Spring 2015

A task-general dynamic neural model of object similarity judgments

Gavin Wesley Jenkins
University of Iowa

A TASK-GENERAL DYNAMIC NEURAL MODEL OF OBJECT SIMILARITY JUDGMENTS

by

Gavin Wesley Jenkins

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

May 2015

Thesis Supervisors: Professor Larissa K. Samuelson
& Professor John P. Spencer

Graduate College
The University of Iowa Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Gavin Wesley Jenkins

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Psychology at the May 2015 graduation.

Thesis Committee: _____
Larissa K. Samuelson, Thesis Supervisor

_____
Thomas A. Farmer

_____
Andrew Hollingworth

_____
John P. Spencer

_____
Teresa A. Treat

To my parents, for their love and support

ACKNOWLEDGMENTS

I express my warm thanks to Larissa Samuelson and John Spencer for their wisdom and guidance as my co-advisors throughout my time at Iowa.

I also thank the remainder of my thesis committee—Teresa Treat, Andrew Hollingworth, and Thomas Farmer—for their advice and guidance in the creation of this work.

This research program was primarily made financially possible by a generous grant by the National Defense Science and Engineering Graduate Fellowship (NDSEG) program.

ABSTRACT

The similarity between objects is judged in a wide variety of contexts from visual search to categorization to face recognition. There is a correspondingly rich history of similarity research, including empirical work and theoretical models. However, the field lacks an account of the real time neural processing dynamics of different similarity judgment behaviors. Some accounts focus on the lower-level processes that support similarity judgments, but they do not capture a wide range of canonical behaviors, and they do not account for the moment-to-moment stability and interaction of realistic neural object representations. The goal of this dissertation is to address this need and present a broadly applicable and neurally implemented model of object similarity judgments. I accomplished this by adapting and expanding an existing neural process model of change detection to capture a set of canonical, task-general similarity judgment behaviors. Target behaviors to model were chosen by reviewing the similarity judgment literature and identifying prominent and consistent behavioral effects. I tested each behavior for task-generality across three experiments using three diverse similarity judgment tasks. The following behaviors observed across all three tasks served as modeling targets: the effect of feature value comparisons, attentional modulation of feature dimensions, sensitivity to patterns of objects encountered over time, violations of minimality and triangle equality, and a sensitivity to circular feature dimensions like color hue. The model captured each effect. The neural processes implied by capturing these behaviors are discussed, along with the broader theoretical implications of the model and possibilities for its future expansion.

PUBLIC ABSTRACT

We compare objects and judge how similar or different they are throughout our daily lives. For example, we judge family relations from the similarity of faces, and we compare the similarity of products as part of our purchasing decisions. Similarity is also critical to specialist and industrial applications like measuring the uniformity of manufactured goods or comparing x-ray images to judge tumor growth.

Psychologists know a great deal about the exact rules that support similarity judgments and the resulting patterns of behavior. For example, we know how differences in features such as color, size, and shape interact with one another to influence similarity judgments. We also know how memories of other objects seen recently and how time pressure or different goals influence similarity judgments.

Less is known, however, about the neural processes behind these similarity judgments. The goal of this dissertation is to fill this gap in knowledge by creating a computer model that can predict and explain the most well-known similarity behaviors, using neurally realistic cognitive processes. This computer model links the neural activity that supports similarity judgments to those previously studied in the context of other cognitive tasks. These connections will allow psychologists to paint a more complete picture of how we process and understand objects in general.

The model also serves an important step toward direct applications of machine simulated artificial similarity judgments between objects. Machine-based similarity may lead to accurate automatic second opinions on medical images or more efficient satellite or surveillance image interpretation.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1

INTRODUCTION

Determining the visual similarity of objects is critical for a wide variety of daily decisions: we use facial resemblance to guess family relationships, we identify similarities between diagrams and scenes to assemble furniture or to navigate unfamiliar neighborhoods, we compare and contrast produce at the grocery store to pick the ripest fruit, and we often express explicit similarity judgments in the course of making decisions about categorization or analogies.

The core ingredients of similarity judgments are well-known: objects are compared according to metric feature dimensions like color, shape, and orientation. The relative differences between compared objects along these dimensions serve as the principal component of similarity judgments, whether those differences are considered as continuous measures (Shepard, 1987), binary match/non-match distinctions (Tversky, 1977), or in terms of number of transformations to close the difference gap between objects (Hahn, Chater, & Richardson, 2002). Similarity judgments have also been shown to be influenced by an array of factors that are independent of metric feature comparisons. For example, the degree of judged similarity between two items can change if an experimenter switches the order the items are mentioned (the "asymmetry" effect, Tversky 1977). Judgments about two objects also depend, in part, on how unique those objects are compared to other previously seen objects, even if those other objects are not currently present (Krumhansl, 1978).

Many formal, computational models exist that are able to quantitatively capture one or more of these similarity judgment behaviors. These models vary from purely abstract

mathematical approaches that explicitly avoid any question of biological implementation (Tenenbaum & Griffiths, 2001) to neurally-inspired connectionist models (Ashby, Paul, & Maddox, 2011; Love, Medin, & Gureckis, 2003). Models also vary from those specifically designed to capture an array of different similarity judgments across tasks (Pothos, Busemeyer, & Trueblood, 2013) to those that implement similarity explicitly, but for a single specific type of task or application of similarity, like categorization (Kruschke, 1992).

The field lacks a model, however, that represents similarity judgment behaviors at the level of neural processes and population dynamics. Although not all models must exist at a neural process level, there are several benefits to capturing behaviors through neural dynamics that have not yet been fully realized by any models in the field. First, behaviors can sometimes originate among interactions at the neural level, and these neural interactions can be difficult to understand without a model that considers this level of detail. For instance, neural models of colorblindness and opposite colored afterimages were useful in clarifying the origin of these phenomena in visual cognition. These are phenomena that do not themselves serve any high level goal of an organism, and they are not easily explained or predicted by an abstract model of color vision. Opponent process theory, however (Hurvich & Jameson, 1957; Hering, 1964), did explain these results when considering color vision from a neural basis of visual receptors sensitive to opposing pairs of colors. In the similarity judgment literature, a neural process model might be equally useful in clarifying the origin of particular biases.

Another reason to pursue a neural model of similarity judgments is that existing models of similarity have not dealt with particular constraints imposed by the neural implementation level. One constraint is that the neural representation of one object must

remain stable in working memory as the second object is perceived and compared. Current models of similarity have not accounted for this level of neural process. For instance, in connectionist models, objects are stored either as pre-consolidated nodes (e.g. Ashby, Paul, & Maddox, 2011) or as vectors of feature dimension nodes (e.g., Kruschke, 1992) which are activated either in parallel or kept abstractly active (i.e., stored in the computer's memory) until the next objects must be processed. In a neural process sense, however, memories are non-trivial to maintain with real time neural dynamics. Thus, it is possible that the mechanistic basis for some similarity behaviors lies in the dynamics involved in maintaining working memories over the interval between perceiving compared objects. For example, Tversky's (1977) findings that similarity judgments can be asymmetric when object order is reversed could plausibly be based on different patterns of memory decay, interaction, or stability at a neural level during the time interval between processing one object and the next. A neural process model is the best method of revealing such possibilities.

A second key constraint imposed by a neural process view is on the nature of the representations that underlie similarity judgments. In the similarity judgment literature, many models posit representations over continuous Cartesian feature spaces. A Cartesian feature space is a representational system where each possible combination of values along feature dimensions (like size, color, or shape) represents a point in a multidimensional space at which an object could be represented. This allows for an intuitive way to think about object comparisons: in such a system, similarity can simply be based on the distance (Euclidean or city-block) between two point objects at different locations in the feature space. A Cartesian feature space quickly becomes neurally implausible, however, when its neural resource demand is considered in naturalistic situations. If comparing objects along

a realistic seven feature dimensions in such a system, for example, with 50 distinguishable steps along each dimension, the memory space would already require at least 50^7 neurons. This is several times more neurons than exist in the brain, which means that a Cartesian feature space is neurally implausible. To date, models of similarity judgments have not tackled this topic in a way that also addresses the representational stability issue discussed above.

A final advantage of neural models is that they allow for an integration of other cognitive functions related to similarity behaviors. Object similarity is directly related to a number of visual cognitive processes, all of which have been investigated at the neural level. Similarity relies on feature perception for the colors, shapes, orientations, etc. being compared (Shepard, 1987; Faubel & Schöner, 2008; Mel, 1997); attention for binding those features to objects (Treisman & Gelade, 1980; Ashby, Prinzmetal, Ivry, & Maddox, 1996; Samuelson, Smith, Perry, & Spencer, 2011; Hommel & Colzato, 2009) or for weighting dimensions (Shepard, 1964; Maunsell & Treue, 2006; Klaus, et al., 2007; Chajut, Schupak, & Algom, 2009); semantic relationships (Recker, Plumert, Hund, & Reimer, 2007); and working memory for remembering two or more of objects long enough to compare them (Johnson, Spencer, & Schöner, 2008; Johnson, Spencer, & Schöner, 2009; Johnson, Spencer, Luck, & Schöner, 2009). Object similarity in turn contributes to various other related downstream behaviors like object categorization (Ashby & Maddox, 2005; Nomura & Reber, 2008) or visual search (Grossberg, Mingola, & Ross, 1994). All of these related processes are understood increasingly at a neural process level, and integrating a neural model of similarity judgments into this picture may help us understand the role that similarity judgments play in broader cognitive processing.

Dynamic Neural Field (DNF) models are neural process models that hold the potential to model a variety of object similarity behaviors. DNF models represent cognitive processes primarily as interactions and activity between and within "neural fields." These fields are arrays of neural units organized by continuous feature dimensions such as color, orientation, or spatial position. Activation within these fields can enter different attractor states that can be stably maintained over short-term delays. In this way, DNF models have explicitly addressed the challenge of representational stability—how an activation pattern can be stably maintained to enable comparisons with other perceived or attended information.

In addition, a recent model has addressed how spatial positions can be used to selectively bind object features together in working memory to support the comparison operations necessary to, for instance, detect changes in object features when they occur (Schneegans, Spencer, & Schöner, in press). Critically, this model addresses the issue of exponentially increasing resource usage implied by Cartesian feature spaces. Because all features of an object representation are 'bound' via a common spatial dimension, this DNF model requires only linearly increasing resources—one additional set of fixed-size feature fields—with each additional feature dimension. Variants of this model have been shown to capture nuances of object representation (Johnson, Spencer, & Schöner, 2008) object binding (Samuelson, Smith, Perry, & Spencer, 2011), word learning (Samuelson, Spencer, & Jenkins, 2013), and object recognition (Faubel & Schöner, 2008).

A final advantage of DNF models is that they have been used to simulate a broad array of cognitive processes and to capture related behaviors, from executive control (Buss & Spencer, 2008) to motor planning (Erlhagen & Schöner, 2002) to spatial cognition (Spencer, Simmering, Schutte, & Schöner, 2007). Most directly relevant to similarity

judgments, DNF models have captured change detection behavior. Change detection is a form of binary similarity judgment, where “different” and “same” correspond directly to “detection” and “non-detection” or “go” and “no-go,” making this a basic starting point for capturing a wider array of other tasks in the similarity judgment literature. It is the change detection version of the model (Schneegans, Spencer, & Schöner, in press; see also Johnson, Spencer, Luck, & Schöner, 2009 for a related model) that will serve as the basis of a similarity judgment model in this dissertation. DNF’s applicability to a range of other cognitive processes, holds the promise of clarifying how similarity processes are related to visual cognition more generally.

A DNF model has not yet been developed to capture specific similarity judgment behaviors. The goal of this dissertation is to adapt the DNF model from Schneegans, Spencer, and Schöner (in press) to explicitly judge similarity and to capture as many canonical behaviors as possible from the object similarity judgment literature, while contributing the unique advantages of neural process modeling to the field.

The first step in achieving this goal is to identify a set of canonical similarity judgment behaviors most meaningful and informative to capture. Specifically, I am interested in task-general behaviors that span similarity contexts. Behaviors that span different similarity tasks are the most likely behaviors to stem from core processes of similarity judgment itself. In the following section, I survey the similarity judgment literature to identify known or potential task-general behavioral effects.

Literature Review of Object Similarity Judgments

I seek to understand and model the cognitive processes that drive object similarity judgments. The most straightforward way to infer the nature of these processes is to examine the behaviors that they drive. Examining patterns of any similarity behaviors

directly lends insight into similarity processes—how systematic, widespread, high or low level they are, etc. For this dissertation, however, behaviors are also specifically needed as targets for initial fits of a task-general, neural process model of similarity judgments.

Task-general behaviors are good targets for modeling work for several reasons. Task-general behaviors by definition are unlikely to be dependent on or easily influenced by task variables, thus these serve as robust modeling targets. They are also efficient targets, because one set of accurate processes in a model allows it to capture that behavior across many contexts. Capturing task-general behaviors also promises the largest number of theoretical connections to existing literature, since task-general behaviors are the most commonly cited and most actively researched. Eventually, a comprehensive similarity judgment model should also be capable of capturing task-specific behaviors, but for initially establishing and fitting a neural process model, specific behaviors are not the most efficient, central, or robust targets.

Prior literature suggests several behavioral patterns that may fit these criteria. I divide my review of the literature into two broad groups of behavioral findings: those based on comparison of the features of objects and those based on factors other than feature comparison. This distinction is intended only as an organizing principle and starting point for approaching the large literature on similarity judgment behaviors.

Feature Comparison

Comparisons between object features are historically the earliest known and the most universally appreciated factor contributing to object similarity judgments. All objects have features, like texture, size, or brightness. The most straightforward way to compare two objects is by comparing these features. Closer matches between features and/or a greater number of feature dimensions along which objects match means higher similarity

between objects. Quantitative features can be precisely compared: an object can be twice as bright as something else, and a 30 degree rotation from one line to another can be said to be twice as much of a difference as a 15 degree rotation. More qualitative features like texture can also be compared but may be restricted to coarser distinctions such as simply match / mismatch evaluations.

All models of object similarity include some version of a feature comparison process. Most models further describe some version of a "feature space" where all or some of these features are compared, especially metric features. One common example of a feature space is a multidimensional Cartesian coordinate system, where each feature is an orthogonal dimension, different values of a feature are points on that dimension, and objects are single points, clouds, or volumes in a space defined often by multiple dimensions. Other types of feature spaces are possible, however. Models that represent objects in feature spaces usually quantify similarity via a distance measure through that space, whether city-block distance, Euclidean distance, or other more complex measures.

Formal, quantitative models of object similarity judgments are more modern, arising in the early to mid 20th century, initially based around feature comparison in rigid, mathematical feature spaces (Richardson, 1938; Torgerson, 1952; Shepard, 1957). The most common type of quantitative feature space was (and may still be) a Cartesian coordinate system as assumed, for example, in an analysis and modeling method called multidimensional scaling (Shepard, 1980). Multidimensional scaling (MDS) is an algorithm that takes dissimilarity values between different possible pairs of objects in a set as input and arranges points in a Cartesian space, each corresponding to an object, such that distances most closely match either the same proportions of the raw dissimilarities

(metric MDS) or the rank-order of the dissimilarities (non-metric MDS).[1] The input values can be derived from any task. The number of dimensions used to arrange the points is specified as an input parameter, with lower numbers yielding simpler models and higher numbers yielding better fits. All such dimensions, however, would still be considered to be orthogonal and Cartesian in an MDS analysis, even if input data may have originated from a task without Cartesian constraints.

An MDS output solution has no fixed axis identities. A three dimensional MDS solution, for example, will place points in three-dimensional space, but the location, rotation, and labeling of the three axes is abstract. Part of the process of interpreting an MDS output is often judging how the output dimensions map onto the input dimensions. For instance, this might include rotating the output of a two-dimensional solution so that the vertical and horizontal correspond to meaningful and visually identifiable axes in the solution. However, regardless of interpretation, an N-dimensional MDS solution always outputs points that fit geometrically into exactly N, Cartesian, orthogonal dimensions.

Because object placements in the MDS algorithm are evaluated by the distance between pairs in the feature space, the algorithm used to calculate distance is another important input parameter. Originally, and still most commonly, Euclidean distance was and is assumed (Richardson, 1938, Hout, Goldinger, & Ferguson, 2013). Early on, however, the alternative of a city-block distance metric was demonstrated for some comparisons (Attneave, 1950). Evidence now suggests that city-block distance is a more appropriate measure when feature dimensions are separable (or “analyzable”) and not

[1] Metric MDS is used if all equal objective intervals in the input data can be trusted to also be psychologically equal in magnitude. Non-metric MDS is used if intervals are not necessarily psychologically consistent, and is the default choice unless measures are carefully controlled.

contingent upon one another, while Euclidean distance is more appropriate for integral dimensions (Arabie, 1991; Shepard 1987; Garner, 1974; Shepard, 1964). For example, color dimensions like color saturation and color hue are generally integral and best measured against one another with a Euclidean metric, while dimensions like size and tilt are analyzable / separable and best measured against one another with a city-block metric. Metrics somewhere in between are also possible for semi-separable dimensions (Shepard, 1964). MDS algorithms traditionally use a Minkowski power formula to compute distance[2], which can technically vary continuously between a power value of 1 (city block) or 2 (Euclidean) or outside of that range.

Which of these parameter values best fits human behavior is a measurable result that can be usefully tested across almost any tasks, since MDS analysis only requires a set of pairwise dissimilarity ratings from any source. In following chapters, I will include MDS as one consistent analysis across empirical tasks and as an analysis of my computational model of similarity judgments. MDS serves as a basis for many other analyses of orderliness of similarity judgments, relative biases between feature dimensions, etc. Number of output dimensions is also a parameter that can be fitted freely to data and to a model and results compared for consistency. The Minkowski distance measure variable, however, was theoretically constrained to 1 (city-block distance) in my analyses due to the separability of the feature dimensions used in my experimental stimuli (described in chapter 2).

[2] The formula is the same as the Euclidean distance formula, but the difference between objects along each dimension is taken to the power of x instead of 2, and the sum is then taken to the power of 1/x instead of 1/2. x is usually set to 1 or 2 for city block or Euclidean distance, but can be any whole or fractional value.

Scaling of distance measures is also a variable in feature comparison. The scaling factor is a function that relates distance in feature space to a final similarity judgment. Scaling is relevant to any distance measure, Euclidean, city-block, or otherwise. Similarity judgments are sometimes treated as scaling linearly or by any other function, although roughly exponential scaling is found to consistently match behavioral ratings across experiments and stimuli (for review, see Shepard, 1987). In an exponential scale, larger and larger distances in feature space have exponentially diminishing impact on similarity judgments, asymptoting toward "completely dissimilar" judgments. As exponential scaling has already been established as task-general and this conclusion has been widely adopted (Nosofsky, 1991; Pothos, Busemeyer & Trueblood, 2013; Kruschke, 1992), it is not necessary to test experimentally in this dissertation, and MDS analysis will assume exponential scaling.

The Cartesian/MDS approach is easy to work with and intuitive, but behavioral similarity is known to not always conform to these assumptions. Tversky (1977) demonstrated that similarity judgments can violate the Cartesian assumption of symmetry, for example. That is, people do not always rate the similarity between A and B as equal to the similarity between B and A. This is impossible in a classic Cartesian space where Euclidean (or city-block) distance is necessarily equal in both directions. Tversky also suggested that other Cartesian assumptions are routinely violated. Violation of the principle of minimality is when two non-identical objects are judged more similar than two identical objects, or when two identical objects are judged "different" or to have any value not on the "most similar" end of a similarity ratings scale. Violation of the principle of triangle equality is also possible. If dissimilarity is imagined as distance in a feature space, then triangle equality holds that two legs of a triangle between three objects cannot add up to

more than the length of the third. Violation of this principle occurs when the *dis*similarity of [objects A and B] plus the dissimilarity of [objects B and C] adds up to more than the dissimilarity of [objects A and C]. This is not necessarily a valid assumption for non-Euclidean spaces, but for a Cartesian feature space coordinate system, violations of triangle equality are damaging evidence.[3]

Evaluating asymmetric similarity judgments is statistically (and for a model, computationally) demanding. Multiple duplicate trials must be run for each pair of presented objects, since pairs can only be compared to themselves in opposite order. I therefore postponed analysis of this behavioral effect for empirical analysis and for the first iteration of my neural similarity judgment model. I did, however, evaluate the principles of minimality and triangle equality across multiple experimental tasks and modeling results. These effects can more efficiently be evaluated across any number of repetitions of each pair of objects.

More recently, the assumption of a Cartesian space has been found to be in conflict with other behavioral findings. People's ability to recognize differences along one dimension improves as two compared objects align along *other* dimensions (Gentner, 1983; Markman & Gentner, 1993; Jameson, et al., 2005). For example, color differences between dog breeds might be noticed more readily than color differences between a typewriter and a dog, even if the absolute difference in color is the same in both cases. The dogs share a number of features that the typewriter does not, like their shape and texture, making dogs more "alignable" with one another. This alignment then makes the remaining

[3] Tversky did not cite data for these violations, implying them to be commonplace. Shepard (1964) had earlier explicitly demonstrated at least triangle inequality in similarity judgments. I was unable to locate explicit published data for violations of minimality, although my own experimental as well as modeling data show these violations.

color difference easier to perceive. This alignability effect violates Cartesian assumptions that dimensions should be orthogonal: distance along color should have a constant contribution to overall similarity in a Cartesian feature space, regardless of what happens in other dimensions, yet behavioral evidence for alignability suggests that this is not the case. Alignability is a straightforward behavioral result that I evaluated for task-generality in empirical analysis.

Circular feature dimensions like color or orientation also violate Cartesian assumptions. A circular feature dimension cannot fit into an orthogonal Cartesian space. The plainly evident fact that that people can perceive and work with circular dimensions (like color hue or line orientation) with little difficulty serves as additional evidence against this classic approach. More broadly, circular dimensions significantly change the mathematics of distance measures in any type of feature comparison. In a linear dimension, moving one equally sized feature step is half the distance as moving two equally sized feature steps. Figure 1 shows how this is not necessarily true along a circular dimension: if distance is calculated as arc lengths around the outside of a circle, then the dimension acts like a linear one, but if distance is calculated as chord lengths across the inside of the circle, then two individually equal feature steps will be perceived as less than twice as different as one step. Shepard (1962, see also Shepard & Farrell, 1985) tested individual pairs of colors and examined their similarity ratios (Figure 1, middle) and demonstrated that most people perceive differences along the circular dimension of color by chord lengths.

Similarity with respect to circular dimensions is further complicated by the fact that even if similarity is perceived as arc distance, it is still ambiguous which direction along a circular dimension an object comparison should consider. In one sense, blue and magenta

are one feature step apart in Figure 1 (right). In another sense, they are five feature steps apart. This ambiguity does not exist in Cartesian spaces, and Cartesian similarity equations cannot function without modification to explicitly resolve this ambiguity. In general, circular dimensions and these associated complications are often overlooked or avoided by similarity models. Nevertheless, the constraints they impose can serve as valuable clues about underlying similarity processes, and a comprehensive model of similarity processes must account for them.

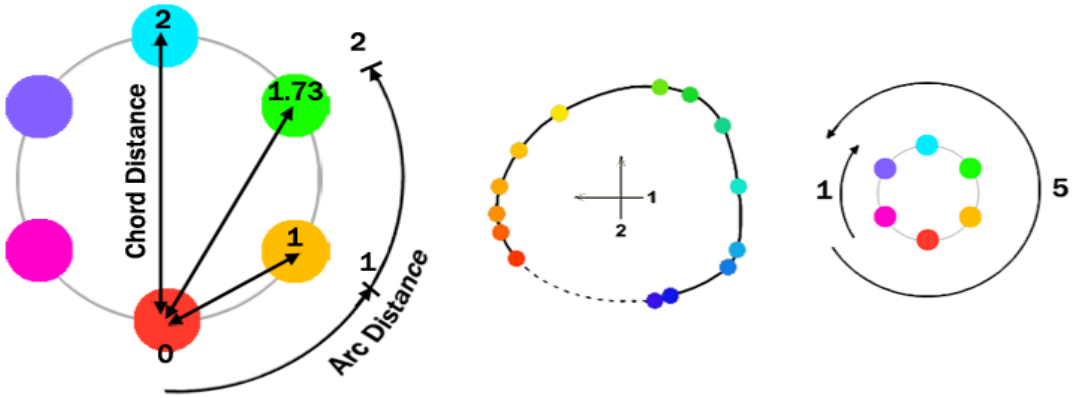

Figure 1: Chord and Arc-Based Circular Distance.

Cartesian feature space suffers still another threat from neural intractability. In a classic Cartesian space, adding dimensions increases the size of the space exponentially per dimension. As discussed above, Cartesian features spaces able to accommodate realistically complex objects require more neurons than are in the human brain. Humans appear to regularly succeed in making relatively high-dimensional comparisons, however.[4] This makes a Cartesian solution not neurally realistic.

[4] Examples of visual features include: color hue, color saturation, brightness, size, texture features such as glossiness, line orientation, monocular depth (lens focus, occlusion, etc.), binocular depth (eye offset), shape features such as intersections, direction and speed of motion, spatial frequency.

Some accounts of similarity deal with these conflicts with Cartesian geometry by abandoning either Cartesian feature space, typical Cartesian object representations, traditional Euclidean/city-block distance measures, or some combination thereof. It is also possible to remain in the realm of feature-comparison-based processes but replace a Cartesian feature space with a different space as the primary solution.

Pothos, Busemeyer, and Trueblood (2013) retained a Cartesian feature space, but proposed a novel object representation and distance measure. They proposed that objects and concepts could be represented by whole lines, planes, surfaces, or volumes in feature space. Then, instead of Euclidean distance, the amount of overlap between those shapes as "seen" (projected) from different perspectives in feature space can serve as a similarity metric. The model is illustrated in Figure 2. Objects can be represented in Pothos and colleagues' model by ranges or volumes of points in feature space, so a pair of objects is represented in the figure by lines instead of points (green and red). This range may perhaps represent a variety of different viewpoints on the objects or different contexts (lighting, state of inebriation) where they might be experienced, and thus would each cover more than one combination of possible feature values. When a similarity judgment is called for, the model chooses a perspective in feature space (eye icons on the left or right of Figure 2). Choice of perspective could change based on factors such as the wording of the question or which object is more attention-grabbing and is attended first. Similarity is then the perceived overlap of the two objects from the relevant perspective. More formally, the distance measure can be described as the size of a spatial projection from one object space to another.

This model can account for asymmetric similarity judgments of the type Tversky (1977) reported. As indicated in figure 2, changing the order of a question about two

objects can lead to asymmetric perceptions of similarity, because the size of the projection between objects changes from different perspectives. Violations of triangle equality can similarly be explained in Pothos and colleagues' model by the details of which order each comparison between objects is presented, or in unguided similarity judgments, which objects a person attends first or finds most salient. Violations of minimality are attributed only to noise.

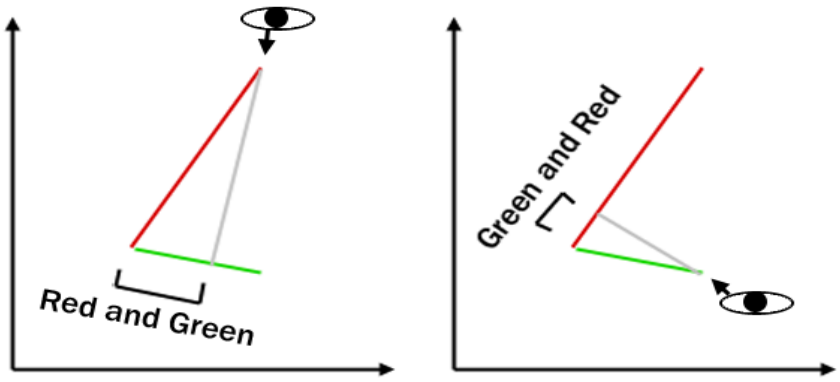

Figure 2: A quantum geometric model of similarity. An overlap-based similarity metric can account for asymmetry in similarity judgments. Green and red vectors represent objects that occupy a range of feature value combinations. Left: The "similarity between red and green" causes the listener to take one perspective (eye icon) and observe a large overlap. Right: The "similarity between green and red" causes the listener to take a different perspective and observe a smaller overlap between the same two lines.

Alternatively, limitations of feature space can be addressed if feature dimensions are represented independently from one another, rather than together in an integrated Cartesian space. Treisman and Gelade (1980) (see also Treisman, 1986) suggest a model like this as an early phase of object processing. According to their feature integration theory, different areas of the brain initially independently process sensory input corresponding to an object's basic feature values along various feature dimensions. For most features, the information is organized in individual spatial maps at this point in

processing. Shortly afterward, featural information that all comes from the same location in space is bound together and stored as an integrated "object file," allowing for both easier tracking with movement and easier comparison to other objects held in memory.

Treisman and Gelade's (1980) theory is described in the context of behavioral effects other than basic similarity judgments, including illusory conjunctions and visual "pop-out" searches, but the architecture of the model still accommodates non-Cartesian similarity effects. The model relies on an alternate feature space, or rather several of them, one per feature dimension. 'Local' similarity can be defined in each individual feature space, *or* similarity could be considered between already-integrated object files. The independent processing of feature dimensions allows Treisman and Gelade's (1980) model to address non-Cartesian similarity. Independent feature dimensions can warp, skew, strengthen, weaken, etc. without affecting other feature dimensions, and in ways that a single integrated Cartesian feature space cannot. As one concrete example, contributions to a similarity judgment from individual feature spaces for circular features like color hue or line orientation could be calculated circularly, without introducing the problems of an integrated Cartesian space.

Dynamic neural field (DNF) theory is based on a similar architecture as Treisman and Gelade's feature integration theory. Each feature dimension is represented independently with its own population of neural units, with values along the dimensions coupled together via a shared spatial dimension to create an 'integrated' object representation. For instance, DNF models may include a "color by space" map, a "size by space" map, an "orientation by space" map, etc., but never a "color by size by orientation..." map as in Cartesian models. Thus, DNF models avoid the exponentially increasing neural cost of a Cartesian feature space and replace it with a linearly increasing

need for resources per feature dimension. Like Treisman and Gelade's model, dimensions in the DNF model are somewhat independent and thus can yield non-Cartesian similarity behaviors once integrated along the shared spatial dimension. It is possible that DNF models could use this dimensional flexibility to specifically explain asymmetry, alignability effects, and circular dimensions, although this has not yet been tested.

Non-Feature Comparison Influences on Similarity Judgments

Similarity is not entirely determined by feature comparisons. Non-feature-based factors like attentional bias or prior knowledge can both directly affect similarity judgments and can indirectly alter judgments by modulating feature comparisons. Non-feature-based effects are more likely to be task-specific than are feature-based effects. While feature comparison is a universal aspect of similarity models and theories, non-feature effects can potentially be isolated and only observed within one context. If so, these task-specific effects may still ultimately be important for fully understanding similarity judgments, but they are not the most efficient targets for initial model building efforts.

An example of a well-known similarity judgment behavior is the "fast-same" effect: in tasks where there is a binary "same" versus "different" response, participants tend to answer "same" trials more quickly (Farell, 1985; Nickerson 1972). This effect is tied to response format more so than feature comparison, and it is task-specific by definition. The fast-same effect cannot transfer very meaningfully to some other similarity tasks like grouping tasks without explicit "same" and "different" answers or high precision reaction time measures. The fast-same effect is only somewhat meaningful in tasks with continuous metric measures like rating scales, and it is not typically studied with these tasks.

Dimensional Attentional Effects. Other non-featural effects on similarity judgments, however, may be task-general. One class of such effects result from participants paying more attention to one feature dimension or one stimulus object than another. Figure 3 shows one way of conceptualizing attentional modulation of similarity: when attending to a given feature dimension, feature space can be thought to expand along that dimension. When ignoring a dimension, feature space can contract along it (Shepard, 1964; Klaus, et al., 2007; Maunsell & Treue, 2006). Attention is often external to object features, due to factors like task instructions or a participant's motivation for reward. Attention can also result from object features themselves: the salience of parts of a scene (Itti & Koch, 2000, 2001; Theeuwes, Kramer, Hahn, & Irwin, 1998) depends on the objects in it and can alter attentional allocation among those objects. However, even though salience can originate from object features, it is not a deterministic or lawful of a result of feature dimensions like direct comparisons of features, and features others than those of the objects being compared matter.

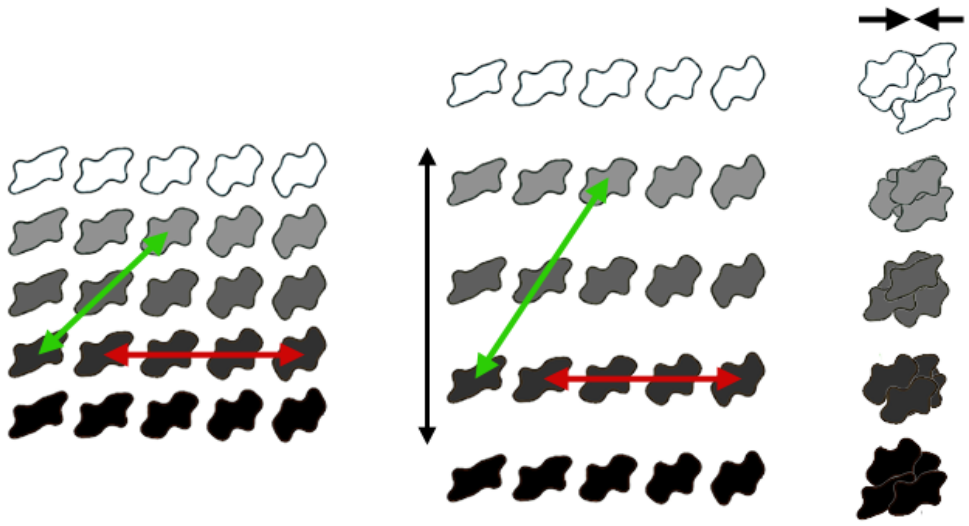

Figure 3: The effects of attentional modulation on similarity judgment. An initial feature space (left) can be attended or ignored more along one dimension than another, effectively compressing or expanding the space (middle) and changing relative object distances/similarities (green vs. red). In an extreme case (right), full dimensional compression can yield clusters that seem "identical" for purposes of a task.

Notably, Tversky (1977) proposed attentional modulation as a solution to the behavioral effects that violate Cartesian assumptions (asymmetry, non-minimality, triangle inequality). The basis of his similarity system was feature-matching: features are listed for two compared objects, and each feature either matches or mismatches (see more recently Navarro & Lee, 2004). This would on its own be essentially a many-binary-features space using a city-block distance measure. Tversky went on, however, to point out that people might attend more to one object in a pair that is more "prominent," whether because it is a more historically important, often encountered, personally preferred, or first mentioned object. Similarity judgments might be weighted more heavily toward the features of this prominent object, whereas features only held by the less prominent object would be de-emphasized in the overall similarity judgment. Features shared by both objects would be treated the same in either direction. This difference explained the asymmetric similarity

relationships Tversky had observed, and thus the critical modulating influence was not itself necessarily based on feature values. Rather, the difference in prominence of objects explained the asymmetric relationship, which may in turn be rooted in non-featural causes.

Others have taken a similar approach. Johannesson (2000) echoed Tversky's feature matching and prominence-based model, achieving somewhat better behavioral fits with a slightly modified version. Nosofsky (1991) also advocated attentional bias to stimuli/objects relative to one another as a solution to problems of asymmetric similarity judgments. Shepard (1964) discussed an alternative but related notion of attending to specific feature dimensions, instead of to objects, unevenly. By shifting dimensional attention rapidly within a task, he proposed, participants could essentially stretch or compact their effective feature spaces and thus show Cartesian violations like triangle inequality throughout the course of an experiment that takes many measurements of similarity. The possibility of biasing a feature dimension rather than an object is a powerful concept for models like Treisman and Gelade's (1980) feature integration model of similarity judgments, or dynamic neural field models, both of which are able to easily modulate feature dimensions independently.

Attention can also account for alignability effects (Goldstone, 1994a; Gentner, 1983; Markman & Gentner, 1993). In this case, when two objects are mostly alignable, attention is drawn away from all of the features (and possibly associated dimensions) that trivially match and toward those few that differ, exaggerating the perceived magnitude of differences. Twin children that differ only in hair color draw attention toward their hair color and encourage observers to perceive them as more different than they perhaps actually are. Alignability is a potentially strong candidate for a modeling target. It is

quantitatively straightfoward and has been replicated and shown to be robust but not tested across an extensive number of tasks.

Neighborhood Density Effects. A consistent factor shown to influence similarity judgments in non-featural ways without reliance on attentional states is neighborhood density: the number of other known objects near in feature space to the objects being compared. A person might know about hundreds of insects that look only slightly different (high neighborhood density), yet not know about anything that looks similar to a giraffe (low neighborhood density). Some researchers have proposed that a feature space might be warped dynamically by neighborhood density, tending to "expand" in regions more densely populated by neighboring exemplars, leading to behavioral ratings of exaggerated dissimilarity (Krumhansl, 1978; Love, Medin, & Gureckis, 2003). An expanded portion of feature space alters relative similarity judgments between objects within that portion of the space. The left half of Figure 4 shows an un-modulated feature space populated by a set of several known objects. Objects on the left of the space have higher neighborhood densities than objects on the right. The same feature space warped by neighborhood density expansion might look like the right hand side of the figure: denser neighborhood objects have expanded away from one another, but less dense neighborhood objects were less affected. This can change relative feature distance/similarity judgments. Neighborhood density represents another straightforward, analyzable choice for empirical testing and as a modeling in the event that it is found to be a task-general behavioral pattern.

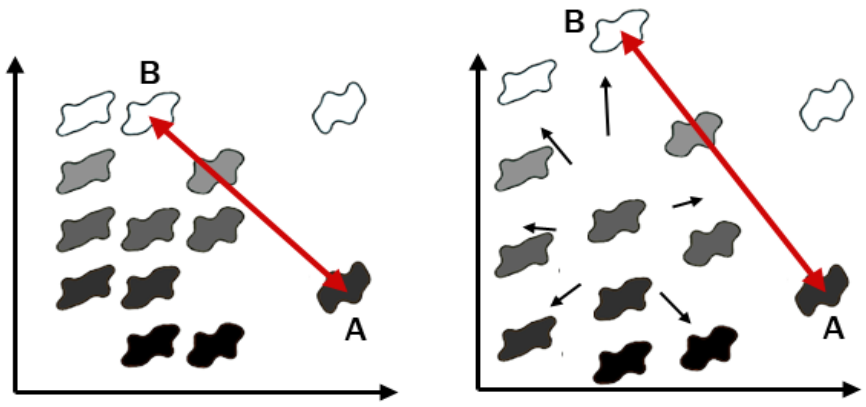

Figure 4. The effects of neighborhood density on feature space. An initial feature space (left) with non-uniform neighborhood density may expand non-linearly in the denser regions (right), such as when a person is reminded of neighborhood density differently by the order of items in a similarity question: "How similar are A and B?" (left) vs. "How similar are B and A?" (right).

Additional Non-Featural Behavioral Effects. A number of additional non-featural behaviors have been demonstrated. For example, gestalt relationships such as arrangements of objects in continuous lines or synchronous movements of objects can emphasize or de-emphasize common features of those objects in a way that affects similarity judgments (Kubovy & van den Berg, 2008). Common category membership can also exaggerate perceived similarity of object pairs. Objects that are similar tend to be members of the same category, however membership in the category itself can exaggerate their perceived similarity even *beyond* the original resemblance. In other words, two similar objects will be judged as even more similar after being placed in a shared category than before, and members of exclusive categories are judged as more different than they otherwise would be (Hund & Plumert, 2003; Hund, Plumert, & Benney, 2002; Recker & Plumert, 2008; Noles & Gelman, 2012). Both gestalt and categorization behavioral effects are task-specific to situations with gestalt object relationships or categorization requirements. Neural similarity judgment processes may be important for fully understanding gestalt perception or categories and vice versa, and these effects are future targets for investigation with the

DNF model. Since gestalt and categorization behaviors do not necessarily derive from core, task-general similarity judgment processes, though, I do not include them in *initial* empirical tests or as initial modeling targets.

Summary of Targets for Modeling

A sizeable set of appropriate behavioral patterns for further empirical and modeling analysis has been identified from the above review of the similarity judgment literature: comparison by feature values, attentional modulation of feature dimensions, perception of circular feature dimensions, violations of minimality, triangle inequalities, neighborhood density effects, and alignability effects. All of these behaviors will be analyzed across multiple similarity tasks to test for likely task-generality. Any behaviors which are found to be task-general across the three representative similarity judgment tasks in this project will be used as targets for neural model fitting and evaluation.

After identifying task-general similarity judgment behaviors experimentally, the DNF model will be fitted to these behaviors and in so doing, will shed light on possible neural level processes underlying similarity behaviors in general. To better establish the degree to which the neural process perspective is unique to the similarity judgment literature, I first survey the set of existing formal, computational models of similarity judgments.

Survey of Computational Models of Object Similarity Judgment

The modeling goal for this project is two-fold: to capture a variety of task-general similarity behaviors computationally, and to do so at a neural process level. These two goals can be conceived as two relevant dimensions of model characteristics, defining a "model space." Figure 5 depicts this model space graphically. Models of similarity judgment can be plotted by their level of implementational abstraction along one axis and

the degree to which they capture task-general similarity behavior versus task-specific behavior on the other axis.

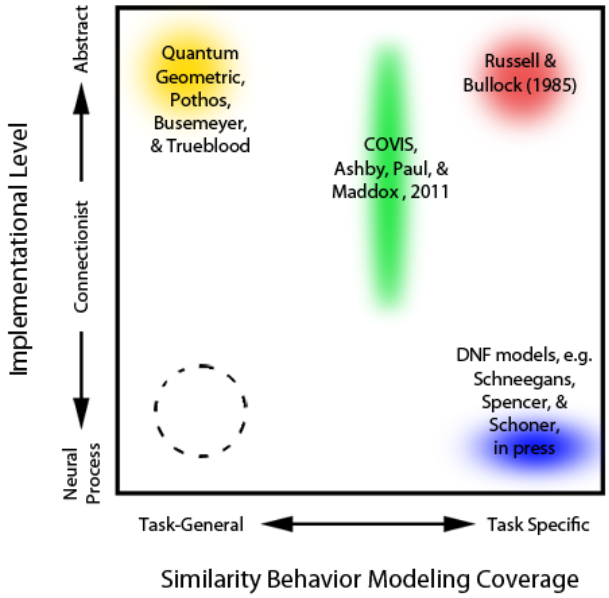

Figure 5: The modeling space of the object similarity judgment literature. Only a representative subset of models are shown here. The dotted circle represents the goal of this dissertation: a task-general neural process model of similarity judgments.

Figure 5 plots only a select few representative models, but many models of similarity exist throughout various regions of this modeling space. Some models are both abstract and address specific behaviors (upper right). A large number of non-neural computational models of similarity judgments have also been developed to account for different combinations of the many commonly observed similarity judgment behaviors (upper left). At the same time, neurally-implemented models that involve similarity exist, but they are primarily designed to capture other behaviors such as categorization. Therefore, these neural models have not sought to capture a wide variety of similarity behavioral effects at once nor to explain task-general similarity processes (mid-figure to lower right).

The combination of characteristics in the lower left of the figure, however—*both* comprehensive capture and neural implementation—has been lacking in models in the object similarity literature. The need for such a model is high: some models of this type are necessary for the most complete and neurally plausible understanding of human similarity judgments and to fully integrate this understanding with related cognitive processes. In seeking to fill this gap, one could consider the alternatives of beginning with models in each of the various populated portions of the Figure 5 modeling space. I discuss these possibilities in the sections that follow.

Task-Specific, Non-Neural Models

Many approaches to similarity actually exist in the upper right quadrant of Figure 5, considering a single task or related tasks where a specific behavior of interest is strongest and/or most convenient to study. For example, Russell & Bullock (1985) addressed facial similarity judgments over developmental time, with tasks altered only minimally for age-appropriateness in order to facilitate comparison of children's and adult's competency. Palmeri (1997) proposed a model specifically focused on explaining the effect of object similarity on learning automaticity of categorization skills over child development. These are important and effective means of advancing and applying knowledge, both before and after consideration of more general similarity processes. These models are furthest from the goal of this particular project, however, to capture task-general neural processes of similarity judgments.

Task-General, Non-Neural Models

Other researchers have applied theories and models of similarity across a broad range of tasks, with an intention of integration across contexts, but without an explicit neural grounding. Kubovy & van Den Berg (2008) reviewed more than a dozen

experiments, and performed three of their own, to converge on an understanding of how perceptually grouped stimuli (such as common movement) affect perceptions of object similarity. Smith and Nelson (1984) performed a battery of diverse similarity judgment tasks to examine differences in children's and adults' perception of holistic versus dimensional similarity.

Gentner and colleagues (Gentner, 1983; Sagi, Gentner, & Lovett, 2012) have applied their Structure-Mapping Engine model of conceptual comparison to a wide array of tasks and behaviors related to similarity, from basic similarity itself to high level analogical reasoning. The structural mapping engine, however, is largely a propositional model couched at an algorithmic level, and it abstracts neural level constraints in a way that does not allow for investigation of the influence of neural processes. Similarly, Tversky's (1977) and Johannesson's (2000) feature matching and object prominence models capture several similarity phenomena but exist at the level of abstract logical set theory.

Some models of similarity explicitly avoid questions of neural process by design. Tenenbaum and Griffiths (2001) implemented a Bayesian model of similarity, which relies on logical, inductive inference according to the Bayes rule as a basis for generalizing known categories to novel objects. The model captures many similarity judgments, but Bayesian processes in general have not been clearly established as neurally plausible (Baddeley, et al., 1997; Brighton & Gigerenzer, 2008; Feldman, 2010), and Tenenbaum's and Griffiths' model in particular makes no attempt at speaking to neural implementation.

Hahn, Chater, and Richardson (2002) proposed a unique model that calculates similarity without strictly quantified feature differences or traditional distances at all. The model is still based on the features of objects and how they compare, but rather than a distance or matching algorithm, the model looks at the number of physical transformations

required to convert one object into another, from a third person perspective. This approach can consider many types of similarity, but is not neurally specified.

Pothos, Busemeyer, & Trueblood (2013) presented a "quantum geometric" model of similarity able to quantitatively capture basic feature comparison effects as well as several classic exceptions to metric similarity, including the "asymmetry" order of presentation effect described above from Tversky (1977). Pothos and colleagues' model is impressive for capturing asymmetry without relying on any modulating factors or parameters outside of their inherent similarity judgment mechanism, but the model is almost entirely mathematically abstract. It employs not only a Cartesian feature space but also a vector projection method for measuring similarity distance in feature space that has no clear neural analogue.

Several models are not neurally plausible, because they implement continuous, orthogonal, Cartesian similarity spaces with exponential resource needs. Pothos, et al.'s 2013 model falls into this category, as do Shepard's (1987), and Krumhansl's (1978). The latter two models introduce modulations and factors to capture more behaviors than early models, but without addressing the problem of neurally implausible resource requirements. Nosofsky's (1986) generalized context model (GCM) is an abstract statistical categorization and similarity model focused primarily on comparisons between whole groups of object exemplars. Nosofsky's model is capable of capturing a number of diverse similarity judgment behaviors, but is still dependent upon an expansive multidimensional feature space.

Task-Specific, Neural Models

Several models address the neural implementation dimension of the modeling space outlined in Figure 5, but do not also achieve comprehensive capture of similarity

behaviors. In particular, a number of proposed connectionist models of object categorization have some degree of neural implementation. These models involve similarity processes to various degrees, but none has been focused squarely on similarity and an intent to capture a comprehensive list of major similarity judgment behaviors. For instance, ALCOVE (Kruschke, 1992) is a three layer connectionist model. The input layer features individual nodes for each considered feature dimension, with activity strength corresponding to feature value. These connect to a hidden layer by dense, pre-determined connections representing a city block similarity rule that scales exponentially. The hidden layer is connected to response category nodes with learned weights between the hidden and response layers. ALCOVE explicitly represents similarity in its first set of connections, but the similarity system is rigidly defined according to a basic, feature-comparison framework. Nuanced behavioral similarity effects like neighborhood density or violations of minimality are not addressed, since the similarity rules instead serve as one of several components used to capture more downstream categorization behaviors. ALCOVE's hidden layer, if scaled to naturalistic stimuli and dimension numbers, may become implausibly large, similar to a Cartesian feature space. This is not as guaranteed as in an explicitly Cartesian model, because heuristics might be employed to keep numbers realistic. Additionally, tying feature value to activity level on a node makes the representation of circular dimensions difficult, since activity cannot "wrap around" from high firing levels to zero in a continuous way.

COVIS (Ashby, Paul, & Maddox, 2011) is an example of a neurally implemented model with partial inclusion of similarity processes that could potentially be adapted to a deeper, comprehensive neural model of similarity judgments. COVIS was designed as a model of object categorization. It is a two part model—one half is not neurally detailed and

performs rule-based, propositional logic based on dividing categories by lines or planes in feature space. The second half is a neurally specified connectionist network that more slowly (but powerfully) learns associations between sensory cortical representations of objects and nodes representing category decision components in the striatum. The model chooses whichever system is more appropriate for a given point in category learning. When the propositional system is used, the model's neural implementation is to this point unclear.

On the whole, COVIS is designed to do a sophisticated job of capturing category learning, and similarity exists in the model as a component toward this end. In the propositional system, objects on the same side of a categorical dividing line are inherently more similar, but this is conflated with category membership. In the associative system, similar things end up in categories together due to sharing features, but there is no account of similarity behaviors in the form of explicit judgments or a clear way to read them off of the system.

Love, Medin, and Gureckis' (2004) SUSTAIN model and Goldstone's (1994a) SIAM models are connectionist models for learning categories and relationships. SIAM focuses on categorization over shorter timescales (individual scenes and small groups of objects), whereas SUSTAIN focuses on longer-term conceptual learning (learning whole taxonomic categories). Both models capture the concept that similarity can depend both on alignability of objects as well as on basic feature matching and comparison. Both models also form category structures "on the fly," recruiting computational resources only as needed, which importantly reduces the danger of neural implausibility that exists in more rigid Cartesian feature space models. Both models also allow for attentional modulation of dimensions, and the authors of SUSTAIN in particular explicitly address this possibility (referring to it as "contortion").

These models are some of the most promising available models as starting points for this dissertation's goal of a neural model that can capture fits of task-general object similarity behaviors. Each captures a set of similarity behaviors, and each as means to potentially avoid the implausible resource demands of Cartesian feature spaces. Ultimately, however neither SUSTAIN nor SIAM extends deeply into neural processing details such as real time memory stability, details of visual object processing.

Harris and Rehder's (2011) KRES model is a connectionist categorization model similar in core architecture to ALCOVE. KRES improves on ALCOVE by allowing for a large number of nodes per feature dimension, enabling it to more easily accommodate circular dimensions in a plausible way. KRES also treats individual dimensions more like multiple single dimensional spaces. This approach plausibly avoids high dimensional space exponential resource requirements in a way similar to DNF models—by adding resources linearly per dimension rather than exponentially. Furthermore, KRES takes into account prior knowledge as a means of distorting new category learning, which could potentially allow it to capture key similarity effects like neighborhood density, which rely on integration of objects recently seen into current similarity judgments. KRES is also more neurally detailed than many of the other models discussed so far in this section, and may be as capable as DNF models in the potential for capturing real time neural dynamic processes, although KRES is not as well inherently connected to other neural level cognitive processes captured by DNF models.

Dynamic neural field (DNF) models provide a good starting point to pursue the goal of a fully neurally implemented model of object similarity judgments. DNF models employ plausible neural representations and neural interactions at all levels, and go beyond connectionist architectures to mirror gross organization of the brain in dorsal and ventral

processing streams and in spatially-based integration across feature dimensions (Spencer, Thomas, & McClelland, 2009). DNF models are able to capture real time processes, which, with the exception of KRES, is not a level of detail other existing similarity models are built to explore. DNF models have not achieved quite the breadth of similarity behavior capture as SUSTAIN, but it has explicitly captured change detection behavior and thus demonstrated feasibility for this project (Schneegans, Spencer, & Schöner, in press). DNF makes up for few prior accounts of similarity behavior with characteristics including neural processing details and the potential ability to integrate similarity processes with other cognitive processes such a perception, attention, and working memory.

Overview of the Dissertation

I have identified a set of canonical, possibly task-general, behaviors from the similarity judgment literature, and outlined reasons why a DNF model is a good candidate to capture these effects at a neural process level. Before the model can be adapted to fit these specific tasks, however, the target behaviors must be tested for their task-generality.

In the following chapter, I outline a set of three diverse similarity tasks, a common set of stimuli to be used across them, a common experimental manipulation primarily aimed toward probing neighborhood density, and a common set of analyses for each task. In the three subsequent chapters, I provide detailed methodologies and results for each of these three tasks in turn. In the final chapter, I will return to the model to adapt, fit, and test it against that subset of behaviors shown to be task-general from the empirical results. I will then discuss theoretical implications in the final chapter.

CHAPTER 2

COMMON METHODOLOGIES

In order to most efficiently and reliably adapt and fit a neural model of object similarity judgments, it is important to know which similarity judgment behaviors are common across a variety of tasks. To this end, I ran three different behavioral similarity judgment tasks: a pairwise ratings task, a speeded pairwise binary same/different task, and a multiple-item spatial arrangement task, on the same set of artificial stimuli. I have analyzed behavior for a set of candidate effects that might plausibly be task general, including MDS analysis for degree of feature-comparison-based judgments, tests for awareness of circular feature dimensions, tests for violations of minimality and triangle equality, analysis of alignability influences, and experimental manipulation and analysis of neighborhood density effects. Effects of any kind that are common to all three tasks are taken as symptoms of task-general processes of object similarity. To some extent, the nature of these processes can be interpreted from behavioral results alone, and this will be discussed in each of the following three experiment chapters. Behaviors that persist across tasks also serve as robust targets for computational modeling, and a comprehensive neural model of object similarity judgments is the primary aim of this dissertation.

In this chapter, I describe the basic structure of each empirical task and how the three tasks work together to cover a range of different similarity judgment contexts. I then introduce the common set of stimuli used across tasks and the common set of analyses applied across tasks.

Three Similarity Tasks

The purpose of conducting three separate tasks was to identify breadth and universality of similarity judgment behaviors across tasks. This required the tasks to be significantly different from one another. For instance, where one task introduces a timing constraint, another should be untimed, and where one task allows extensive deliberation, another task should make slow and/or conscious strategies difficult. As much variation in task characteristics should be employed as possible while still allowing tasks to be easily compared to one another, analyzed in the same ways, and relevant to real world similarity judgments and existing literature.

In Table 1, I outline key characteristics along which my tasks differed. The pairwise ratings task is the most commonly used task in the similarity literature. In this task, participants rated each possible pair of objects in a set by their similarity, one at a time, on a scale from 1-9. The simplicity and open-endedness of the task make it widely applicable, easy to implement, and therefore ubiquitous. Pairwise ratings are also almost guaranteed to show several known behavioral similarity judgment effects, since the task has been used in canonical studies in the field. This includes key test behaviors in this project (Tversky, 1977; Shepard, 1987; Shepard 1964; Krumhansl, 1978).

Table 1: Characteristics of the three behavioral tasks. By varying along a number of dimensions, the tasks provide strong tests of generalization. Effects of similarity that persist across all three, despite their many differences, are effects that are probably driven by general, underlying processes of similarity.

Task Characteristic	Pairwise Ratings	Pairwise Same/Different	Spatial Arrangement Method
Participant's ability to intentionally strategize	Moderate	Very Low	High
Time penalty	No	Yes	No
Number of objects visible at a time	2	2	16
Are there correct answers?	No	Yes	No
Constraints on geometry of responses	None	None	Must fit 2-D workspace
Do participants choose judgment order?	No	No	Yes
Precision of responses per judgment	9 ratings steps	2 options, same/different	up to 707 pixels

As shown in Table 1, the pairwise ratings task was not extreme in any task characteristic relative to the other two comparison tasks chosen for this investigation. It served here as a baseline task, matching one other task or falling moderately in between the other tasks along the relevant characteristics. Relative to the pairwise ratings task, the speeded pairwise same/different task was faster paced, coarser grained, and more spontaneous. The spatial arrangement task, by contrast, was slower paced, finer grained, more deliberate and strategic with a greater awareness of context. More detailed descriptions and discussions of all three tasks and their unique characteristics are provided in the following sections.

Pairwise Ratings Task

Pairwise comparison using a ratings scale is by far the most commonly used similarity task, to the point of serving as an obligatory benchmark against which to validate other measures of similarity (Perry, Cook, & Samuelson, 2015; Hout, et al., 2013;

Goldstone, 1994b; Lee and Navarro 2002). Isolated pairs of objects are shown to participants, who rate them based on their similarity using a provided numeric scale, in my case a 1-9 scale labeled “least similar” to “most similar.” In my task, participants clicked a scale on a computer monitor. All possible pairs of individual stimuli in a given test set were presented at least once to accumulate an overall picture of perceived similarity relationships.

The pairwise ratings task is capable of capturing a wide variety of similarity effects, and it is not known for producing characteristic, task-specific effects. The task is also unconstrained in terms of possible responses—participants are free to show any pattern of ratings across pairs, from rating every pair identically, to complete randomness, to showing asymmetry effects and other violations of Cartesian feature spaces, to showing perfect Cartesian organization, to anything in between. Participants are usually not (and were not in my version) given any feedback or information about what patterns they will be shown or suggestions about what patterns they should judge similarity by, other than a description of the ratings scale. The generic nature of the task makes for an excellent starting point in searching for task-general similarity effects by confirming that the effects can be seen with the chosen stimuli and by confirming that the analyses used can successfully detect them under basic conditions.

Although the task is relatively unconstrained, it does impose *some* minor constraints that could influence participants’ behavior. These can be effectively controlled, however. For example, only one pair of objects is viewed and judged at a time, and thus participants are not able to decide in what order they sample the stimuli, yet they may have been influenced by the order in which the pairs were presented. This constraint can be reduced by random object order between participants and by repeated testing of each object

pair—any extreme ratings due to the random order of pairs is then softened by averaging them with duplicate trials that appear in a different position in the order. Due to being a pairwise task, another constraint is that the task lacks an immediate reference to the full stimulus set of other objects being tested. This can constrain the information available to participants in the first few trials of an experiment, since participants may not have had an appropriate reference frame or sense of scale yet for the stimuli in the experiment on which to base similarity judgments. In my particular instance of the pairwise ratings task, I address this problem by exposing participants initially to a sample of all of the items in the stimulus set. I included an exposure to each item as an initial phase in all three tasks used in this dissertation.

Overall, the pairwise ratings task provided a neutral task environment along all parameters included in Table 1. Participants had a reasonably precise ratings scale to express relative similarity, while the other two tasks used rougher or finer-grained responses. The pacing of the task was moderate, with no time limits but also short, simple trials that moved quickly. The task also invited participants to spend some effort deliberating, due to the untimed trials and many ratings steps available, but calculated patterns of judgments were limited by seeing only two items at a time. The task was largely unconstrained, with no requirements placed on participants' responses such as rules for correct answers or any instructions or feedback about relationships between different pairs' ratings.

Speeded Pairwise Same/Different Task

Binary same/different tasks are common in the literature (Belke & Meyer, 2002; Bindra, et al, 1968; Farrell, 1985; Johnson, Spencer, Luck, & Schöner, 2009) and superficially similar to the ratings scale task described above, but instead of a continuous

scale of different ratings, participants respond with only two options, “same” or “different.” In my version of the same/different task, participants responded using computer keys. There was a correct answer to every trial in this task. “Same” pairs of objects were those that were the same in any way, and “different” pairs had to be different in every way. The task included feedback to reinforce these rules after every trial.

To compensate for lower statistical power of a binary response as opposed to a 9-level rating, more trials of the same/different task were necessary compared to the pairwise ratings task. Participants in this experiment saw between five and six trials of each pair (depending on counterbalancing details covered in chapter 4). Instead of analyzing participants’ direct ratings as in the ratings task, in the same/different task, a participant’s “rating” of how similar two objects are was taken to be the percentage of the duplicate trials for each pair that the participant answered as “same” (regardless of the correct answer). For instance, a pair with four “sames” out of six repetitions was one the participant perceives as more similar than a pair with two “sames” out of six repetitions.

A consequence of defining a participant’s rating across multiple repeated trials (spread randomly through the experiment) was that it was almost impossible in this task for participants to intentionally influence their pattern of results beyond a single trial. Trying to remembering answers to matching trials that occurred a hundred trials ago while also remembering the intervening hundred answers is not feasible. Therefore, any patterns of results from the pairwise same/different task can be interpreted as unintentional, cognitively low-level effects. This differs from the pairwise ratings task, where ratings were transparent and participants could more readily apply an experiment-wide pattern to their answers for the entire stimulus set.

Another consequence of needing to run more trials of the same/different task for statistical power was that trials needed to be faster to fit into the same experiment duration. For this reason, as well as to add additional diversity between tasks, I included a time pressure component to this task. The same/different task was naturally somewhat faster than its ratings equivalent, due to the keyboard input and fewer options, but to ensure quick responses, participants heard an annoying buzzing sound if they took too long on a trial (longer than 1500ms). A speeded component further distinguished this task from the less-pressured ratings task, and reduced participants' ability to consider patterns of responses across trials.

Correct and incorrect answers were an important variation from the pairwise ratings task. The existence of a correct answer is necessary in same/different tasks, due to there existing two prominent interpretations of binary similarity. "Conjunctive" similarity is when objects must match in every way to be called "same" and all other objects are "different." "Disjunctive" similarity, which I chose for this task, is when pairs are the "same" if they match along *any one* dimension and "different" only if they differ in *every* respect. Without instructions to rate according to one rule or another, participants might randomly decide, and rather than a continuous distribution of individual differences like in a ratings task, results might show a bimodal distribution across participants and/or trials. Not only would it be difficult to know which rule a participants was using at any given time, but conjunctive and disjunctive results cannot simply be inverted and collapsed together. They require different analyses and predict different known similarity judgment behaviors (Farrell, 1985). Thus, not defining a correct answer would complicate analysis and further reduce statistical power, since each group would need to be analyzed separately. Correct answers also simply serve as another source of variety between tasks. I

chose a disjunctive rule over a conjunctive one, because disjunctive similarity provides a more even mixture of “same” and “different” pairs in a medium or large stimulus set with few feature dimensions, like the one used in the present experiments.

Overall, the speeded pairwise task served as a faster paced, cognitively lower-level version of the basic pairwise ratings task. Time pressure was higher, precision and need to dwell on each pair was lower, and the instructions specified correct answers. The fact that each rating was distributed over many trials also limited the influence of any intentional patterns of responses other than attempting to score correct answers and constrained the task in a way the ratings task was not. Thus, data from this task can be used to identify effects that might be specific to slower, deliberate, better specified tasks.

Spatial Arrangement Method (SpAM)

My third task was a spatial arrangement method where participants visually indicated their similarity judgments using distance in space as a metaphor similarity. Several objects were presented at once, and participants placed them into a two dimensional workspace such that shorter distances between any two pairs corresponded to greater similarity. Figure 6 shows a SpAM trial. Sixteen items appeared at once on the sides of a computer monitor in two rows of eight. A square workspace in the center served as a spatial metaphor for a two-dimensional feature space. Participants dragged each item into the workspace in any order until they were satisfied that the distance between each pair of objects represented the relative similarity of those objects, with closer pairs being more similar.

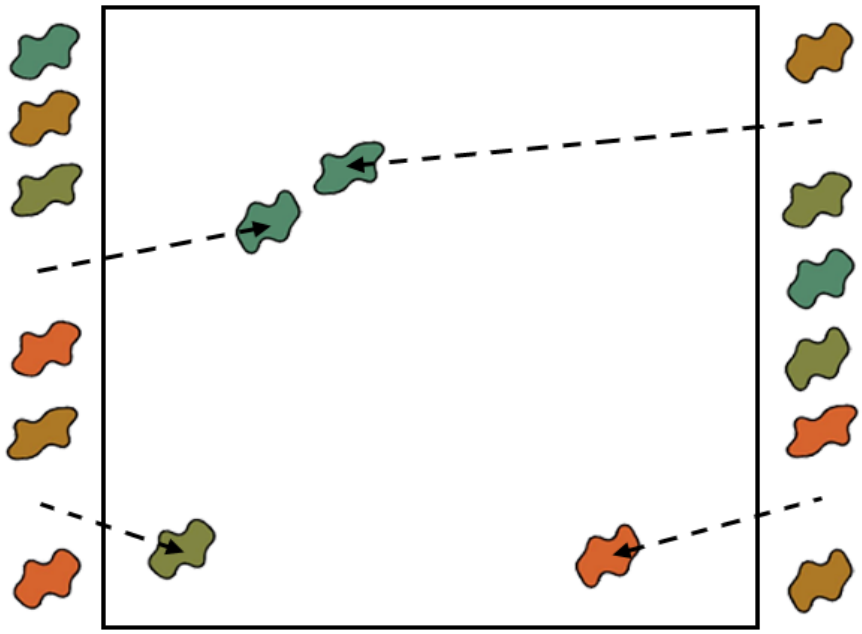

Figure 6: A SpAM trial. Sixteen items are initially arrayed along the sides of the central workspace in random order in two item banks. The participant here has moved four items into the workspace, using space as a metaphor for similarity. Note that arrangements must be two dimensional, and even though only four items have been moved, six pairwise relationships have been defined so far.

Goldstone (1994b) developed SpAM for efficient collection of similarity judgments, since each object placement (beyond the second) implies multiple similarity judgments. The task is commonly used (Perry, Samuelson, Malloy, & Schiffer, 2010; Hout, Goldinger, & Ferguson, 2012; Kriegeskorte & Mur, 2012; Jenkins, Samuelson, Smith, & Spencer, 2015), but it is not as popular as pairwise tasks. SpAM is useful for the present research project, because it offers unique task characteristics to help identify general, underlying similarity effects.

Unlike pairwise tasks, participants could see the full stimulus set at all times in SpAM. The time pressure was also the lowest of the three tasks, precision the highest (individual pixels of movement), and the visual complexity and multi-step process invited careful contemplation. Together, these factors place SpAM as opposite the same/different

task along the continuum of intentional, strategic behavioral tasks. SpAM is also unique in that responses are constrained to a particular geometry.

In both pairwise tasks, participants were free to give responses that imply any feature space, Cartesian or otherwise (such as objects pairs that either do or do not yield equal similarity judgments depending on order).[5] In SpAM, however, judgments must physically fit a Cartesian 2-dimensional space that literally exists on the computer monitor. This provided some task variety, but interpreting it requires caution. Some effects like triangle inequalities are mathematically *impossible* to observe in SpAM, due to the mandatory 2-dimensional spatial arrangement. On one hand, this does mean that triangle inequality cannot be a strictly task-general similarity effect, since SpAM is an established similarity task and cannot always exhibit triangle inequality. On the other hand, this type of mandatory 2-dimensional constraint is difficult to imagine in many natural similarity judgments in the real world, and the task-specific impossibility of the behavior may therefore not be theoretically important or imply anything significant about similarity judgment processes in the brain. By contrast, other between-task distinctions like fast-paced versus slow-paced tasks or deliberative versus pressured tasks are more useful: both describe natural contexts for similarity judgments and neither mathematically determines anything about which behavioral effects will or will not generalize. For these reasons, I still evaluate violations of minimality and triangle inequality in upcoming chapters and in the neural model, despite the caveat that they are not absolutely task-general.

[5] In the same/different pairwise task, correct and incorrect answers heavily discouraged judgments that implied non-Cartesian, odd or disorganized feature spaces, but such judgments were still *possible*.

Another potential complication of SpAM is that it has a task-specific tendency to emphasize conceptual and semantic relationships over perceptual ones, unlike pairwise tasks (Goldstone, 1994b). This issue was fully avoided, however, in my version, since the stimuli used (described in the following section) were novel, artificial stimuli that had no semantic or conceptual content.

Common Stimuli and Stimulus Sets

To quantitatively compare participants' similarity judgments in any given task to the other tasks, to results from the literature, or to modeling outputs, it was necessary to use a stimulus set with quantifiable feature values. An object like a dog might have dozens of feature dimensions—fur pattern, height, weight, breed, friendliness, running speed, etc.—which are hard to explicitly control and quantify. Thus, I used artificial stimuli instead, which varied along well-controlled feature dimensions. The dimensions used—specific types of color hue and shape—have both been studied psychometrically and at the neural level using fMRI, and mathematical differences between feature steps are known to be well-matched to psychologically perceived differences.

The full set of stimuli is shown in Figure 7. The first dimension was a trigonometric outline shape defined by a single angle parameter (Drucker & Aguirre, 2009). The second was fill color, which only varied by hue according to CIE l*a*b color space. Both dimensions are circular, but I sampled only half the perceivable range for each. This allowed me to test for circular dimensional awareness in similarity judgments. At the same time, this ensured that opposite corner stimuli (as in opposite corners of Figure 7) still remain more perceptually distant than any other pairs, which avoids the ambiguity of participants potentially comparing pairs in two different "directions" around a circular feature dimension. In a 180-degree sample, there is only one clear directional relationship

for every pairwise comparison. I sampled five steps along each dimension, for a total of 25 stimuli.

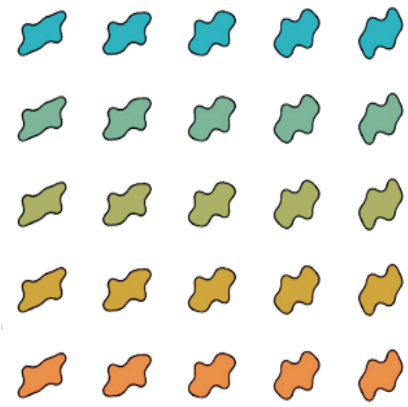

Figure 7: Stimuli used in behavioral experiments. Stimuli vary by a one-dimensional shape parameter and by color hue. Both dimensions are sampled across 180 degrees of their full circular dimensions.

All participants were introduced to the full set of 25 stimuli in an introductory exposure phase of each of the three experiments, but no participants worked with the full set of stimuli in test trials. Instead, each participant made similarity judgments for sets of 16 items. Which 16 items a participant worked with depended on which of two experimental conditions they were assigned to.

The two conditions manipulated the degree of neighborhood densities of objects. Although a fully represented, square grid of stimuli is a common stimulus setup in the similarity literature (Hout, Goldinger, & Ferguson, 2013; Kriegeskorte & Mur, 2012; Little, Nosofksy, Donkin, & Denton, 2013; most experiments utilizing Gabor patches), grids do not offer interesting variations in neighborhood density.

Therefore, I divided participants in all three tasks into two stimulus conditions that changed neighborhood densities: a “square” condition where participants saw stimuli from a typical grid in feature space and an “L” condition where participants saw stimuli from an

"L" shaped pattern in feature space. Figure 8 illustrates the two conditions. Objects in the "L" condition had on average lower neighborhood density than in the square condition due to an "L" being a longer, thinner shape. In order for both to have 16 items, the square condition samples only a 4x4 grid.

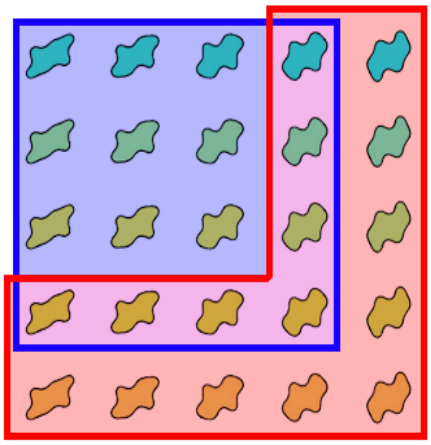

Figure 8: Experimental conditions. Square condition participants worked with stimuli from the blue region of the full stimulus set. "L" condition participants worked with stimuli from the red region. Some stimuli (purple) were seen by participants in both conditions. Both conditions include 16-item subsets of stimuli.

The conditions also had the side effect of testing separation of dimensions. The "L" shape separates out the two feature dimensions along the two "arms" of the "L." One arm varies mostly along color, the other along shape. The square condition does not separate or highlight either feature dimension in any subset of its stimuli. This could have various effects, which were difficult to predict, but would likely show in MDS visualization. One particular possibility was that the manipulation would highlight alignability effects, since objects in one arm of the "L" are not very alignable with objects in the other arm. There is no sharp cutoff in alignability, however, in groups of objects in the square condition.

The two feature dimensions used in this stimulus set are cognitively separable (Shepard, 1987). Because of this, a city-block formula was used in all situations where

distances along both dimensions were integrated into one overall distance for analysis (Shepard, 1987).

Common Analyses

A fundamental requirement of testing for the task-generality of behaviors is the ability to analyze the data and detect similarity in behaviors across all three tasks in the same way. Below, I first explain how each task's output was standardized to a common format, then I outline a series of analyses that I applied to the results of each task.

Standardizing Results

In the pairwise ratings task, the native answer format is a continuous similarity rating. This only needed to be inverted to obtain the standardized dissimilarity output format that is most useful for MDS and other analyses. (10 – [a 1-9 similarity rating] = [a 1-9 dissimilarity rating]). In the pairwise same/different task, I repeated each pair at least 5 times for each participant. I obtained a participant's rating of dissimilarity for a pair by taking the proportion of repeat trials for that pair to which they answered "different." For instance, if a participant saw the same pair five times and answered "different" twice, I took his dissimilarity rating for that pair to be 0.4. This conversion was based on the intuition that a participant's indecision is proportional to the strength of his or her similarity perceptions. In SpAM, I directly interpreted Euclidean distances between pairs of placed objects in pixels as dissimilarity scores.

Within each of the three tasks, after standardizing raw responses using the above methods to all measure continuous dissimilarity, I then independently scaled the responses of individual participants within each task such that all participants within that task had the same average pairwise dissimilarity score. This was necessary for each participant to contribute equally to group averages. I did not scale the three tasks to be equal to each

other, however. Tasks were only compared to one another on measures of internal ratios and relative relationships. The units of measure in the tasks were therefore unimportant analytically and also held no external theoretical meaning (e.g., pixel distance in SpAM is only informative with relation to other pixel distances).

Between-Subject Multidimensional Scaling (MDS) Analysis

MDS is a common analysis of similarity data used to visualize patterns of similarity between sets of objects and to determine the number of dimensions participants are using to make their similarity judgments. MDS takes as input a set of behavioral dissimilarities for each pair of items in a set. Since my scores were standardized to dissimilarity ratings in all tasks, MDS analysis was identical for all three tasks.

The MDS algorithm begins with the same number of items as it is given in random positions in an output space. It then moves items small amounts in random steps, checking after every step to see if the change improved or hurt overall fit to data and throwing the change out if it hurt fits. The exact details of the step change and fitting algorithms are beyond the scope of this document, except to say that the version of MDS I used employed non-metric, rank-order fits and used a city-block distance measure, and an exponential scaling. Non-metric MDS is used when different intervals in behavioral responses cannot be guaranteed to be perceptually equal (Borg & Groenen, 2005). Although my stimuli *were* controlled to have perceptually equal steps, my tasks' response formats were not, so I chose a non-metric MDS analysis. City-block distance is appropriate for my stimuli's relevant feature dimensions—while color features (hue, saturation, lightness) are confusable with one another, and geometric features (orientation, shape) may be confusable, my dimensions of hue and shape can be well separated from one another (Shepard 1984). Exponential scaling is a default choice after Shepard (1987) unless otherwise specially indicated.

An MDS algorithm must be told how many dimensions to use for its output space. Conventionally, the valid choice for a number of dimensions is determined by a "scree plot." The algorithm is run several times using different output spaces with different numbers of dimensions, and the stress value—the MDS measure of badness of fit of a solution—is calculated for each result. The stress values are plotted against the number of output dimensions to form a line graph like the one in Figure 9. The valid choice for the final MDS is the number of dimensions where the line shows a clear "elbow" or sudden and unique change in slope (e.g., Wickelmeier, 2003). An additional criterion is that MDS results should differ significantly from those that would be obtained from random data with the same number of input objects, in order to ensure that patterns observed are not primarily or entirely due to noise. Since MDS operates on rank orders and relative distances, random data for any type of experiment with a given number of items is identical, and standard fit values and variances are known for a set of a given size, such as my 16-item sets (Spence & Ogilvie, 1973).[6]

[6] I additionally verified all numbers derived from Spence and Ogilvie's tables using Matlab's MDS functionality.

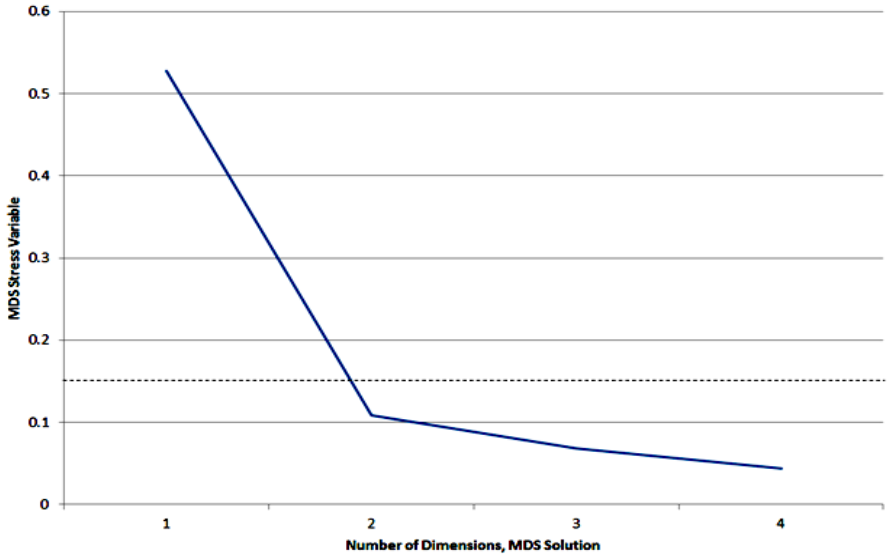

Figure 9: An example scree plot. Here, a set of dissimilarities has been input into an MDS algorithm four times, with different numbers of output dimensions specified each time (1-4 output dimensions). The algorithm outputs a stress value, plotted here on the Y-axis. The valid number of dimensions is where an elbow is formed by the graph.

A possible caution regarding MDS with my stimulus set is that since MDS assumes a Cartesian feature space, it is not ideal to provide data from stimuli with two circular dimensions (Borg & Groenen, 2005). Circular dimensions cannot be represented in a Cartesian space. Ambiguity regarding which "direction" around a circle to compare stimuli would be particularly problematic in MDS. However, since my feature dimensions are both sampled only from 180 degrees of a full circle, and one direction in each dimension is clearly implied between all objects pairs, the MDS analysis can treat both dimensions as if the semicircular samples are "bent" slightly into straight lines, and any effect on results should be minimal. This being said, there is precedent for using MDS on circular dimensions even with a full 360 degree sampling (Shepard, 1962), and there are theoretical reasons to believe that circular dimensions may result in very specific MDS patterns if perceived circularly. The resulting MDS patterns may not directly represent accurate psychological feature space representation, but *can* confirm that dimensions are being

perceived circularly (Shepard, 1985). I will discuss this latter point in greater detail in the results of the first MDS analysis in the following chapter.

Individual Multidimensional Scaling Analysis

In addition to MDS solutions for average dissimilarity across participants, I also analyzed individuals' MDS solutions. Ultimately, similarity judgments occur at the individual level, and a general model of similarity needs to capture the range of common individual behaviors, not just an averaged group result. Average/group MDS solutions may be representative of a single strategy shared by every individual, but more likely, they represent a blend of two or more distinct individual patterns. These individual trends should be captured by similarity models and may even be more theoretically informative than the group result.

Two tests were performed on individual MDS results, one subjective, one automated. First, general "orderliness" of each individual MDS result was judged by raters. Four raters blind to task type evaluated each MDS output for the pattern of items output by the algorithm. Raters quantified overall orderliness of each pattern on a scale from 1-5 corresponding to random/undecipherable up to clear, distinct patterns with no objects out of place. Raters were told to not expect or prefer any particular type of pattern (grids, lines, triangles, clusters, etc.), but only rate patterns on their intentionality, consistency, and orderliness. The results of these ratings were used to drop from analysis those participants who all four raters unanimously judged as having "undecipherable" judgment patterns. Across all experiments, only four participants were dropped in this way.

The automated portion of analysis was performed to determine the degree to which individual participants skewed their ratings of similarity by specific feature dimensions. Both the square and "L" experimental conditions were symmetrical by feature dimension,

with equal number of steps and distribution of stimuli along each dimension. Thus, any difference in final judgments when collapsing data across one dimension or the other equates to a dimensional bias that may indicate attentional modulation of feature dimensions. The mere existence of attentional modulation is a behavior I wish to test for task-generality, but if it does exist, I also aim to quantitatively measure its magnitude and use this to fit or improve modeling results. The measure of attentional bias starts with the MDS solution, and then determines the distance between each immediately adjacent pair of stimuli in feature space—for instance, two stimuli that differ by a single feature step in the shape dimension, but match in the color dimension. By definition, all adjacent stimuli like this vary by a fixed amount in one dimension (one step) and not at all in the other. Attentional bias was calculated as the sum of dissimilarities between pairs differing in color divided by the sum of dissimilarities between pairs differing in shape, or vice versa. A ratio of 1 or 0 would describe completely ignoring one dimension, while a ratio of 0.5 would be no attentional bias at all.

MDS solutions in general (group or individual) allow excellent insight into participants' similarity perceptions. Solutions are required to fit into a valid Cartesian pattern, however, and they may therefore be misleading if participants provide non-Cartesian responses, which we expected them to do (Goldstone, 1994b, Tversky, 1977). There is no possible way to display, for example, triangle inequalities in an MDS plot. No graphical representation of a leg of a triangle can be longer than the other two legs. It is therefore important to test for similarity judgment behaviors like triangle inequality using tests other than MDS. This is also true of circular feature dimensions perception, neighborhood density, the principle of minimality, and alignability effects.

Circular Dimension Awareness Test

Participants may have perceived arc-based distances for circular dimensions, or chord-based distances. Figure 1 shows this distinction within each dimension. If chord-based measure fit better to similarity judgment behavior, this is an indication that participants are aware of the circularity of feature dimensions and that this circularity affects similarity judgments.

This question of circular dimension perception was tested outside of the context of MDS, based instead on the standardized dissimilarity scores taken directly from each task. I compared behavioral pairwise dissimilarities with predicted dissimilarities using root mean square error. I performed this test first for arc-based city block distance, then for chord-based city-block distance, and identified the better fitting method.

Specifically, arc-based city-block distance was predicted as [number of feature steps difference in color + number of feature steps difference in shape]. Chord-based city-block distance was the same, but with chord lengths replacing raw number of feature steps. For feature dimensions like mine with 5 feature steps along a 180 degree semi-circle, feature step differences of [0, 1, 2, 3, 4] correspond to chord lengths of [0, 1, 1.83, 2.40, 2.56]. Better fits to chord than to arc-based dissimilarity was taken to imply an awareness of circularity of feature dimensions. Better fits to arc-based dissimilarity would be ambiguous between circular versus linear dimensional perception.

Tests for Tversky Violations

For each task, I ran a set of tests to search for violations of Cartesian feature space assumptions, specifically those of triangle equality and minimality (such as discussed in Tversky, 1977). Triangle inequality is when object pair (A and B) is judged more dissimilar than the sum of dissimilarities of (B and C) + (A and C). In Cartesian space, one leg of a

triangle defined by three points can never be longer than the sum of the other two legs, yet sometimes, participants will judge objects this way when able, such as in pairwise tests. Inequalities can be detected by considering every set of three items in a stimulus set and measuring each for whether one of the dissimilarities is larger than the sum of the other two. The result of this test is a count of triangle inequalities, considered in the context of the total number of object triplets involved. It is impossible for SpAM to show triangle inequalities, but the effect was analyzed in both pairwise tasks, and this provided some evidence of task-generality.

Violation of minimality is most strictly when an object is judged more dissimilar to itself than to a non-identical objects.[7] For this test, I analyzed the standardized dissimilarity judgments of every identical pair of objects and searched for every other pair including that object to see if any such pairs were judged more similar (less dissimilar). The result of the test is a count of instances of minimality violation, in the context of the total number of similarity judgments.

Neighborhood Density Analysis

I tested for neighborhood density behavioral influences by first assigning a neighborhood density to each object: I counted immediate neighbors horizontally or diagonally to each object in feature space for a given experimental condition (square or "L"), and the sum of these neighbors became the neighborhood density of that object in that condition. I then correlated these densities with and the difference between observed and predicted dissimilarity scores.

[7] Minimality violation can also arguably include an identity pair of objects not being rated as similar as possible, but I did not analyze for this type of evidence, since it is potentially ambiguous with other causes, like people simply not using the top end of a similarity ratings scale, for example.

Standardized behavioral dissimilarity data alone was expected to correlate uninterestingly with neighborhood density, simply because neighborhood density was, for example, lower near the corners of a grid of stimuli, and higher near the middle, and these positions were also related to average distances to other objects. By calculating the *difference* between behavioral similarity ratings and predicted similarity ratings instead,[8] predictable factors like position in stimulus feature space cancel out. All that is left in the difference is the *extra* dissimilarity above and beyond that predicted from feature space distances alone. It is this exaggeration (or suppression) of dissimilarity beyond Cartesian feature space predictions that has been shown in past literature (Krumhansl, 1978; Love, Medin, & Gureckis, 2003).

Alignability Analysis

Using the same method as neighborhood density, I correlated the alignability of object pairs with the difference between predicted and observed dissimilarity scores. Alignability I operationalized as the lower of the two numbers of feature steps along color and shape dimensions, or the difference along the most similar dimension. If objects are two steps apart in color and three in shape, alignability is two. If objects match on either dimension, alignability is zero, and so on. Again uninterestingly, if two objects have fewer feature steps between them, they will tend to be judged more similar purely due to being closer in stimulus feature space. However, by using difference between predicted and observed similarity, the uninteresting feature space prediction from input stimulus differences alone cancels out before a correlation is determined.

[8] The predicted values for a given task were simply the city-block distances between objects in the input stimulus feature space.

Limitations to Theoretical Interpretation of Results.

For some of the above analyses described in this chapter, results superficially contradicted the findings of established effects in the literature. It is important to note that the tasks I am using were not intended to rigorously test or challenge specific behavioral effects from any specific previous experiments. For example, none of my tasks matched the procedural details of those used in the past to detect alignability effects. Typical alignability experiments involve explicitly writing down a list of differences in words, whereas all of my tasks used non-verbal similarity judgments. I still analyzed alignability effects in all three tasks, but this served as a test of task-generality, not as a critical replication of, for example, Gentner's (1983) theory or behavioral data. Non-replication in a task that is different from the original task is, however, valid evidence that an effect is *not task-general*, and testing task-generality is the primary goal of the tasks and analyses in the empirical portion of this thesis.

In the following three chapters, I describe three experiments, one that uses each of the three behavioral tasks introduced here. I then evaluate the evidence or lack thereof for the set of similarity judgment behaviors tested by the list of analyses above, and I discuss the theoretical implications of these results in terms of task-general similarity judgment processes. The behaviors found to be task-general will serve as a basis for modeling in chapter 6.

CHAPTER 3

EXPERIMENT 1 – PAIRWISE RATINGS TASK

The pairwise ratings task served as the first of three tasks to be compared to one another to determine task-general patterns of similarity judgments. In the pairwise ratings task, participants were shown all possible pairs of objects one at a time and were asked to provide a 1-9 similarity rating of each pair. Figure 10 depicts the decision portion of a single trial of the task. Stimuli to compare appeared above a visible ratings scale on a computer monitor. Participants clicked a location on the ratings scale to respond.

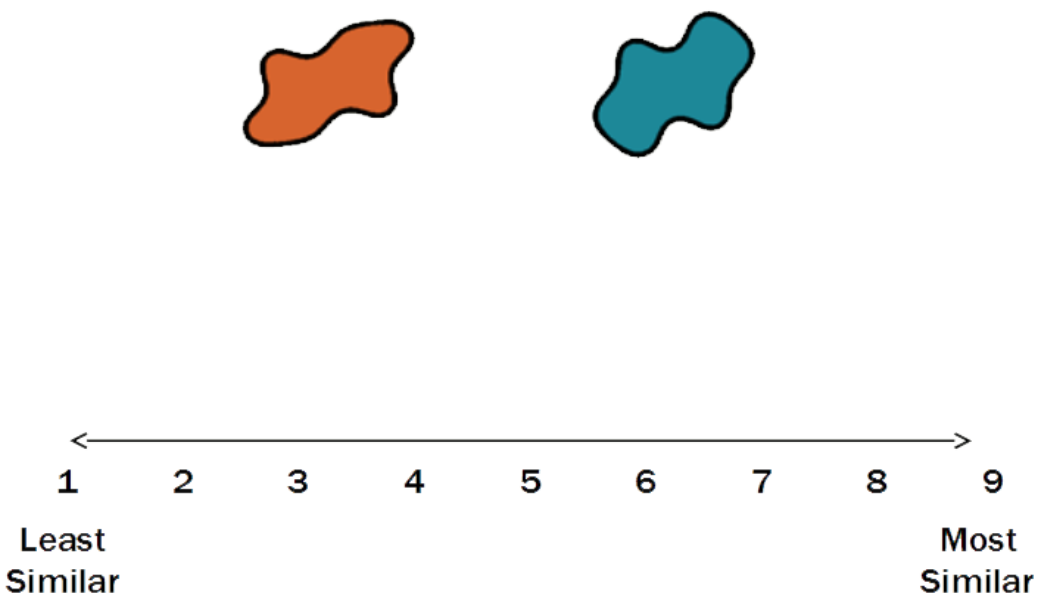

Figure 10: The decision portion of a pairwise ratings trial. Two objects to be compared appear in the top portion of the screen, and a labeled ratings scale appears below. Participants have as long as they desire to click on one of the numbers on the scale. In between object displays, a fixation cross appears in the space between the objects in this figure for 500ms.

Pairwise ratings tasks like this one are the most commonly used tasks in similarity research. Pairwise ratings are simple and straightforward to explain to participants; they are

technically easy to implement in the laboratory; and the task forces minimal constraints on participants' answers. For all of these reasons, pairwise ratings tasks are ubiquitous, and there are more known behavioral similarity effects that have been discovered using this task than any other. This makes a pairwise ratings task the richest opportunity for replicating several meaningful patterns of similarity judgments and verifying that stimuli and analyses used were sufficient to capture known behavioral results, before continuing to test for task-generality in the other two, less standard tasks.

Methods

Participants

Twenty participants were recruited from the pool of an introductory psychology course in a Midwestern town. Three participants were dropped, one participant because the MDS algorithm was unable to converge on a solution to his individual ratings, and two participants because all four raters unanimously judged their individual MDS ratings to have the lowest allowed rating for meaningful pattern or organization of judgments.

Stimuli

This experiment used the set of 25 stimuli described in chapter 2. All participants were exposed to the full set as a preview at the start of the experiment. Participants were then divided into the two conditions described earlier—square and "L" subsets—for test trials. Each participant therefore saw pairs during test from within a 16 item subset for their condition, equaling 136 possible pairs. Participants were shown each pair twice, for a total of 272 trials per participant. The order of the whole set of 272 trials and the order of objects within pairs were randomized for each participant and presented in a single block. Introduction to stimuli at the start of the task, randomization of trials, and redundancy of trials were all implemented to reduce possible bias due to the order of trials.

Procedure

After giving informed consent, participants were directed to a computer terminal with a mouse, keyboard, and Dell 12”x15” 60Hz (1280x1024 resolution) monitor about 20 inches from their seat. All subsequent instructions were text on-screen.

Participants were first instructed to watch the screen passively while the full set of stimuli was shown as a preview at a rate of one second per stimulus, in the center of the screen. Participants were then instructed,

> “In this section of the experiment, you will be shown pairs of objects from the set you were shown at the start of the experiment. Please look at the black + in the middle of the screen when it appears. You will see a numerical scale 1-9, with 1 being least similar and 9 being most similar. Please click on a number to indicate how similar you think each pair of objects is, according to this scale.”

At the start of each trial, a central, black fixation cross was displayed for 500ms. The cross was then removed, and the two stimuli were displayed to the left and right of the cross’ previous position, separated by a total of 7.9 degrees of visual angle of white space. Participants responded by clicking on the rating scale at the bottom of the screen. The rating scale was labeled with “least similar” and “most similar” at the 1 and 9 ratings endpoints throughout the experiment.

Analysis

The analyses described in chapter 2 were applied to this task. To standardize the pairwise ratings data into the cross-task format of a list of pairwise dissimilarities, each rating from the raw data was subtracted from 10. Higher ratings correspond to greater similarity on a 1-9 scale, so the formula (10 – rating) yields a 1-9 *dis*similarity scale instead that is easier to analyze and required for MDS.

Results

Group Multidimensional Scaling

Group MDS solutions used the average dissimilarity ratings of each pair of objects, across all participants, and across both presentations of each pair of objects per participant. These were converted into rank order before the MDS was solved, as part of the non-metric MDS algorithm used.

MDS analysis provides a visualization of the closest fit of pairwise data into a Cartesian space with a specified number of dimensions. The specified number of dimensions needs to be determined before results can be interpreted. This depends first on a "scree plot" of the stress values of MDS solutions with various numbers of dimensions. The scree plot for all three experiments (pairwise ratings, pairwise same/different, and SpAM) is shown in Figure 11. The appropriate number of dimensions is where the scree plot shows an "elbow." This point is where the largest gain in fit is achieved for a given cost in parsimony (higher number of dimensions fitted). Beyond the point of the elbow, additional parsimony yields disproportionately diminishing returns in goodness of fit.

All conditions of all tasks show an elbow at two dimensions, with possibly one subjective case for the Experiment 2 square condition. However, even in this case, there is no *better* elbow, only a potential lack of an elbow. Flatter curves in MDS solutions imply a higher rate of noise, not any conclusions about dimensionality (Spence & Ogilvie, 1973). Thus, the Experiment 2 square condition MDS solution should also be assumed to best fit 2-dimensions: if there is any elbow in its scree plot, it is at two dimensions, and if there is not, then the two-dimensional solutions of the other plots are the best evidence from which to infer a two-dimensional solution for the final condition as well. Overall, then, all MDS algorithms were run here as two-dimensional ones for analysis.

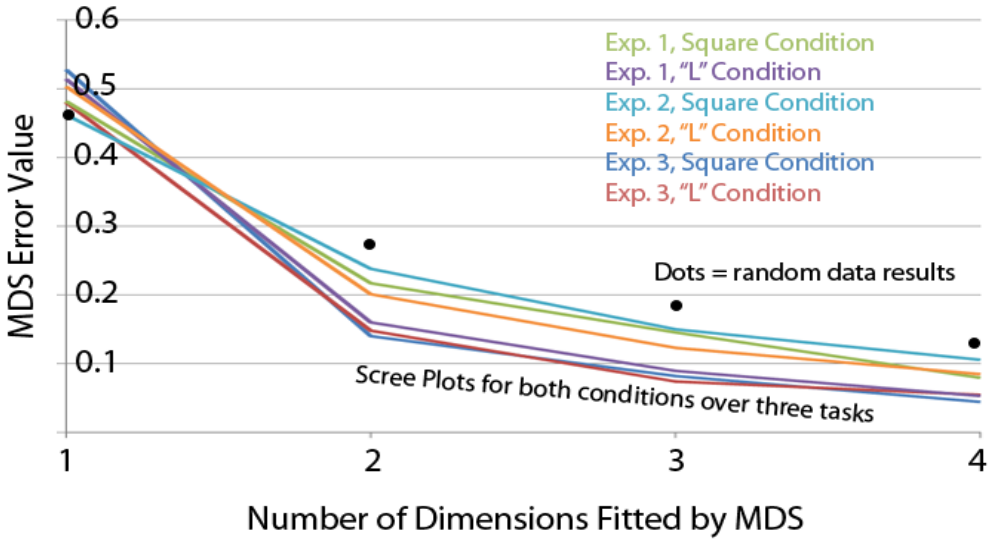

Figure 11: Scree plot for group MDS analysis. This plot includes all tasks and conditions. All plots show a clear elbow at two dimensions, indicating a two-dimensional best blend of parsimony and data fit. All conditions are significantly different than random data fits at two dimensions.

A second check to perform on MDS analysis before interpreting is to compare fits to those of randomly generated data. If a solution is no better than that for random data, then any patterns seen in the MDS solution would probably only be artifacts. Thus, if any MDS solution does not fit actual data better than random data, MDS is not a good analytical tool for that data. To decide on an appropriate statistical test, I considered whether each individual condition's MDS solutions were significantly different from random fits or not, and only considered them analyzable if they were. Also, since a two-dimensional MDS solution was already decided for all conditions, only two-dimensional results need to be compared. Finally, MDS is based on a random starting set of positions for objects, and every run of the algorithm can settle into different local minima: MDS solutions for one set of data can be considered as a group of data points with a variance. Given these conditions, a set of t-tests was considered most appropriate, comparing each set of 2-D MDS solutions on behavioral data to its corresponding set of 2-D MDS solutions

on random data, with any individual condition needing to reject the null hypothesis for MDS analysis to be applied to that condition. All six t-tests (with 50 runs of behavior and random data each) rejected this null hypothesis at $p < 0.0001$. MDS analyses of these tasks and conditions can therefore be assumed to be showing meaningful patterns from similarity judgments.

Once the appropriate number of dimensions was determined and checked against random data fits, the best-fitting MDS solutions with those numbers of dimensions were chosen out of 50 runs. These are shown in Figure 12. Green lines connect objects that share a color, and red lines connect objects that share a shape. Blue lines in the "L" condition indicate the two end pairs of the "arms" of the "L"—one would normally be red and one would be green.

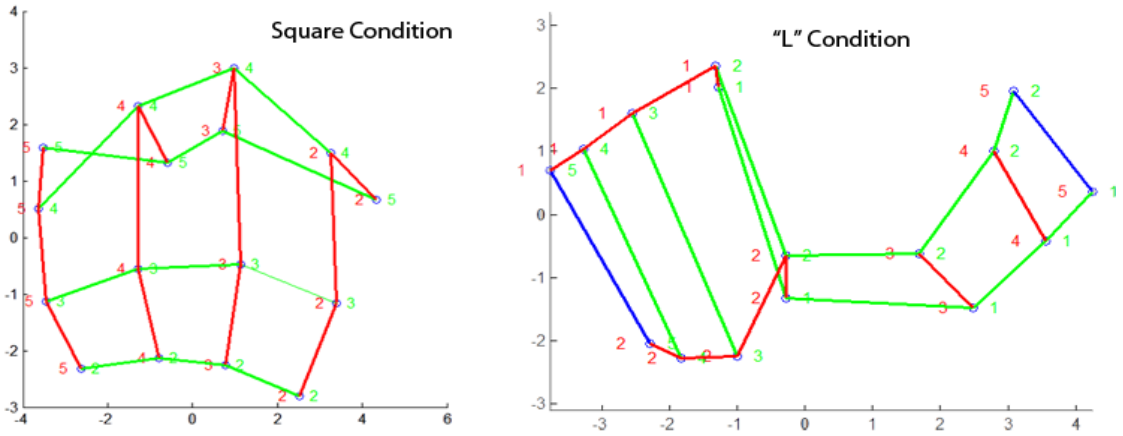

Figure 12: Group MDS solutions for the pairwise ratings task. The square condition shows a clear grid except for two of the color values being confused with one another (green lines flipped or overlapping). The "L" condition shows an "L" shape that has slight curvature, a more obtuse angle than the expected "L," and an attentional shape bias (longer green than red lines).

As seen on the left in Figure 12, in the square condition the results of MDS analysis roughly fit a square shape, indicating that participants were largely driven by direct, classic feature comparison in their judgments. Recall that both feature dimensions of the stimuli are circular, so the curvature seen in this pattern (and the "L" pattern) is not unexpected, although the formal test for curvature for these tasks was not based on MDS results. Colors 4 and 5 appear to have been often confused by participants, causing the top two green lines to wave in and out of one another, but otherwise, the other 19 feature value comparisons were judged in the order expected. In the "L" condition, results are much more distorted, although in two specifically identifiable ways. First, the "L" has been bent to become more obtuse, which means that participants saw the arms of the "L" as artificially more different than is implied by the raw feature values alone. This would be consistent with seeing each arm as an ad-hoc category and accordingly inflating the distance between categories (Hund & Plumert, 2013; Hund, Plumert, & Benney, 2002), but there is not enough evidence from

the data to conclude this for sure. The second major distortion is that both arms of the "L" are noticeably wider along the green lines. This leads to the fat and skinny look of the two arms and can be a result of simply participants being more attuned to shape differences than to color ones. Why such a bias would only appear in the "L" condition and not the square condition is unclear. Overall, participants showed some minor distortions—curvature, a widened "L", and a shape bias in the "L" condition, but otherwise performed according to the mathematical metrics of the input feature values.

Individual Multidimensional Scaling

Individual MDS solutions served two major functions. It is possible that group MDS results show unrealistic patterns that arise purely from averaging other patterns. Thus, the first function of individual MDS solutions was to test whether the overall results fit any individual's judgments. This was done by looking at individual's MDS solutions and verifying that at least some participants' judgments showed similar patterns to those of the overall group. In this experiment, there was a high correspondence between individual MDS solutions and the group average MDS solutions. Figure 13 shows example individual MDS solutions for two of the participants in the task, one from each condition. The patterns match those of the corresponding overall group MDS outputs remarkably well, except that this particular "L" condition participant did not show a shape bias like that seen in the averaged data. It is clear from these results that the overall group results in the previous section can be interpreted as behaviorally robust and not an unrealistic artifact of averaging.

However, not *all* participants match the group patterns. The second application of individual MDS analysis was to allow examination of minority patterns of similarity judgments that differ from the overall group pattern but that are still systematic and reflect

a general model of similarity that may still be valuable to capture in the model. In the "L" condition, the individual shown in Figure 13 matched the group solution in every respect except as strong of a shape bias, and good matches in general were found between individual and group results. In the square condition, three pattern types were observed. The first was the roughly square match to the feature dimensions used as input, seen in Figure 13.

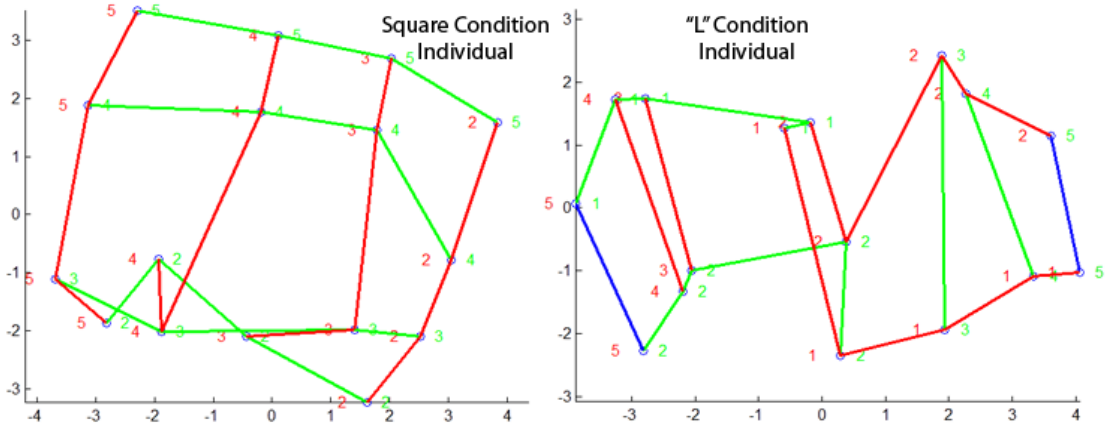

Figure 13: Two individual MDS solutions. One is from a participant in each condition, who correspond to the same patterns as the group MDS solutions in Figure 12. Not every participant matched group results like these, but these data suggest that the group results are behaviorally realistic and are not artifacts of averaging.

The second and third patterns are shown in Figure 14. One is simply disorganized. Participants like these were rare and were dropped from analysis if four out of four raters indicated a score of 1 out of 5 on a scale of intentional-looking organization to MDS outputs for individual solutions. Ten percent of total participants were dropped for this reason—two from the pairwise ratings square condition. The other pattern resembles a thick line or rope. This is the result of extreme dimensional bias. In the example in Figure 14, green lines are stretched much further and show more consistency in relation to one another than do the shorter, haphazard red lines. This participant judged similarity as if objects were grouped almost entirely by their shapes, and almost completely disregarded

color, both in meaningful pattern and in raw amount of impact on similarity judgments. This suggests that the participant attended more to the entire shape dimension than to the entire color dimension.

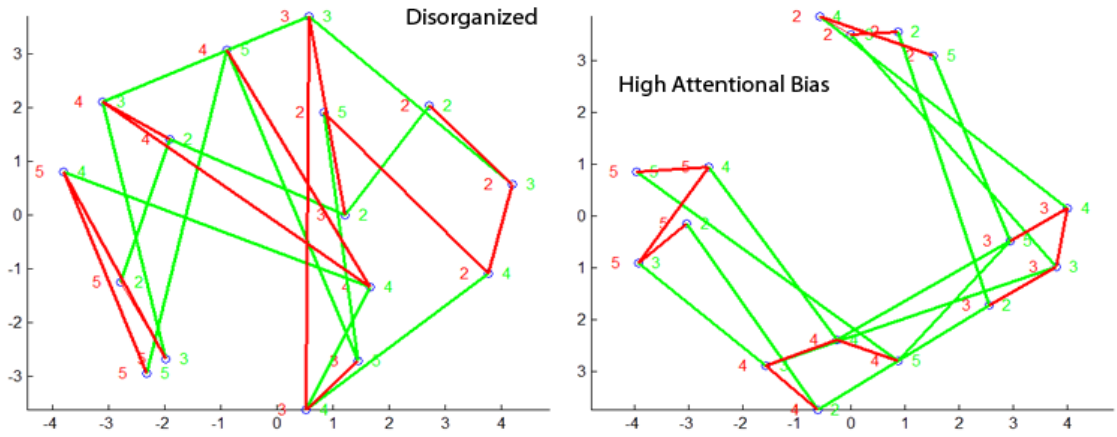

Figure 14: Two additional individual MDS solutions. Both are incidentally from the square condition, demonstrating the two major non-Cartesian MDS individual solution patterns. On the left is an example of a disorganized pattern that was dropped from analysis. On the right is an example of a highly attentionally modulated pattern—this participant is showing a strong bias toward judging objects by their shapes and mostly ignoring their colors.

The latter variable was quantified more formally by comparing the average length of green lines to the average length of red lines as a ratio for each participant. In the square condition, the geometric mean ratio was 0.39 red length : green length (geometric standard deviation 5.33), representing a bias toward shape being influential for similarity judgments (1.0 would be equally balanced). The most extreme bias in a single participant in the square condition was a 0.083 ratio. In the "L" condition, geometric mean bias was 0.38 (geometric standard deviation 2.85), which was also biased in the same direction, and the most extreme individual ratio was 0.064. Some participants did show color biases, the largest of which was a 1.56 ratio (or 0.64 if measured in the opposite, green : red ratio).

Circular Dimension Awareness Test

Participants performing the pairwise ratings task could have judged similarity according to either an arc-based or chord-based distance measure because both dimensions used in these tasks are circular. To test each of these possibilities, a simulation was run to predict the judgments of a hypothetical participant rigidly employing each type of possible distance measure. Each pairwise distance was calculated from the raw numerical feature distances using arc or chord assumptions. Actual behavioral data was then compared to each of these different predictions using root mean square errors by object pair. Group averaged data were used for these comparisons, not individual data. Data were in the same standardized format used as an input to the MDS analysis (analysis was *not* based on the output of the MDS analysis).

Results were in favor of the chord-based metric by a large margin in both conditions. RMSEs in the square condition were 1.4 for chord-based and 1.79 for arc-based fits. RMSEs in the "L" condition were 0.87 for chord-based and 1.70 for arc-based fits. The best fits to behavioral data for both square and "L" conditions were simulations using chord distances within feature dimensions. The fact that chord-based distances were used suggests that participants recognize the circularity of the circular dimensions.

Tests for Tversky Violations

Behavior was tested for triangle inequalities and violations of minimality, as outlined in chapter 2. Triangle inequalities were detected in the pairwise ratings task. This task included 560 different combinations of three objects among the 16 objects in each condition. Each of these triplets of objects can be tested for triangle inequality. Each triplet A, B, C has three distances, AB, BC, and AC. Triangle inequality is when one of those distances is greater than the sum of the other two. Out of the 560 triplets in the group data,

zero inequalities were found for the square condition—all triplets of objects could form geometrically possible triangles. Fifteen inequalities were found for the "L" condition—in other words, fifteen of the triplets of objects had pairwise ratings that could not form any geometrically possible triangle, due to one leg being longer than the sum of the other two legs. Across all triplets of all trials of individuals, there were still zero inequalities found in the square condition, and 805 inequalities found in the "L" condition (out of 5,600 triplets).

Violations of minimality were defined as any pair of non-identical objects that was rated more similar than trials with the identical pairs of either of those objects. Group data in this task showed no violations of minimality. Individual data in total showed 50 violations in the square condition across all trials of all subjects (1,904 total) and 16 violations in the "L" condition (out of 2,720 trials).

The fact that both types of measured violations of Cartesian assumptions were observed in this task implies that feature comparison was not the sole basis of similarity judgments. These results also imply either a non-Cartesian representational space for objects or an external modification of the inputs or outputs of that space, such as an influence of object salience in boosting object representations unevenly.

Neighborhood Density Analysis

Densities moderately correlated with differences between observed and expected similarity judgments. Neighborhood density for each object was the number of neighboring objects immediately adjacent to that object vertically, horizontally, or diagonally in a flat conception of feature space for each condition. An object in the corner of the square condition, for example, has 3 neighbors, one to each of two sides, and one diagonal. An object in the center of the square condition has 8 neighbors, one in each surrounding direction. Density for a pair was the sum of densities of the objects in the pair. Pair density

was then correlated with the difference between predicted and observed dissimilarity for the same object pairs. Predicted dissimilarity was based on the chord-city-block distance based on the circular analysis in the results section above. Correlations showed an r = 0.248 for the correlation between densities and divergences for the square conditions and r = 0.288 for the "L" condition. Both correlations were significant ($p < .05$). Thus, object pairs with more near neighbors in feature space tended to have their ratings of dissimilarity exaggerated compared to pairs with one or both members in sparser areas of feature space. This finding suggests an expansion of feature space in dense feature space neighborhoods.

Alignability Analysis

Alignability effects were weaker. Alignability is the degree to which two objects match on their most similar feature dimension. Non-alignability, then, is the minimum value between color distance and shape distance. Difference between observed and expected similarity judgments is the same as in the neighborhood density test above. Non-alignment correlated with observed-expected differences in the square condition at r = -0.07 and in the "L" condition at r = -0.06. Both correlations were weak but still significant due to high Ns (2176 in square and 2448 in "L" conditions). The negative coefficients mean that more aligned objects lead to exaggerated dissimilarity ratings compared to equivalent pairs with the same distances but less alignment. In other words, two objects at X distance where X is mostly composed of distance along the color dimension (more horizontal or vertical in feature space) will tend to be rated as more dissimilar than two objects, also at X distance in feature space, but with X composed of equal distance along both color and shape (more diagonal in feature space).

Discussion

All tested effects showed meaningful patterns in this task, providing an excellent baseline for establishing potential task-general similarity judgment behaviors over all three tasks in this dissertation.

Group MDS analysis indicated a strong core reliance on basic feature-comparison as a factor in similarity judgments across most subjects and in group means. This was modulated by attention to individual feature dimensions (not just to feature values) in several, but not all, subjects, with a shape attention bias more common than a color attention bias. The "L" condition showed a more obtuse angle between the "arms" of the "L" than is suggested by feature values alone. The cause of this effect is unclear, but it can be tested across tasks for potential task-generality.

Individual MDS analyses corroborated the group results. Some specific participants showed the same patterns as the group averages, meaning that the group results are not merely an artifact of averaging but represent individual behavioral patterns as well. Three types of patterns seen in the group results, none of which are surprising: disorganized patterns were likely from noisy or inattentive participants and were dropped from analysis, feature-comparison patterns closely matched input feature values, and high attentional modulation resulted in some participants having clustered patterns, with one feature discriminated in similarity judgments but not the other. Attentional modulation of an entire dimension could be due to a task being perceived as too difficult. If a participant is overwhelmed, due to judging two circular dimensions at once, for instance, he or she could simply give up on considering one of those dimensions. Attentional modulation, especially when not at an extreme ratio, could also indicate mere personal preference or temporary

salience of a dimension in the task, with participants still recognizing both dimensions but being influenced more by one than the other.

Circular awareness test results were meaningful, though not surprising. Previous studies have shown that people employ chord-based distance measures when working with uni-dimensional stimuli (Shepard, 1962). Here, I verified that this perceptual bias toward chord distance extends to stimuli composed of two circular dimensions at once, in the pairwise ratings task.

Violations of minimality and triangle equality were more surprising. These violations are typically discussed in the context of more complex stimuli than I used, for example, semantically meaningful stimuli, and/or in difficult, high pressure tasks. This task used relatively simple, artificial, non-semantic stimuli. There should also have been no strong bias based on the right or left placement of the objects on the screen (plus, these positions were randomly assigned), and the task context was stable over the experiment. The exact cause of the inequalities is thus unclear, although interestingly, all triangle inequalities occurred in the "L" condition. This effect may be related to the warping of the "L" judgments to exaggerate differences between far ends of the two arms of the "L" as seen in the MDS results. The "L" condition is not simply less "accurate" compared to Cartesian predictions overall, either, because the square condition showed a noticeably higher rate of violations of minimality.

Neighborhood density effects were weakly to moderately strong. Greater neighborhood density correlated with a tendency to exaggerate dissimilarity, the equivalent of "expanding" feature space more in densely represented regions. These results are consistent with neighborhood effects previously observed in the literature and models of

similarity that account for neighborhood density (Krumhansl, 1978; Love, Medin, & Gureckis, 2003).

Alignability effects were very weak, with r values less than 0.1 in each condition, but the effects were statistically significant and in line with known effects in the literature. As objects become more alignable (more closely matching in both *or* just one feature dimension), participants tend to exaggerate their dissimilarity compared to other pairs at the same distance in feature space but not matching as well along the most similar feature dimension (less alignable).

Overall, data from the pairwise ratings task replicated several known behavioral effects in similarity judgments and established new trends as well. To some extent, these findings already reveal clues about general similarity processes. Most notably, the results of the current experiment suggest that the representations of the stimuli underlying participants' judgments include a strong feature-comparison component; that participants are generally competent at dealing with circular dimensions; and that Cartesian assumptions are violated in pairwise similarity tasks. Ultimately, however, the primary goal of the experimental portion of this project is to test which of these behavioral results are *consistent* across task contexts, and therefore which should be taken as the most informative indicators of underlying general similarity processes. Only by comparing to additional tasks addressed in the following two chapters can we establish which of the current results are most relevant to a general model of similarity.

CHAPTER 4

EXPERIMENT 2 – PAIRWISE SAME/DIFFERENT TASK

The speeded pairwise same/different task resembled the pairwise ratings task in that pairs of items were displayed at a time for similarity judgments. Instead of rating similarity on a 1-9 scale, however, participants were given only two response options: "same" or "different." Also unlike in the ratings task, the answer choices ("same" and "different") were explicitly defined for participants, creating an objectively correct and incorrect answer on each trial. Participants were reminded of this by feedback after every trial. The existence of correct/incorrect answers and feedback reduced the likelihood that participants would consciously react to patterns like circular dimensions or neighborhood density, since those variables are irrelevant to correct answers. The same/different task is therefore biased toward showing automatic, low-level similarity effects more so than the pairwise ratings task.

Same/different trials were faster than ratings trials. Participants used the keyboard rather than the mouse, and they were given explicit time pressure in the form of a buzzing sound for taking too long on a trial. Trials on average in this task were about three times faster than in the ratings task (mean 0.84 seconds versus 2.57 seconds per trial). This speed difference served as a further test of generalizability between tasks.

Methods

Participants

Twenty-two participants were recruited from the pool of an introductory psychology course in a Midwestern town. They were randomly assigned to the same two conditions as the prior experiment. Four participants were dropped: one per condition for

failing to meet a pre-determined accuracy cutoff of 70% correct trials, and one per condition due to all four raters unanimously judging their individual MDS ratings to have the lowest allowed rating for meaningful pattern or organization of judgments.

Stimuli

The same set of 25 total stimuli was used as in the previous task, and the square and "L" conditions were the same, still consisting of 16 item subsets. Each pair of objects appeared between five and six times for each participant. This is a higher number of identical trials compared to the ratings task (which had 2 identical trials per pair). This is due to the fact that similarity judgments in the same/different task are computed as a single ratio of "same" : "different" responses across all repeat trials for any given pair. Therefore, having two identical trials for a given pair would be analogous to a 1-3 ratings scale, three identical trials would be analogous to a 1-4 ratings scale, and so on. Five to six identical trials was used in this experiment for a number of reasons. First, five to six identical trials is analogous to a 1-6 or 1-7 ratings scale, which allows more precision in analysis. Five to six identical trials per pair also resulted in the highest number of total trials that participants were able to reliably complete in a half-hour experimental session. This is the unit by which compensation was awarded in the laboratory, and hour-long sessions for a similar task in previous studies proved too long for subjects to remain attentive and accurate.

Overall, each participant completed 728 trials (square condition) or 740 trials ("L" condition). The inconsistent number of trials between conditions was due to the constraints of the same/different task. In this task, similarity had to be defined in order for there to be a correct and unambiguous answer for each trial. A disjunctive definition of similarity was used in this task, which means that "same" is effectively defined as "matching along at

least one dimension." Specifically, the instructions given to participants were, "'Different' pairs are pairs where both objects are different in EVERY way. 'Same' pairs are the same in ANY way." The numbers of "same" and "different" responses under this definition are not equal. Figure 15 shows why this is the case graphically. Depending on the number of stimuli and their arrangement in feature space, the ratio of "same" to "different" answers changes. In the square stimulus set to the left of Figure 15, a given object (black cell) is the "same" as six other objects (green cells) and "different" from nine other objects (red cells). In an "L" stimulus set on the right of Figure 15, this ratio changes. Objects in the "joint" portion of the "L" have a higher ratio of "same" matches, and objects in the arms of the "L" have a lower ratio of "same" matches.

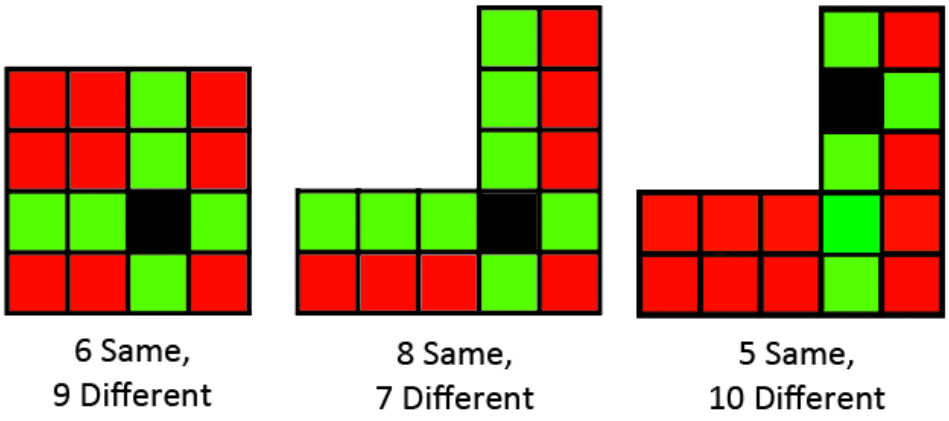

Figure 15: Allocation of same and different answers. Depending on the shape of the stimulus space, there are different numbers of "same" and "different" answers between a given stimulus and all the other stimuli. In order for the task to have the same number of overall "same" and "different" answers, different numbers of extra same trials are needed.

Specifically, for the square pattern, there were 64 "same" pairs and 72 "different" pairs overall. For the "L" pattern, there were 62 "same" pairs and 74 "different" pairs. An equal number of each type of correct answer is conventional in any forced-choice task, however, to reduce possible confounds and biases. Participants therefore received extra

"same" trials until the rate of each type of response was equal for their condition. Starting with five exposures to each pair of objects, this required five (exposures per pair) times eight (to correct from 64 to 72) trials for the square condition = 48 extra same trials and five times twelve (to correct from 62 to 74) = 60 extra same trials for the "L" condition. Five exposures to each possible pair between 16 stimuli yields 680 trials, and when the extra same trials are added, the total becomes either 728 or 740 by condition. The particular pairs that received sixth exposures rotated between subjects.

The order of presentation for the full set of either 728 or 740 trials was randomized for each participant and presented in a single block.

Procedure

The computer station, seat position, stimulus size, and 25-item pre-exposure phase were the same as in the pairwise ratings task. Participants then received the following set of instructions:

> "In this section of the experiment, you will be shown pairs of objects from the set you were shown at the start of the experiment. Your job is to decide whether each pair is a 'same' or 'different' pair. 'Different' pairs are pairs where both objects are different in EVERY way. 'Same' pairs are the same in ANY way.
>
> The 'A' key on the keyboard means 'same.'
> The 'L' key on the keyboard means 'different.'
>
> It is important that you provide your answers as QUICKLY as possible while still being accurate (a check mark or 'X' will tell you if you are correct).

Please look at the black + in the middle of the screen when it appears.

Always answer quickly. You will receive a warning sound if you are slow in answering."

On each trial, a central black fixation cross was presented for 500ms before each trial, as in the pairwise ratings task. The cross then disappeared, and stimuli were presented to either side. Participants responded by keyboard. Which key corresponded to which answer was counterbalanced across participants. If the participant did not answer within 1500ms, a loud, annoying buzz sounded, but the trial still continued until participants responded to ensure data was collected for every trial.[9]

Participants were given feedback at the end of each trial in the form of a large red "X" or a green check mark, centered on the screen, presented for 500ms. Feedback was provided so that the participants would not stray from the instructed definitions of "same" and "different" into other possible definitions—most notably from a disjunctive definition to a conjunctive definition. Feedback also encouraged participants to continuously concentrate on accuracy, helping to ensure that any similarity effects outside of following task instructions were not intentional.

Analysis

Analysis followed the same pattern as the pairwise ratings task. To standardize the same/different data into a list of pairwise dissimilarities, the proportion of "different"

[9] The buzz was created by using Matlab's (version 2009a; MathWorks, Inc.) soundsc() function with the array argument equal to the tangents of 1-500 and with a sampling rate of 5,000.

answers out of all repeat trials with a pair of objects was used as a similarity rating. For example, if a given pair of objects appeared six times for a given participant, and he or she responded “different” on 4 of those trials, the dissimilarity rating for the pair would be scored as 4/6 = 0.67. The ratio ignored the correct answers for a trial, since a participant’s actual response was considered indicative of their perception of similarity, regardless of whether the answer was correct. The maximum similarity score for a pair was therefore 0 and the maximum dissimilarity score for a pair was 1. This conversion captures the intuition that that participants who are unsure about whether a pair is similar or different are more likely to be inconsistent in their judgments as the trial is repeated throughout a half hour task, just as an unsure participant would click on a mid-range value in the pairwise ratings task.

Results

Group Multidimensional Scaling

Group MDS analysis used standardized input. This was a list of converted dissimilarity ratings by object pair, averaged across all participants.

As described in chapter 3, all conditions of all tasks show an elbow in their MDS scree plots at two dimensions, and all conditions significantly differ from patterns derived from random input at this number of dimensions. Thus, analysis for MDS results in this task continued at the two-dimensional level. Figure 16 shows group MDS results.

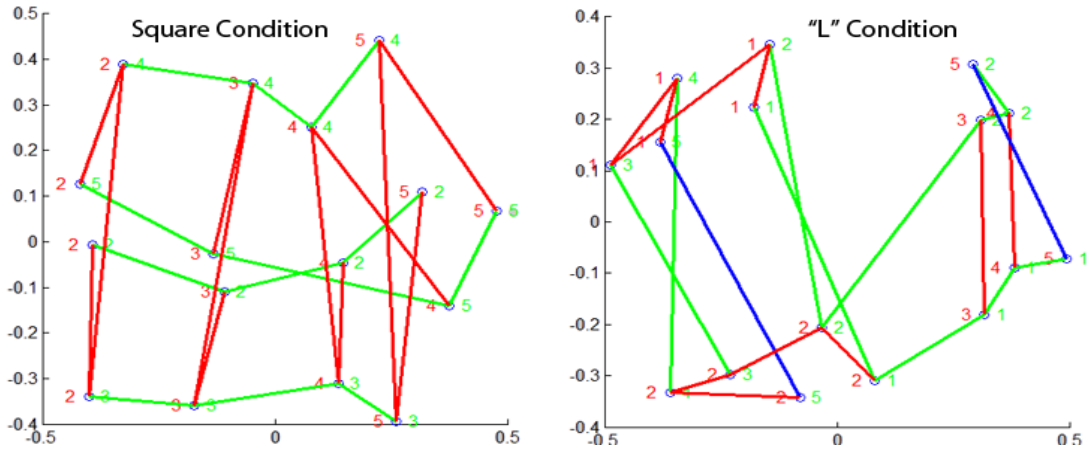

Figure 16: Across-subject MDS solutions for Experiment 2. The square condition shows a grid with confused color values. The "L" condition shows a noisier version of the pattern seen in the ratings task.

It is immediately apparent in the figure that MDS results for the same/different task are noisier and less organized than in the ratings task. This is likely due to two factors. First, the statistical power of the task is lower: five to six repetitions of a binary measure is not as precise as two repetitions of a nine-level measure. Second, the measure of similarity in this task was a non-conscious one, derived from data spread out over hundreds of trials for every data point and dependent upon participant errors to show meaningful patterns. This last point is illustrated in Figure 17 which shows the MDS two-dimensional fit to a hypothetical participant in the same/different task with perfect accuracy.

To the left of Figure 16, participants' judgments from the square condition still fall somewhat in an orderly grid, but with slight confusion between shapes 4 and 5, and much confusion between all colors. The fact that all green lines are mostly horizontal and all red lines mostly vertical implies participants still perceived feature dimensions as orthogonal, but they did not judge similarity according to the intended order of feature values as well as in the pairwise ratings task. The "L" results to the right of the figure resemble those from

the ratings task, but with generally more noise. The “L” shape is still perceptible, and the green lines are still longer on average than the red lines, implying a shape bias. The angle of the “L” no longer appears obtuse, but closer to the originally expected right angle, suggesting a more feature-comparison-driven similarity judgment process on this count.

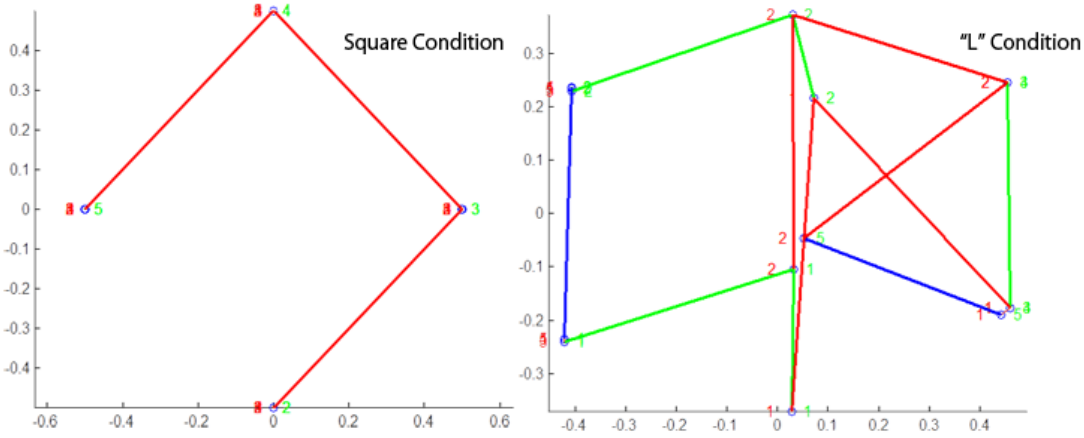

Figure 17: MDS solution to perfect accuracy. The patterns resulting from following task instructions do not resemble the actual MDS solutions, those from the ratings task, or traditional Cartesian predictions.

Individual Multidimensional Scaling

Analysis of individual MDS plots serves as a test of whether the group MDS solutions are realistic at the individual-participant level or artifacts of averaging a set of many other patterns. Figure 18 shows example individual plots from each condition. The square condition participant in the top left of the figure did not show equally grid-like patterns in MDS results as the group solution. Instead, the majority of solutions for the square condition were as in the figure, attentionally-modulated toward one dimension or the other. This is still evidence for feature-comparison based judgment, but not of both dimensions at once as the group solution suggests. Rather, the aggregate grid-like group solution is likely deriving organization in each dimension from different individuals attending to each dimension separately. The “L” condition also showed attentionally

modulated solutions, but a higher proportion of organized individual solutions. The individual to the top right of Figure 18 shows a solution matching the general shape but actually more organized than the group result and closer to the solutions from the ratings experiment. There is a clear division between arms of the "L," a shape bias, and a degree of curvature. The two participants in the bottom half of Figure 18 show other individual representative solutions *not* chosen to most closely fit the group results.

One participant in each condition again was rated as maximally disorganized by all raters and was dropped. In general, no new patterns were observed compared to the ratings task. Fewer, but still some, well-organized participants were found, as many disorganized participants, and a larger number of attentionally modulating participants.

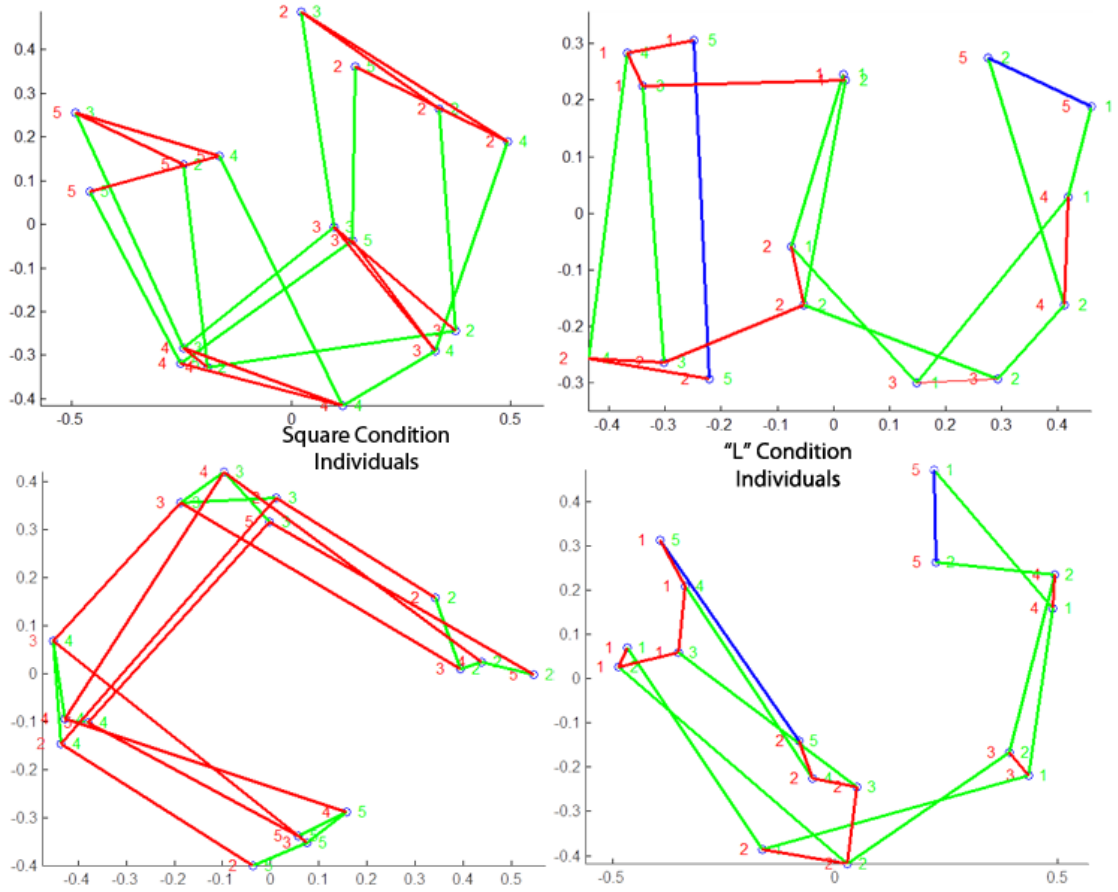

Figure 18: Individual MDS solutions. The top row shows individual solutions closest to group solutions. The bottom row shows two other less closely matching individual solutions. The left two participants are from the square condition, and the right two are from the "L" condition.

Degree of attentional modulation was again tested quantitatively. In the square condition, the geometric mean ratio was 1.84 red length : green length (geometric standard deviation 2.98), representing a bias toward *color* being influential for similarity. The most extreme bias in a single participant in the square condition was a 17.86 ratio, biased toward color. In the "L" condition, geometric mean bias was 0.71 (geometric standard deviation 3.81), biased toward shape, and the most extreme individual ratio was 0.072 toward shape (another participant was biased at 10.75 toward color).

Circular Dimension Awareness Test

This test was performed the same as in the pairwise ratings task. Results were once again in favor of the chord-based metric, though by a lesser margin. RMSEs in the square condition were 0.20 for chord-based and 0.23 for arc-based fits. RMSEs in the "L" condition were 0.16 for chord-based and 0.21 for arc-based fits. The magnitudes of all numbers were lower than in the ratings task, due to the standardized dissimilarity scores in the same/different task being ratios of "different" responses out of one, rather than ratings out of nine. RMSEs have no inherent units and derive magnitude from the format of the data. Again, in this task, participants showed strong evidence of awareness of the circularity of the feature dimensions.

Tests for Tversky Violations

Recall that triangle inequalities were detected in the pairwise ratings task. As in that task, the same/different task again included 560 different combinations of three objects among the 16 objects in each condition, despite the different answer format. Out of all triplets in group data, zero inequalities were found for the square condition, and zero were also found across individual data. Fifteen inequalities were found for the "L" condition group data—in other words, fifteen of the triplets of objects had pairwise ratings that could not form any geometrically possible triangle, due to one leg being longer than the sum of the other two legs. Across all individual data in the "L" condition, there were 901 triangle inequalities (out of 5,600 triplets). These results match those of the ratings task, with no inequalities for the square condition, and inequalities in a moderate portion of triplets for the "L" condition.

Group data in this task showed two violations of minimality in the square condition and none in the "L" condition. Individual data in total showed 12 violations in the square

condition across all trials of all subjects (5,824 total) and 50 violations in the "L" condition (out of 7,400 trials). This is a lower proportion of minimality violations in either condition than in the ratings task. Violations of Cartesian assumptions still suggest, however, that judgments in the same/different task were not dependent upon or consistent with an unmodified Cartesian feature space.

Neighborhood Density Analysis

Neighborhood densities correlated with difference between predicted and observed similarity judgments about as strongly as in the pairwise ratings task. Neighborhood density for each object was the number of neighboring objects immediately adjacent to that object vertically, horizontally, or diagonally in a flat conception of feature space for each condition. Density of a pair was the sum of the densities of the two objects. This was correlated with the difference between the observed group dissimilarity rating for that pair and the predicted dissimilarity using the best fitting distance measure (in this case, chord city block distance). Neighborhood density and observed minus predicted difference correlated at $r = 0.117$ for the square condition and $r = 0.276$ for the "L" condition. Both correlations were significant. This is consistent with denser regions of feature space having exaggerated dissimilarity scores as if they were perceptually expanded compared to sparse regions of feature space.

Alignability Analysis

Alignability effects were moderately strong in this task. Non-alignment is the predicted distance between two objects along a feature dimension for whichever feature dimension is most similar between them, i.e., the lesser of the two distances along color and shape dimensions between a pair of objects. Differences between predicted and observed similarity judgments were the same as in the neighborhood density test. Non-

alignment correlated with differences in the square condition at r = 0.114, and at r = 0.343 in the "L" condition. Positive correlations mean that pairs which align less (greater non-alignment) have exaggerated dissimilarities compared to well aligned objects. These correlations were many times stronger and opposite in direction to both those from the pairwise ratings task and predictions from prior literature (Gentner, 1983).

Discussion

As expected, the pairwise same/different task provided mixed results in terms of matching the findings of the pairwise ratings task. Triangle inequalities, minimality violations, neighborhood density effects, and fits to chord-based distance measures in circular dimensions were consistent between tasks. MDS results were somewhat consistent, still showing good evidence of feature-comparison-based judgments, but less strongly than in the ratings task. The shape bias remained in the "L" condition, but the attentional bias flipped to a color bias in the square condition compared to the ratings task. Regardless of direction of attentional modulation, however, many participants in both tasks were still prone to judging similarity along only one dimension at a time even when two are available and easily discernible. A greater number of participants overall showed strong attentional bias MDS solutions. Solutions were generally noisier and less organized, although this is likely due in large part to the lower statistical power of the task, and to the fact that participants were trying to answer items correctly, which, if successful, would result in solutions that conflict with traditional feature comparison predictions (Figure 17).

Alignment effects reversed in direction and increased in magnitude compared to the pairwise ratings task. This result represented a complete break from the ratings task and ruled out alignability effects as task-general for purposes of this project. The reason for the reversal is unclear—alignability is not traditionally tested using stimuli like these or flat

rating type responses. Degree of alignability may become unstable or ambiguous in comparisons with only two simple dimensions, since how well aligned stimuli are can only be defined along one dimension while affecting judgments through the other. Regardless, this behavioral effect's reversal indicates it is not an ideal target for initial modeling.

The tendency to sometimes collapse across one feature entirely and respond along the other (clustering patterns) is an MDS pattern that seems so far to be common to similarity judgments across tasks and which has a particularly strong effect, despite a variable direction of effect. Attentional modulation in the "L" conditions, in particular, seems to affect MDS solutions almost as much as feature-comparison itself.

It is especially noteworthy that several tested similarity judgment behaviors persisted in the same/different task despite the fact that unlike in the open-ended ratings task, these behaviors *lead to wrong answers*. The instructions for the same/different task indicate a rule for similarity judgments that demands symmetry across dimensions, that does not tolerate neighborhood density warping, and that does not even reinforce every aspect of basic Cartesian-type feature comparison. Showing any of these similarity judgment behavioral patterns in this task could only possibly lower accuracy in comparison to an attentionally balanced treatment of both dimensions that ignored neighborhood density, etc. The persistence of almost every behavior except alignability effects despite the consequences of decreased accuracy in a task with feedback after every trial is impressive evidence of task-generality.

Overall, results of the two experiments so far suggest several ways in which similarity judgments are consistent across the specifics of the task in which the judgments are made. The final task, covered in the next chapter, tested the opposite extreme of task characteristics from the same/different task (Table 1). Whereas the same/different task is

fast and discourages any similarity judgment patterns aside from the instructed correct answer, spatial arrangement tasks are slow, deliberative, and open-ended. Spatial arrangement therefore offers a stringent test of task-generality compared to the same/different task.

CHAPTER 5

EXPERIMENT 3 – SpAM

SpAM is a method of collecting similarity judgments between many objects at once, using space as a metaphor for similarity. Figure 6 depicts a trial of SpAM in progress. Participants were given 16 objects at once for each experimental condition (square and "L") arranged on the sides of a computer screen. They freely moved all of the objects into a central workspace so that distances between each pair corresponded to similarity, with more similar objects closer together.

SpAM was developed by Goldstone (1994b) to be more efficient than pairwise comparisons. Every object a participant places after the second one implies multiple similarity judgments at once, since it has a distance to every other object already placed in the workspace. For example, the ninth item placed in a workspace implies eight similarity judgments at once, since it has a distance to each of the other eight items already placed. As many similarity judgments are implied by 16 object placements in SpAM as by 136 pairwise trials.

The characteristics of SpAM are dramatically different than the pairwise ratings task and especially differ from the fast-paced same/different task. In general, SpAM allows for relatively high participant awareness of the context and consequences of every similarity judgment compared to the other two tasks, and it encourages slow, intentional, structured similarity judgments. SpAM also eliminates any need to rely on memory of the features of a stimulus set as a whole or to rely on memories of previous similarity judgments earlier in the experiment. This is because all placements made during the task are visibly apparent to the participant in the workspace in front of them at all times. For the

same reason, relationships between sets of objects are easier to spot than in pairwise tasks and therefore have a better chance to influence participants' actions.

In addition, every action in SpAM encourages broader consideration of the full set than in pairwise tasks, because each placement has more immediate implications than in the pairwise tasks—moving the ninth item implies eight distances to other already-placed objects at once, and participants were explicitly instructed to consider these multiple relationships. Although some participants may have ignored these instructions, as a group they showed evidence of taking longer to think about SpAM actions, of considering context more, and of employing deeper strategies: raters of individual participant placements' organization did not classify any patterns as indistinguishable; time spent per object placement was longer than pairwise trials (3.45 seconds per placement versus 0.84 and 2.57 seconds per trial in same/different and ratings pairwise tasks); and in several exit debriefings, participants reported explicit logic and "strategy" in their set of placements, unlike in the other two tasks.

SpAM has the unique property of forcing all judgments to fit a Cartesian feature space, which means that SpAM has a more constrained set of possible responses than pairwise tasks. This constraint also makes it difficult to test some violations of Cartesian feature space. For example, triangle inequalities are impossible in SpAM for any given individual, because every triplet of objects must form a literal geometric triangle in the spatial workspace.

Methods

Participants

Twenty-three participants were recruited from the pool of an introductory psychology course in a Midwestern town. One participant was dropped for failing to move any of the stimuli from their starting positions during the task.

Stimuli

The same set of stimuli was used as in the two previous tasks, and participants were assigned to the same square and "L" conditions as in the prior experiments with the same 16-item subsets of stimuli. Participants saw all 25 objects in an introductory phase that matched that of the two pairwise tasks. Participants then saw each of the 16 objects for their condition once, simultaneously, in a single SpAM trial. In SpAM, all stimuli are presented at once, and it is the nature of the task that all object pair relationships are judged with a single pattern of placements. Thus, more than one trial was unnecessary. More than one SpAM trial would also have included identical stimuli, and many participants would likely have noticed this and simply repeated their previous placements.

The initial starting positions of the stimuli along the sides of the SpAM workspace were randomized per participant into two vertical rows of eight objects. Order of initial object display positions did not significantly correlate with final placement position or average distances to other objects in final placement positions.

Procedure

The computer station, seat position, stimulus size, and pre-exposure phase were the same as in the previous two tasks. After pre-exposure, participants were instructed to move objects into positions such that more similar objects were placed closer together and more dissimilar objects were placed further apart. Participants were explicitly told not to pay

attention to pairs in isolation, “All distances matter. Please do NOT just consider two objects together and then ignore the distance from those objects to all the other ones in the box.”

Participants were then given a workspace with two banks of eight scrambled objects along the sides of the screen. Participants dragged objects with their mouse until satisfied with object positions. They could re-position objects as many times as desired. Participants then hit any key on the keyboard twice to indicate completion.

Analysis

Data from SpAM took the form of pixel distances between each pair of final object placement positions in the workspace. City block distance was used as the method of measuring pairwise pixel distances, for the same reason city-block distance was used in MDS analysis, circular dimension awareness analysis, and elsewhere in this project: the dimensions of color hue and shape are separable and thus indicate the appropriateness of city block distance measurements (Shepard, 1987). These dissimilarity judgments were then standardized by scaling each participant’s distance judgments such that each participant’s average pixel distance across pairs was equal. Analysis then followed the same pattern of tests as in the previous two tasks, with some modifications.

Individual MDS analyses were dropped from analysis in SpAM and replaced by analysis of scaled final placement positions. The purpose of MDS is to fit objects into a Cartesian geometric space in a way that matches dissimilarity input as closely as possible. Since Cartesian object placements are already mandatory in SpAM, an MDS analysis at the individual level would be redundant and could add nothing to analysis other than potential error. Instead, participants’ raw placement patterns filled the role of individual MDS solutions, and were subjected to the same analysis procedures as individual MDS solutions

were in the prior two studies. This included the same human ratings of organization of each individual "MDS result." When rating SpAM task participants' placements, raters were simply given the actual placements from SpAM, but presented graphically the same way as in the previous two experiments, as if they were MDS results. Raters were not aware of this underlying difference between data sources between tasks (they were not in fact aware of any specific differences between tasks at all). *Group* MDS solutions proceeded as in the prior two studies, because after averaging across many individuals, pairwise dissimilarities are no longer guaranteed to perfectly fit into a Cartesian space, and thus an MDS algorithm is not redundant for averaged group SpAM data.

Triangle inequalities and violations of minimality were dropped from analysis in SpAM, due to being impossible a priori to observe in a task that is constrained to a Cartesian answer space. These analyses were still included in the project overall (in the previous two tasks), because the enforced Cartesian constraint on SpAM is not itself a naturalistic task characteristic in most real life situations. Since I found both effects in the first two pairwise experiments where they are possible to observe and where the constraints on these effects in particular are more naturalistic, I consider these behavioral effects to already be sufficiently task-general to serve as computational modeling targets.

Results

Group Multidimensional Scaling

Group MDS analysis used the standardized input, which in this case was the same as raw behavioral placement distances, scaled for equal average distances as described above. As the scree plot from chapter 3 indicated, a two-dimensional fit was most appropriate for SpAM group MDS, as it was for the previous two tasks. The MDS group solutions are shown in Figure 19. The square condition solution resembles that of the

pairwise ratings task: is a grid-like shape with some noisy local confusion between neighboring feature values. The "L" condition shows a right angle between arms of the "L," unlike the more obtuse warped angles in the first task. The solution looks more like a cross than an "L" due to some confusion in the order of feature values in the shape dimension (red lines). In other words, the shapes of the shape-matching arm of the "L" were treated as if they were in the middle of the shape dimension rather than on one end of it. Color values show similar confusion of order. Neither condition group solution suggests any obvious overall attentional bias to either dimension.

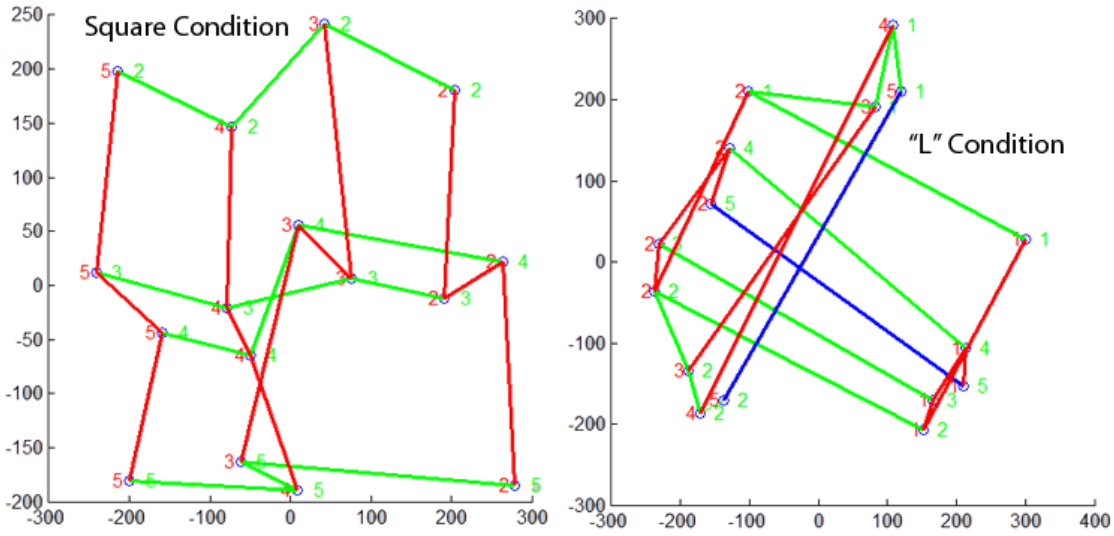

Figure 19: Group MDS solutions for SpAM. The square condition shows a noisy grid shape with minor to moderate local confusion of the order of feature values. The "L" condition shows a less noisy solution, but with more severe confusion of feature value order. The two feature dimensions appear as almost entirely orthogonal, unlike in the pairwise ratings task. No group attentional bias is clear from the group MDS solutions.

Individual Placement Analysis

Instead of individual MDS solutions, participants' direct placements in the SpAM workspace were analyzed as if MDS solutions. Since SpAM constrains answers to a two-dimensional, Cartesian answer space, just like a two-dimensional MDS solution, MDS analysis would be redundant for individual SpAM data. Just like the individual MDS

solutions of the pairwise tasks, raters coded direct placements of objects from SpAM for their level of organization, but were not made aware of the difference between SpAM and the pairwise tasks (SpAM placements were presented to raters as if MDS solutions graphically). Raters did not unanimously agree on any single participant's solution as deserving the lowest possible organization rating, therefore no participant's data was dropped as had been the case in the previous two experiments. This is not surprising, since participants could view their entire set of similarity judgments on screen at one time, unlike in the other tasks.

Patterns of individual placements trended more toward feature-comparison-driven patterns than in previous experiments. Some participants still also showed highly attentionally modulated patterns of placement but fewer than in pairwise tasks. Figure 20 shows two individual patterns most closely matching the group results. The square condition individual, like in the group solution, shows mostly grid-like placements with some minor local confusion of feature value order. The "L" condition individual pattern shown here is not the most organized pattern seen, only the one closest to the appearance of the group solution—specifically, it shows ratings that place differences along each feature dimension mostly at right angles, but which confuses the order of feature values liberally along both dimensions. Some participants individually showed much more organized solutions. Figure 21 depicts the placements of an individual in the "L" condition who placed objects almost perfectly in line with the predictions of a Cartesian model using only feature-comparison methods of similarity judgment.

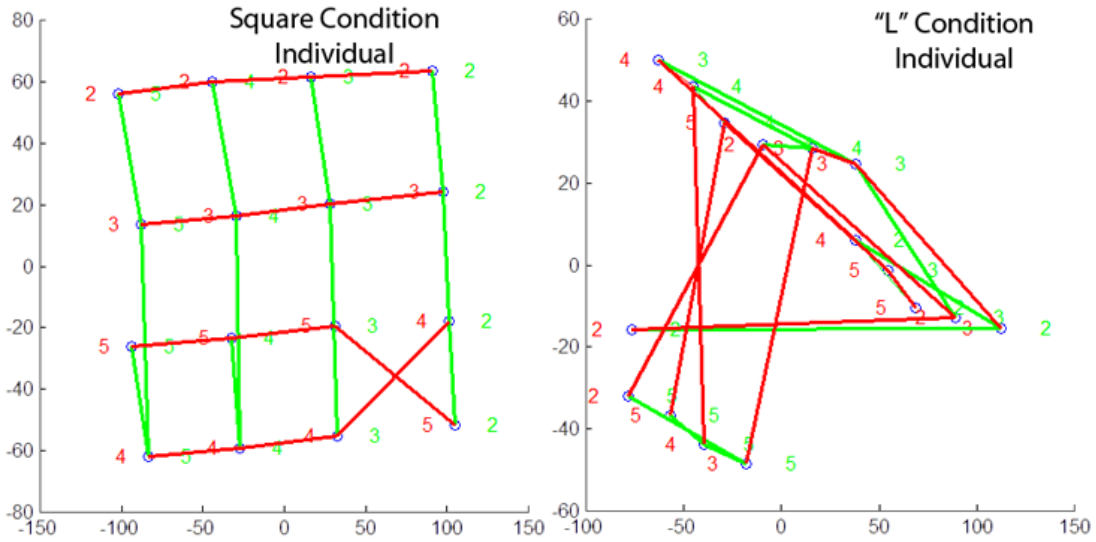

Figure 20: Individual solutions for SpAM. The square condition participant shows a grid like the group solution, but with one row swapped along part of its length. This is more orderly but in the same trend as the group results. The "L" condition participant fits the group "L" pattern of mostly orthogonal feature dimensions, little obvious attentional bias, and many cases of confused order of feature values.

Degree of attentional modulation was tested quantitatively, as in the previous two tasks. In the square condition, the geometric mean ratio was 0.99 red length : green length (geometric standard deviation 3.22), representing no dimension bias. The most extreme bias in a single participant in the square condition was a 10.31 ratio biased toward color. In the "L" condition, geometric mean bias was 0.55 (geometric standard deviation 4.68), also biased toward shape, and the most extreme individual ratio was 0.02 toward shape (another participant was biased at 3.92 toward color).

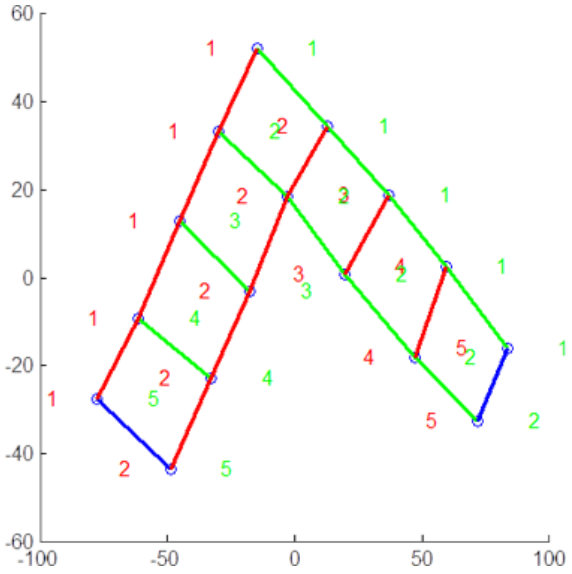

Figure 21: A more orderly participant. This participant was from the "L" condition.

Circular Dimension Awareness Test

This test was performed the same as in the pairwise tasks. Results were mixed. RMSEs in the square condition were 624 for chord-based and 520 for arc-based fits, the first case of a closer fit to an arc-based distance measure. RMSEs in the "L" condition were 100 for chord-based and 107 for arc-based fits, matching the previous two tasks. This task did not confirm that participants were aware of the circular nature of the dimension in the square condition, but confirmed awareness of this in the "L" condition. The arc-based fit in the square condition does not contradict circular dimension awareness. An "arc" based fit could result from perception of distances across an arc of a circular dimension, or it could result from perception of straight line distances as if the dimension were non-circular. Still, chord-based distance measures were eliminated as a central modeling target, at least for square condition stimuli. I will still report results of a test of the model on arc-based and chord-based distance measures, but the ideal result is not completely clear from these empirical results.

Neighborhood Density Analysis

Neighborhood density effects were consistent with the previous two tasks. Summed neighborhood densities of the objects in a pair correlated with the difference between predicted and observed dissimilarity of that pair moderately and positively in both conditions. In the square condition, density and divergence correlated at r = 0.180, and in the "L" condition it correlated at r = 0.343. Both correlations were significant. Again, this is consistent with denser regions of feature space having exaggerated dissimilarity ratings as compared to equivalent pairs in sparser regions of feature space.

Alignability Analysis

Effects were again inconsistent relative to previous tasks. Non-alignment between objects in pairs (minimum distance along either one dimension between objects) correlated with the difference between predicted and observed dissimilarity ratings at r = 0.119 for the square condition, and at r = -0.121 for the "L" condition. These results imply that as objects become more alignable, participants in SpAM exaggerate their similarities in the square condition, but exaggerate their differences in the "L" condition. This is different from the results of either the pairwise ratings task (both very weak, negative correlations) or the pairwise same/different task (both weak/moderate, positive correlations).

Discussion

SpAM is at the opposite extreme in object similarity judgment task characteristics compared to the pairwise same/different task (Table 1). It is slower, more deliberate, and rich in context compared to the pairwise tasks. In addition, SpAM provides persistent perceptual feedback about similarity judgment patterns that the pairwise tasks lack. Because of this, SpAM was expected to rule out some behavioral similarity effects as task-general. This was the case for the chord-based distance measure fitting best in the square

condition. SpAM also showed no disorganized individual judgment patterns like the previous two tasks, although this trend was not a modeling target. SpAM further disconfirmed alignability effects as a task-general behavioral pattern. This task also was unable by definition to show triangle inequalities or violations of minimality, which as previously discussed, is taken as a reason to place less emphasis on these behavioral patterns as modeling targets but not to rule them out of consideration completely.

SpAM also reinforced several effects as task-general, however. Table 2 summarizes the results of all three tasks. In each task, a strong underlying influence of feature value comparison was detected. Even when some individual neighboring feature values were confused in order, the overall patterns of judgments showed feature dimensions being treated separately and most values being correctly distinguished as separate and related to overall similarity judgments. The "L" condition group MDS results in the pairwise ratings task showed an obtuse angle for the "L" condition that implies some degree of conflation between the two feature dimensions. However, dimensions within each individual arm of the "L" were still treated as orthogonal, and the overall angle of the "L joint" was still less than 180 degrees, suggesting that differences between the feature dimensions were still appreciated as well. Feature comparison was expected across all three tasks, since it is common to all prior models of similarity judgments.

Attentional modulation of whole dimensions was observed across all conditions of all three tasks, most notably in individual MDS analysis. The tendency for some participants to consider only one feature dimension at a time when making decisions is a consistent theme across all three diverse tests of similarity judgments, and is a known behavioral pattern in similarity judgments from prior research (Shepard, 1964; Treisman & Gelade, 1980).

Neighborhood density showed moderate but consistent correlations with an exaggeration of dissimilarity over all three tasks. When several objects in a set take up a region of feature space near one another, perception of dissimilarity is exaggerated beyond the amount implied by the basic number of feature steps between each of the objects of a pair. This is also consistent with findings in the literature (Krumhansl, 1978; Love, Medin, & Gureckis, 2003)

A final factor that was implicit in the discussion of all three tasks but is important to explicitly state with regard to the neural model is that there were several clear differences between the square and the "L" conditions across tasks. These differences extended beyond merely the feature value patterns of stimuli in each condition in feature space. The two conditions showed differing levels of minimality violations and triangle inequalities, different types of non-feature-comparison effects in MDS results, different biases on average toward attention to one feature dimension or another, and so on. Especially important is that these differences occurred even in the pairwise task, where the experimental condition was only apparent in the context of a range of trials. This implies that participants were remembering the objects and/or judgments across many trials. For a neural model, this implies the need for a long term memory system to keep track of patterns of stimuli, pairs, and/or judgments over time.

Table 2: A summary of empirical results from Experiments 1-3.

	Pairwise Ratings		Pairwise Same/Different		Spatial Arrangement Method	
Test	Square	"L"	Square	"L"	Square	"L"
Group MDS	Systematic, some curvature, no clear dimension bias, some value confusion	Systematic, some curvature, shape bias, values in order	Somewhat noisy, inconsistent curvature, no strong dimension bias, some value confusion	Noisy, slight curvature matching ratings task, shape bias, several value confusions	Noisy, no curvature, no clear dimension bias, several value confusions	Systematic, no curvature, some shape bias, many value confusions
Individual MDS	Two disorganized participants, average shape bias (0.39 ratio color : shape influence)	No disorganized participants, average shape bias (0.38 ratio)	One disorganized participant, average color bias (1.84 ratio)	One disorganized participant, average shape bias (0.71 ratio)	No disorganized participants, no dimension bias (0.99 ratio)	No disorganized participants, average shape bias (0.55 ratio)
Circular Dimension Measure	1.40 chord RMSE 1.79 arc RMSE	0.87 chord RMSE 1.70 arc RMSE	0.20 chord RMSE 0.23 arc RMSE	0.16 chord RMSE 0.21 arc RMSE	624 chord RMSE 520 arc RMSE	100 chord RMSE 107 arc RMSE
Triangle Inequalities	0% of triplets	2.7% of triplets	0% of triplets	2.7% of triplets	not applicable	not applicable
Violations of Minimality	2.6% of trials	0.6% of trials	0.2% of trials	0.7% of trials	not applicable	not applicable
Neighborhood Density	r = 0.248	r = 0.288	r = 0.117	r = 0.276	r = 0.180	r = 0.343
Alignability	r = -0.07	r = -0.06	r = 0.114	r = 0.343	r = 0.119	r = -0.121

CHAPTER 6

A DYNAMIC NEURAL FIELD MODEL OF SIMILARITY

A central goal of this dissertation is to develop a DNF model that captures a broad set of similarity judgment behaviors with an emphasis on the task-general behaviors observed in Chapters 3-5. In this chapter, I introduce the DNF similarity judgment model in detail, including its architecture and the process by which it simulates similarity judgment tasks. I demonstrate the model's ability to replicate each of the behaviors that were task-general across my three empirical tasks: an influence of feature comparison, attentional modulation by dimension, and a sensitivity to neighborhood density. The model also captured meaningful differences between square and "L" experimental conditions, a task-general pattern that arose from the empirical analyses.

The model also shows violations of minimality, triangle inequalities, and a best fit to chord-based feature distance for circular feature dimensions. These behaviors were observed frequently in Experiments 1-3 but not shown to be entirely task-general.

The DNF model presented here uses a neural architecture that does not consume exponential resource usage as feature dimensionality increases; it captures the real time neural process dynamics of behavioral tasks as they unfold; and it theoretically relates the process of similarity judgments to related processes in visual cognition and categorization that have been the focus of previous DNF models.

Architecture

The DNF model consists of a number of neural fields, 1- and 2-dimensional fields of organized neural units. There are two feature dimensions—shape and color for this project's stimuli—and a spatial dimension representing the spatial position of objects in the task space. Fields were 50 units in size along each dimension, with the exception of a single one-dimensional decision field that was 200 units in size.

Each individual unit is receptive to values along the dimension of the field to which it belongs, with maximal receptivity at a particular value per dimension. Figure 22 shows a two-dimensional example field and the receptive field of a single neural unit within it. This unit is in a two-dimensional color by space field and is most receptive to "green" and "left" feature and space values in that field (see Gaussian receptive fields along each axis). A unit next to this unit might be maximally receptive to "very slightly yellow-green", and "very slightly further to the left," and so on across the field in either dimension. Every unit is sensitive to values other than its maximum 'preferred' value(s), but decreasingly so across a dimension (see black receptive curves, top and left of figure). How quickly this sensitivity falls across units follows a Gaussian curve, with a width controlled by model parameters. Along the color dimension, for example, the neural unit in Figure 22 is maximally receptive to green stimuli, but will still respond weakly to stimuli as far along the color dimension as orange or cyan.

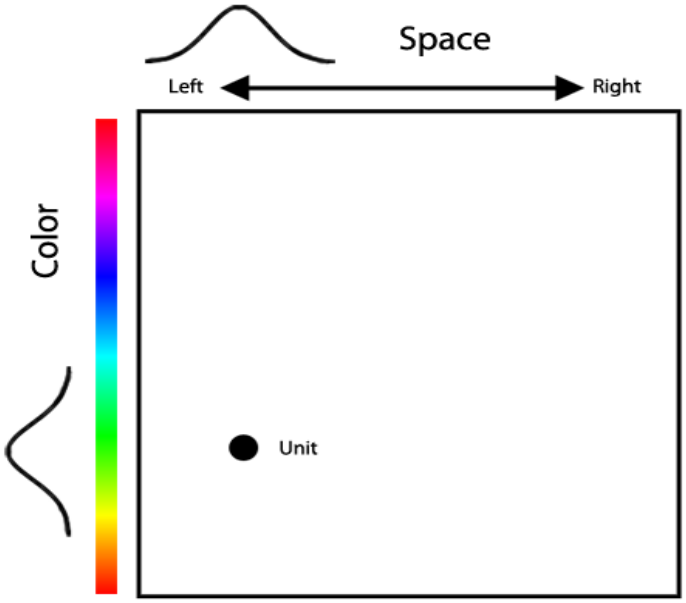

Figure 22: An abstract unit in a neural field.

Figure 23 shows a set of fields and associated activation values from a portion of the DNF model. The large square field is a neural visual field where objects first enter the model. As in Figure 22, the square visual field in Figure 23 is organized by the dimensions of retinal space (horizontal axis) and a feature dimension, in this case color (vertical axis). The colors seen in the figure represent the levels of activation of individual units in the field. The mottled blue background is the resting level of the field (with noise). The circles of light blue to orange are “peaks” of activation, where neural units are being driven by perceptual input, in this case from two objects, at A and B. The horizontal position of each peak represents where that object falls on a simplified one-dimensional retina, and the vertical position of each peak represents the color of that object. Objects A and B are therefore appearing at different locations in space and have different colors.

Input to the model assumes that objects have exact feature values corresponding to particular units in the visual neural field. The two peaks of activation in Figure 23 do not

appear as single units of activity, however. This is because many nearby units have receptive fields that overlap the input values, with diminishing strength further from the input location and color.

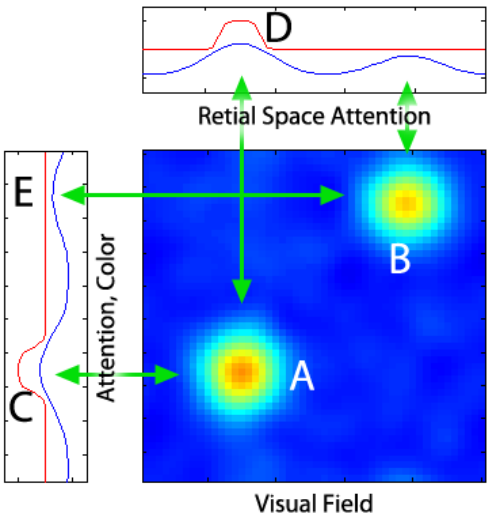

Figure 23: Visual and attentional fields in the DNF model.

Activation in the model passes between fields across shared dimensions. The two white boxes in Figure 23 (C and D) depict one-dimensional attention fields. Activity in these fields corresponds to degree of attention allocated to either a particular color (in the color attention field) or a particular position in space (in the space attention field). Blue lines indicate the excitation levels of the neural units. The red lines represent a threshold of activation—units that are above this threshold send outputs to any fields they are connected to.[10]

The visual and attention fields are interacting with one another in Figure 23. Each unit in the attentional fields is receiving summed input from the row or column of units in

[10] The two-dimensional visual field works the same way with regard to thresholds of activation, but the activity and output are not both visualized in two-dimensional fields in the figures, only the activity level.

the visual field next to it, and each attentional unit is also sending back its own output distributed to every unit along those same rows and columns. Each unit in the color attention field is receiving maximum input from a row of units in the visual field, and each unit in the retinal space attention field is receiving input from a column of units in the visual field. These activation projections decrease strength over distance, according to a Gaussian profile.

All fields in the model—as shown in Figure 23 and subsequent figures—also feed into themselves. Figure 24 shows the two types of feedback a field receives from itself. Self-feedback is only sent for above-threshold activation, just like output to other fields. When threshold activation is reached, a field sends self-excitation to itself (green, Figure 24), and it also send a *broader* but *weaker* pattern of lateral inhibition to itself (red, Figure 24). The result of both excitation and inhibition together is a "Mexican hat" shaped pattern of influence (black, Figure 24), with heightened activation at the site of the original activity, but a trough of inhibition surrounding that area. This pattern of feedback allows for stable, persistent neural activation. The self-excitation can maintain the pattern of activity, while the lateral inhibition stops the pattern from growing out of control. The exact strength and shape of the excitation versus inhibition can be tuned to make a field self-sustaining (with strong neural interactions) or primarily input-driven (with weak neural interactions).

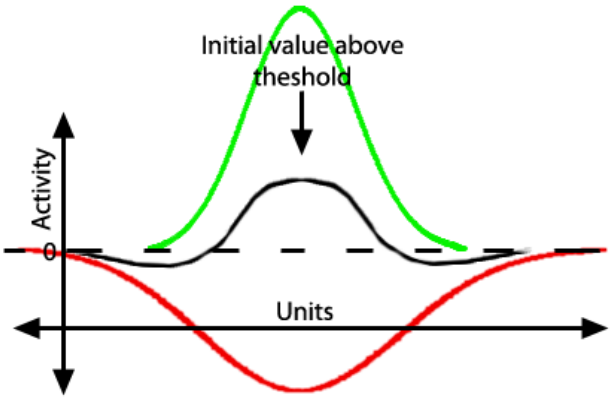

Figure 24: Self-feedback unit dynamics.

The final type of input to a regular field is global inhibition (not directly shown in figures). When any part of a field reaches threshold levels of activation, it can feed back inhibition to *all* of its own units regardless of their receptive fields. Global inhibition is used for fields that require competitive peak formation. Attentional fields are examples of fields with strong global inhibition. Only one feature value or spatial location is typically attended at once, and this is achieved with global inhibition. Once one peak forms, representing attention to a feature or location, global inhibition suppresses any other peaks from forming. Local self-excitation, however, can still maintain the first peak despite its own global inhibition. Thus, the field functions as a first-winner-takes-all competitive system. The first object attended will suppress attention to any other objects until the attention field is somehow destabilized (via outside inhibition, removal of input, etc.).

Returning to Figure 23, we can now understand the full range of dynamics displayed. Input from the visual system outside the model is driving two peaks in the visual field, representing two objects, A and B. One of the objects is currently being attended (A). This object was likely attended due to random fluctuations: background noise pushed its peak's activity slightly higher than the other object's peak at first. Since both peaks in the

visual field are driving both attention fields at corresponding values (green arrows), this slight boost would have allowed one object to reach attentional threshold first. The peaks at that location and color in the attentional fields then won the attentional competition and suppressed any other peaks from forming. Reverberating activity between space and color synchronize the attentional fields to both attend to the same object if they did not do so originally. The input from the un-attended object (B) to the attentional fields is also visible, but is being pushed below threshold by global inhibition (E). The attentional fields are also connected positively back to the visual fields (green arrows). This is causing the attended object's peak at A in the visual field to become stronger and redder—remember, only above-threshold activation sends output, so the suppressed object at B is not strengthened.

Full Model

Figure 25 depicts the full DNF model of object similarity judgments. Figure 26 shows the model with all connections between model components indicated. Some parts of the model shown in Figure 25 are familiar from the introductory example above. The same visual field is visible at A(color), and the same two objects can be seen at B(color) and C(color). The same attentional fields are also visible at D(color) and at E. There is now another visual field and attentional field below the first, at A(shape). There are still only two objects, however: points B(shape) and C(shape) represent the same two objects as B(color) and C(color), but the shape values of these objects are indicated in the new visual shape field. The purpose of two rows of fields is to represent the stimuli's two features. In general, any field in the model that is receptive to a feature dimension would be replicated once in the model for every relevant feature dimension. Importantly, this adds neural resources linearly, but not exponentially, for high dimensional object representations. Since my stimuli from Experiments 1-3 had two feature dimensions, I have two rows of feature

fields. Each object is thus represented as two peaks, both at the same retinal spatial location but at different feature values. In Figure 25, the two visible objects match in shape, but differ in color.

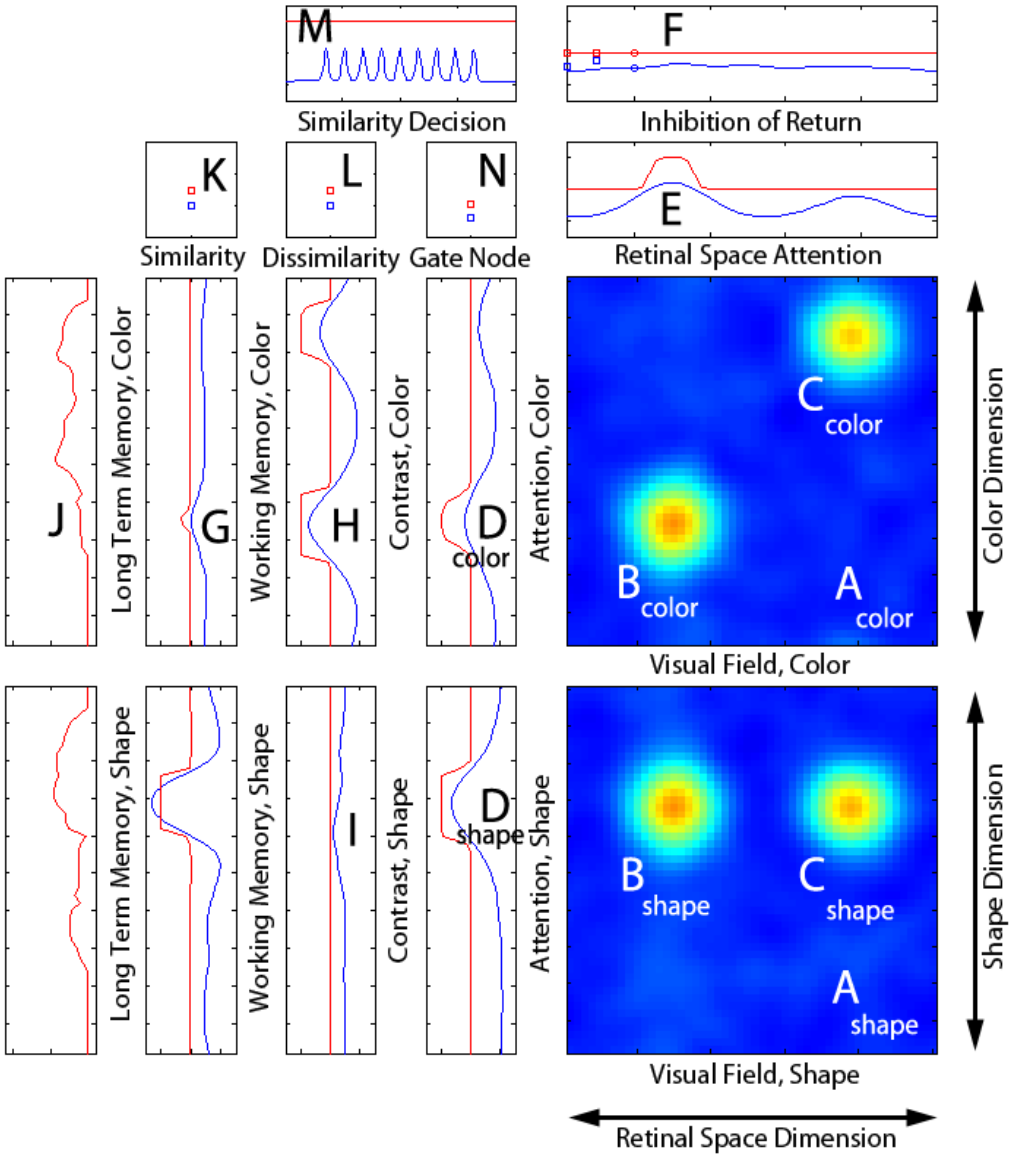

Figure 25: The full DNF model.

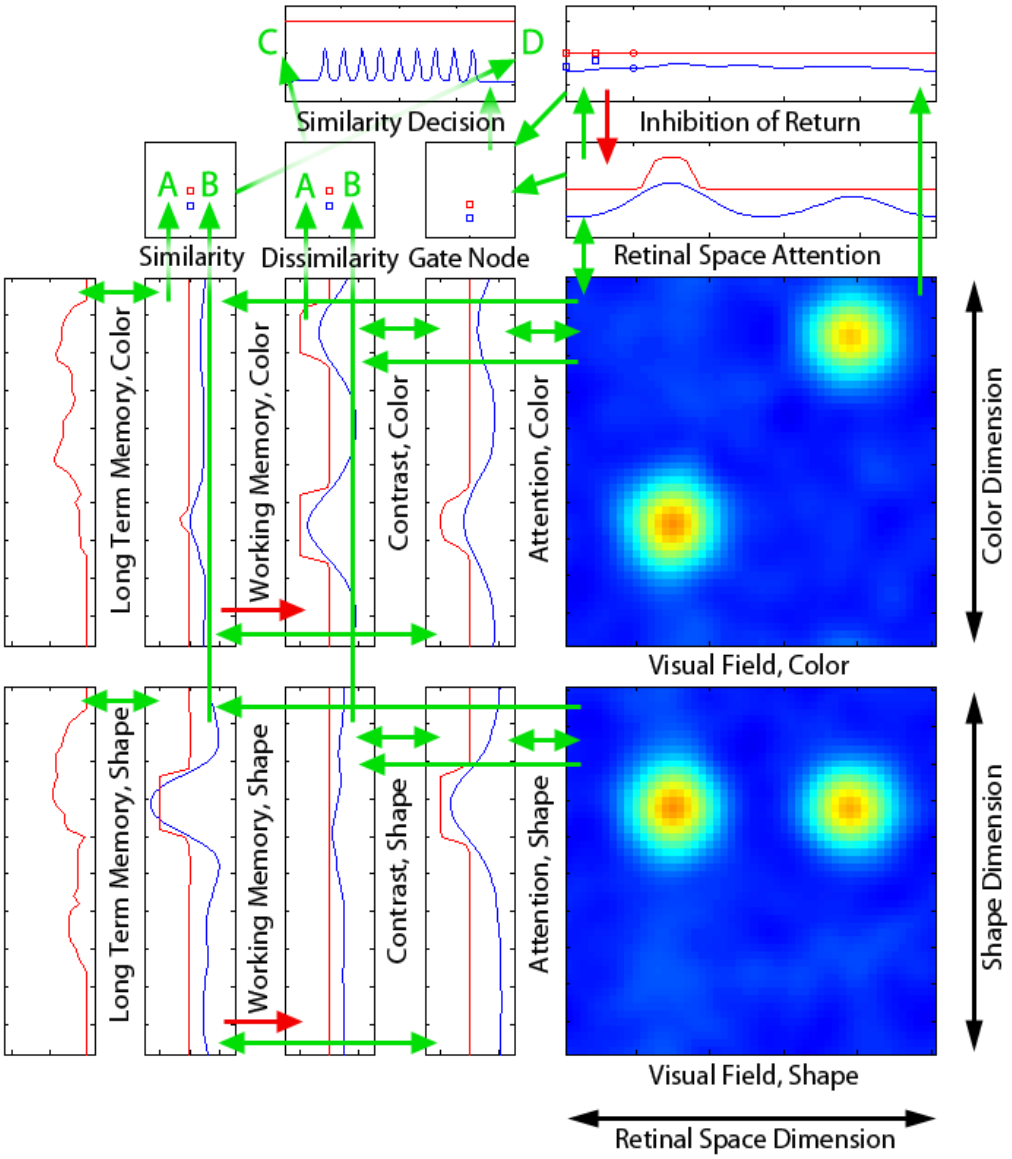

Figure 26: The full DNF model with connections.

The field seen at F in Figure 25 is an “inhibition of return” (IOR) field. Unlike attention, the IOR field has weak global inhibition and can sustain several peaks at once. Peaks form here partially based on input from the visual field, but primarily due to input from special “peak detector” nodes. These are visualized as small points to the left of the IOR field box in the figures, though they are not part of the IOR field. Peak detectors fire

when an object's features have been consolidated in working memory. Once this occurs, the baseline level of the IOR field is raised, and a peak builds in the IOR field at the location that is the current focus of attention. This IOR peak actively suppresses the associated peak in the spatial attention field, releasing the model from the current focus of attention. IOR, thus, keeps attention shifting to new stimuli as soon as stimuli are consolidated in working memory.

At position G in Figure 25, a feature (in this case color) working memory field is shown. This field stores the colors of recently seen objects. In the DNF model of Schneegans, et al. (in press), objects are represented in a two-dimensional color by space working memory field that binds features together in the spatial frame of a scene. This avoids confusion if, for instance, the system had to remember different objects of the same color. For the current model, I used a simplified one dimensional working memory because the model only had to compare two objects on any given trial. With the simplified model, similarity judgments occur after one object is stored in working memory and the model shifts attention to the second object to engage in a comparison of the two items. The working memory fields are weakly self-sustaining. Working memory consolidation in Figure 25 along the shape dimension has occurred faster than in the color dimension, due to the objects sharing a shape and sending stronger, overlapping activity across the attended value of the shape fields.

A one-dimensional color contrast field is seen in panel H. A contrast field detects novelty. As indicated by arrows in Figure 26, the color contrast field receives excitatory input from the visual field and color attention fields. Thus, currently attended objects in the visual field strongly excite the contrast field. The contrast field is inhibited, however, by working memory, inhibiting peaks for already-stored feature values. The result is that the

contrast field only builds peaks temporarily for *new* feature values, seen but not yet held in memory. As soon as objects with those features are fully stored in memory, the novelty activation is suppressed. At point I in Figure 25, this can be seen in action: since the two objects overlap in shape, the working memory peak from the first attended object has already destroyed any contrast peak in the shape dimension. There are no longer any novel shapes in the visual field.

The final fields in Figure 25 not specific to similarity judgment are long term memory fields, such as the one in panel J. These fields show only a red line. They do not operate under the same dynamics as the other neural fields. Long term memory fields operate under principles of Hebbian learning, gradually building up representational strength over many trials, which then decays even more slowly if not reinforced. In Figure 25, the long term memory activity shown had already built up over 40 trials before the two objects in this trial were presented. Long term memory receives input only from above-threshold units in the working memory fields. It provides feedback exclusively to the working memory fields. There are no Gaussian receptive fields to the connections between long term and working memory fields; the relationship is 1:1 between exactly matching units only. Long term memory in this model amplifies any working memory signal that occurs at a spot with frequent recent working memory activity.

Similarity Components. Thus far, all fields described are common to the DNF model proposed by Schneegans et al. that was used to capture processes of visual working memory and change detection. The remaining nodes and fields are those specific to similarity judgment in the current model.

Figure 25 shows the addition of a few single unit nodes and a one-dimensional field for capturing explicit similarity judgments in the model, at K, L, M, and N. At the core of

the similarity portion of the model are the single unit “Sim[ilarity]” (K) node and “Dissim[ilarity]” (L) node. These are driven by the sum of all activation in the fields below them. The dissimilarity node is driven by summed activity over the feature contrast fields. Contrast fields detect novelty, so whenever they are highly active, the currently attended object must be different than anything already stored in working memory.

The similarity node is driven by summed activation of the working memory fields. When objects in a scene match in a feature, their activity overlaps and is exaggerated in working memory. This is the same effect that caused the shape memory field in Figure 25 to form a peak faster than the color memory field. Summed working memory activity therefore serves as a relative indicator of similarity strength, since more matching features adds to more total activity in feature working memory fields.

The similarity and dissimilarity nodes themselves feed into opposite sides of the “similarity decision” field (M), which is shown here as it would be in a pairwise ratings 1-9 choice task. Dissimilarity projects a very broad peak of activation to the left side of the decision field, and similarity projects a broad peak of activation to the right. Depending on the nodes’ relative strengths, anywhere from one end to the middle to the other end of the decision field will receive the maximal input. Sub-threshold peaks exist all along the scale field, representing 1-9 similarity ratings (driven by instructions or viewing a ratings scale on a monitor, for example). Depending which part of the field receives the most activation, any one of these sub-threshold peaks will reach threshold first. The decision field has high global inhibition, so possible answer responses compete until only one wins and self-excites strongly enough to trigger a numerical similarity judgment.

This decision competition cannot happen too soon, however. The contrast and working memory fields begin to become active via input from the first attended object, so

without any additional control, the decision field would decide immediately after attending to the first item in a trial. To avoid this, the decision field is kept at a low resting level, with similarity and dissimilarity signals insufficient by themselves to push it to a decision. In addition, the decision field also requires input from a "gate" node (N in Figure 25). This gate node receives input from the IOR field and from spatial attention, and requires input from both of these fields to activate the gate node and trigger a decision. This means that the model actively attends to the first object, consolidates these features in working memory, and then shifts attention to the second object. At that point, the gate node is engaged and the model is ready to compare the objects and make a decision based upon the activation of the similarity and dissimilarity nodes.

The Remaining Time Course of a Pairwise Ratings Trial. At this point, all of the fields and processes relevant to a single pairwise ratings similarity judgment trial have been described. Figures 27 and 28 show these processes unfolding late in an example experimental trial.

In Figure 27, the full working memory trace has established itself across both feature dimensions, unlike in Figure 25 (A). Color working memory was slower to establish, since colors did not match and thus overlap activation along the color dimension, unlike along the shape dimensions. The newly established peak in working memory has also suppressed the peak at the same color value in the color contrast field at Figure 27 B. Moreover, the consolidation of an item into working memory along both feature dimensions has triggered the peak detector system and initiated inhibition of return to the first attended object (C), which has inhibited return of attention to the spatial location of that object (D). With attention broken to the first object, the system is beginning to attend to the next object (E). The combined activity from the IOR field and from attention to a

new object (C and E) is almost enough to trigger a similarity judgment via activation of the gate node (F). The decision field is rising higher and closer to threshold (G). Similarity and dissimilarity signals have been feeding into the similarity and dissimilarity nodes at all times, and the decision field is already slightly biased toward the similarity side.

In Figure 28, the model has progressed further. Only the top few non-feature-specific fields are shown here. The gate node (A) has now received enough activation from IOR and spatial attention (C and D) for long enough to build activation and push the similarity decision field (B) to threshold. Responses are competing through global inhibition. The model appears to be favoring an answer of 7 out of 9 similarity for these two objects (B). Note that as the model cycles through additional time steps, eventually a single rating will be selected via global competition in the decision field. The IOR field is now inhibiting return of spatial attention to both objects because they have both been attended.

The 7 out of 9 rating is driven by strong activation to the similarity node due to the overlapping features in shape projecting robust activation to working memory. At the same time, any contrast peaks were destroyed in shape, lowering the dissimilarity signal. Some dissimilarity signal persists, though, due to the remaining contrast detection in color. The model can also capture the subtler case of *partial* overlap between objects in a feature dimension. If feature values are close but not identical, the working memory boost due to overlap is weaker, and the destruction of contrast field peaks is only partial, due to the Gaussian shape of interactions within and between fields.

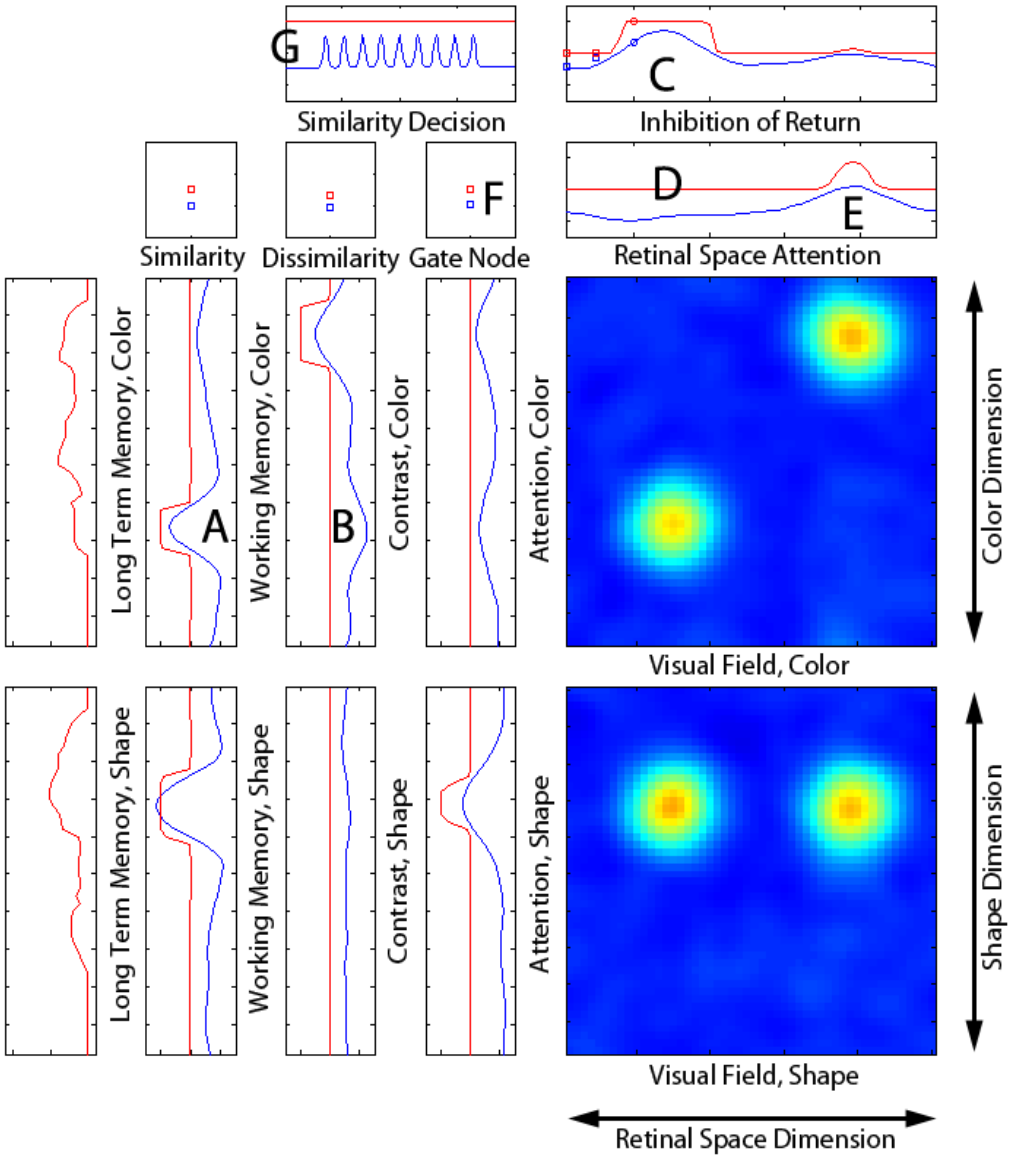

Figure 27: The DNF model after peak detection. A full working memory trace has been established for the first attended object, IOR has activated and suppressed spatial attention to the location of the first item, and the system is beginning to attend to the second object.

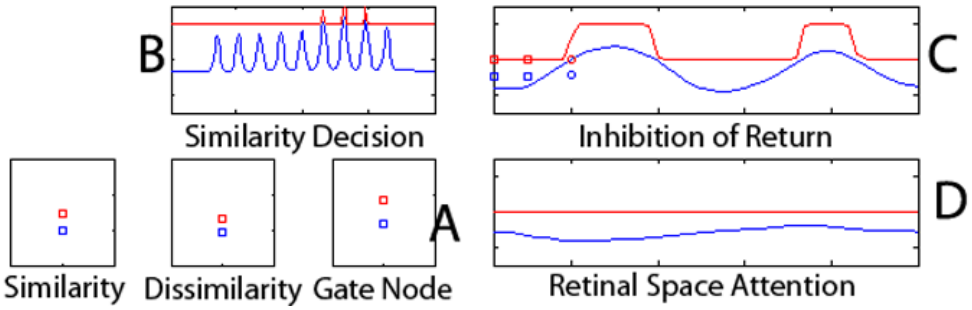

Figure 28: The DNF model at decision. Return of spatial attention is inhibited to both object locations, and the gate node is driving a competition between possible ratings in the similarity decision field.

Experimental Simulation

Pairwise Ratings Task

Figures 25-28 depict the entire process of a pairwise ratings trial in the DNF model. The model simulates actual, individual participants by being given each trial that each participant received, in the same order the participant experienced. Objects were positioned at different coordinates along the color and shape dimensions in the visual field according to the original input feature values used in the construction of the stimulus feature space.

Each participant's attentional bias toward each of the two feature dimensions was also provided to the model. If, for example, a given participant was twice as influenced by color differences in their similarity judgments than by shape differences in individual MDS analysis, then in the model, the influences of the working memory and contrast shape fields on the similarity and dissimilarity nodes (connections marked A in Figure 26) were reduced by half, while the influences of the color fields (B in Figure 26) were untouched. The model used exact ratios of attentional bias between feature dimensions from each human participant's personal MDS analyses to simulate that participant. All of these ratios for individual participants are listed in Table 3 at the end of this chapter.

The model then fit the relative influence of the same and different nodes on the decision field (connections C versus D in Figure 26). The strength of these connections relative to one another was varied over nine simulations of each participant (all combinations of three strengths each of connections C and D in Figure 26), and the best fit between model output and behavior was used for each simulated participant. This accounted for individual participant differences in biases toward one end of the ratings scale versus another, or different interpretations of what "similar" meant, such as a more conjunctive versus disjunctive definition of similarity. The best fitting weights of connections to same and different nodes from model simulations of each participant are listed in Table 3. Responses were recorded for the model as whichever response peak first reached an activity level of 8, which was sufficient to indicate a clear winner from global inhibitory competition.

The model's other parameters were then tuned to best fit overall performance over the four simulated conditions—pairwise ratings square and "L" conditions and pairwise same/different square and "L" conditions. A combination of matching means and correlations between empirical and model results for individual trials was used to evaluate the fit of different parameter sets. The same and different node weights were fitted automatically per participant for every set of parameters tested, not only in the final simulation. Only one set of model parameters was used for both conditions of both tasks. The only variables between simulations over these conditions were the participant-specific attentional biases, the participant-specific order of trials matched by the model, and the switch between nine and two output options in the decision field. The full set of best fitting model parameters used for all conditions is provided in Table 4 at the end of this chapter. Table 4 also indicates parameters changed from Schneegans, et al. (in press) in bold.

Pairwise Same/Different Task

The pairwise same/different task was modeled in the same way as the ratings task, but with only two sub-threshold peaks in the similarity decision field at positions corresponding to 1 and to 9 in the ratings task. These two endpoints become "different" and "same" responses, respectively. This was the only change to the model architecture. The model also received the appropriate trial order and attentional biases of participants in the same/different task while simulating this task.

The fact that the same/different experiment had an objective correct answer is ignored by the model. The distinction between disjunctive and conjunctive similarity rules *could* be hard-coded into the model as a different basic ratio of weights between similarity and dissimilarity nodes into the decision field. A conjunctive rule requires identical matches to respond "same," so this could involve stronger projections from the dissimilarity node. This would make a single difference strong enough to overwhelm the system and force a "different" response, even if other features matched. A disjunctive rule requires only one matching feature to respond "same," so the projections from the similarity node could be stronger to push the decision field to a "Same" answer, even with only one match's input. I did not need to explicitly implement these ratios based on instructions, however, because this ratio was already being freely fit by the model to individual participants' behavior.

SpAM

SpAM was not modeled in this dissertation. SpAM involves sixteen objects instead of two; constant switching between two different task spaces (the item banks and the placement workspace) instead of one; two relevant spatial dimensions per spatial field instead of one; and a more complex, two-dimensional response format. All of these

requirements are possible to meet in a future, expanded version of the DNF model that integrates the scene representation model of Schneegans, et al. (in press). Two task spaces can be achieved with two different spatial feature fields. This is an architecture that was temporarily limited from the Schneegans, et al. for quicker simulations but will be re-introduced for capturing SpAM results. One field organizes object representations by retinal space and feature, while the second field organizes objects by scene or task (allocentric) space and feature, regardless of the current retinal view. The two spaces are linked in DNF by transformations that represent proprioceptive knowledge of body, head, and eye position.

The two-dimensional workspace (for perception and responses) of SpAM can be accommodated in DNF using three-dimensional space-space-feature fields. These fields still use realistic numbers of neural units and account for both dimensions of retinal space or SpAM workspace.

For the DNF model to maintain multiple objects per trial requires larger fields and different neural dynamics to allow for narrower stable object representations. The dynamics of feature comparison's impact on similarity judgments would also need to change from the pairwise task model. Feature comparison itself as a dynamic between perception, working memory, and long term memory would remain the same, but instead of leading to decisions through "same" and "different" nodes, SpAM placements would need to be decided in spatially organized fields. The Schneegans, et al. (in press) model included spatial working memory and spatial contrast fields analogous to the feature fields in the present model. These were unnecessary for simulating pairwise tasks, but in SpAM, these would form the basis of object placement decisions. While feature fields determine the similarity of objects through overlap of features, spatial fields would similarly

determine valid object locations by overlap of spatial locations. New placements directly on top of existing objects are discouraged, but placements nearby similar objects are encouraged. This can be accomplished with a reversed version of the "Mexican hat" Gaussian field comparison shown in Figure 24. Instead of local, strong excitation added to wide, weak inhibition, SpAM placements can depend on local, strong inhibition and wide, weak excitation. An upside-down "Mexican hat" pattern would discourage placement of objects on top of one another, but encourage placement of objects near one another. Combined with feature comparison in the feature fields, the two systems can satisfy the rules of object placements in SpAM.

Results

Analyses for all simulated data were identical in form to the analyses run in the experimental conditions, except with yoked model ratings substituted for human ratings on a trial-by-trial basis. I review these results individually in this section, and a summary of all modeling results is also provided in Table 5 at the end of this chapter.

Group Multidimensional Scaling

Figures 29 and 30 show 2x2 cell comparisons of behavioral and modeling group MDS results. The left column of each figure is the square condition, and the right column is the "L" condition. The top row of each figure shows behavioral results, while the bottom row shows modeling results. Figure 29 depicts pairwise ratings group MDS solutions, and Figure 30 shows pairwise same/different group MDS solutions.

The fits by the model in all cases are very close to the corresponding behavioral solutions. In both square conditions, the model struggles somewhat with achieving the correct dimensional attentional bias. The ratings square solution shows the more orderly fit, and the same/different square solution shows the more disorganized fit, including

confusion between shape values, as in the human MDS solution. In the square conditions of both tasks, the behavioral solution is less organized, with one versus zero feature value confusions in the ratings task and two versus one feature value confusions in the same/different task, but the relative organization between tasks is consistent between behavior and modeling results.

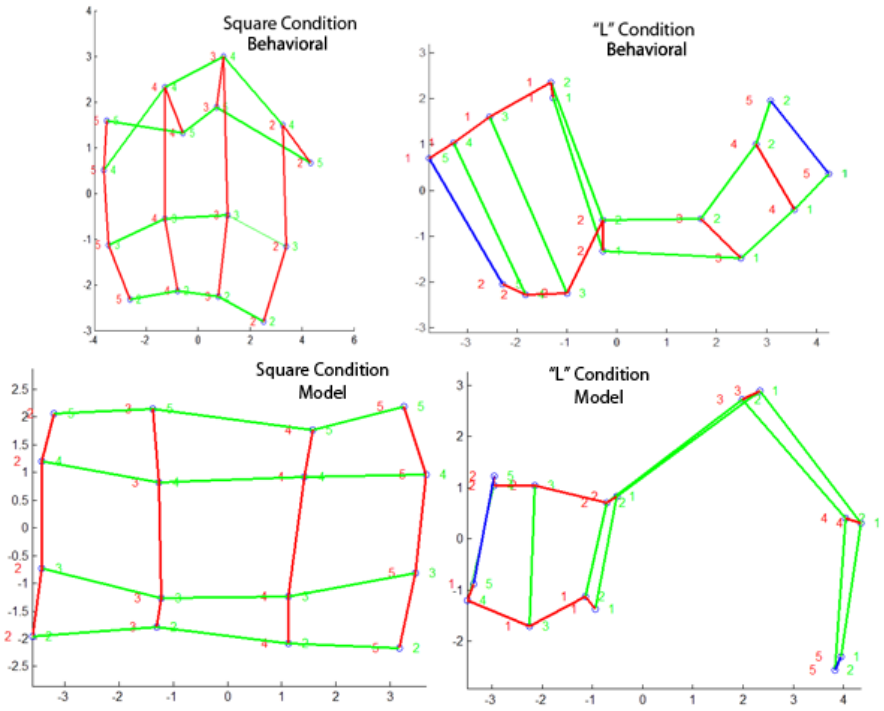

Figure 29: Ratings task MDS fits.

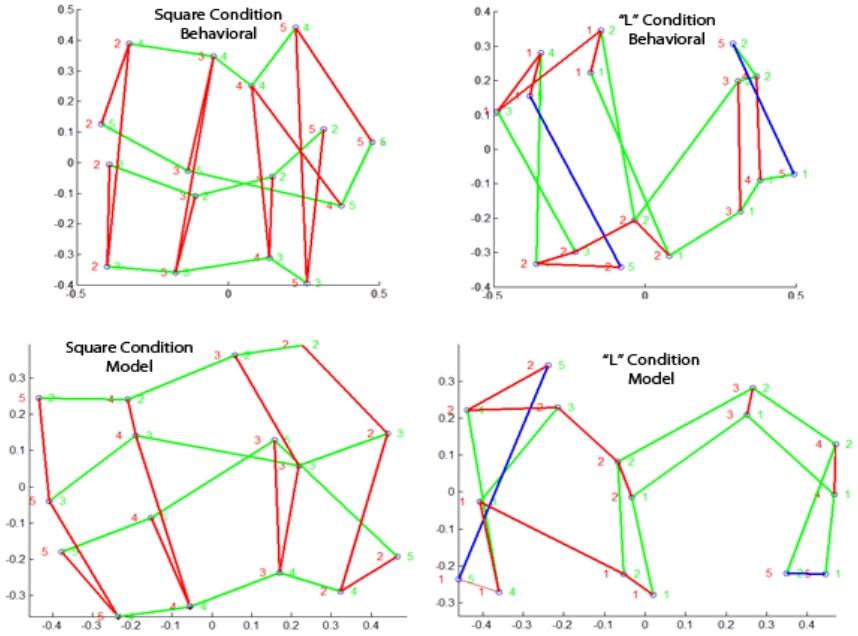

Figure 30: Same / Different task MDS fits.

In the "L" condition, the modeling fits are closer to behavioral results. The model shows all of the key features of the behavioral solution in the ratings task: an attentional shape bias, an obtuse angle to the "L" shape, and good organization otherwise for all solution placements. The model also captures behavioral differences when progressing to the same/different "L" solution: higher disorganization, less of an overall attentional bias, and a greater number of feature value confusions.

Individual Multidimensional Scaling

Group MDS results may be artifacts of averaging, and especially in the case of an artificial model, this is an important possibility to test. Overall, individual MDS solutions fit corresponding behavioral data as well if not more closely than the group results. Figure 31 shows a representative selection of several individual MDS solutions in the square ratings condition. Other conditions fit similarly, but the range of individual patterns is easier to appreciate with examples from a single condition and task.

For almost every participant, the model captures individual MDS solution shape, the correct attentional bias to individual feature dimensions, and even level of organization of different subjects' solutions. This last factor, well-orderedness of solutions, is surprising. An example can be seen in the distinction between the first and second rows of Figure 31. This distinction must be the result of overall feature dimension bias (fed directly to the model), same/different bias (the fitted parameter), or the fact that the same random order of trials was given to the model as presented to subjects. None of these should obviously be predictive of organization of similarity judgments overall.

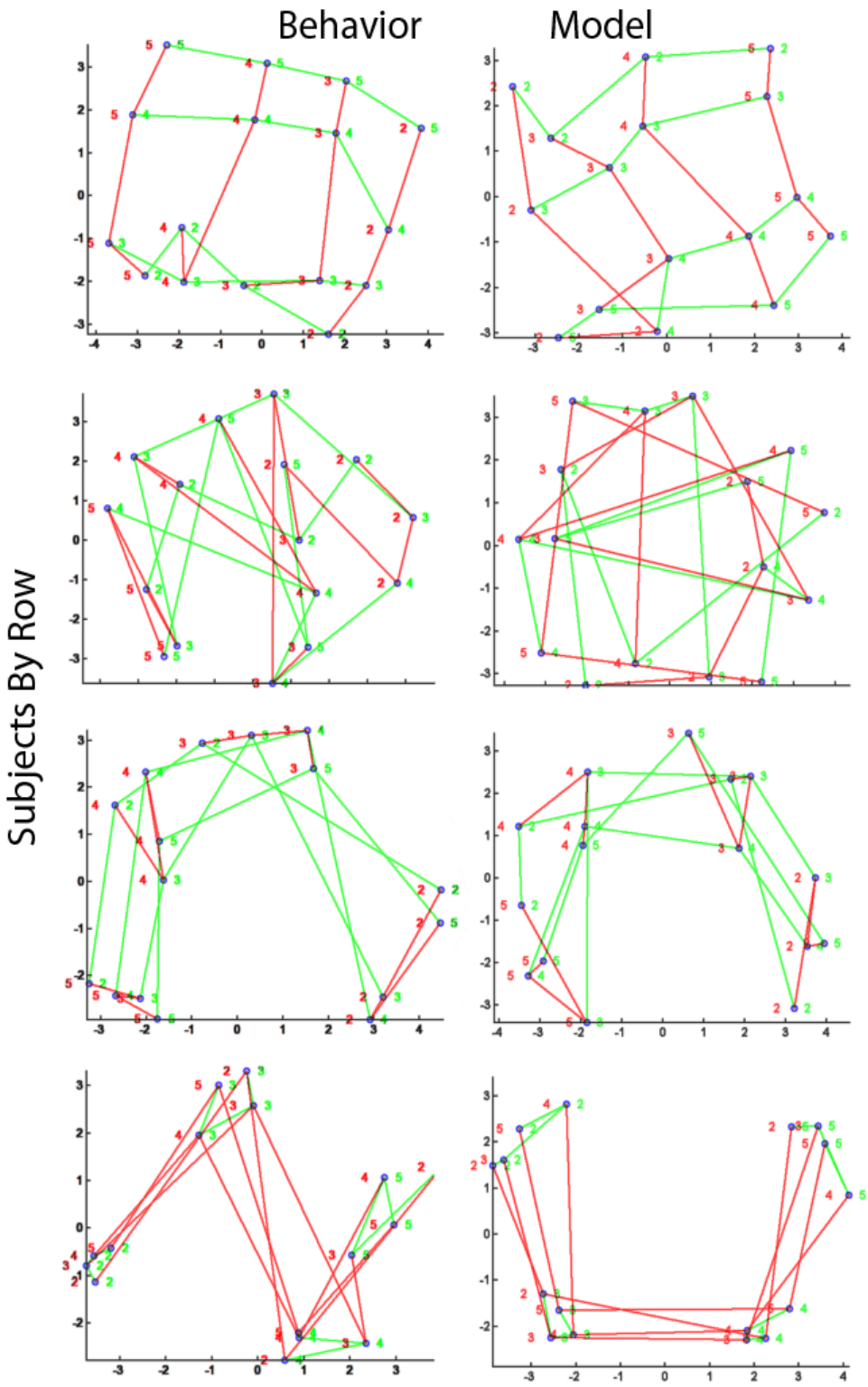

Figure 31: Individual MDS model fits.

Feature Comparison

Both group and individual MDS solutions demonstrate the model's strength in utilizing feature comparison as a foundation of its similarity judgments. Feature comparison is ubiquitous in behavioral patterns, and capturing this factor is a necessary achievement of any similarity model. The neural processes that support feature comparison in the DNF model are those involved in the perception and memory representation of objects across the visual, contrast, working memory, and long term memory fields. When features overlap between objects, peaks in the working memory field form with higher overall strength than when features do not overlap. This can be seen in action in Figure 27. The peak at point A in the color working memory field is weaker than the peak in the shape working memory field below it, because the color peak is driven by one object's feature activation, and the shape peak is driven by two object's activations. Activation in the working memory field thus serves as a similarity signal.

Long term memory also contributes to a similarity signal. Repeated activity at a feature value in working memory builds a long term memory trace that feeds back into working memory in future trials. This creates a higher similarity signal for objects with features that match those seen in previous trials.

A dissimilarity signal is also present in the model in the overall activation of the contrast fields. Contrast fields are excited by visual input and inhibited by working memory activation. Thus, they show activation for feature values that have not previously been perceived during a trial, which correspond to the differences between a pair of objects.

Across all of these feature comparison processes, strength, width, and timing of peak activity contribute to final similarity and dissimilarity signals. Differences in strength between features seen in one versus both objects support the information in the similarity signal. The width of activation allows for more nuanced similarity judgments than binary distinctions. If peaks are very narrow, then they will only interact when features identically match. Peaks with width, however, can overlap by varying amounts, allowing for more informative intermediate similarity and dissimilarity signals when features only approach one another. The timing of activation is involved in similarity signals through accumulating long term memory activation. Timing is also involved in dissimilarity signals: the same signal will generate a dissimilarity signal when seen for the first time and a similarity signal when seen for the second time in a trial.

Dimensional Attention Modulation

The model successfully captured the degree of color and shape bias in similarity judgments in its simulations of most individual participants. Several processes were tested to drive this behavior. The most successful process, included in the final version of the model, controlled dimensional attention by changing the relative weights of connections to the "same" and "different" nodes in the model (K and L in Figure 25) from different feature fields. For example, the connection from the color working memory field to the "same" node might be strengthened relative to the connection from the shape memory field to the "same" node (and likewise for the contrast to "different" node connections). If so, the model will be influenced more by color than by shape feature comparisons. This process occurred only in the explicit similarity decision portion of the model. Therefore, object perception, feature comparison, and memory representation were unaffected by

dimensional attention, since it only applied in the explicit similarity decision portion of the model.

An alternative process placed attentional dimension earlier than feature comparison, in the low-level visual processing outside of the model. Dimensions that participants neglected as less influential were re-scaled to show fewer distinctions between objects. All features of all objects in less influential dimensions for a participant were placed closer together in the initial input to the model. A completely neglected dimension would treat every object as if identical in the neglected dimension. This led to overall higher similarity signals for subjects with strong attentional biases, but the model compensated for this by automatically fitting stronger weights to the "different" node compared to the "same" node in participants with strong dimensional biases. The result was that neglected dimensions carried no differential similarity information and thus did not contribute to overall patterns of similarity judgments. This process implies that the features of objects in neglected dimensions never get loaded into working memory or long term memory distinctly, which provides a means of empirically testing the difference between this dimensional attention process versus the node-weight process. The low-level visual version of dimensional attention qualitatively fit participants' behaviors, but did not quantitatively fit as well as the node-weight process model.

Circular Dimension Awareness Test

Chord and arc-based distance measures were tested for goodness of fit to raw modeling output (not MDS results), as in the empirical analyses. In all cases, the chord-based distances fit better than arc-based distances. This reflects the model's sensitivity to circular feature dimensions. Circular sensitivity is not surprising in this model, because I

used circular fields. That is, each feature field had neural connectivity as if it wrapped into a cylindrical field.

Model results for the square condition showed an RMSE fit between modeling output and arc-predicted responses for the ratings task of 1.08 (chord) versus 1.48 (arc), and for the same/different task of 0.146 (chord) versus 0.164 (arc). Model results for the "L" condition showed an RMSE fit between modeling output and arc-predicted responses for the ratings task of 1.20 (chord) versus 2.07 (arc), and for the same/different task of 0.093 (chord) versus 0.147 (arc).

Tests for Tversky Violations

Violations of minimality and triangle inequalities did not fit the condition-specific patterns from behavioral data (Table 2). However, both were detected by the model, and the model demonstrated a tendency to show both types of effect without any explicit fitting to do so. Three violations of minimality were detected in group data in the same/different square condition, and 186 triangle inequalities were detected in the same/different "L" condition.

The means by which the model captures these behaviors is unclear process-wise. Neural noise may account for some of the results. The fact that triangle inequalities were only found in the "L" condition, however, while violations of minimality were only found in the square condition, suggests a source other than just noise. Long term memory activity can also potentially account for these effects. Long term memory boosts the similarity signal by exciting working memory. Therefore, an identical pair of items seen in an early trial in the experiment may generate a weaker similarity signal than an almost identical pair of items seen later on. Feature comparison between just the objects in each trial will generate a stronger similarity signal for the identical items, but the additional activation

from accumulated long term memory in the later trial may outweigh this difference. This would lead to a violation of minimality. A similar process may explain triangle inequalities. One leg of a triangle of objects might be seen in an early trial before long term memory accumulates, and another leg might be seen later after long term memory has accumulated.

Neighborhood Density Analysis

All conditions of the modeling results showed effects of neighborhood density. In the ratings task square condition, neighborhood densities of object pairs correlated with the difference between predicted and observed ratings at $r = 0.093$, using the same analysis method as in Experiments 1-3. In the ratings task "L" condition, this correlation was $r = 0.221$. In the same/different task square condition, the correlation was $r = 0.153$. In the same/different task "L" condition, the correlation was $r = 0.289$.

Although the ratings task square results were more weakly correlated than expected, all results were significant at a $p < 0.05$ level and were in the correct direction compared to behavioral results. Magnitudes of correlations overall were also comparable to the behavioral magnitudes. These results suggest that the model accurately exaggerates dissimilarities for objects with many nearby neighbors in feature space, similar to behavioral participants. This is only possible due to long term memory layers in the model, which retain memory traces of the features of recently seen objects on earlier trials.

Differences Between Square and "L" Conditions

This was not a quantitative test performed on behavioral results, but it was a task-general trend that the two experimental conditions led to consistent qualitative differences in modeling results, as in behavior. For example, the model successfully captured the obtuse angle of the "L" in MDS results in the ratings task, which is not directly explained

by input feature differences between conditions. The number of Tversky violations also varied in the model unevenly by experimental condition.

The model is able to remember and is affected by the influence of patterns of stimuli across trials (like square versus "L" patterns) due to its long term memory layers. The long term memory trace in the square condition is evenly distributed over many trials. This evenly distributed activation sends back evenly distributed activation boosts to the working memory field across trials. In the "L" condition, however, one arm of the "L" is unevenly dense for each of the two feature dimensions (a different arm for each dimension). Thus, the long term memory traces are lopsided toward one extreme of values. These values get boosted much more often by long term memory than activation elsewhere in the feature dimension, and this can lead to characteristic differences between experimental condition results in the model.

The Fast-Same Effect

The fast-same effect is a tendency in explicit similarity pairwise tasks—primarily same/different tasks—for "same" responses to have lower reaction times than "different" responses. Especially, *identical* objects have the most significant drop in reaction times (Nickerson, 1972). I passed over testing for this effect in empirical data and did not explicitly fit the DNF model to it for several reasons, including it seeming too task-specific as mentioned previously. Additionally, the effect shows best with conjunctive similarity rules and high-dimensional stimuli, not disjunctive rules and two-dimensional stimuli like in Experiment 2 (Farell, 1985). However, I did look for the basic foundation of this effect in modeling processes and results. All tasks and conditions in the model show faster reaction times (model simulation steps before a decision) when more "similar" answers are given. In the ratings task, square condition, similarity rating and reaction time correlated r

= -0.25. In the ratings task, “L” condition, r = -0.30. In the same/different task, square condition, r = -0.29. In the same/different, “L” condition, r = -0.30.

These results overshoot the fast-same effect’s extent in the literature, but the underlying process exists to drive a more realistic fast-same effect with fine tuning. Features that match overlap activation in the model, from the visual field onward. Since all fields have self-excitation feedback loops, higher initial activation from overlapping object features accelerates all peak formation downstream from the visual field. Peaks form more quickly, are detected more quickly, and raise the IOR peaks and attention shifts that trigger a decision more quickly.

Discussion

Overall, modeling results captured nearly every target effect. A notable exception was the distribution of Tversky violations of minimality and triangle inequality among tasks and conditions. These did not fit the pattern of behavioral results, but the model was shown capable of demonstrating both types of violations.

The successful fitting of other effects of MDS group and individual patterns, circular dimension sensitivity, neighborhood density effects, and qualitative differences between experimental conditions, all indicate that the neural processes implied by the architecture of the DNF model are plausible, underlying mechanisms for a variety of object similarity judgments behaviors, generalizing across at least two types of judgment tasks.

Importantly, the model achieved its fits without relying on a neurally implausible multidimensional Cartesian representation. The DNF model instead employs an architecture that requires additional resources only linearly per each dimension added. The model also captured behavioral results while accounting for the stability of object representations as the system autonomously attended to each object. In some cases,

working memory stability was central to the mechanism of similarity judgments. For example, when the model builds working memory peaks, these inhibit and destabilize peaks in the contrast fields, thus changing the relative signals to similarity and dissimilarity nodes and tipping the balance towards a similarity judgment.

All matches to behavior were also detected in an architecture with minimal changes from the model proposed by Schneegans et al. (in press). The only additions to the model were the addition of similarity nodes, the gate node, and the decision field. This implies rich integration of visual cognitive processes with processes of similarity judgments. I expand upon these ties in the conclusions chapter below.

Table 3: Empirical dimension bias and fitted same/different parameters by participant

Participant			
Ratings Square	Color : Shape Ratio	Similarity Node	Dissimilarity Node
1	0.08	0.0875	0.0625
2	0.18	0.1125	0.0500
3	0.09	0.1125	0.0625
4	0.90	0.1125	0.0625
5	0.13	0.1125	0.0625
6	0.71	0.1000	0.0750
7	12.60	0.0875	0.0625
Ratings “L”			
1	0.36	0.1125	0.0750
2	0.12	0.1000	0.0625
3	1.58	0.1125	0.0625
4	0.06	0.1125	0.0625
5	1.17	0.1000	0.0500
6	0.10	0.1000	0.0500
7	0.56	0.1000	0.0500
8	0.32	0.1000	0.0750
9	0.61	0.1125	0.0750
10	1.14	0.0875	0.0750
Same/Diff Square			
1	0.81	0.1000	0.0625
2	4.46	0.0875	0.0500
3	0.89	0.1000	0.0625
4	0.83	0.0875	0.0625
5	1.57	0.1125	0.0750
6	17.86	0.1125	0.0750
7	0.55	0.0875	0.0625
8	3.14	0.0875	0.0500
Same/Diff “L”			
1	0.07	0.1000	0.0750
2	2.11	0.1000	0.0750
3	0.57	0.0875	0.0625
4	10.75	0.0875	0.0500
5	2.26	0.1125	0.0625
6	0.22	0.1125	0.0500
7	0.31	0.1125	0.0625
8	0.37	0.1000	0.0500
9	0.62	0.1000	0.0625
10	1.02	0.1125	0.0625

Table 3, Continued.

SpAM Square	Color : Shape Ratio
1	10.31
2	0.17
3	2.58
4	0.41
5	1.12
6	0.21
7	1.04
8	1.06
9	2.52
10	0.75
SpAM "L"	
1	0.22
2	0.02
3	0.95
4	3.92
5	1.37
6	0.75
7	0.15
8	4.75
9	0.38
10	0.17
11	1.80

Table 4: DNF model parameters.

Global Param		Fields with Global Inhibition	
Sim Tau	5	attention retinal	**-0.8**
Tau Build	500	attention feature	-0.5
Tau Decay	800	attention scene	**-0.1**
kernel cutoff	3	pairwise decision	-3.5

	Initialization Parameters		Connections (from field left of table to field top of table, "[sigma]/[strength]") 1st row = self excitation, 2nd row = self inhibition													
Field	Rest Level	Beta	Attn Retinal	Attn Feat	Attn Scene	IOR	Visual	Contr	WM	Peak	COS	Sim	Dissim	Gate	Pair Decis	LTM
Attn Retinal	-5	2	*4/8* *8/**-1***		2/0.4	4/.5	4/1.5									
Attn Feat	-5	2		*4/**5***	4/2		4/**0.2**	4/**4**	2/**2.6**							
Attn Scene	-5	**2**			*4/**4***			4/**-.75**		**1.3**						
IOR	-5	4	4/-10			*4/20* *8/-17*										
Visual	-5	2	4/0.8	4/**1.25**		4/0.15	*4/7.5* *8/-7.5*	4/**1.25**	**2/0.25**							
Contrast	-5	4		4/1.75				*4/**18*** *8/-20*					fitted			
WM	-5	4		4/1	4/1			**3/-16**	***2/21*** *8/-25*			fitted				1:1
Peak	-5	4								*4*	1/**5**					
COS	-5	4		1/-4		1/**2.6**					*4*					
Sim	-2	1										*2*			10	
Dissim	-2	1											*2*		10	
Gate	-6	1												*0.2*	30	
Pair Decis	-46	1.5													*3/75*	
LTM	0								4/3							

self-connection
changed from Schneegans, et al.

Table 5: A summary of modeling results.

	DNF Model – Ratings		DNF Model – Same/Different	
Test	Square	"L"	Square	"L"
Group MDS	Systematic, no curvature, some shape bias, many value confusions	Systematic, no curvature, slight shape bias, no value confusions	Somewhat noisy, no curvature, no obvious dimension bias, some value confusion	Noisy, curvature, shape bias in one arm, several value confusions
Individual MDS	Same ratios as in empirical data			
Circular Dimension Measure	1.08 chord RMSE 1.48 arc RMSE	1.20 chord RMSE 2.07 arc RMSE	0.146 chord RMSE 0.164 arc RMSE	0.009 chord RMSE 0.147 arc RMSE
Triangle Inequalities	0% of triplets	0% of triplets	0% of triplets	33.2% of triplets
Violations of Minimality	0% of trials	0% of trials	0.05% of trials	0% of trials
Neighborhood Density	r = 0.093	r = 0.221	r = 0.153	r = 0.289

CHAPTER 7

CONCLUSIONS

This dissertation aimed to fill the need in the object similarity judgment literature for a broadly applicable neural-level model of core similarity processes. I first determined a set of similarity judgment behaviors most likely to be indicative of core similarity processes via a set of three representative and diverse similarity judgment tasks and a commons set of stimuli. I examined data from all three tasks for the presence of similarity effects commonly reported in the literature such as attentional tuning and the triangle inequality, and for signatures of the specific stimuli used. Behaviors common to all tasks were most likely to derive from processes performed universally across similarity judgment contexts.

A DNF model of change detection (Schneegans, Spencer, & Schöner, in press) adapted to simulate similarity judgments succeeded in capturing the target task-general behaviors: feature comparison effects, dimensional attention, neighborhood density effects, and qualitative differences between square and "L" conditions. In doing so, the model demonstrated the plausibility of a set of core neural level processes underlying object similarity judgment. In the model, similarity and dissimilarity signals are inherent byproducts of the dynamics of creating and maintaining working memory representations of objects.

The DNF model also captured a number of non-task-general behaviors to which it was not specifically fitted, including violations of minimality, triangle inequalities, better fits for chord-based distance measures, and, tentatively, the fast-same effect. Several possible processes behind these model results were discussed above, but these fits may also imply a relationship or continuum between "task-specific" and "task-general" behaviors. In

this chapter, I address theoretical implications of the empirical results themselves, the importance of the processes and architecture used by the model to best fit those results, and directions for future research.

Implications

Empirical Implications

The three empirical tasks in this dissertation served to establish the task-generality of a variety of similarity judgment behaviors. Task-general behaviors are the most likely indicators of core similarity processes, those relevant to similarity regardless of response format, specific variants of similarity definitions, or situational factors like time pressure. Task-general processes do not solely define similarity judgments and are not necessarily the largest contributors to any given similarity judgment, but they are critical starting points for understanding the processes that underlie similarity judgment, because they unify and connect research efforts across different similarity judgment contexts. Task-general processes in explicit similarity judgments are also likely to be broadly involved in implicit similarity judgments as components of tasks like categorization or word learning.

Overall, the empirical data presented here did reveal a number of common behavioral effects across the three different similarity judgment tasks. The strongest task-general factor in similarity judgments observed across Experiments 1-3 was the influence of feature comparison in similarity judgments. Objects were consistently judged in similarity at least somewhat proportionally to their distances from one another along relevant dimensions in feature space. Previous work has established feature comparison as a task-general factor, in series of experiments with common stimuli and analyses designed to examine this effect (Hout, et. al, 2013; Goldstone, 1994b), and the results of the present experiments replicated these findings. All participants in all tasks and conditions showed

featural organization, aside from four participants who showed disorganized patterns of judgment. Thus, all patterns of similarity judgment behavior in the empirical data correlated at least partially with feature distances, suggesting that feature comparison is an inherent and unavoidable step in object comparison.

Unlike feature comparison, other behaviors tested had not been previously established as task-general, but were observed across the experiments in this dissertation to show consistent involvement in similarity judgment. The most consistent of these were attentional modulation of feature dimensions, neighborhood density effects, and systematic differences between judgments in the square versus "L" conditions. Some degree of consistency was also observed in measures of awareness of circular dimensions, violations of minimality in similarity judgments, and triangle inequalities.

Dimensional attention biases were consistently observed across many individuals in each task tested in this dissertation. Many participants were individually biased to weight either color or shape feature distances as a larger contributor to final similarity judgments, as measured by ratio of average distances between single-feature-step neighbors along one dimension versus the other, taken from the MDS solutions. These results indicate that differences in the amount of influence a particular feature dimensions has in similarity judgments is not exclusively a result of biases in task instructions or a side effect of response type, but rather is a fundamental factor in comparing objects across various task contexts. Despite showing consistent dimensional biases, however, participants were not exclusively biased to either the color or shape dimensions. This suggests that dimensional attention is not only a task-general effect, but that it may not be specific to particular dimensions, either. Given that only two dimensions were tested here, however, data involving a larger number of dimensions is necessary to further explore the generality of

this attentional effect. In particular, the current experiments tested only separable, circular dimensions. This should ideally be expanded to integral and non-circular dimensions in future studies.

Neighborhood density effects were found to completely generalize across tested tasks and conditions, strongly implying that this is a core behavioral pattern in similarity judgment. Neighborhood density is an emergent product of the featural relationships between a large set of objects in a stimulus set because on any given trial of either the pairwise or same/different task only two stimuli were presented. This suggests that long term memory is a strong and consistent influence across many types of similarity judgments. In SpAM, long term memory was not necessarily implicated in the observed neighborhood density effect: since all objects were visible on screen at once; enabling relationships between objects to also have been driven perceptually or through working memory. SpAM does not contradict the definite importance of long term memory in the other tasks, however, and the results do not necessarily imply that long term memory might not have been used in SpAM.

In importance of long term memory also suggested by the fact that the square and "L" conditions affected similarity judgments in all tasks, beyond the differences predicted by the raw feature differences in these two conditions. The "L" condition showed more distortions overall from Cartesian predictions, including noisier fits, feature dimensions that were not always shown by MDS as being orthogonal to one another, and more confusions between neighboring feature values. The two conditions also showed dramatic differences in the numbers of different types of Cartesian violations (minimality and triangle equality) and in shape versus color dimensional attention bias between them. Differences between conditions in the two pairwise tasks necessarily require that

participants were influenced by long term memories, since the difference between the two conditions (beyond raw feature distances) was only detectable across many trials. Again, in SpAM, working memory and ongoing perceptual information may or may not have replaced the role of long term memory in driving these differences between conditions. Additional manipulations would be required to resolve this ambiguity, such as only subsets of objects being arranged in each of several restricted size SpAM trials.

Participants showed awareness of circular dimensions in the pairwise tasks from Experiments 1 and 2.Participants also showed evidence of being aware of circular dimensions in the "L" condition of the SpAM task. These findings suggest that circular dimensions are processed and/or represented in ways that are fundamentally unique compared to linear dimensions. The exact nature or extent of this difference is difficult to assess, however, with data drawn exclusively from circular dimensions. Future studies using identical tasks and analyses but one or more linear dimensions are necessary to better understand the nature and extent of circular dimension processing.

Despite being commonly cited behaviors, triangle inequalities and violations of minimality have not been systematically demonstrated across similarity judgment tasks, and in fact, published evidence of these effects occurring at all is scarce. Shepard (1964) tested for triangle inequality across three tasks, but did not control for common stimuli and suggested that the difference in stimuli may have been the reason for finding inequalities only in his third experiment. The present experiments, however, provide a number of quantitative observations of triangle inequalities and violations of minimality across both pairwise tasks. SpAM, with a Cartesian answer space, could not show these effects within a single trial. Task-generality of these effects remains ambiguous but is a promising possibility. The implications for cognitive processing are less clear than with other

behavioral effects, however. Violations of minimality and triangle inequalities could result purely from noise and can also be explained by a diverse array of cognitive factors suggested by previous similarity models, from salience (Nosofsky, 1991) to distance measures (Pothos, et al., 2013) to mere order of presentation (Tversky, 1977).

Alignability was found to be mostly task-*specific* to the original effects derived from the literature (Markman & Gentner, 1993). Alignability behavior did not consistently generalize to the rated similarity judgment task and was inconsistent between conditions in SpAM. The effect may depend on the higher dimensional stimuli and possibly on tasks involving listing differences, like in the source literature. Regardless of the exact factor(s) missing from Experiments 1 and 3, results suggest that alignability effects in the DNF framework would be at least partially dependent upon a layer of processing beyond the core dynamics between attention, working memory, and long term memory. Just as the pairwise tasks rely on a set of nodes and a decision field for ratings-type responses, verbal tasks like listing differences would likely depend on separate, verbal response fields whose dynamics may drive alignability effects. Listing differences also requires identifying a specific feature value, not just adding across all activation of fields, and the different types of peak detector and localization mechanics involved in this may also contribute to the alignability effect.

Modeling Implications

A primary contribution of the current work is in specifying a neural process model of object similarity judgments. The processes of the model and the parameters that most successfully captured similarity data have theoretical implications for the psychological processes underlying human similarity judgments and related cognitive abilities.

Core Similarity Processes of Feature Comparison. The goal of choosing task-general similarity judgment behaviors as initial modeling targets was to increase the likelihood of discovering "core" similarity processes. A core set of processes was suggested by the DNF model. Feature comparisons most fundamentally drove similarity judgments in the model. This is consistent with existing empirical evidence. Feature comparison is ubiquitous across all models of similarity judgments, even those that share almost no other characteristics. From Pothos, et al.'s (2013) abstract projection model in multidimensional feature space to Kruschke's (1993) connectionist model to Hahn, Chater, and Richardson's (2003) object transformation model, metric differences between features ultimately drive the basis of similarity across the literature. Additionally, the particular dynamics of overlapping activation peaks in the DNF model are consistent with the exponential scaling of similarity as a function of feature distance, which is well-known empirically (Shepard, 1987; Nosofsky, 1991). The density of activity under a Gaussian curve scales exponentially with distance from the center of the peak, so if feature comparison derives from the amount of overlapping neural activity along feature fields, as in the DNF model, this can explain the often observed exponential relationship between raw feature distance and perceived similarity.

In the DNF model, feature comparisons were similarly central to similarity judgment, because they were captured by basic interactions between long term memory, working memory, and perception of features. Overlapping features lead to strong similarity signals, captured in the working memory fields. Features present only in the visual field and not in working memory lead to contrast peaks that signal dissimilarity. The strength of peaks across these fields, the widths of peaks, and the time course of peak formation and decay were all observed to be important to similarity judgments in the model.

This set of processes is remarkable in that it implies that similarity and dissimilarity signals may be automatic and integral to the processing of objects in general. Visual, contrast, and working memory feature fields were not fields unique to this similarity model or the change detection DNF model (Schneegans, Spencer, & Schöner, in press; Johnson, et al., 2009), but are instead part of the basic DNF architecture for object processing. Therefore, similarity may be involved as an automatic byproduct of *any* object-based task like categorization, word learning, or analogical reasoning. This may even be relevant for tasks where featural similarity is not explicitly involved or is distracting to a behavioral goal. For example, participants could be explicitly instructed and rewarded for categorizing objects by their semantic relationships, such as dogs being categorized with leashes and bones and *not* with other animals that look like dogs. Superficial featural similarities and differences between objects could then align with the correct answers or not, and their tendency to facilitate or interfere with performance would indicate the degree of automaticity of feature comparison in non-feature-based object tasks.

Dimensional Attention Process. As discussed in the previous chapter, it is unclear which potential process in the DNF model drives modulation of attention to feature dimensions. The best fits suggest that dimensions are modulated separately from feature comparison and working memory formation, occurring as changes in weights in the connections from the object processing fields to the similarity and dissimilarity decision nodes. If so, this would further support the concept of the feature comparison processes in the DNF being a core, initial process of similarity, only modulated downstream by dimensional biases. This possibility only fits somewhat better than the alternative of dimensional attention occurring at the level of early visual processing, before working memory and feature comparison processes. This distinction is important in establishing the

sequence of processes in object similarity judgments.

It is possible to empirically distinguish the two different dimensional attention processes of node weights versus early visual processing biases. Participants could be run on a similarity judgment task like the pairwise ratings task, then tested for their memories of feature values along different feature dimensions. If attentional modulation of dimensions occurs *before* basic object processing and memory loading, then participants who show strong biases to consider one dimension like shape in similarity ratings task should show the same biases in their memories of the relative scaling of feature distances along different dimensions. If attentional modulation occurs *after* basic object processing, then even participants with strong similarity judgment biases should not show the same biases in their memories of features observed.

Long Term Memory. Long term memory in the DNF model is capable of explaining differences between the square and "L" experimental conditions that go beyond the different sets of pairwise feature comparisons in these two object distributions. In the "L" condition, long term memory can accumulate especially large amounts of activation at one end of each feature dimension that is associated with the arm of the "L" perpendicular to that dimension. If one arm of the "L" is mostly blue objects, then the blue end of the color long term memory field will build up more activation than the orange end. In the square condition, long term memory activation builds evenly on average.

The buildup in the "L" condition can explain incidental differences between the two experimental conditions, but a deeper theoretical possibility is that this uneven long term memory buildup could be the initial basis for passive category formation. Even in situations where categories are not specified, named, or relevant to a task, like Experiments 1-3 here, irregularities and clusters of objects can still exist. When activity clusters in one

part of the long term memory field it can boost and exaggerate similarity ratings of subsequent matching items. This acts like a rudimentary category, implying higher similarity for new matches to the cluster than short term feature comparison alone would predict. This process could be related to the statistical learning known to play a role in, for example, early word learning (Saffran, Asline, & Newport, 1996; Kloos & Sloutsky, 2008).

Consistency with Other Models. The present DNF model of object similarity judgments employs processes consistent with previous DNF models of other cognitive tasks of change detection, executive control, and category learning. The DNF similarity model also relates to processes and architecture from non-DNF models like the feature-integration theory (Treisman & Gelade, 1980) and KRES (Harris & Rehder, 2011). This provides convergent evidence for at least the core object perception and memory interactions suggested critical to similarity judgments in the model.

The object similarity model here is most related to the change detection DNF models from which it was derived (Johnson, et al., 2009; Schneegans, Spencer & Schöner, in press). Although long term memory and the similarity decision nodes and fields were added, and although a number of parameter values were changed in magnitude, no other changes were made to the original architecture and connectivity of the change detection model. The largest parameter change by a large margin was to the strength of the feedback from featural attention to the visual perception field, which was lowered from 1.25 to 0.20. This parameter was lowered to allow more control over attention from the spatial inhibition of return field. However, the feedback connection is still strong enough to also perform the role it did in the change detection model: synchronizing initial attention to a consistent object, rather than spatial and featural fields each being able to attend to different objects. Thus, all of the fundamental steps of object processing remain the same between the two

models, and the working memory and contrast fields effectively code for similarity and dissimilarity, respectively, in both cases.

Buss' (2013) DNF model of executive control in the dimensional change card sort (DCCS) task also shows some of the same processes of attentional modulation and long term memory as the DNF model of object similarity judgments. In the DCCS task, participants sort cards with two-dimensional stimuli according to one or the other of the two dimensions. The rules for which dimension matters switch during the task. Three-year-old, but not 4-year-old, children tend to perseverate on the previous rule after the rule switches. The DCCS model addresses dimensional bias by modulating the overall resting level of space / feature fields. The model boosts the field corresponding to the feature of the current sorting rule. This boosts the likelihood of sorting by that dimension. This is analogous to the attentional mechanism in the present similarity model, where all activity—similarity and dissimilarity signals—in the participant's preferred dimension are boosted relative to that in the non-preferred dimension. The DCCS model is then also influenced by remaining memory traces left over from recent trials. Memories of sorting according to a different rule may or may not outweigh the boost given by the current instructions, and this fact can be leveraged to allow children to overcome their rule perseveration in the DCCS task with memory manipulations (Perone, Molitor, Spencer, & Samuelson, 2014). This is the same process by which long term memory boosts the similarity signal in the present model, and thus is the process by which the square and "L" conditions lead to qualitatively different outcomes from lingering memory traces.

A previous DNF model of taxonomic word learning by (Jenkins, Samuelson, & Spencer, 2015) matches the similarity portion of the feature comparison dynamics in the current model. The word learning model does not explicitly simulate a contrast field and so

does not include an inherent dissimilarity signal, but neural dynamics in working memory function similarly to the current model. The word learning model was developed to capture a behavior called the "suspicious coincidence effect" where a novel category label learned with one exemplar is generalized more broadly than the same label learned with three simultaneously presented exemplars (Xu & Tenenbaum, 2007). The reverse effect occurs for three sequentially presented exemplars, leading to narrower generalization (Spencer, Perone, Smith, & Samuelson, 2012). The DNF model of the suspicious coincidence effect captured the behavior as a function of changes in width between visual and working memory fields. Simultaneously presented objects that match in features interact dynamically via lateral inhibition (see Figure 24) and narrow one another's perceptual representations (see also Schneegans, et al., in press; Johnson, et al., 2009). The narrower projection to working memory then overlaps with fewer features of the test items presented for generalization. The exemplars and test objects are therefore judged to be less similar, and generalization of the novel label does not occur.

The DNF is also consistent with models outside of the DNF literature. Treisman and Gelade's (1980) feature-integration model predicts a system where different separable feature dimensions are represented independently in early visual processing. Most feature dimensions are represented in Triesman and Gelade's model as maps of values of those feature at different points in space. This is a very similar system to feature-space fields in the DNF architecture, although with less potential cross-talk between feature dimensions and fewer dimension-specific resources like individual feature dimension attention fields. Both models achieve the property of avoiding implausible exponential resource usage by treating dimensions in parallel in this way.

KRES (Harrris & Rehder, 2011) also shares a number of architectural similarities to

the present DNF model. KRES is oriented toward capturing categorization behavior, but it simulates neural units with dimensional interaction dynamics similarly to a DNF 1-dimensional neural field. A fully connected dynamic feature dimension allows KRES, like DNF models, to fluidly blend and overlap representations of objects in both strength and shaped patterns of neural activation, providing convergent evidence for the importance of strength/width dynamics in object comparison.

Limitations of Empirical and Model Implications

Stimulus Limitations. In order to isolate the variable of similarity judgment task, stimuli were held constant across all three tasks in this dissertation. The stimuli used were organized by hue and shape. Both of these dimensions were perceptually circular. Both were also separable from one another (Shepard, 1964). It is not a possibility that the behavioral effects are only observed in these dimensions: all of the behaviors studied have previously been observed more generally across feature dimensions. Neighborhood density effects were shown with Morse code, letter glyphs, and musical intervals early in the similarity literature (Krumhansl, 1978); feature comparison is observed across feature dimensions; alignability effects have been observed with stimuli ranging from cartoon creature drawings varying along body part shapes and textures (Goldstone, 1994a) to complex stimulus comparisons like atoms aligned with solar systems based on functional dimensions (revolving motions, Gentner, 1983). These behaviors are therefore not feature dimension specific in any narrow way, such as appearing for color but not orientation.

However, the possibility remains for different *classes* of feature dimensions to be potentially relevant to similarity judgment processes. Researchers often intentionally seek separable dimensions in a task for easier isolation in analysis, so the majority of the previous literature and diversity of dimensions mentioned above has used separable

dimensions within tasks. Integral dimensions tested simultaneously have a strong possibility of revealing different and theoretically important patterns, contradicting the task-generality of specific behaviors, or suggesting new cognitive processes. An example of a set of integral dimensions is color hue and color saturation.

Behavioral similarity judgments along integral dimensions are known to affect similarity judgments in some respects, such as fitting better to Euclidean than to city-block feature distance measures (Shepard, 1964; Shepard, 1987). Integral dimensions also pose theoretical difficulties for existing similarity models of several types. Models with Cartesian feature spaces generally assume that all feature dimensions are orthogonal to one another, but integral dimensions are not completely independent, introducing geometric complications. Similarly, models that tally features (Tversky, 1977; Johannesson, 2000) implicitly assume independent dimensions and do not easily account for integral dimensions.

In the DNF model, each feature dimension is assumed to have a set of attention, contrast, and working memory fields, and feature dimensions are bound across space. Integral dimensions do not clearly fit into this architecture. Different possibilities exist for how to treat integral dimensions. Feature dimensions like hue and saturation may have their own sets of fields, but share additional connections not shared by other feature dimensions. Integral dimensions may also be conflated at an early visual perceptual level, prior to the types of neural processes captured in this DNF model. Alternatively, some small sets of feature dimensions might form higher dimensional fields in the DNF architecture, such as hue by saturation by space fields. These may require millions of neural units rather than thousands, but this is still neurally plausible as long as integral dimensions cluster together in small groups, and the exponential resource usage stops at

exponents of two or three. Addressing these issues will require additional testing using integral feature stimuli, while also controlling for task differences and using common analyses for behavioral effects, as was done here. Modeling efforts can then suggest which of the various neural process possibilities achieves the best fits with the least complexity.

Another limitation based on stimuli is that it is unclear to what extent circular dimensions influenced the outcome of this work, most importantly for the circular dimension awareness tests. Non-circular dimensions like brightness, line thickness, or spatial frequency would be informative as a baseline of comparison in conjunction with or to the exclusion of circular dimensions like hue and shape used in this project. Hout, Goldinger, and Ferguson (2013) used two stimulus sets, one that varied along two linear dimensions and another that varied along a linear and a circular dimension. The feature dimensions were not guaranteed to be perceptually equal, however. This is critical for circular dimension awareness analysis, because if a dimension is not perceptually controlled, feature steps may follow chord-based distances coincidentally for other reasons than circular dimension awareness. Thus, it is important to test feature dimensions that are both varied in circularity and all psychometrically controlled for this analysis.

SpAM Limitations. SpAM limited some of the analyses in this project as a result of its Cartesian constraints. In particular, chord-based distance measures, violations of minimality, and triangle inequalities were impossible to observe in any individual SpAM solution. The Cartesian constraints are not necessarily inherent to SpAM in all implementations, however. If not all objects are presented simultaneously in a single trial, SpAM is capable of showing non-Cartesian similarity judgments within the data from a single participant. Kriegeskorte and Mur (2012) describe a detailed quantitative methodology for approaching this methodology. First, a larger set of objects is chosen than

sixteen. Then a number of subsets of the large object set are presented on different SpAM trials. After each trial, software analyzes the placements online and chooses the next set of stimuli in order to expand upon ambiguous or densely clustered groups of placed items in previous trials. This allows object pairs that were previously underspecified to be more of the direct focus of the next trial, mathematically approaching the least strained overall solution. Kriegeskorte and Mur propose this method as merely an efficient and better fitting means of data collection on large stimulus sets, but the minimally strained final solution space is also ideal for the current project. Minimal strain most closely approaches the unconstrained characteristic of the pairwise similarity judgment tasks. Since objects are distributed over trials in different combinations, this also allows for the same pairs to conflict with themselves or change in rating over time, making behaviors possible like asymmetry effects, triangle inequalities, violations of minimality, and best fits to chord-based distance measures in a single participant's data.

Future Directions

The current DNF model captures a variety of core similarity judgment behaviors. A number of opportunities exist for future improvements and expansion of this core to explain a more comprehensive array of similarity behaviors from a neural process perspective. Both additional empirical data and modeling analyses and processes are implied by the results of the findings from this dissertation.

Task Specific Effects

The current model was adapted to initially capture task-general effects of object similarity judgments, due to task-general effects being most likely to be robust effects driven by core processes appropriate for the first iteration of a neural process model. Task-specific behaviors are equally important for understanding similarity, however, and with a

model anchored to plausible core processes, capturing more nuanced and contextual behaviors is a feasible next step. There are many task-specific similarity behaviors in the literature. Here, I focus on one example mentioned previously in this dissertation: the alignability effect. The empirical data is well-known, and future steps would involve an attempt to capture these effects directly in the DNF model of similarity developed here.

Alignability Effect. The alignability effect (Gentner, 1983; Markman & Gentner, 1993) was inconsistent in the experimental data from chapters 3-5, but remains a consistent and important effect in it's the effect's original context: listing differences between complex naturalistic stimuli. The DNF model could potentially capture the effect consistently under these conditions. If the DNF model were given complex stimuli that are not alignable, it would form a large number of contrast peaks. In the case of complex stimuli, several contrast peaks might form even within single feature dimensions. A dog and a toaster are not just different in texture; they each involve multiple textures, none of which are shared by the other. A texture contrast field may therefore have ten peaks forming at once while comparing these objects. Those peaks can blur and blend, and this can make it difficult to distinctly pull out any one of them or its maximal value cognitively. This would interfere with the task of listing explicit differences.

Objects that are alignable, though, like a hydrogen atom and a solar system, would create far fewer contrast peaks. The ones that *did* still form, like a difference in size, would be sparser and less confused with other peaks as in the non-alignable object example. These differences might therefore be easier to explicitly list due to the better isolation of peaks. This hypothesis could be quantitatively tested with an appropriate motor output simulation for "explicitly naming differences" in the DNF model.

Developmental Trajectories

Dynamic neural field models have a history of capturing developmental trajectories of cognitive phenomena (Buss, 2013; Spencer, et al., 2007; Schutte & Spencer, 2009; Perone, Simmering, & Spencer, 2011; Thelen & Smith, 1994; Thelen & Ulrich, 1991). A possible route for future research in similarity, then, is to apply the present DNF model to investigating how object similarity judgments unfold over development at a neural process level.

Buss (2013), in the DCCS DNF model described above, suggested that the ability to suppress the effect of long term memory activity changes over development, and that this explains an age-based effect in perseverating in the dimensional change card sort task when sorting rules are switched. If the communication between long term memory fields and working memory fields is difficult to suppress in young children, then it is also likely that the communication between working memory fields and contrast fields or visual fields and contrast or working memory fields is more or less difficult to suppress by age. These connections exist along the same axis of communication in the model as long term to working memory connections. The experiments suggested above—designed to look for feature comparison effects in tasks that don't require featural similarity—should also show particularly strong developmental trajectories if this hypothesis is correct. Young children, compared to adults, should show stronger influences of featural similarity in tasks that do not require feature comparison.

The ability to suppress activity between feature fields in the model is also likely to develop gradually over development, since I observed evidence in adults of difficulty suppressing long term memory activation (in a more difficult task than the basic DCCS). In the same/different similarity judgment experiment, participants were influenced by long

term memories of previous trials, as evidenced by some behavioral differences between square and "L" conditions and by neighborhood density effects. Both experimental condition and neighborhood density are irrelevant to task instructions: participants can ignore both and still achieve perfect accuracy in the task. These factors are therefore at best distracting to good performance. The fact that these behavioral patterns were observed in spite of instructions suggests that even the adult participants in Experiment 2 were unable to fully suppress long term memory to working memory connections. Any developmental trend, therefore, is probably not sudden or absolute, and should follow a gradual trajectory.

The spatial precision hypothesis (Spencer, et al., 2007; Schutte & Spencer, 2009) is another developmental concept derived from DNF models that may be relevant to similarity judgments. The hypothesis is that children tend to form and maintain wider, less distinct, less precise peaks of activation, while adults can form either wide or narrow, strong, and precise peaks by comparison. In terms of similarity judgments in the present model, wider peaks should lead to more gradual changes in similarity judgments as objects move further apart in feature space. Very narrow peaks quickly stop overlapping with only small distance from one another along a feature dimension, leading to more of a binary signal. Wide peaks continue to overlap even at large feature value differences, but gradually less so with distance. Wider peaks should also lead to clearer exponential scaling of similarity judgments as feature value distance changes, since exponential scaling is likely due to the dynamics of overlapping Gaussian activity patterns. Wider peaks may also imply more holistic similarity judgments that take into account all feature dimensions at once, because with wider peaks, it is more difficult for two objects to not influence similarity or dissimilarity signals in a meaningful way along every relevant feature dimension: even changes in very different features are likely to overlap in neural fields and

still lead to changes in similarity judgments. A holistic to dimensional shift in similarity judgments over development is a known behavioral effect (Smith & Kemler, 1977), and future modeling work can establish whether spatial precision may explain such a developmental trend from a neural process level. Another possible contributing factor is dimensional attention processes (Perry & Samuelson, 2013), already captured in the DNF model.

Rule-Based Similarity

Sometimes, similarity is defined by explicit rules. The conjunctive and disjunctive definitions of same/different pairwise similarity are examples of explicit similarity rules. In conjunction, similarity is defined as an identical match only, and in disjunction, it is defined as a match along one or more dimensions. The DNF model captures this type of rule with different relative weightings of the connections between neural fields for different features and the similarity and dissimilarity decision nodes. A heavy similarity node weighting allows any one matching dimension to force a "same" decision, representing disjunctive similarity. A weakly weighted similarity node requires the combined activation of matches along all dimensions to drive a "same" decision, representative conjunctive similarity.

Other types of rule-based similarity are possible and common. Categories and taxonomies are often defined and compared according to rules that draw distinctions at specific feature *values*. This is more complex than a dimension-based rule like in the DCCS model (Buss, 2013). Instead of raising the resting level of an entire dimension, only a subset of values must be emphasized to the exclusion of others within a dimension. A "chair" might fall on one side of a width feature dimension, while a "bench" might fall on another, and the two could be judged to be dissimilar even if they match along many other

dimensions like material and color. The differences between categories like benches and chairs could be learned in the model over many individual learned exemplars. The method for representing rules in the DNF as, for example, spoken instructions acted on seconds later, however, is less clear. The resting level of half of a field could be raised, but the neural connectivity may be implausible for this solution. More likely, dynamic activity from a field related to propositional interpretations or relationships could temporarily drive heightened activation in one portion of a feature field but not another.

Overall Conclusions

This dissertation presents the first neural process model of object similarity judgments. The model met its initial goals of demonstrating neural processes underlying object similarity judgments while using plausible amounts of neural resources, capturing real time memory dynamics, and suggesting theoretical connections to related neural process work. In the course of developing and analyzing this model, some broader questions have also been raised about theoretical issues underlying object similarity judgment.

Despite the large amount of empirical and theoretical work in the field, a clear definition of what psychological similarity *is*, exactly, is still elusive. Similarity, even among objects in particular, takes many surface-level forms, from a component of categorization decisions to explicit rating judgments to a basis of known object identification. Currently perceived objects can be similar to one another, as can a perceived object and a remembered one or two or more objects from memory with no perceptual input. Similarity could theoretically be a set of behaviors that result from related but not identical processes across different tasks, or similarity assessment could be a central phenomenon that occurs prior to any task-specific cognitive processes.

The current DNF model offers new evidence as to the exact extent and meaning of object similarity. The model suggests that feature comparison processes are fundamental to object perception itself. Similarity signals in the model result from basic dynamic interactions between attention, working memory, and long-term memory that occur as any perceived objects are loaded into memory. Therefore, a feature-comparison component of object similarity may be central to all behaviors involving object comparison and can potentially be considered a core aspect of object similarity in general.

A full account of object similarity, however, also includes processes more specific to certain types of tasks and contexts. The alignability effect, violations of minimality, triangle inequality, and different means of measuring similarity within circular dimensions are examples of behavioral patterns that seem to arise in only a subset of similarity judgment situations. The distinction between task-general and task-specific processes is not a stark, binary one, however. Most likely, processes exist along a continuum of task-generality. The fact that the DNF model captured a number of similarity behaviors to which it was never fitted suggests that many of these behaviors are closely related to or derivative from core processes like feature comparison and dimensional attention modulation.

Overall, the current investigation suggests that object similarity is a phenomenon that has deep, fundamental roots in object cognition in general, but also still an extensive and diverse set of specialized machinery and processes for specific needs and applications. Future research and modeling work, especially work to capture more task-specific behaviors in the DNF model, will reveal further details about the relationships, shared components, or connections between core and peripheral object processes and the depth, extent, and consistency of similarity perception in the brain.

The DNF similarity model also relates to more general theoretical themes from across cognition, several of which have been revealed by a shared neural implementation between models. Similarity may show broad developmental trajectories as predicted by the spatial precision hypothesis; similarity may rely on executive control and conscious suppression (or lack thereof) of automatic processes like in the task switching literature; and similarity may be implicated in or conflated with statistical learning processes. As the present neural process model of similarity is further developed, it will continue to benefit from and offer insights into a growing unified neural account of featural cognition.

REFERENCES

Ashby, F. G. & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149-178.

Ashby, F. G., Paul E., & Maddox, W. T. (2011). COVIS. In Pothos, E. M. & Wills, A. J. (Eds.), *Formal approaches in categorization* (274-298). New York, NY: Cambridge University Press.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103*(1), 165-192.

Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist? Psychometrika, 56(4), 567-587.

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology, 63,* 516-556.

Baddeley, R., Abbot, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. (1997). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London: Biological Sciences, 264,* 1775-1783.

Belke, E. & Meyer, A. S. (2002). Tracking the time course of multidimensional stimulus discrimination: Analyses of viewing patterns and processing times during "same"-"different" decisions. *European Journal of Cognitive Psychology, 14 (2),* 237-266.

Bindra, D., Donderi, D. C., & Nishisato, S. (1968). Decision latencies of "same" and "different" judgments. *Perception & Psychophysics, 3,* 128-136.

Borg, I. & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications (2nd ed)*. New York, NY: Springer Science + Business Media.

Brighton, H. & Gigerenzer, G. (2008). Bayesian brains and cognitive mechanisms: Harmony or dissonance? In N. Chater & M. Oaksford (Eds.) *The Probabilistic Mind: Prospects for Bayesian Cognitive Science* (pp. 189-208). Oxford, England: Oxford University Press.

Buss, A. T. (2013). *Closing the developmental loop on the behavioral and neural dynamics of flexible rule-use* (Doctoral dissertation). University of Iowa, Iowa City, Iowa.

Buss, A. T. & Spencer, J. P. (2008). The emergence of rule-use: A dynamic neural field model of the DCCS. In *Proceedings of the Thirtieth Annual Conference of the Cognitive Science Society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Chajut, E., Schupak, A., & Algom, D. (2009). Are spatial and dimensional attention separate? Evidence from Posner, Stroop, and Eriksen tasks. *Memory & Cognition, 37*(6), 924-934.

Drucker, D. M. & Aguirre, G. K. (2009). Different spatial scales of shape similarity representation in lateral and ventral LOC. *Cerebral Cortex, 19*(10), 2269-2280.

Erlhagen, W. & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological Review, 109,* 545-572.

Farell, B. (1985). "Same"-"different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin, 98 (3),* 419-456.

Faubel, C. & Schöner, G. (2008). Learning to recognize objects on the fly: A neurally based dynamic field approach. *Neural Networks, 21,* 562-576.

Feldman, J. (2010). Ecological expected utility and the mythical neural code. *Cognitive Neurodynamics, 4,* 25-35.

Garner, W. R. (1974). *The processing of information and structure.* Oxford, England, Lawrence Erlbaum.

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

Goldstone, R. L. (1994a). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20(1),* 3-28.

Goldstone, R. L. (1994b). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26(4),* 381-386.

Grossberg, S., Mingolla, E., & Ross, W. (1194). A neural theory of attentive visual search: interactions of boundary, surface, spatial, and object representations. *Psychological Review, 101*(3), 470-489.

Hahn, U., Chater, N., & Richardson, L. B. (2003). Similarity as transformation. *Cognition, 87,* 1-32.

Harris, H. D. & Rehder, B. (2011). The knowledge and resonance (KRES) model of category learning. In Pothos, E. M. & Wills, A. J. (Eds.), *Formal approaches in categorization* (274-298). New York, NY: Cambridge University Press.

Hering, E. (1964). *Outlines of a theory of the light sense.* Cambridge, Mass: Harvard University press.

Hommel, B. & Colzato, L. S. (2009). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition, 17,* 120-140.

Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient, spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General. 142(1),* 256-281.

Hund, A. M. & Plumert, J. M. (2003). Does information about what things are influence children's memory for where things are? *Developmental Psychology, 39(4),* 939-948.

Hund, A. M., Plumert, J. M., & Benney, C. J. (2002). Experiencing nearby locations together in time: The role of spatiotemporal contiguity in children's memory for location. *Journal of Experimental Child Psychology, 82,* 200-225.

Hurvich, L. M. & Jameson, D. (1957). An opponent-process theory of color vision. *Psychological Review, 64*(6), 384-404.

Itti, L. & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Itti, L. & Koch, C. (2001). Computational modeling of visual attention. *Nature reviews Neuroscience, 2,* 194-203.

Jameson, J., Gentner, D., Day, S., Cristie, S. Colhoun, J., & Bartels, D. (2005). Clarifying the role of alignability in similarity comparisons. *Proceedings of the Twenty-Seventh Annual meeting of the Cognitive Science Society,* 1048-1053.

Jenkins, G. W., Samuelson, L. K., Smith, J. R., & Spencer, J.P. (2015). Non-Bayesian noun generalization in 3- to 5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science.*

Jenkins, G. W., Samuelson, L. K., & Spencer, J. P. (2014). Learning words in space and time: A dynamic neural field account of the suspicious coincidence effect. *Manuscript submitted for publication.*

Johannesson, M. (2000). Modelling asymmetric similarity with prominence. *British Journal of Mathematical and Statistical Psychology, 53,* 121-139.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A dynamic neural field model of visual working memory and change detection. *Psychological Science, 20*(5), 568-577.

Johnson, J., Spencer, J. P., & Schöner, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. *New Ideas in Psychology, 26,* 277.

Johnson, J. S., Spencer, J. P., & Schöner, G. (2009). A layered neural architecture for the consolidation, maintenance, and updating of representations in visual working memory. *Brain Research, 1299,* 17-32.

Klaus, G., Toellner, T., Krummenacher, J., Eimer, M., & Muller, H. J. (2007). Brain electrical correlates of dimensional weighting: An ERP study. *Psychophysiology, 44*(2), 277-292.

Kloos, H., & Sloutsky, V. M (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General, 137,* 52–72.

Kriegeskorte, N. & Mur, M. (2012). Inverse MDS: Inferring dissimilarity structure from multiple item arrangements. *Frontiers in Psychology: Perception Science, 3, Article 246,* 1-13.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial data. *Psychological Review, 85(5),* 445-463.
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 1,* 22-44.
Kubovy, M. & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review, 115(1),* 131-154.
Lee, M. D. & Navarro, D. J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review, 9,* 43–58.
Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 801-820.
Love, B. C., Medin, D. L., & Gureckis, T. M. (2003). SUSTAIN: A network model of category learning. *Psychological Review, 111,* 309-332.
Markman, A. B. & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology, 25,* 431-467.
Maunsell, J. H. R. & Treue, S. (2006). Feature-base attention in visual cortex. *Trends in Neurosciences, 29*(6), 317-322.
Mel, B. W. (1997). SEEMORE: Combining color, shape, and texture histogramming in a neurally inspired approach to visual object recognition. *Neural Computation, 9,* 777-804.
Navarro, D. J. & Lee, M. D. (2004). Common and distinctive features in stimulus similarity: A modified version of the contrast model. *Psychonomic Bulletin & Review, 11(6),* 961-974.
Nickerson, R. S. (1972). Binary-classification reaction time: A review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements, 4*(17), 275-318.
Noles, N. S. & Gelman, S. A. (2012). Effects of categorical labels on similarity judgments: A critical analysis of similarity-based approaches. *Developmental Psychology, 48*(3), 890-896.
Nomura, E. M., & Reber, P. J. (2008). A review of medial temporal lobe and caudate contributions to visual category learning. *Neuroscience and Biobehavioral Reviews, 32*(2), 279-291.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: general, 115*(1), 39-57.
Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23,* 94-140.
Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(2), 324-354.
Perone, S., Molitor, S. J., Buss, A. T., Spencer, J. P., & Samuelson, L. K. (2014). Enhancing the executive functions of 3-year-olds in the dimensional change card sort task. *Child Development,* doi: 10.1111/cdev.12330.
Perone, S., Simmering, V. R., & Spencer, J. P. (2011). Stronger neural dynamics capture changes in infants' visual working memory capacity over development. *Developmental Science, 14,* 1379-1392.
Perry, L. K., Cook, S. W., & Samuelson, L. K. (Unpublished Manuscript). An exploration of context, task, and stimuli effects on similarity perception.
Perry, L. K. & Samuelson, L. K. (2013). The role of verbal labels in attention to dimensional similarity. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuch (Eds.), *Proceedings of the Thirty-fifth Annual Conference of the Cognitive Science Society.*
Perry, L.K., Samuelson, L.K., Malloy, L.M., & Schiffer, R.N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science, 21*(12), 1894-1902.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. *Psychological Review, 120(3),* 679-696.
Recker, K. M. & Plumert, J. M. (2008). How do opportunities to view objects together in time influence children's memory for location? *Journal of Cognition and Development, 9(4),* 434-460.
Recker, K. M., Plumert, J. M., Hund, A. M., & Reimer, R. (2007). How do biases in spatial memory change as children and adults are learning locations? *Journal of Experimental Child Psychology, 98*(4), 217-232.
Richardson, M. W. (1938). Multidimensional psychophysics. *Psychological Bulletin, 35,* 659-660.
Russell, J. A. & Bullock, M. (1985). Multidimensional scaling of emotional facial expressions: Similarity from preschoolers to adults. *Journal of Personality and Social Psychology, 48*(5), 1290-1298.
Sagi, E., Gentner, D., & Lovett, A. (2012). What difference reveals about similarity. *Cognitive Science, 36(6),* 1019-1050.
Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926-1928.
Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *PLoS ONE* 6(12): e28095. doi:10.1371/journal.pone.0028095.
Samuelson, L.K, Spencer, J. P., & Jenkins, G. W. (2013). A dynamic neural field model of word learning. In Gogate & G. Hollich (Eds.), *Theoretical and computational models of word learning: Trends in psychology and artificial intelligence.* Hershey, PA, US: Information Science Reference/IGI Global.
Schneegans, S., Spencer, J. P., & Schöner, G. (in press). Integrating "what" and "where": Visual working memory for objects in a scene. In J. P. Spencer & G. Schöner (Eds.), *Dynamic thinking: A primer on dynamic field theory.* New York, NY: Oxford University Press.
Schutte, Anne R. and Spencer, John P. (2009). Tests of the Dynamic Field Theory and the Spatial Precision Hypothesis: Capturing a Qualitative Developmental Transition in Spatial Working Memory. *Faculty Publications, Department of Psychology.* Paper 496. http://digitalcommons.unl.edu/psychfacpub/496.
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. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function, part II. *Psychometrika, 27(3),* 219-246.
Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54-87.
Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science, 210,* 390-398.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820). 1317-1323.
Shepard, R. N. & Farrell, J. E. (1985). Representation of the orientations of shapes. *Acta Psychologica, 59,* 103-121.
Smith, L. B., & Kemler, D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology, 24,* 279-298.
Smith, J. D. & Nelson, D. G. K. (1984). Overall similarity in adults' classification: The child in all of us. *Journal of Experimental Psychology: General, 113*(1), 137-159.
Spence, I. & Ogilvie, J. C. (1973). A table of expected stress values for random rankings in nonmetric multidimensional scaling. *Multivariate Behavioral Research, 8*(4), 511-517.
Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2012). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science, 22*(8), 1049-1057.

Spencer, J. P., Simmering, V. R., Schutte, A. R., & Schöner, G. (2007). What does theoretical neuroscience have to offer the study of behavioral development? Insights from a dynamic field theory of spatial cognition. In Plumert, J. & Spencer, J. P. (Eds.), *The Emerging Spatial Mind* (pp.320-361). Oxford, UK: Oxford University Press.

Spencer, J. P., Thomas, M. S. C., & McClelland, J. L. (2009). *Toward a unified theory of development.* Oxford, England: Oxford University Press.

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

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9,* 379-385.

Thelen, E. & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge: MIT/Bradford.

Thelen, E. & Ulrich, B. D. (1991). Hidden skills: A dynamic systems analysis of treadmill stepping during the first year. *Monograms of the Society for Research in Child Development, 56*(1), 1-98.

Torgerson, W. S. (1952). Multidimensional Scaling: I. Theory and method. *Psychometrika, 17,* 401-419.

Treisman, A. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-138.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American, 255*(5), 114B-125.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84(4),* 327-352.

Wickelmeier, F. (2003). An introduction to MDS. *Reports from the Sound Quality Research Unit, 7.*

Xu, F. & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245-272.