column 1 or row 1, and hence is equivalent to the 14 possible comparisons examined in full in Experiment 1.

In order to find the simplest transformation, we conducted an exhaustive search and selected the set with the smallest number of operations. This almost always led to a unique transformation. I note that the search space for these items is small. There are typically very few ways to conceptualise transforming one pair into the other, and the simplest is also, in our experience, the most natural. For example, in the case where dimension 1 = AB/BA and dimension 2 = AB/BA, one could either a) swap on two dimensions or b) by a whole object swap. The latter is not only simpler, but also more intuitive and hence likely to be found first. A reader should be able to verify these claims by picking a few pairs and trying to apply the transformations themselves.

### **Appendix A.3 - Model predictions for SA**

#### *SME*

Model predictions for SME followed directly Larkey and Markman (2005). The similarity measure for SME was based on MIPS only. The exclusive focus on MIPS stems from SME's strict adherence to the *one-to-one* mapping constraint, which states that features and/or relations cannot form multiple alignments. Strictly speaking this constraint governs the mapping process, and it would be possible to impose similarity measures on the final mapping that also took into account MOPs though this has not been common practice in the past.

The way mappings were determined is illustrated in the following example: In the comparison - black square, white triangle vs. black triangle, black triangle – there are a number of mapping conflicts. The black square maps onto both triangles on the basis of colour but under the constraint can only form one-to-one alignments. The final mapping would involve 2 MIPS, black square to black triangle (colour MIP), and



white triangle to black triangle (shape MIP). The ‘leftover’ mappings are the matches that do not correspond (i.e., MOPs). This would be the black square to the second black triangle (colour MOP).

Note that this approach does not code MIPs on the basis of location, i.e., left of/right of. A MIP model that referred to feature locations was also tested where features were only aligned if they occupied the same relative position in a pair. These two alternative approaches make quite distinct predictions, correlating only at  $r=.463$  for Experiment 1. The original version used by Larkey and Markman provides considerably better data fits, so reporting in the manuscript is restricted to that scheme.

### *SIAM*

The simulations used the original code developed by L. Larkey for Larkey and Markman, (2005) which was generously made available to us. As in that paper, all simulations used the default model values.