

UC Riverside

UC Riverside Electronic Theses and Dissertations

Title

What's in a Rule: Two-Dimensional Rule Use in Category Learning

Permalink

<https://escholarship.org/uc/item/70x2x6wk>

Author

LaShell, Patrick J.

Publication Date

2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

What's in a Rule: Two-Dimensional Rule Use in Category Learning

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Patrick Jonathon LaShell

March 2010

Dissertation Committee:

Professor Michael Erickson, Chairperson
Professor Steven E. Clark
Professor Curt Burgess

Copyright by
Patrick Jonathon LaShell
2010

The Dissertation of Patrick Jonathon LaShell is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

The credit for completion of this work and any future successes that I may obtain lies within those many individuals that have supported me.

First and foremost among them is my advisor and colleague, Michael Erickson, whose continued belief and intellectual standards allowed me the time and vision to progress this far.

My academic committees, in all their various iterations, were vital in keeping me progressing. Gratitude is due to John Andersen, Dale Barr, and Curt Burgess. A special note of thanks for Steven Clark, who like Michael, was there for the entire ride.

I also owe many thanks to my peers and research fellows, Kayumari Sanchez, Justin Estep, Elizabeth Melton, Seth Wagerman, and Jesse Breneman.

Appreciation is also due to the research assistants in our laboratory, your contributions are deeply valued.

Yet another special thanks is reserved for the office staff at the University of California, Riverside, without the assistance of the staff, lead by Dianne Fewkes and Faye Harmer, things would have been much more difficult.

Credit is also goes to my undergraduate mentors, Mark Hatala, Michael Seipel, and Karen Smith. Without them, I would not have known where to go. Even if it took me so long, I finally made it.

Not to forget my beginning, I thank my parents, Leo and Barbara LaShell. They brought me into and raised me in this world, without them I would truly be nothing.

Thanks, everyone.

Patrick LaShell

To humanity and science.

ABSTRACT OF THE DISSERTATION

What's in a Rule: Two-Dimensional Rule Use in Category Learning

by

Patrick Jonathon LaShell

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, March 2010
Professor Michael Erickson, Chairperson

Current theories of rules in category learning define rules as one-dimensional boundaries. However, recent evidence by Yang and Lewandowsky (2004) and Lewandowsky, Roberts, and Yang (2006) suggests that rules may also be two-dimensional boundaries. Four experiments are presented that test for two-dimensional rule use in categories with stimuli composed of integral or separable and commensurate or noncommensurate dimensions. Participant categorization behaviors were organized into groups based upon displayed strategies. These groups were modeled by three models of category learning, ALCOVE, ATRIUM, and a version of ATRIUM modified to use two-dimensional rules. Evidence was found supporting two-dimensional rule use in categories containing stimuli with commensurate dimensions.

Contents

List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Dissertation Overview	2
2 Rules in Categorization	3
2.1 Rules in Category Learning	5
2.1.1 ATRIUM	5
2.1.2 COVIS	6
2.1.3 RULEX	7
2.2 A Definition of a Rule	7
2.2.1 Boundaries in Psychological Space	8
2.2.2 Dimensional Interactions: Separable or Integral	9
2.2.3 Rule Verbalizability	10
2.3 Two-Dimensional Rules	13
2.4 Comparisons of One- and Two-Dimensional Rules	15
2.5 Research Questions	18
2.5.1 Predictions of Theoretical Accounts of Rule Use in Categorization	19
2.6 Experiment Design	20
2.7 Model Testing	22
2.8 Overview	23
3 Experiments	24
3.1 Experiment 1: Separable and Commensurate Dimensions	24
3.1.1 Methods	24
3.1.2 Results and Discussion	33
3.2 Experiment 2: Separable and Noncommensurate Dimensions	40
3.2.1 Methods	40
3.2.2 Results and Discussion	43
3.3 Experiment 3: Integral and Commensurate Dimensions	46
3.3.1 Methods	46
3.3.2 Results and Discussion	50

3.4	Experiment 4: Integral and Noncommensurate Dimensions	53
3.4.1	Methods	54
3.4.2	Results and Discussion	55
3.5	Evaluation of Procedure Changes	61
3.6	Summary	61
4	Cluster Analysis	62
4.1	Cluster Analysis Method	63
4.2	Experiment 1: Separable and Commensurate Dimensions	66
4.3	Experiment 2: Separable and Noncommensurate Dimensions	70
4.4	Experiment 3: Integral and Commensurate Dimensions	74
4.5	Experiment 4: Integral and Noncommensurate Dimensions	77
4.6	Overall Experiment Cluster Analysis	79
4.7	Summary	82
5	Mathematical Modeling	84
5.1	ALCOVE	85
5.2	ATRIUM	87
5.3	ATRIUM-DR	89
5.4	Modeling Details	90
5.4.1	Relative Salience	90
5.4.2	Rule Bias Parameters (β_1 & β_2)	90
5.4.3	Search Methods	90
5.5	Experiment 1: Separable and Commensurate Dimensions	91
5.6	Experiment 2: Separable and Noncommensurate Dimensions	101
5.7	Experiment 3: Integral and Commensurate Dimensions	109
5.8	Experiment 4: Integral and Noncommensurate Dimensions	117
5.9	Summary	124
6	General Discussion	126
6.1	Two-Dimensional Rules or Predecisional Integration	127
6.2	Dimensional Attention with Two-Dimensional Rules	133
6.3	The Continuum of Separability and Integrality	134
6.4	Two-Dimensional Rule Commensurability	136
6.5	Extensions to the Current Research	136
6.6	Two-Dimensional Rules and Knowledge Partitioning	137
6.7	Summary	139
	References	140

List of Figures

2.1	Example Categories Using One-Dimensional Rule Boundaries	16
2.2	Examples of Fitness Categories	17
2.3	The Rule and Exception Category Structure	21
3.1	Experiment 1: Participant Learning Criterion	25
3.2	Examples of Experiment Stimuli	26
3.3	Experiment 1: Physical Dimensions of Stimuli	27
3.4	Abstracted Rule and Exception Category Structure	28
3.5	Rule and Exception Category Structures	30
3.6	Category Structures for All Phases of Experiment 1	31
3.7	Experiment 1: Learning	35
3.8	Transfer Predictions for Exemplar-, Rule- and Exception-based, and Mixtures of Exemplar- and Rule-based Accounts of Categorization	38
3.9	Experiment 1: Transfer Responses	39
3.10	Experiment 2: Participant Learning Criterion	40
3.11	Experiment 2: Physical Dimensions of Stimuli	42
3.12	Experiment 2: Learning	45
3.13	Experiment 2: Transfer Responses	47
3.14	Experiment 3: Participant Learning Criterion	48
3.15	Experiment 3: Physical Dimensions of Stimuli	49
3.16	Experiment 3: Learning	52
3.17	Experiment 3: Transfer Responses	53
3.18	Experiment 4: Participant Learning Criterion	54
3.19	Experiment 4: Fourier Descriptors	56
3.20	Experiment 4: Physical Dimensions of Stimuli	57
3.21	Experiment 4: Learning	58
3.22	Experiment 4: Transfer Responses	60
4.1	Examples of Individual Participant Transfer Performance	65
4.2	Experiment 1: Cluster Analysis Plots	67
4.3	Experiment 1: Average Participant Transfer Performance by Cluster	68
4.4	Experiment 2: Cluster Analysis Plots	70
4.5	Experiment 2: Average Participant Transfer Performance by Cluster	71
4.6	Experiment 3: Cluster Analysis Plots	75

4.7	Experiment 3: Average Participant Transfer Performance by Cluster	75
4.8	Experiment 4: Cluster Analysis Plots	77
4.9	Experiment 4: Average Participant Transfer Performance by Cluster	78
4.10	Combined Experiments: Cluster Analysis Plots	80
4.11	Combined Experiment: Average Participant Transfer Performance by Cluster	81
5.1	Experiment 1: Rule-Generalizers: 1-D Conditions Profile Plots	94
5.2	Experiment 1: Rule-Generalizers: 2-D Conditions Profile Plots	95
5.3	Experiment 1: Exception-Generalizers: 1-D Conditions Profile Plots	97
5.4	Experiment 1: Exception-Generalizers: 2-D Conditions Profile Plots	98
5.5	Experiment 2: Rule-Generalizers: 1-D Conditions Profile Plots	104
5.6	Experiment 2: Rule-Generalizers: 2-D Conditions Profile Plots	105
5.7	Experiment 2: Exception-Generalizers: 1-D Conditions Profile Plots	107
5.8	Experiment 2: Exception-Generalizers: 2-D Conditions Profile Plots	108
5.9	Experiment 3: Rule-Generalizers: 1-D Conditions Profile Plots	112
5.10	Experiment 3: Rule-Generalizers: 2-D Conditions Profile Plots	113
5.11	Experiment 3: Exception-Generalizers: 1-D Conditions Profile Plots	114
5.12	Experiment 3: Exception-Generalizers: 2-D Conditions Profile Plots	115
5.13	Experiment 4: Rule-Generalizer: 1-D Conditions Profile Plots	120
5.14	Experiment 4: Rule-Generalizer: 2-D Conditions Profile Plots	121
5.15	Experiment 4: Exception-Generalizers: 1-D Conditions Profile Plots	122
5.16	Experiment 4: Exception-Generalizers: 2-D Conditions Profile Plots	123
6.1	An Information Processing Model of Categorization	128

List of Tables

2.1	Overview of the Requirements for Valid Rules	8
2.2	Predictions of Theoretical Accounts for Evidence of Two-Dimensional Rule Use	20
3.1	The Breakdown of Participants by Condition and Learning Criterion for Experiment 1	24
3.2	The Breakdown of Participants by Condition and Learning Criterion for Experiment 2	41
3.3	The Breakdown of Participants by Condition and Learning Criterion for Experiment 3	47
3.4	The Breakdown of Participants by Condition and Learning Criterion for Experiment 4	55
4.1	Division of Participants into Clusters in Experiment 1	66
4.2	Division of Participants into Clusters in Experiment 2	72
4.3	Division of Participants into Clusters in Experiment 3	74
4.4	Division of Participants into Clusters in Experiment 4	78
4.5	Division of Participants into Clusters Across Experiments	80
5.1	Summary of ALCOVE Parameters	87
5.2	Summary of ATRIUM Parameters	89
5.3	ALCOVE Fits for Experiment 1	91
5.4	ATRIUM Fits for Experiment 1	92
5.5	ATRIUM-DR Fits for Experiment 1	93
5.6	ALCOVE Fits for Experiment 2	101
5.7	ATRIUM Fits for Experiment 2	102
5.8	ATRIUM-DR Fits for Experiment 2	103
5.9	ALCOVE Fits for Experiment 3	110
5.10	ATRIUM Fits for Experiment 3	110
5.11	ATRIUM-DR Fits for Experiment 3	111
5.12	ALCOVE Fits for Experiment 4	118
5.13	ATRIUM Fits for Experiment 4	118
5.14	ATRIUM-DR Fits for Experiment 4	119
5.15	Summary of Model Fits	125

Chapter 1

Introduction

Categorization is the process of assigning group membership to stimuli. This process allows organisms to react to stimuli based on previous experiences with all group members, rather than only the immediate stimulus. Thus when a child with a dislike of green vegetables encounters asparagus for the first time, the child is able to act upon their knowledge of green vegetables and reject it immediately, rather than having to taste the new food. While problematic for parents in this instance, categorization gives people the ability to distinguish edible from poisonous, friendly from dangerous, and boys from girls, and it is vital for survival.

Current theories of categorization hold that categorization is best explained by multiple systems of categorization working together rather than a single monolithic system (Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994). One proposed system of categorization is rule-based categorization (Bruner, Goodnow, & Austin, 1956; E. E. Smith & Medin, 1981; Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994). In the following dissertation I address what *rule-based* means within perceptual category learning research and attempt to expand its definition to account for new phenomena while retaining a strict and clear meaning.

Current theories of categorization that incorporate rules (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994) are heavily influenced by the classical theory of categories (Bruner et al., 1956; E. E. Smith & Medin, 1981). The classical theory defines rules as propositions operating on single features, creating well defined category boundaries, and requiring explicit awareness. These propositions, that relate category membership to a single feature along a psychological dimension, will be referred to as *one-dimensional rules*.

Recent research has provided evidence that the rules used in category learning may not be

restricted to one-dimensional rules (Yang & Lewandowsky, 2004; Lewandowsky et al., 2006). In these category learning studies, a model using *two-dimensional rules*, rules incorporating information from different dimensions simultaneously rather than sequentially, were better able to explain their data than a model using *one-dimensional rules*, which incorporate information from multiple dimensions in a sequential fashion. This finding suggests that current accounts of rules in category learning, which rely upon one-dimensional rules, may be insufficient in explaining category learning behaviors. An update to rule-based theories that allows for the use of two-dimensional rules may be necessary to account for these findings.

1.1 Dissertation Overview

This dissertation addresses the nature of rules in categorization. Specifically, I address whether two-dimensional rule use is possible and the conditions that allow for it. This dissertation is organized as follows. Chapter 1 is this short introduction. Chapter 2 addresses the circumstances under which two-dimensional rule use may be possible. In Chapter 2 two-dimensional rules are introduced, compared, and contrasted against one-dimensional rules. Two factors that have been shown to influence category representations, the type of relationship between the stimulus dimensions and verbalizability of the stimulus dimensions, are introduced and discussed. In addition, current theories of rule use in category learning are described and predictions are formed to distinguish the theories with respect to dimensional separability and verbalizability. Chapter 3 presents experiments that test the predictions made in Chapter 2. Chapters 4 and 5 cover the clustering and computation modeling of the experiments. Lastly, Chapter 6 summarizes the results of the experiments, includes my closing arguments for the expansion of the term *rule* to include two-dimensional rules, and posits theoretical implications resulting from this expansion.

Chapter 2

Rules in Categorization

Categorization is the mental process by which stimuli are assigned membership to a group with other stimuli. Making category assignments allows for the generalization of category properties to the new stimuli and increases the available information for further inference generation. Decisions can then incorporate information from previous experiences with all group members, rather than a single potentially unfamiliar stimulus. Contemporary researchers have suggested that human categorization behaviors are a combination of different systems of categorization (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994). Systems based on rules, exemplars, and procedures have all been proposed to account for categorization behaviors. Distinguishing between these different systems relies on having well defined theories that are continually informed and updated by empirical findings.

Recently, researchers have questioned the assumptions of current rule-based theories of categorization (Yang & Lewandowsky, 2004; Lewandowsky et al., 2006). Rule-based theories of categorization describe rules as providing necessary and sufficient conditions for category membership. Examples of typical rules are: *animals that have wings are birds*; *men over six feet tall are tall*; and *little green men are Martians*. Current theories of categorization that use rules have also defined rules as specifying potential category membership on the basis of *individual* stimulus features (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994). Recent evidence has been provided that people may be using rules that combine information from *multiple* stimulus features (Yang & Lewandowsky, 2004; Lewandowsky et al., 2006).

In Yang and Lewandowsky (2004), participants were given a category learning task devised to demonstrate *knowledge partitioning*. Knowledge partitioning is the idea that people do not have a unified homogeneous understanding of the world, but instead possess distinct packets of

information. Furthermore, these packets are difficult to access when not the focus of our cognitive efforts. Thus, our knowledge of fruit pies is isolated from our knowledge of π . While knowledge partitioning does not play an important role in this dissertation, the experiments and category structure used by Yang and Lewandowsky (2004) are relevant.

Yang and Lewandowsky's (2004) experiments used a two-dimensional category structure with diagonal boundaries separating the members of different categories. These boundaries separated the categories, not along a single dimension, but along a combination of two dimensions. Furthermore, the category structures were composed of two separate substructures which were distinguished by a context cue. During the transfer phase, participants displayed two different strategies. Knowledge partitioning participants were sensitive to the context cue and categorized stimuli according to the substructure that matched the context cue. True-boundary participants were not sensitive to the context cue and categorized stimuli according to the nearest substructure (ignoring the context cue).

Exemplar accounts of category learning can readily account for category structures with diagonal boundaries (Kruschke, 1992; Nosofsky, 1986). However, an exemplar model of category learning (ALCOVE: Kruschke, 1992), was unable to produce the knowledge-partitioning pattern of behavior. ALCOVE was unable to learn to use the context cue to differentiate the substructures. This was due to ALCOVE only learning when cues are predictive of category membership. The context cue was not predictive of category membership, because each category response was equally likely to occur with either context cue. In contrast, a rule using model (ATRIUM: Erickson & Kruschke, 1998) modified to use rules that operate on two dimensions, instead of only one, was able to explain these data.

To further challenge the assumptions of current rule-based theories, Lewandowsky et al. (2006) reused the category structure from Yang and Lewandowsky (2004) with two diagonal boundaries to further explore knowledge partitioning. In two experiments, Lewandowsky et al. (2006) tested the ability of participants to form knowledge partitions in category structures using stimuli that varied on integral, separable, verbalizable, and nonverbalizable dimensions. Stimuli possessing these types of dimensions have been found to change how participants learn to categorize these stimuli. These types of dimensions and their effects upon category learning will be discussed in more depth later on, for now keep in mind that they have been shown to change category learning behaviors. Lewandowsky et al. (2006) found that participants were able to form knowledge partitions regardless of the type of underlying dimension. They also found that the participants' categorization behaviors were similar those of the participants from Yang and Lewandowsky (2004),

which were found to be best fit by a modified model that conflicts with current theories of rule-based categorization.

In summary, current theories of rule-based categorization state that rule-based category representations operate on single features. When rules operate on multiple features, each feature is processed individually by a one-dimensional rule, and then the result of each one-dimensional rule is combined to make a final decision. Yang and Lewandowsky (2004) found that their data was better explained by a model incorporating two-dimensional rules rather than the one-dimensional rules supported by current models of categorization. Further experiments by Lewandowsky et al. (2006) using different stimuli were also found to support two-dimensional rule use. The findings of Yang and Lewandowsky (2004) and Lewandowsky et al. (2006) contrast with current theories of rule-based categorization that only support one-dimensional rules. In light of these findings, direct and more extensive tests of rule use are appropriate, namely whether people can use two-dimensional rules and under what conditions.

2.1 Rules in Category Learning

Current theoretical accounts of rules in categorization are based on the classical theory of rules developed by Bruner et al. (1956). In the classical theory, rules are characterized as necessary and sufficient conditions for category membership and are a result of explicit reasoning and hypothesis testing. The classical theory of rules in categorization has been demonstrated to be insufficient to account for many phenomena of category learning such as: fuzzy categories based on typicality (Rosch & Mervis, 1975), prototypes effects (Posner & Keele, 1968, 1970), and family resemblance overriding explicitly given rules (Allen & Brooks, 1991). However, some successful multiple-system models of category learning have included rule-based subsystems (e.g., ATRIUM: Erickson & Kruschke, 1998; COVIS: Ashby et al., 1998; RULEX: Nosofsky et al., 1994). To evaluate the current state of rules in categorization, I will characterize rule-based theories on the basis of rule use in these three models.

2.1.1 ATRIUM

ATRIUM (Attention To Rules and Instances in a Unified Model: Erickson & Kruschke, 1998) is a hybrid rule- and exemplar-based model of category learning instantiated in a connectionist model framework. ATRIUM uses a mixtures-of-experts approach to category learning. It develops different representations, experts, modules, or subsystems to master particular types of

stimuli and then learns to weight the contributions of each representation appropriately for the various stimuli. The ability of the model to match the optimal representation for a specific type of category structure is termed *representational attention*. Additionally, the model possesses *dimensional attention* (Kruschke, 1992; Nosofsky, 1986) which allows it to learn which dimensions are relevant for categorization. Learning in the model is *error-driven* with incorrect responses causing changes in representations.

ATRIUM contains two types of modules, an exemplar module, specifically a version of ALCOVE (Kruschke, 1992), and at least one rule module. Erickson and Kruschke (1998, p. 107) constrain rules in ATRIUM to “a boundary that is orthogonal to a [single] psychological dimension.” Typically there is one rule module for each of the dimensions in the category structure. When a stimulus is presented to the model, it is processed simultaneously by both rule- and exemplar-based modules. Each of the modules produces a category response, and then a gating module combines each of the other modules’ responses to make a final category response. When feedback is given, the modules learn based upon the contribution of each module to the final response and the difference between the final response and the correct response. The more a module contributes to the final response, the greater the amount of learning that occurs in the module.

Thus, rules in ATRIUM are defined as one-dimensional boundaries in psychological space perpendicular to a single psychological dimension. There are no further assumptions underlying their representation in ATRIUM. Yang and Lewandowsky (2004) modified ATRIUM to use a two-dimensional rule, however this is not part of the original model.

2.1.2 COVIS

COVIS (COmpetition between Verbal and Implicit Systems: Ashby et al., 1998) is a hybrid model of categorization combining an explicit rule hypothesizing system, relying upon executive functions, with an implicit procedural system. Rules in COVIS are verbalizable rules. Ashby et al. (1998, p. 446) define verbalizable rules as strategies that can be verbally described that define category membership on the basis of an attentionally attendable stimulus property possessing a semantic label. Ashby et al. (1998) accept the difficulty in defining the precise set of verbalizable rules, and characterize unidimensional rules as a large subset of possible verbalizable rules. A unidimensional rule is a “rule that uses a decision bound that is orthogonal to some stimulus dimension” (Ashby et al., 1998, p. 446).

The implicit system in COVIS is also represented by boundaries in psychological space. However the implicit system is not restrained to simple one-dimensional rule boundaries as the explicit system is restrained. Instead the implicit system is allowed to create boundaries as needed.

In conclusion, rules in COVIS are defined as one-dimensional boundaries in psychological space orthogonal or perpendicular to a single psychological dimension. Furthermore, they are required to be verbalizable.

2.1.3 RULEX

RULEX (RULE-plus-EXception model: Nosofsky et al., 1994; Nosofsky & Palmeri, 1998) is a hybrid model of categorization combining a rule-based system with an exemplar-based system. RULEX attempts to learn categories using hypothesis testing. RULEX searches for an optimal rule and then learns exceptions to that rule. Rules in RULEX are one-dimensional boundaries in psychological space that are perpendicular to a single psychological dimension.

To learn a category structure, RULEX first attempts to find a single one-dimensional rule that perfectly solves a category structure. If RULEX fails to find this perfect rule, RULEX then attempts to find a one-dimensional rule with accuracy above a performance criterion. If no one-dimensional rule reaches the performance criterion, RULEX then switches to testing conjunction rules. After finding a rule with accuracy above the performance criterion or failing to find a good rule after exhaustively testing all conjunction rules, RULEX then switches to storing items that are not classified correctly as exceptions.

Thus, rules in RULEX are defined as one-dimensional boundaries in psychological space perpendicular to a single psychological dimension. Additionally, rules in RULEX are the result of explicit hypothesis testing.

2.2 A Definition of a Rule

Combining how ATRIUM, COVIS, and RULEX instantiate rules provides the basis of a definition of rules in current accounts of category learning. Table 2.1 contains this definition of a rule. While the models represent and use rules in different ways, they are consistent on what constitutes a rule. They define a *rule* as a single proposition relating potential category membership to a single feature or dimension; that is, *rules* are *one-dimensional rules*.

According to these models, a one-dimension rule is a proposition that assigns a category label in relation to a single psychological dimension. Furthermore these models also require that

Table 2.1: Overview of the Requirements for Valid Rules

Requirements for Valid One-dimensional Rules	
One-Dimensional	The rule forms a boundary operating along a single dimension.
Separable Dimension	The relevant dimension is separable from other stimulus dimensions.
Verbalizable	The rule fulfills the three requirements for rule verbalizability.
Requirements for Rule Verbalizability	
Separable Dimension	The relevant underlying dimension is separable.
Semantic Label	The relevant underlying dimension has or can have a valid semantic label.
Semantic Relationship	The boundary is formed according to a relationship that has or can have a valid semantic label.
Requirements for Valid Two-dimensional Rules	
Two-Dimensional	The rule operates on two dimensions.
Separable Dimensions	The relevant dimensions are separable.
Verbalizable	The rule fulfills the three requirements for rule verbalizability.
Commensurate	The dimensions are measured in the same units.

this dimension must be attentionally selectable or separable from other dimensions. Additionally, COVIS has a third requirement for rule membership: rules must be verbalizable (e.g., Ashby et al., 1998). RULEX also requires rules to be formed as a result of hypothesis testing, and thus may also implicitly require the commensurability assumption.

2.2.1 Boundaries in Psychological Space

As defined thus far, *one-dimensional rules* are a subset of decision-bound theories. They can be modeled in psychological space as linear boundaries, orthogonal to a psychological stimulus dimension. Furthermore, one-dimensional rules partition psychological space into category response regions.

Decision-bound theories of categorization represent categories as response regions in a multi-dimensional psychological space (Ashby & Gott, 1988; Ashby & Townsend, 1986). Cat-

egory relevant features are used to form a psychological space that is partitioned into different response regions by decisional boundaries. When an item is presented, it is mapped to a specific location in psychological space corresponding to a response region and a category label is retrieved. One-dimensional rules are a subset of decision-bound theories in which rule boundaries must be orthogonal to a single psychological dimension. Thus in a typical two-dimensional category structure representation, with the underlying psychological dimensions as the axes, rule decision boundaries are represented by either horizontal or vertical lines.

2.2.2 Dimensional Interactions: Separable or Integral

Dimensional interaction refers to the type of relationship that exists between two stimulus dimensions (Lewandowsky et al., 2006), either integral or separable (Ashby & Maddox, 1990; Ashby & Townsend, 1986; Garner, 1970, 1974; Nosofsky, 1986; Shepard, 1957, 1964). This type of relationship exists only when relating one dimension to another dimension; it does not exist in isolated dimensions. Separable dimensions are dimensions that can be perceived and operated on independently of each other. Prototypical examples of separable dimensions are brightness and size (e.g., Garner, 1977). Integral dimensions are dimensions that are difficult or impossible to attend to separately. Prototypical examples of integral dimensions are saturation and brightness (e.g., Garner, 1970). Dimension pairs can be classified as separable or integral using multi-dimensional scaling (MDS; Dunn, 1983; Erickson & Kruschke, 1998; Shepard, 1957), filtration and condensation tasks (Nosofsky & Palmeri, 1997), speeded classification tasks (Ashby & Maddox, 1990), and several other methods (for a more detailed review of this literature, see Maddox, 1992).

Stimuli that vary along separable and integral dimensions have been found to generate different types of categorization behaviors (Ashby, Ell, & Waldron, 2003; Nosofsky, 1986; Waldron & Ashby, 2001). For example, Shepard and Chang (1963) found that performance in category learning of stimuli with integral dimensions could be predicted from confusions in identification, whereas the performance in the category learning of stimuli with separable stimuli could not. Similarly, Nosofsky and Palmeri (1996) performed an experiment based on the seminal task and category structure developed by Shepard, Hovland, and Jenkins (1961) and found that the relative difficulty ranking of category structures varied depending on whether the category relevant dimensions were integral or separable.

For a dimension to be used in a one-dimensional rule, it must be a valid source for an orthogonal boundary. This requires the dimension to be perceived independently of other dimensions, and therefore must be a separable dimension. Such separable dimensions can then used as

the origin of the orthogonal rule boundary. In contrast, integral dimensions are not perceived independently of each other. The psychological representation of a specific value along one integral dimension changes with changes in the second integral dimension. Therefore, representationally sensible orthogonal boundaries in psychology space cannot occur. Thus one-dimensional rules require separable dimensions.

2.2.3 Rule Verbalizability

The last property of a rule, that rules must be verbalizable, is not supported by all three representative models. COVIS and RULEX state that rules are the result of explicit hypothesis testing, while ATRIUM remains agnostic. If rules are verbalizable and the result of an explicit reasoning processing, they must be, and contain, concepts and relationships that can be verbalized.

Evidence that supports rule use requiring explicit reasoning comes from a variety of neurological, behavioral, and self-report sources. Evidence for rule use occurring with explicit reasoning comes from self-report data in categorization tasks. People are able to describe the rules that they used while doing the task. Likewise, when people are given rules to follow, categorization behaviors are better described by rule theories than exemplar theories (e.g., Nosofsky, Clark, & Shin, 1989; although see Allen & Brooks, 1991 for an example of family resemblance overriding rule use). Additionally, people can follow rules even without reinforcement, whereas performance in non-rule-based tasks have been found to be impaired with delayed or absent reinforcement (Maddox, Ashby, & Bohil, 2003).

Further evidence that supports rule use requiring explicit reasoning comes from Waldron and Ashby (2001). Waldron and Ashby (2001) found that rule-based category learning tasks are more susceptible to disruptions caused by increased demands on working memory and executive attention, than similar non-rule-based category learning tasks. Waldron and Ashby (2001) had participants engage in a two-part experiment. In the first part, the practice phase, participants were taught two variations of two different categorization tasks. The tasks consisted of learning either a one-dimensional rule-based category or a three-dimensional non-rule-based category using items that were composed of four binary features.

The one-dimensional rule-based category was similar to the Type I category first tested by Shepard et al. (1961). For the one-dimensional rule-based categories, items were classified based on the presence of a single feature. If an item had value A on a feature, it was a member of category A. If an item had value B on a feature, it was a member of category B. The three-dimensional non-rule-based category was similar to the Type IV category first tested by Shepard et al. (1961). For

the three-dimensional non-rule-based categories, items were classified based on the presence of two or three features that matched a three dimensional category prototype (one dimension was set as irrelevant).

One week after the training phase, participants were given the testing phase. In the testing phase, participants were given four new variations of the rule- and non-rule-based categories to learn (two of each). In the control condition, participants were not given any additional tasks. In the experimental condition, participants were given a numerical Stroop task to perform simultaneously with the category learning task. In the numerical Stroop task condition, two numbers were displayed along side the category item for the first 200 ms of each trial. Participants were to remember the numerical value and size of the two numbers. After classifying the category item, participants are prompted with either *size* or *value*, signaling which number they were to recall.

Performance in the one-dimension rule-based category learning task was found to be more impaired by the Stroop task than the more difficult three-dimensional non-rule-based category learning task. This suggests that the mechanisms underlying rule-based category learning are the same mechanisms involved in the numerical Stroop task. These mechanisms have been hypothesized to be part of brain structures that have been linked to selective attention and working memory functions, the same functions that underly executive reasoning (Ashby et al., 1998). This finding suggests that rule-based category learning, regardless of the difficulty of the rule, requires explicit reasoning resources. Furthermore, rule-based category learning requires more explicit reasoning resources than non-rule-based category learning.

Additional evidence that rule use requires explicit reasoning comes from studies with people who have Parkinson's disease and from studies with people who have amnesia. Ashby, Noble, Vincent, Waldron, and Ell (2003) found that participants with Parkinson's disease were highly impaired when learning categories defined by a rule. Parkinson's disease has been found to impair explicit reasoning. However, these individuals with Parkinson's disease did not perform differently than age-matched control participants on non-rule-based tasks. This suggests that rule-based category learning performance is impaired by a loss of explicit reasoning. In contrast, people with amnesia caused by damage to the medial temporal lobe, but possessing functional working memory and executive attentional systems, display the opposite pattern. They have been found to perform normally on rule-based categorization tasks while being impaired on exemplar-based categorization tasks (Janowsky, Shimamura, Kritchevsky, & Squire, 1989; Leng & Parkin, 1988). This suggests that rule-based category learning performance is not as reliant upon explicit memory systems. Together, these examples suggest that rule use requires explicit reasoning and is less

dependent upon long term memory.

Accepting these studies as providing sufficient evidence for the requirement of rule verbalizability, the next step is to specify the requirements for a verbalizable rule. Ashby and Maddox (2005) claim that for a rule to be verbalizable, three properties are required. First, the rule operates on a psychological stimulus feature that possesses (or has the ability to possess) a valid semantic label. Second, the stimulus feature can be attended to in isolation (i.e., the feature can be represented as a point along a single separable psychological dimension). Third, the relationship relating category membership to a stimulus feature must contain an operator that must also possess (or have the ability to possess) a valid semantic label.

An example of a rule that is verbalizable would be: *tall men are more than six feet in height*. First, the stimulus feature that the rule operates on is *height* and it is a readily understood concept that possesses a valid semantic label, thus the stimulus feature is verbalizable. Second, the stimulus feature, again *height*, can be attended to independently from other dimensions, thus it is also a separable dimension. Third, the relationship determining category membership *more than* is a readily understood concept that possesses a valid semantic label, hence it too is verbalizable. Therefore, according to Ashby and colleagues (Ashby et al., 1998; Ashby & Maddox, 2005), this rule qualifies as a verbalizable rule.

However, the verbalization requirement is not without contention. For example, one objection to the verbalizability requirement for rules comes in the vein of the criticisms of the Sapir-Whorf hypothesis: Do people's rules follow the form of their underlying representations, or do people's underlying representations follow the form of their rules? In other words, is the semantic system the source of rules and thus passes on its limitations to rules, or does another system create the rules which are then moderated by the semantic system? If the semantic system only moderates the rules that people use, then it is possible that rules are not required to be verbalizable. It is necessary to verify this assumption.

One way to verify the verbalizability assumption requires characterizing the full sets of all categories with verbalizable (and potentially verbalizable) rules and of all categories with nonverbalizable rules. The categories that people can and cannot learn may provide evidence on the nature of the relationship of people's rules, semantic systems, and representations. Finding categories with nonverbalizable rules that people can learn would support representations guiding rules, whereas failing to find learnable categories with nonverbalizable rules would support rules guiding representations. This search must be exhaustive because it may be the case that only some categories with nonverbalizable rules are learnable by people whereas others are not. Likewise, it may be the case

that some categories with verbalizable rules are not learnable by people, in which case there may be even more limitations on the human categorization system than previously thought.

However, this is not a feasible solution to characterizing learnable rules because an exhaustive search of this nature may not be possible. It is unlikely that there is a way to be certain that the entire set of all dimensions and rules have been found and accurately divided into verbalizable and nonverbalizable. Thus an exhaustive search of this nature would be an inefficient use of researchers' time and resources.

Two more feasible approaches to the problem of identifying verbalizable rules comes from Ashby et al. (1998) and Erickson and Kruschke (1998). Ashby et al. (1998) acknowledge the difficulty in defining the complete set of verbalizable rules and take the tactic of identifying subsets of verbalizable rules and nonverbalizable rules. In this approach, experiments are designed using two types of categorization structures. The first type of structure can be represented using verbalizable rules. The second type of structure is assumed to be very unlikely or impossible to represent using verbalizable rules, and instead is represented by a nonverbal system. In COVIS this system is procedurally based, while its counterpart in ATRIUM is exemplar based, and its counterpart in RULEX is the storage of exceptions (which are exemplar-like, but not complete exemplars). These two types of category structures can then be used to compare and contrast human performance on verbalizable and nonverbalizable rule structures.

Erickson and Kruschke (1998) provide an alternative solution to the problem of identifying verbalizable rules. Erickson and Kruschke (1998) explored other facets of category learning while limiting their experiments to category structures that could be represented using verbalizable rules, thereby avoiding the necessity to validate the verbalizability assumption. This is a conservative approach that does not take on the extra assumption of restricting rules to verbalizable concepts and relationships as does Ashby and colleagues' approach. The results of Erickson and Kruschke's approach should apply regardless of the eventual resolution of rule verbalizability requirements.

2.3 Two-Dimensional Rules

In contrast to the volume of research on one-dimensional rules in category learning models (Ashby et al., 1998; Erickson & Kruschke, 1998; Nosofsky et al., 1994), only Ashby and colleagues (e.g., Ashby et al., 1998; Ashby, Ell, & Waldron, 2003) have proposed tentative requirements for two-dimensional rules. The requirements for two-dimensional rules are extensions of the requirements for one-dimensional rules. Like one-dimensional rules, two-dimensional rules are proposi-

tions that assign category membership in relation to psychological dimensions, must be composed of separable dimensions, and be verbalizable. Additionally, Ashby and Maddox (2005) proposed two fundamental differences between one- and two-dimensional rules: (1) two-dimensional rules combine information from two dimensions, and (2) the dimensions used in two-dimensional rules must be *commensurate* (i.e., be measured in the same units).

The first difference between one- and two-dimensional rules proposed by Ashby et al. (1998) is that the combination of information from the two underlying dimensions in two-dimensional rules occurs in a predecisional information integration stage. In this stage, information from two psychological dimensions are combined. The decision criterion operates on this combination of values instead of the individual psychological dimensions. This predecisional information integration stage occurs before the decision rule is applied. Later stages may or may not have access to the underlying dimensions in isolation. That is to say, the combination of the two underlying dimensions may create a new dimension and prevent further independent processing of either of its two component dimensions.

The second difference between one- and two-dimensional rules proposed by Ashby et al. (1998) is the additional requirement of *commensurate* dimensions for two-dimensional rules to be verbalizable. For two dimensions to be combined in a two-dimensional rule, it is necessary that that the dimensions must be measured in the same units or be *commensurate*. To further illustrate the idea of commensurate dimensions, consider the following examples. The rule *a rectangle is in category A if it is at least twice as tall as it is long* is a valid verbalizable two-dimensional rule according to Ashby et al. (1998). This rule incorporates the separable dimensions of height and width. The rule incorporates a verbalizable relationship between the dimensions, *at least twice as tall as*. Lastly, since both dimensions are measures of distance, they are commensurate. This fulfills all of Ashby et al.'s (1998) requirements for a valid two-dimensional rule.

In contrast, according to Ashby et al. (1998) the rule *a rectangle is in category B if it is twice as red as it is long* is not a valid verbalizable rule. This rule incorporates the separable dimensions of height and color. The rule also incorporates a verbalizable relationship between the dimensions, *at least twice as tall as*. However, since these two dimensions are measured in different types of units, distance and hue, they are not commensurate. This fails to fulfill all of the requirements for a valid two-dimensional rule (Ashby et al., 1998), and thus cannot be used to form a valid two-dimensional rule.

A later account by Ashby, Ell, and Waldron (2003) provides an alternative to the original account by Ashby et al. (1998). They argue that rules can also be created by abstracting stimulus

dimensions, which satisfies the requirement of commensurate dimensions. Ashby, Ell, and Waldron (2003) provide the example rule, “respond A if the stimulus is small on dimension x and small on dimension y”(p. 1115). In this example rule, the two new psychological dimensions that are formed are abstractions of the original stimulus dimensions. So for example, a dimension that represented *redness* of a stimulus could be created measuring the saturation level of the color red in the stimulus. Pink stimuli would have a small value of *redness*, while blood red stimuli would have a large value of *redness*. These dimensions would be standardized, and therefore unitless. By allowing for the creation of new dimensions, the category structure’s representation in psychological space is converted to a new psychological representation. This allows for two-dimensional rules because these new unitless dimensions are necessarily commensurate. In allowing the creation of new psychological dimensions, the verbalizability requirement of rule dimensions is relaxed by allowing many (if not all) dimensions to be converted to commensurate dimensions.

To summarize, there seems to be a consensus for a definition of one-dimensional rules. One-dimensional rules are single, potentially verbalizable propositions relating potential category membership to a single underlying verbalizable and separable dimension. In contrast, only Ashby and colleagues (e.g., Ashby et al., 1998; Ashby, Ell, & Waldron, 2003) have proposed a definition for two-dimensional rules. Ashby and colleagues propose that two-dimensional rules are single verbalizable propositions relating potential category membership to a combination of two underlying verbalizable (i.e., commensurate), and separable dimensions.

2.4 Comparisons of One- and Two-Dimensional Rules

Given these definitions of one- and two-dimensional rules, the following are some examples of typical category structures to help clarify how rules may be represented and how one- and two-dimensional rules may be differentiated. An example of a category structure formed using a one-dimensional rule is categorizing Cardinals from a set of birds containing Cardinals and Blue Jays (see Figure 2.1a). In this example, birds are mapped onto a psychological space with the dimensions of size and color. The category Cardinal is formed by the one-dimensional rule: *Cardinals are red*. Color is selected as the relevant psychological stimulus dimension and a linear bound is placed orthogonally to the color dimension, separating red from blue. When a target bird is presented, the only dimension or feature that matters in the categorization decision is the bird’s color. The particular bird’s color dimension is evaluated according to the rule, *are red*. This evaluation is

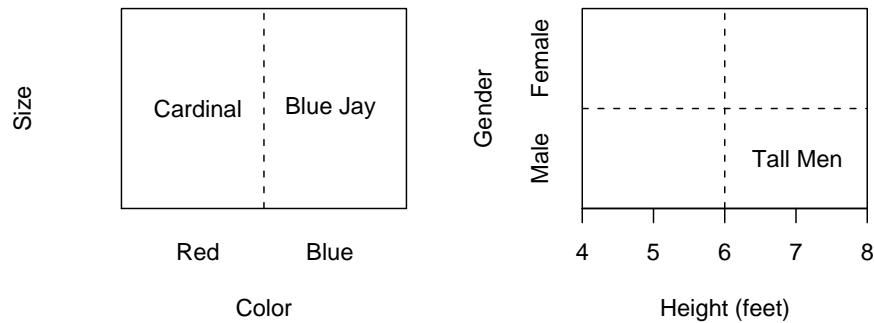


Figure 2.1: Example categories using one-dimensional rule boundaries. (a) 1D rule. (b) Two 1D rules (Conjunction).

then passed on to a final decision process. Other features, such as size, orientation, or number of legs, have no effect on the categorization decision.

One-dimensional rules can also be used to form more complex categories, such as tall men. In this example, people are mapped onto a psychological space with the dimensions of gender and height. To form the category tall men, one-dimensional rules can be combined to form more complex rules. In this case a conjunction rule can be used to form the rule: *tall men have a height over 6 feet and are male* (see Figure 2.2b). To evaluate the rule, boundaries are placed on the relevant dimensions of height and gender. To be a member of the category a target person's dimensions are evaluated by the rules: *is a male and is over six feet in height*. The results of each of these individual evaluations are then combined to make a final categorization decision. In all instances of one-dimensional rules, the decision on each dimension is made separately before being combined to make an overall categorization decision. This process results in a series of applied rule boundaries, with each rule orthogonal to its own relevant psychological dimension.

Now consider another typical example category: an overweight person. In this example, people are mapped onto a psychological space with the two dimensions of weight and height. The problem is that the physical fitness classification of a person depends on more than a single dimension (in this simplified case), it depends on both the height and weight of an individual. One solution to this category problem is to use the Body Mass Index (BMI) to judge fitness (see Figure 2.2a). The BMI is a ratio of mass to squared height. Classification decisions can then be made using both dimensions simultaneously, not individually. People can be classified on this ratio; for

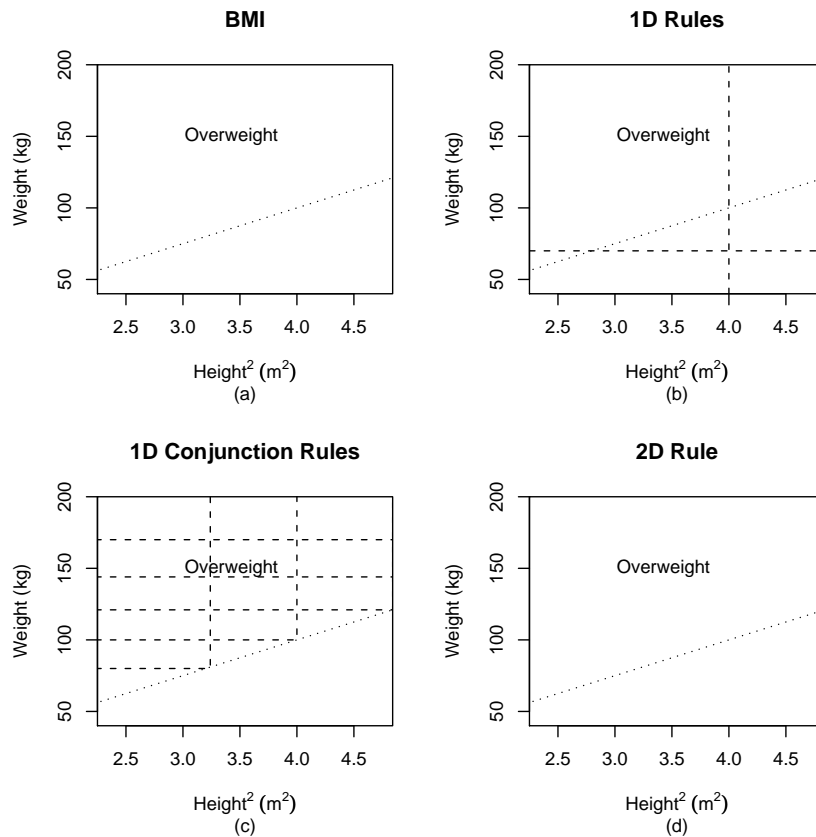


Figure 2.2: Examples of Fitness categories. (a) BMI solution. (b) Two one-dimensional rules. (c) Multiple one-dimensional conjunction rules. (d) Two-dimensional rule.

example, overweight people have BMI scores over 25. Thus the BMI can be used to categorize any combination of weight and height, and its classifications match human classifications.

The problem is that one-dimensional rules cannot reasonably duplicate the category structure formed using the BMI. Unlike the previous cases it is not sufficient to describe the category, an overweight person, using rules operating on single dimensions such as *an overweight person weighs over 70 kilograms* or *an overweight person is less than 2.0 meters tall* (see Figure 2.2b). Nor is it plausible to describe the category in terms of conjunctions of single dimensions such as *an overweight person weighs over 80 kilograms and is under 1.8 meters tall*. It is possible to define this category as a series of one-dimensional conjunction rules such as “*an overweight person weighs over 80 kilograms and is under 1.8 meters tall OR over 100 kilograms and is under 2 meters tall OR over 120 kilograms and is under 2.2 meters tall OR ...*” (see Figure 2.2c). While this set of rules would be able to describe the category, depending on the degree of precision, such a series of

iterative rules would be extremely lengthy, inefficient (many areas are redundantly classified), and require more processing to use than a typical one-dimensional conjunction rule. Given people's tendency to create categories using very simple representations rather than complicated representations (Garner, 1974), it is unlikely that such a process occurs.

An alternative solution is to use two-dimensional rules. An example of a two-dimensional rule that can match the BMI's classifications is *an overweight person is someone whose weight divided by their squared height is greater than 25* (see Figure 2.2d). In this case weight and height are combined to form a new dimension upon which an orthogonal rule boundary is placed. This boundary is orthogonal to an underlying rule dimension, but because this new dimension is a combination of other dimensions, the result is a diagonal boundary relative to the original dimensions in psychological space.

To reiterate the differences between one- and two-dimensional rules, in the case of a two-dimensional category structure, one-dimensional rules evaluate relevant psychological dimensions separately and then combine evaluations to form a single category decision. Thus one-dimensional rules always form rule boundaries in psychological space that are orthogonal to the relevant psychological dimensions. In contrast, two-dimensional rules combine information from relevant dimensions before an evaluation is performed. The underlying dimensions are then evaluated simultaneously and the result of that evaluation is then used to make a single category decision. This results in boundaries orthogonal to a combination of the underlying relevant psychological dimensions in psychological space (i.e., diagonal boundaries).

2.5 Research Questions

This research addresses the issue of two-dimensional rule use in human category learning. Two factors have been identified that may influence the ability to use two-dimensional rules: the relationship between the category relevant stimulus dimensions (i.e., separable versus integral), and the necessity of rule verbalizability (i.e., commensurate dimensions). This results in three questions that guide this dissertation.

1. Are rules limited to operating on one dimension at a time?
2. Are rules always open to explicit awareness and verbalization?
3. Are psychologically distinct dimensions necessary for the formation of rules?

It should be noted that the purpose of this research is not to demonstrate that people *always* use two-dimensional rules, but instead to explore when people *can* use two-dimensional rules.

2.5.1 Predictions of Theoretical Accounts of Rule Use in Categorization

Table 2.2 shows the predictions of different theoretical accounts of two-dimensional rule use. Predictions were made about the ability of people to use two-dimensional rules when the stimuli are composed of integral, separable, verbalizable, and nonverbalizable dimensions. These predictions will be used to evaluate the results of the experiments later discussed in this dissertation.

Current models of category learning (e.g., ATRIUM, COVIS, and RULEX) state that people can only use rules that align with a single dimension. These models predict that two-dimensional rule use should not be possible, under any circumstance. If evidence is found that people can use two-dimensional rules, then a complete theory of category learning will need to be able to account for the circumstances that allow for two-dimensional rule use.

Ashby et al. (1998) proposed a set of guidelines for two-dimensional rule use. While these guidelines are not instantiated within COVIS, the model could be altered to use two-dimensional rules (likewise, RULEX and ATRIUM could also be altered to use two-dimensional rules). Under Ashby et al.'s (1998) account, two-dimensional rules are formed using hypothesis testing and require underlying dimensions to be separable and commensurate. This account predicts that two-dimensional rule use will be found in tasks with separable and commensurate dimensions. An example rule of this type would be: *Rectangles that are taller than they are wide are members of the category.*

Ashby, Ell, and Waldron (2003) proposed a relaxed account for two-dimensional rules. It allows for separable perceptual dimensions to be abstracted into new unitless psychological dimensions. These dimensions are then necessarily commensurate with other abstracted psychological dimensions and can be used to create two-dimensional rules. This account predicts that evidence for two-dimensional rules may be found in tasks with separable dimensions, regardless of whether those dimensions are verbalizable or nonverbalizable. An example rule of this type would be: *Rectangles that are redder than they are wide are members of the category.*

The last account suggests that rule-based systems of categorization do not rely on reasoning or executive functions, and that the ability to describe the verbal rules used in a task is metaknowledge that does not reflect the underlying system. This view would predict that two-dimensional rules could be formed on integral or separable dimensions and verbalizable or nonver-

Table 2.2: Predictions of Theoretical Accounts for Evidence of Two-Dimensional Rule Use

Account	Separable		Integral	
	Verbalizable	Nonverbalizable	Verbalizable	Nonverbalizable
Current Rule Models (e.g., ATRIUM, COVIS, RULEX)	No	No	No	No
Two-dimensional rules can be generated from commensurate and separable dimensions (Ashby et al., 1998)	Yes	No	No	No
Two-dimensional rules can be abstracted from separable dimensions (Ashby, Ell, and Waldron, 2003)	Yes	Yes	No	No
Rules only describe what an underlying system is doing	Yes	Yes	Yes	Yes

balizable dimensions. This view is only provided as a null-hypothesis account of two-dimensional rule use, I do not know of any support for it.

2.6 Experiment Design

The accounts of two-dimensional rule use were evaluated with a series of experiments devised to investigate two-dimensional rule use using separable or integral dimensions and verbalizable or nonverbalizable dimensions. Four experiments were performed testing the use of two-dimensional rules. In each experiment participants were presented with stimuli created with a unique combination of integral or separable dimensions and verbalizable or nonverbalizable dimensions.

Each experiment used the same category structure, a rule and exception category structure (see Figure 2.3). This structure allows for the discrimination among the different theories of two-dimensional rule use in category learning. Performance on training items provides information on category mastery and is used to discriminate between the types of rules used by participants in the task. Performance on transfer items (items not shown during training) provides information on the types of representations used in learning the rule and exception category structure.

If participants generalize based on an *exemplar* representation, transfer items will be classified according to the exception. If participants generalize based on a *rule* representation, transfer items will be classified according to the rule. Therefore, transfer item performance will be informative for determining whether individual participants are using exemplar-based or rule- and exemplar-based categorization strategies, while training item performance will provide information on what type of rule-based strategies participants are using.

In each experiment there were two conditions that affected how rule items could be classified. In one set of conditions, participants were presented a one-dimensional rule-based category structure that could be optimally solved using a one-dimensional category bound. In the other set of conditions participants were presented a two-dimensional rule-based category structure that could be optimally solved using a two-dimensional bound. This two-dimensional structure was generated by rotating the one-dimensional structure 45 degrees and was counterbalanced with clockwise and counterclockwise rotations. In both the one-dimensional and two-dimensional conditions, correctly classifying exception items required information from both dimensions. These four experiments complete an overall between-subjects experimental design of 2 (integral vs. separable) x 2 (verbal vs. nonverbal) x 2 (one-dimensional optimal bound vs. two-dimensional optimal bound).

2.7 Model Testing

The categorization behaviors of participants reaching a learning criterion were used to group participants using similar strategies (Lee & Webb, 2005). After grouping, each group was evaluated using three models of categorization: an exemplar-based model, ALCOVE (Kruschke, 1992); a hybrid-system model using both rule-based and exemplar-based systems, ATRIUM (Erickson & Kruschke, 1998); and a modified version of ATRIUM that uses diagonal rules, ATRIUM-DR. Testing with ALCOVE was used to verify the conditions for one-dimensional rule use. Only stimuli that allow for rule use in the one-dimensional conditions should be predicted to have the possibility of two-dimensional rule use in the two-dimension conditions.

The type of category representation used by participants was assayed by determining the best fitting model. Participants using an exemplar-based representation were best fit by ALCOVE. Participants using one-dimensional rules were best fit by ATRIUM. Participants using two-dimensional rules were best fit by ATRIUM-DR. The experimental conditions with participants best fit by ATRIUM-DR are the conditions that allow for two-dimensional rules. These conditions can be used to evaluate the predictions made by the accounts guiding this research (see Table 2.2).

2.8 Overview

Yang and Lewandowsky (2004) and Lewandowsky et al. (2006) have provided evidence that people may be able to use two-dimensional rules in categorization. This is in contrast to current rule-based theories of categorization (e.g., ATRIUM, COVIS, RULEX) that only support one-dimensional rules in categorization. Two factors have been identified that may influence people's ability to use two-dimensional rules, the verbalizability of the stimulus dimensions and the dimensional relationship (separable or integral) of the relevant stimulus dimensions. Predictions on two-dimensional rule use from four accounts of category learning have been derived to provide hypotheses for experiments testing the effects of these factors.

In the following chapters, a series of experiments using a rule plus exception category structure are reported testing these hypotheses. The results of each experiment are analyzed independently before being combined to make an overall evaluation. Participant categorization behaviors from these four experiments are partitioned into groups of participants using similar strategies. These clusters are then evaluated using three different computational models of category learning, ALCOVE, ATRIUM, and ATRIUM-DR. The results will provide new information characterizing the use of two-dimensional rules in category learning, which is necessary for a complete theory of category learning.

Chapter 3

Experiments

3.1 Experiment 1: Separable and Commensurate Dimensions

3.1.1 Methods

Participants

The participants were 99 students from the University of California, Riverside enrolled in an introductory psychology class who participated to fulfill a class requirement. Participant learning performance during the last transfer phase was evaluated to select only participants who learned the category structure for analysis. Participants were required to achieve above chance levels of accuracy for exception items (22%) and rule items (38%). Figure 3.1 shows participant performance with respect to these criteria. Applying these criteria resulted in the exclusion of data from 19 participants and retaining data from 80 participants (see Table 3.1). The proportion of participants lost from each condition was not significantly different, $\chi^2(df = 3, N = 99) = 0.371, p = .94$.

Table 3.1: The Breakdown of Participants by Condition and Learning Criterion for Experiment 1.

Learning Criterion	Horizontal	Vertical	Negative Diagonal	Positive Diagonal	Total
Achieved	18	22	20	20	80
Failed	8	5	3	3	19
Total	26	27	23	23	99

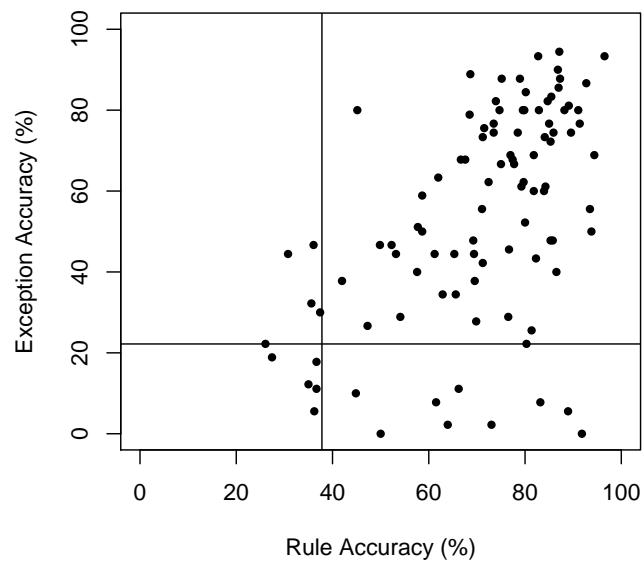


Figure 3.1: Participants from Experiment 1 are plotted by their accuracy on rule and exception items during the last transfer phase. Transfer items were not included in these calculations. The horizontal line is the rule accuracy criterion of 38%. The vertical line is the exception item accuracy criterion of 22%.


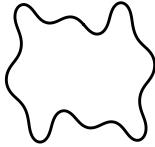

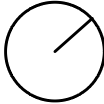
	Commensurate	Noncommensurate
Integral		
Separable		

Figure 3.2: Examples of the stimuli used in the experiments. Lower left: Experiment 1 used rectangles varying in height and line segment position, which have commensurate and separable dimensions. Lower right: Experiment 2 used circles varying in size and position of an internal radial line segment, which have noncommensurate and separable dimensions. Upper left: Experiment 3 used rectangles varying in size and height, which have commensurate and integral dimensions. Upper right: Experiment 4 used Fourier descriptors varying in initial phase and amplitude of a sine wave component, which have noncommensurate and integral dimensions

Apparatus

The experiment was performed in dimly lit individual booths on computers using the E-Prime programming environment. The stimuli were displayed on 17" monitors set at a resolution of 1024 x 768. Sound was played through headphones. Up to five participants could perform the experiment simultaneously.

Stimuli

The stimuli were rectangles that varied in height and the position of an internal vertical line segment (See examples in Figure 3.2). These dimensions have been previously found to be separable (Erickson & Kruschke, 1998). Additionally, because the dimension are both measured in units of distance, the dimensions are commensurate. Each dimension had eight possible values,

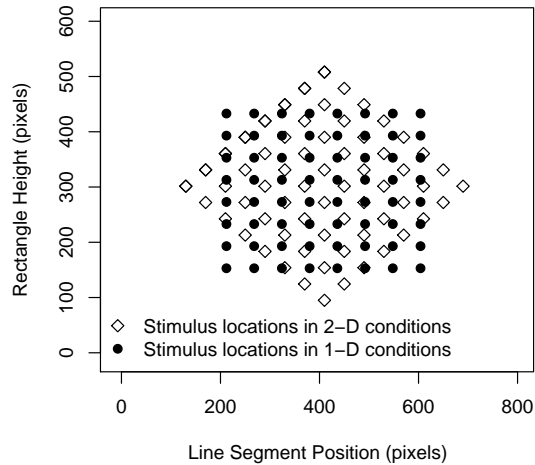


Figure 3.3: The physical dimensions of the stimuli in Experiment 1.

resulting in 64 possible stimuli, of which each participant only saw 40 (see Figure 3.4). The mapping of stimuli to line segment position and rectangle height is shown in Figure 3.3.

For the one-dimensional conditions, the rectangle height ranged from 153 pixels to 433 pixels in steps of 40 pixels. The position of the line segment ranged from 212 pixels to 606 pixels from the left side of the rectangle in steps of 56 pixels. The positions of the two dimensional condition stimuli were created by rotating the positions of the one-dimensional stimulus by 45 or -45 degrees around the center of the category structure. For the two-dimensional conditions, the rectangle height ranged from 95 pixels to 499 pixels in steps of 59 pixels. The position of the line segment ranged from 130 pixels to 688 pixels from the left side of the rectangle in steps of 80 pixels.

Category Structures

All stimuli were drawn from a four-label rule-plus-exception category structure, an abstracted example of this type of category structure is shown in Figure 3.4. In this example, correct classification of both types of training items (i.e., rules and exceptions) can be achieved by attending to the primary and secondary dimensions. Rule items (labeled as A & B in Figure 3.4), can be accurately classified by attending only to the primary dimension. If an item possesses a dimension value less than the midpoint on the primary dimension, the item is classified as a member of category A, if greater than the midpoint, as a member of category B. In contrast, correct classifi-

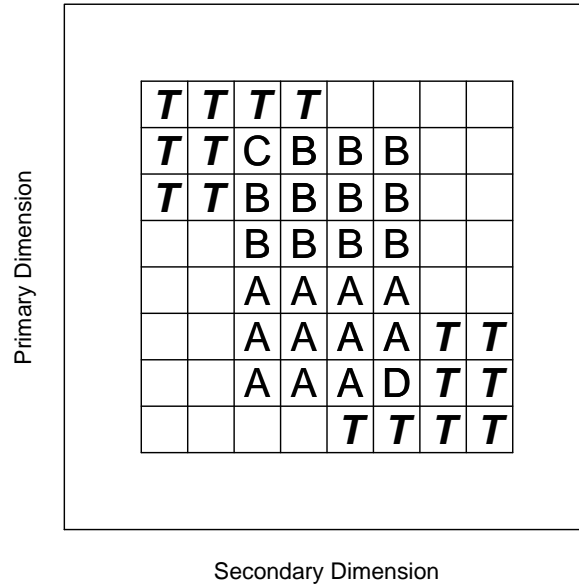


Figure 3.4: The abstracted rule and exception category structure used in the experiments. Rule items are labeled A and B. Exception items are labeled C and D. Transfer items are labeled T.

cation of exception items (labeled as C & D in Figure 3.4) requires attending to both primary and secondary dimensions. The two exception items are identified by a unique combination of features on the primary and secondary dimensions. The remaining items, the transfer items (labeled as T in Figure 3.4), were used to test representations used by participants. These items were not presented during the training phases. From this rule and exception category structure, four different category structures were created, two one-dimensional category structures and two two-dimensional category structures.

For the one-dimensional conditions, the primary and secondary dimensions corresponded to the height of the rectangle and the position of an internal vertical line segment. The boundary between rule items was orthogonal to the primary dimension. The *horizontal condition* (see Figure 3.5a) corresponded to a category structure with rule items that could be classified using a boundary orthogonal to the height dimension (i.e., all rectangles shorter than 300 pixels were in category A, and all rectangles taller than 300 pixels were in category B). The *vertical condition* (see Figure 3.5b) corresponded to a category structure with rule items that could be classified using a boundary orthogonal to the line segment position dimension. These conditions served to counter

balance the assignment of height or line segment position as the primary dimension for the one-dimensional category structures.

For the two-dimensional conditions, the primary and secondary dimensions again corresponded to height and line segment position, and were counterbalanced across participants. In these conditions, however, the category structure was rotated 45 degrees around the center of the category structure. This leads to a boundary separating the rule items with either a positive 45 degree slope, the *positive diagonal condition* (see Figure 3.5c), or with a negative 45 degree slope, the *negative diagonal condition* (see Figure 3.5d). To learn these two-dimensional category structures, information from both dimensions is necessary for accurate classification of both rule and exception items. However, unlike the one-dimensional category structures, in these structures the primary dimension does not correspond to a single dimension, but to a linear combination of the two relevant dimensions.

The experiment was composed of five phases (see Figure 3.6 for the phases and presentation frequencies). In the first phase, the *rule training phase*, participants were exposed only to rule items. Each of 20 rule items, 10 from each category, were presented once per block for 7 blocks, resulting in a total of 140 trials. In the second phase, the *rule and internal exception training phase*, participants were exposed to rule and internal exception items. Each of the 22 rule items were presented once per block and each of the 2 exception items were presented 6 times per block, for 6 blocks, resulting in a total of 204 trials.

In the third phase, the *transfer test with internal exceptions phase*, participants were exposed to rule, exception, and transfer items. Each of the 22 rule items were presented twice, each of the exception items was presented 15 times, and each of the transfer items were presented 3 times. In addition, the most diagnostic items were shown more frequently. The extreme corner transfer items were shown 9 times each and the rule items adjacent to the transfer items were shown 5 times each (Figure 3.6 displays item frequencies). This resulted in a total of 152 trials for phase 3. In the fourth phase, the *rule and external exception training phase*, participants were exposed to rule and exception items. Each of the 22 rule items were presented once per block and each of the 2 exception items were presented 6 times per block, for 14 blocks, resulting in a total of 476 trials.

In the fifth phase, the *transfer test with external exceptions phase*, participants were exposed to rule, exception, and transfer items. Each of the 22 rule items were presented twice per block, each of the exception items were presented 15 times per block, and each of the transfer items were presented 3 times per block. In addition, the most diagnostic items were shown more frequently. The extreme corner transfer items were each shown 9 times per block and the rule items

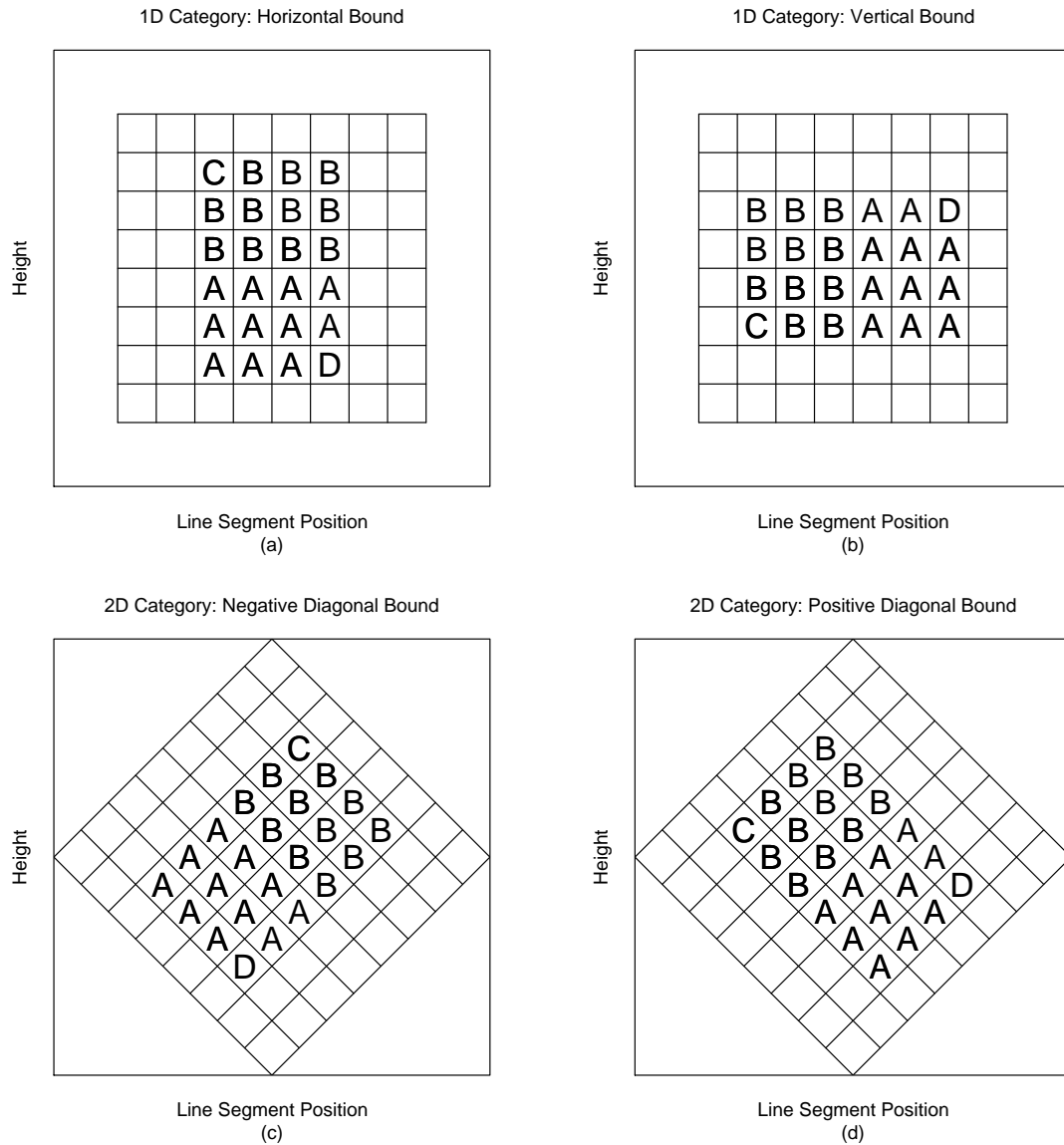


Figure 3.5: The four rule and exception category structures used in Experiment 1. Rule items are labeled A and B. Exception items are labeled C and D. Transfer items are not displayed, but can be inferred from Figure 3.4. The figure is organized as follows: A. Horizontal bounded category structure. B. Vertical bounded category structure. C. Negative diagonal bounded category structure. D. Positive diagonal bounded category structure.

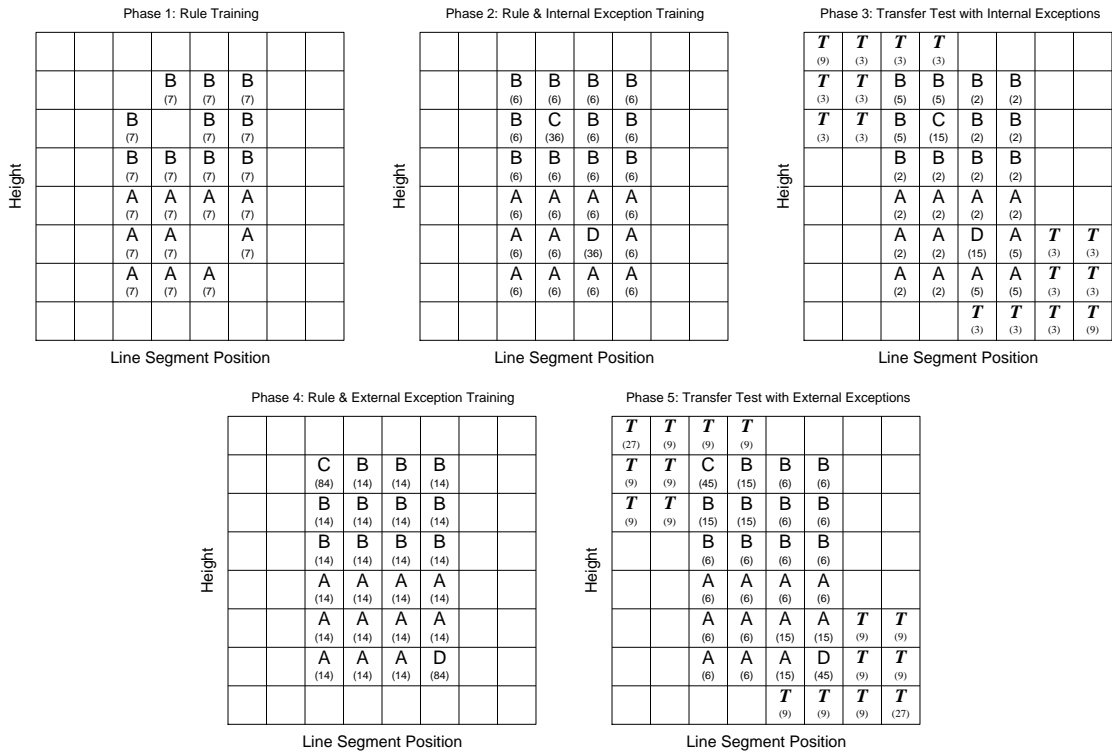


Figure 3.6: Category structures for each phase of Experiment 1. The letter represents the category label for that stimulus. Rule items are labeled A and B. Exception items are labeled C and D. Transfer items are labeled T. The number represents how many times each stimulus was shown per phase.

adjacent to the exception items were each shown 5 times per block (Figure 3.6 displays items frequencies). With 3 blocks, this resulted in a total of 456 trials for phase 5. Overall, there were 1428 total trials in Experiment 1. Trial order was randomized within each block for each participant and the assignment of category label to category was randomized for each participant.

The category structure was designed with two types of exception items, internal and external exceptions (compare phase 2 with phase 4 in Figure 3.6). In the first three phases, the internal exception items were surrounded by rule items. In the later phases, the exception items were shifted one unit on both dimensions further from the center of the category structure to the edge of the training items to become the external transfer items.

The purpose of the internal and external exception items was to encourage a more uniform representation of exception items among participants and then to allow for the differentiation of different theories of categorization. Internal exception items are used to convey the notion of exceptions as single items as uniformly as possible among the participants. If only external exception items are studied, participant representations are more likely to be divided between representing exceptions as single items and representing exceptions as groups of items. Initial training on internal exceptions increases the likelihood of participant representations of exceptions as single items. The purpose of the external exception items was to allow for the differentiation of participants using single strategies from those using multiple strategies of categorization using ATRIUM and ALCOVE. With internal exceptions ALCOVE and ATRIUM make the same predictions and cannot be differentiated, however with external exceptions the models make different predictions. Hence the shift of internal to external exception items in the experiment.

Procedure

At the beginning of the experiment, participants were presented with instructions and the cover story for the experiment. Participants were told they would be learning to identify four different types of space shuttle schematics. On each trial, participants were presented with a stimulus and asked to classify each as a member of one of four categories by pressing one of four keys on a keyboard (D, F, J, K). The instructions emphasized that the stimuli could be accurately identified, and that progression through the experiment could be facilitated by responding as quickly and as accurately as possible.

In a trial, participants were presented with a blank black screen for 250 ms followed by the stimulus. The stimulus remained on screen until a response was made by pressing one of the four keys (D, F, J, or K). Following a response, participants were given feedback and the correct

category label. The correct category label remained on screen until the participants pressed the spacebar. This allowed the participant the opportunity to study the correct stimulus and category label pairing. Participants who misidentified the stimulus heard a low pitched tone for 500 ms and were forced to remain at the feedback screen for a minimum of 1500 ms. Participants who correctly identified the stimulus did not hear a tone and were able to continue to the next trial immediately after pressing the spacebar.

In the transfer phase, participants were given transfer items to classify. On trials with transfer items no useful feedback was given after participants made a category response. The feedback screen only reported that the response was recorded. Participants were then able to continue to the next trial immediately after pressing the spacebar.

3.1.2 Results and Discussion

Learning

Averaged participant learning curves per condition can be seen in Figure 3.7. All conditions show declines in performance in blocks 8, 14, and 27. In block 8, the exception items were first introduced, which increased the number of valid category labels to four. Blocks 14 and 27 were the start of the transfer phases, which added trials with unfamiliar transfer stimuli that were not followed by correct category labels nor accuracy feedback.

On the last block of training, participants averaged 75.6% rule accuracy. A one-way ANOVA revealed that the final training block rule accuracy was not equal across conditions, $F(3, 76) = 5.56, p = .001$. Participants in the vertical condition had the highest accuracies on rules items (83.9%), followed by the horizontal condition participants (75.5%), the positive diagonal condition participants (74.3%), and the negative diagonal condition participants (68.2%). Tukey's HSD tests found that participant rule accuracy in the vertical condition (83.9%) was significantly higher than in the negative diagonal condition (68.2%), $p = 0.0007$, but no other significant differences were found. Additionally, a one-way ANOVA found that participants in the one-dimensional conditions were more accurate on rule items (80.1%) than participants in the two-dimensional conditions (71.3%), $F(1, 78) = 9.35, p = .003$.

On the last block of training, participants averaged 73.8% exception accuracy. A one-way ANOVA revealed no significant differences between the conditions, $F(3, 76) = 0.47, p = .70$. Participants in the positive diagonal condition had the highest accuracies on exception items (78.0%), followed by the negative diagonal condition (75.0%), the vertical condition (72.3%), and the hor-

izontal condition (69.4%). Additionally, a one-way ANOVA did not find significant differences between participants in the one-dimensional and two-dimensional conditions on exception item accuracy, $F(1, 78) = 1.14, p = .29$.

To evaluate differences between rule and exception accuracy during the last training block by participant, a *rule-advantage* score was calculated by subtracting average exception accuracy from average rule accuracy. The resulting measure refers to the performance of the participant on one type of category structure relative to the other. Positive rule-advantage scores indicate that participants were more accurate on rule items, a rule-advantage. Negative rule-advantage scores indicate that participants were more accurate on exception items, an exception advantage. Low or zero rule-advantage scores indicate that participants were equally accurate on exception and rule items.

Overall, participants averaged a 1.9% rule-advantage score. Participants tended to perform better on rule items than exception items. A one-way ANOVA revealed a significant difference between the conditions, $F(3, 76) = 3.51, p = 0.02$. Participants in the vertical condition had the highest rule-advantage score (11.5%), followed by the horizontal condition (6.0%), the positive diagonal condition (-3.6%), and the negative diagonal condition (-6.8%). Tukey's HSD tests found a significant difference between participants in the vertical condition (11.5%) and the negative diagonal condition (-6.8%), $p = 0.026$, but no other significant differences between the conditions were found. Additionally, a one-way ANOVA found that participants in the one-dimensional conditions had higher rule-advantage scores (9.1%) than participants in the two-dimensional conditions (-5.2%), $F(1, 78) = 9.72, p = .003$.

Finally, t-tests were performed on the rule-advantage scores to determine if they were different than 0. Only participants in the vertical condition (11.5%) were found to have an average rule-advantage score significantly different from 0, $t(21) = 2.47, p = .02$. This suggests that participants in the vertical condition were more accurate on rule items than exception items.

Transfer

Participant performance during the transfer phase was analyzed to characterize differences between the conditions and to provide initial evidence for the categorization strategies involved. Participants could classify items in four different ways. They could classify items in a consistent fashion, either as an *exception-consistent* item according to the nearest exception item, or as a *rule-consistent* item according to the boundary used to create the category structure. They could also categorize items in an inconsistent fashion, either as an *exception-inconsistent* item according to the

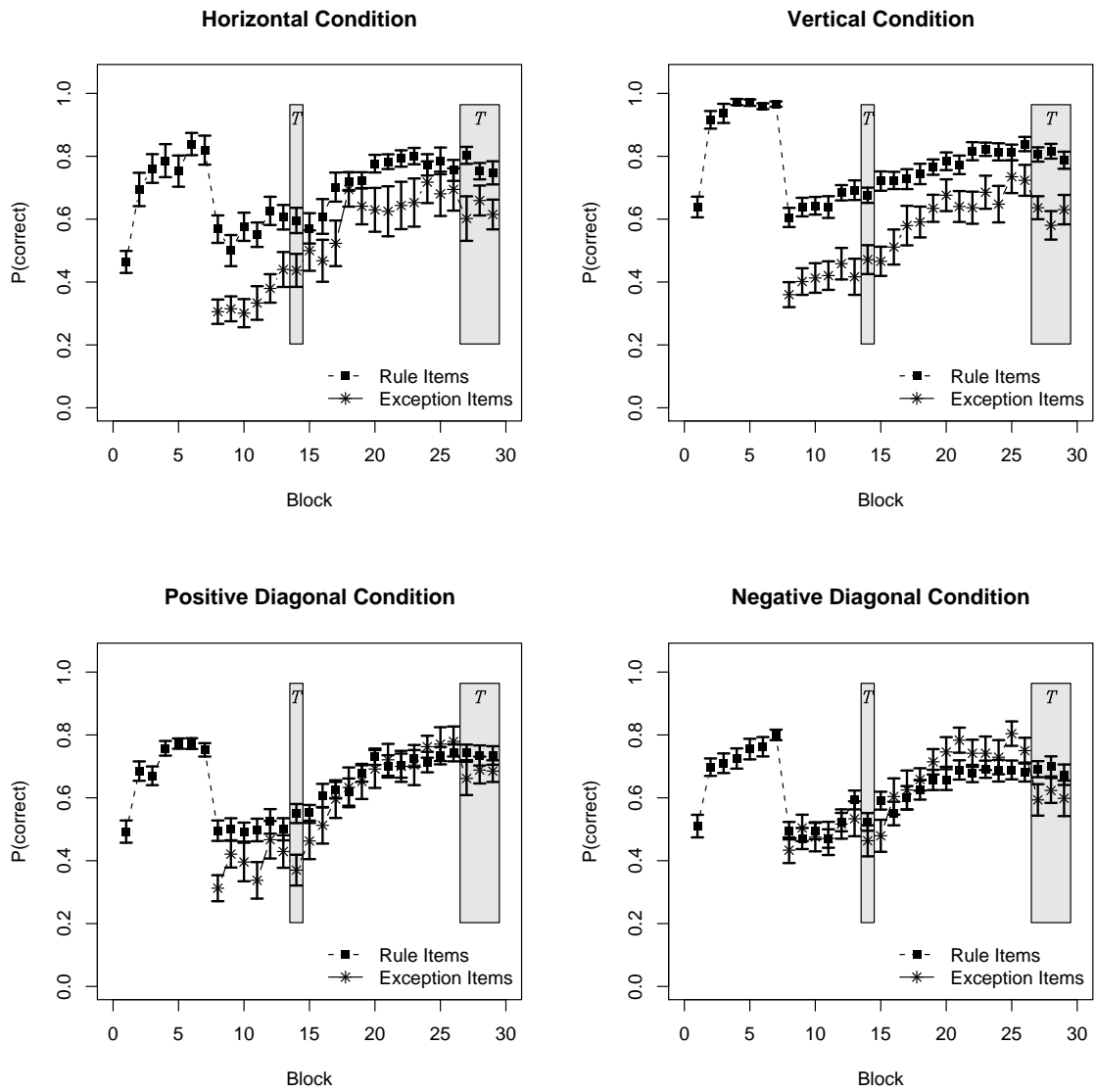


Figure 3.7: Learning curves for Experiment 1. The points are the mean participant accuracies per block by condition. The error bars are the standard errors of average participant accuracy for each block.

exception item on the opposite side of the category structure, or an *rule-inconsistent* item according to the rule category on the opposite side of the boundary used to create the category structure.

Rule- and exception-consistent responses measure how well participants learned the category structures. These responses can also be used to identify the strategies underlying participant behavior. In contrast to consistent responses measuring how well participants learned the category structures, inconsistent responses measure how poorly participants learned the category structures. Rule-inconsistent responses indicate an uncertainty as to location of the boundary between the rule categories. High rates of rule-inconsistent responses indicate a category structure where participants confused the two rule categories. Low rates of rule-inconsistent responses indicate a category structure where participants were able to distinguish the two rule categories.

Exception-inconsistent responses are less likely to occur than the other types of responses if participants learned the category structures. To make an exception-inconsistent response, participants must classify an item according to the exception training item on the opposite side of the rule boundary. If participants display high rates of exception-inconsistent responses they most likely have not learned the category structure, and they will not be useful for understanding how people who have learned to categorize behave.

Responses to transfer items can be used to characterize participant performance and the underlying categorization strategies used in the task. There are two types of transfer items, the 2 items at the extreme corners of the category structure (one on each side of the category structure), hereby referred to as *extreme transfer* items, and the 14 remaining *transfer items* that are closer to the training items, including the exception training items. Performance on these items can differentiate the strategies of categorization used to classify the items due to the interaction of the exception items and rule items on transfer performance.

When evaluating the responses to the transfer items, the extreme transfer items are most diagnostic for identifying strategies. The extreme transfer items are farthest from the training items and are the purest measure of participant strategies. The transfer items that are closer to the training items are more likely to be influenced in different ways by different participant strategies. Participant responses to these items are informative, but because information is combined over many items, the strategies involved may be mixed and therefore difficult to interpret. The predictions made by various strategies will therefore be characterized by specific types of extreme transfer item performance and by a wider range of non-extreme transfer item performance. Figure 3.8 displays examples of these strategies, including performance on the less informative (non-extreme) transfer items. Performance on both extreme and (non-extreme) transfer items is provided to allow for

direct comparisons between these examples and later figures displaying the performance of the participants.

Exemplar-based strategies of categorization predict high rates of exception-consistent responses on extreme transfer items and either high rule, high exception, or a mixture of high rule and high exception responses on transfer items. Performance on transfer items, those items nearest to the exception item, show the strength of generalization from the exception item. High exception-consistent response rates on transfer items indicate conditions in which participants have strong exception item influence and weak rule item influence. In contrast, high rule-consistent response rates on transfer items indicate conditions with weak exception item influence and strong rule item influence.

Rule- and exemplar-based strategies of categorization predict high rates of rule-consistent responses on extreme transfer items and either high rule, high exception, or a mixture of high rule and high exception responses on transfer items. Performance on transfer items, those items nearest to the exception item, show the strength of generalization from the exception item. High exception-consistent response rates on transfer items indicate conditions in which participants have strong exception item influence, most likely due to exemplar strategies. In contrast, high rule-consistent response rates on transfer items indicate conditions with weak exception influence, and low influence from the exemplar strategies.

Mixtures of participants using exemplar-based and rule- and exception-based categorization strategies produce response patterns on transfer and extreme transfer items with large variances and a mean dependent upon the proportion of participants with each strategy, as shown on the final row of Figure 3.8. Note that none of the predictions in Figure 3.8 would result from participants using a guessing strategy or participants pressing a single key in response to a transfer item. Participants using a guessing strategy would have performance for transfer and extreme transfer items characterized by consistent responses rates of 25% and low variance. Likewise participants who press a single key in response to transfer items would be characterized by having high inconsistent responses rates of 25% and high variance.

Transfer and extreme transfer item performance in Experiment 1 can be seen in Figure 3.9. Participants in the horizontal, vertical, and negative diagonal conditions were found to have no significant differences in transfer and extreme transfer item performance. This pattern of behavior is consistent with mixtures of participants using different strategies. These data will be explored in more detail in the model fitting chapters, however at this point these conditions tentatively support the idea that participants can solve these categories a variety of different ways. As this supports the

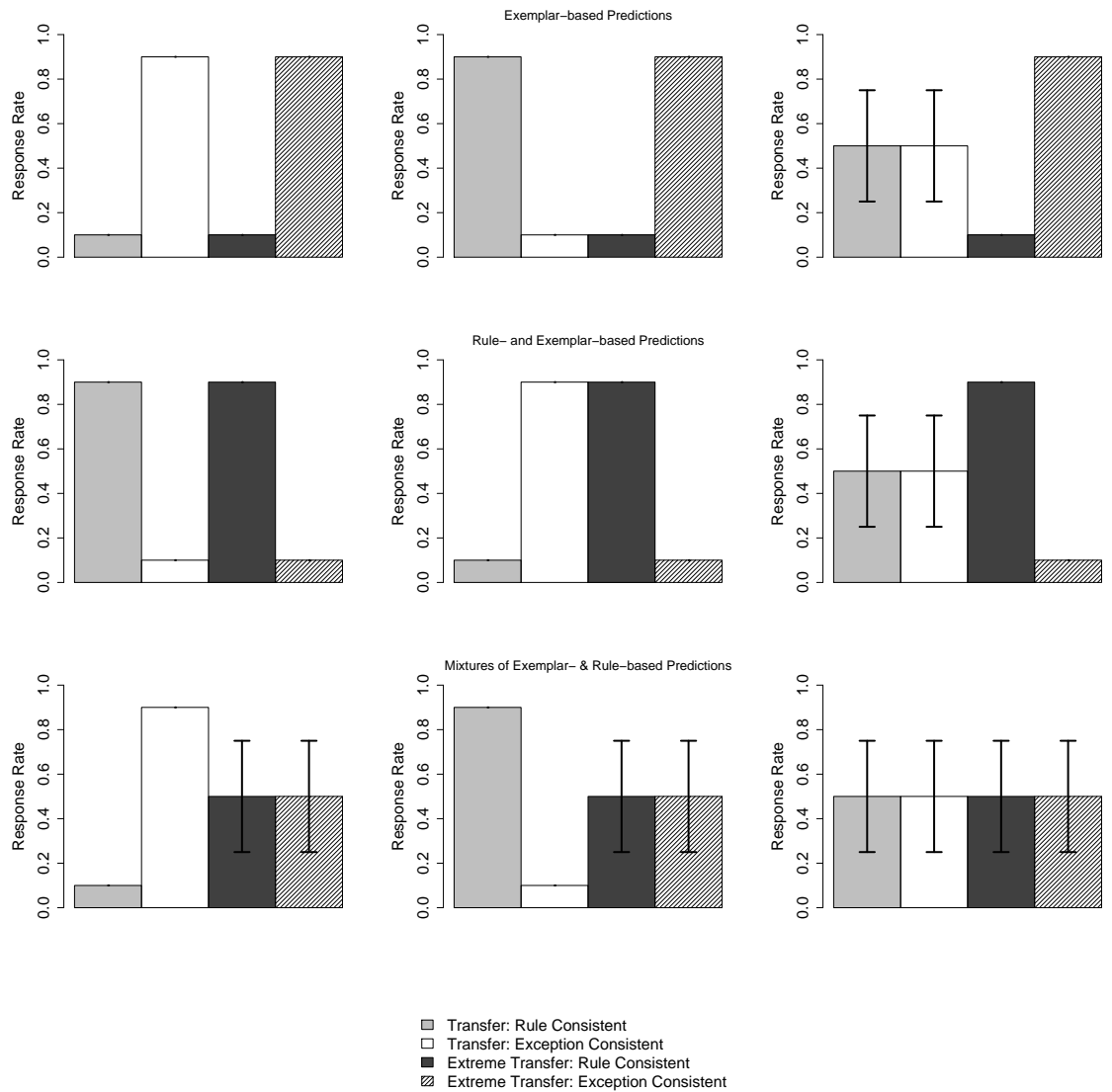


Figure 3.8: The transfer pattern predictions based on the underlying categorization strategy. The top row shows three different predictions of exemplar-based accounts. The middle row shows three different predictions of rule- and exception-based accounts. The bottom row shows three different predictions of mixtures of participants with rule- and exemplar- based accounts.

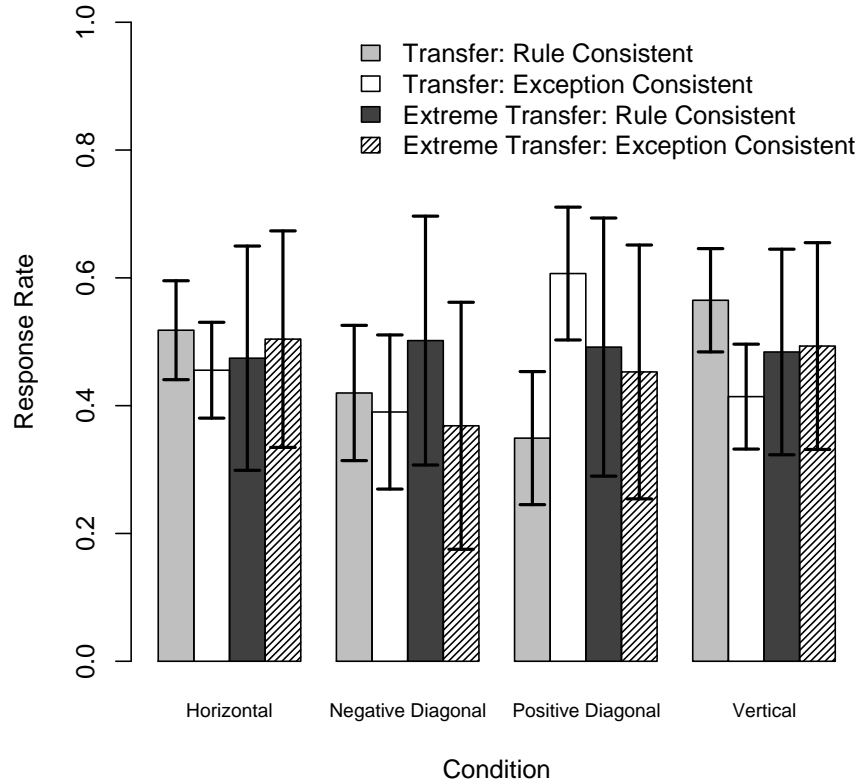


Figure 3.9: Responses to transfer items in Experiment 1 by condition. The extreme transfer items are the two stimuli on the outside corners of the category structure. The transfer items are the remaining 14 non-extreme transfer items. The error bars are the standard errors of average participant response rates.

use of rule-based categorization, this also tentatively supports the use of rules in the one-dimensional conditions and in the two-dimensional negative diagonal condition.

In contrast to the other conditions, participants in the positive diagonal condition had significantly more exception responses on transfer items than rule responses $t(19) = 2.82, p = .01$, but no differences in response rates for extreme transfer items. This pattern of performance is consistent with rule- and exception-based strategy responses. This condition also provides evidence of two-dimensional rule use. The results of the transfer analysis for categories using separable and commensurate dimensions provide support for one-dimensional and two-dimensional rules.

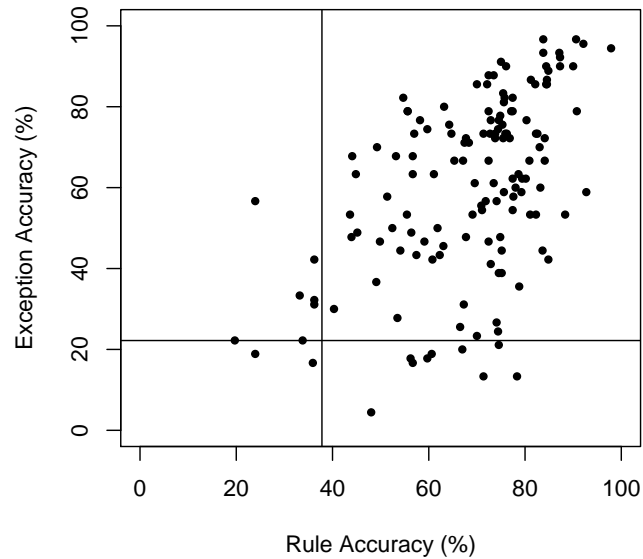


Figure 3.10: Participants from Experiment 2 are plotted by their accuracy on rule and exception items during the last transfer phase. Transfer items were not included in these calculations. The horizontal line is the rule accuracy criterion of 38%. The vertical line is the exception item accuracy criterion of 22%.

3.2 Experiment 2: Separable and Noncommensurate Dimensions

Experiment 2 was very similar to Experiment 1. The major changes were new stimuli composed of separable and noncommensurate dimensions, and a series of procedural changes designed to increase participant performance (which is described later).

3.2.1 Methods

Participants

The participants were 144 students from the University of California, Riverside enrolled in an introductory psychology class who participated to fulfill a class requirement. Applying the criteria from Experiment 1 resulted in the exclusion of data from 18 participants and retaining data from 126 participants (See Figure 3.10). The proportion of participants lost from each condition were not significantly different, $\chi^2(df = 3, N = 144) = 2.17, p = .54$ (see Table 3.2).

Table 3.2: The Breakdown of Participants by Condition and Learning Criterion for Experiment 2.

Learning Criterion	Horizontal	Vertical	Negative Diagonal	Positive Diagonal	Total
Achieved	29	29	32	36	126
Failed	5	6	5	2	18
Total	34	35	37	38	144

Stimuli

The stimuli were circles with radial line segments that ranged on the separable and non-commensurate dimensions of circle size and angle of the radial line (See examples in Figure 3.2). Circle size and angle of orientation of a radial line are classic examples of separable dimensions (Ashby & Maddox, 1998). Additionally, because the dimensions are measured in different units, angle degree and distance, they are not commensurate. Each dimension had eight possible values, resulting in 64 possible stimuli, of which each participant only saw 40 (see Figure 3.4). The mapping of stimuli to the circle size and angle of the radial line segment is shown in Figure 3.11.

It should be noted that the definition of these stimuli as separable and noncommensurate is only valid for two-dimensional rules. The dimensional characteristics of separability and commensurability are based upon the relationship of two dimensions. One-dimensional rules are defined by a single dimension, so these characteristics are not meaningful when applied to one-dimensional rules.

For the one-dimensional conditions, the angle of the radial line segment ranged from 18 degrees to 105 degrees in steps of 12 degrees. The circle radius ranged from 64 pixels to 180 pixels in steps of 16 pixels. The positions of the two-dimensional condition stimuli were created by rotating the positions of the one dimensional stimulus by 45 or -45 degrees around the center of the category structure. For the two-dimensional conditions, the angle of the radial line segment ranged from 0 degrees to 122 degrees in steps of 8.75 degrees, and the circle radius ranged from 40 pixels to 203 pixels in steps of 11.66 pixels.

Procedural Changes

Previous studies with these types of stimuli and categories with the student population found that a 90-minute experimental session was necessary for participants to learn the category structure and then test transfer performance. To increase participant interest and arousal during this task, some modifications were made to the procedure to make the task seem more like a game.

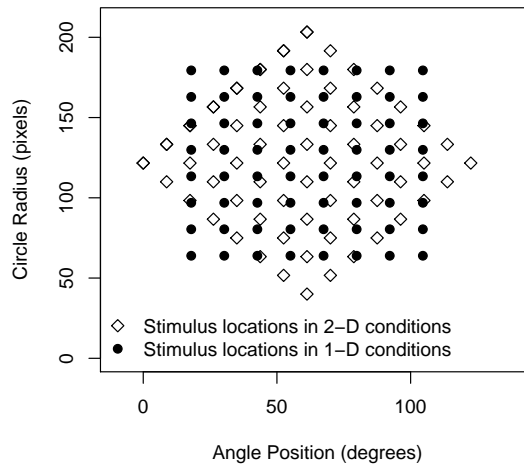


Figure 3.11: The physical dimensions of the stimuli in Experiment 2.

In this version of the task, participants were instructed that they were taking on the role of a trainee spaceship captain whose task was to defend the Earth against the evil space aliens. As a trainee, participants were to learn to categorize four types of alien ships which would allow them to use the correct weapons to defeat the aliens.

During the task, the participants were presented a display with the view from a space ship radar. The display was broken up into two sections by a vertical white line. The long-range radar portion of the screen consisted of the right 25% of the screen. The remaining 75% of the display was the short-range viewing screen and was used to present the stimuli.

In a trial, participants were shown a simulation of an encounter with a hostile alien ship. At the start of a trial a red dot, representing the alien, approached a blue dot, representing the participant's ship, on the long-range radar. The approaching sequence took 250 ms during which the short-range viewing screen was blank. Once the alien ship was in range, the stimulus appeared on the short-range viewing screen. The stimulus remained on screen until a response was made.

Following a response, an animation (lasting 35 ms) was shown on the long-radar screen, in which the participant's ship fired upon the alien ship. The shot fired by the participant's ship was accompanied by a laser sound (lasting 200 ms). Each of the four key responses fired a unique colored shot (green, blue, yellow, or red) and a unique laser sound. Sounds and colors were randomized for each participant.

If the participant correctly identified the stimulus, the alien ship on the long-range radar screen was covered with a randomized star-like polygon representing an explosion of the same color as the fired shot. Additionally, an explosion word (*BOOM*, *KAPOW*, *BANG*, *ZAP*, or *SPLAT*) was randomly selected to appear in the center of the explosion. If the participant misidentified the stimulus, the alien ship would return fire (a yellow line appeared connecting the two ships on the radar display with no animation), the message *Enemy Returning Fire!!* was displayed, and an additional failure message was randomly selected to appear in the center of the radar screen (*Shields Held!*, *Dodged!*, *Missed!!*, *Evaded!!*, or *Avoided!!*).

After the firing animation sequence, the participant was given feedback and the correct category label. The correct category label remained on screen allowing the participant the opportunity to study the correct stimulus and category label pairing. To continue to the next trial the participant pressed the spacebar. Participants who misidentified the stimuli heard a low pitched tone for 500 ms and were forced to remain at the feedback screen for a minimum of 1500 ms, while participants who correctly identified the stimulus were able to continue to the next trial immediately after pressing the spacebar. No feedback was given for transfer items, the screen reported that the alien ship had *warped out*. After a transfer item, participants were able to continue to the next trial immediately after pressing the spacebar.

In addition to these animations, sounds, and a more fantastic cover story, a running score of performance was kept. The score was visible during the feedback portion of a trial at the top of the screen. Participants earned 10 points for correct responses and lost 5 points for incorrect responses.

3.2.2 Results and Discussion

Learning

Averaged participant learning curves per condition can be seen in Figure 3.12. All conditions show declines in performance in blocks 8, 14, and 27. In block 8, the exception items were first introduced, which increased the number of valid category labels to four. Blocks 14 and 27 were the start of the transfer phases, which added trials with unfamiliar transfer stimuli that were not followed by correct category labels nor accuracy feedback.

On the last block of training, participants averaged 76.2% rule accuracy. A one-way ANOVA revealed that rule accuracy in the final training block was not equal across conditions, $F(3, 122) = 2.80, p = .004$. Participants in the positive diagonal condition had the highest accuracies

on rules items (81.3%), followed by the vertical condition participants (75.5%), the negative diagonal condition participants (73.7%), and the horizontal condition participants (73.2%). Tukey's HSD tests found differences between participants in the positive diagonal condition (81.3%) and in the horizontal condition (73.2%), $p = .06$, and between participants in the positive diagonal condition (81.3%) and the negative diagonal condition (73.7%), $p = .08$. No other differences approached significance. Additionally, participants in the one-dimensional conditions were not found to be less accurate on rule items (74.3%) than participants in the two-dimensional conditions (77.7%) with a one-way ANOVA, $F(1, 124) = 2.04, p = .16$.

On the last block of training, participants averaged 80.4% exception accuracy. A one-way ANOVA revealed significant differences between the conditions, $F(3, 122) = 5.30, p = .002$. Participants in the positive diagonal condition had the highest accuracies on exception items (88.7%), followed by the negative diagonal condition (81.5%), the horizontal condition (76.7%), and the vertical condition (72.4%). Tukey's HSD tests found differences between participants in the positive diagonal condition (88.7%) and in the vertical condition (72.4%), $p = .001$, and between participants in the positive diagonal condition (88.7%) and the horizontal condition (76.7%), $p = .03$. No other differences approached significance. Additionally, a one-way ANOVA found that participants in the one-dimensional conditions (74.5%) were lower in exception accuracy than participants in the two-dimensional conditions (85.3%), $F(1, 124) = 11.93, p = .0008$.

To evaluate differences between rule and exception accuracy during the last training block by participant, the rule-advantage score was calculated by subtracting average exception accuracy from average rule accuracy. Overall, participants averaged a -4.2% rule-advantage score, participants tended to perform better on exception items than rule items. A one-way ANOVA did not reveal significant differences among the conditions, $F(3, 122) = 2.49, p = 0.06$. Participants in the vertical condition had the highest rule-advantage scores (3.1%), followed by the horizontal condition (-3.5%), the positive diagonal condition (-7.3%), and the negative diagonal condition (-7.8%). Additionally, a one-way ANOVA found that participants in the one-dimensional conditions had higher rule-advantage scores (-2.0%) than participants in the two-dimensional conditions (-7.6%), $F(1, 124) = 5.41, p = .02$.

Lastly, t-tests were performed on the rule-advantage scores to determine if they were different than 0. Participants in the negative diagonal condition (-7.8%) were found to have an average rule advantage score less than 0, $t(31) = 2.61, p = .013$, as well as participants in the positive diagonal condition (-7.3%), $t(35) = 3.05, p = .004$. This suggests these participants were more accurate on exception items than rule items.

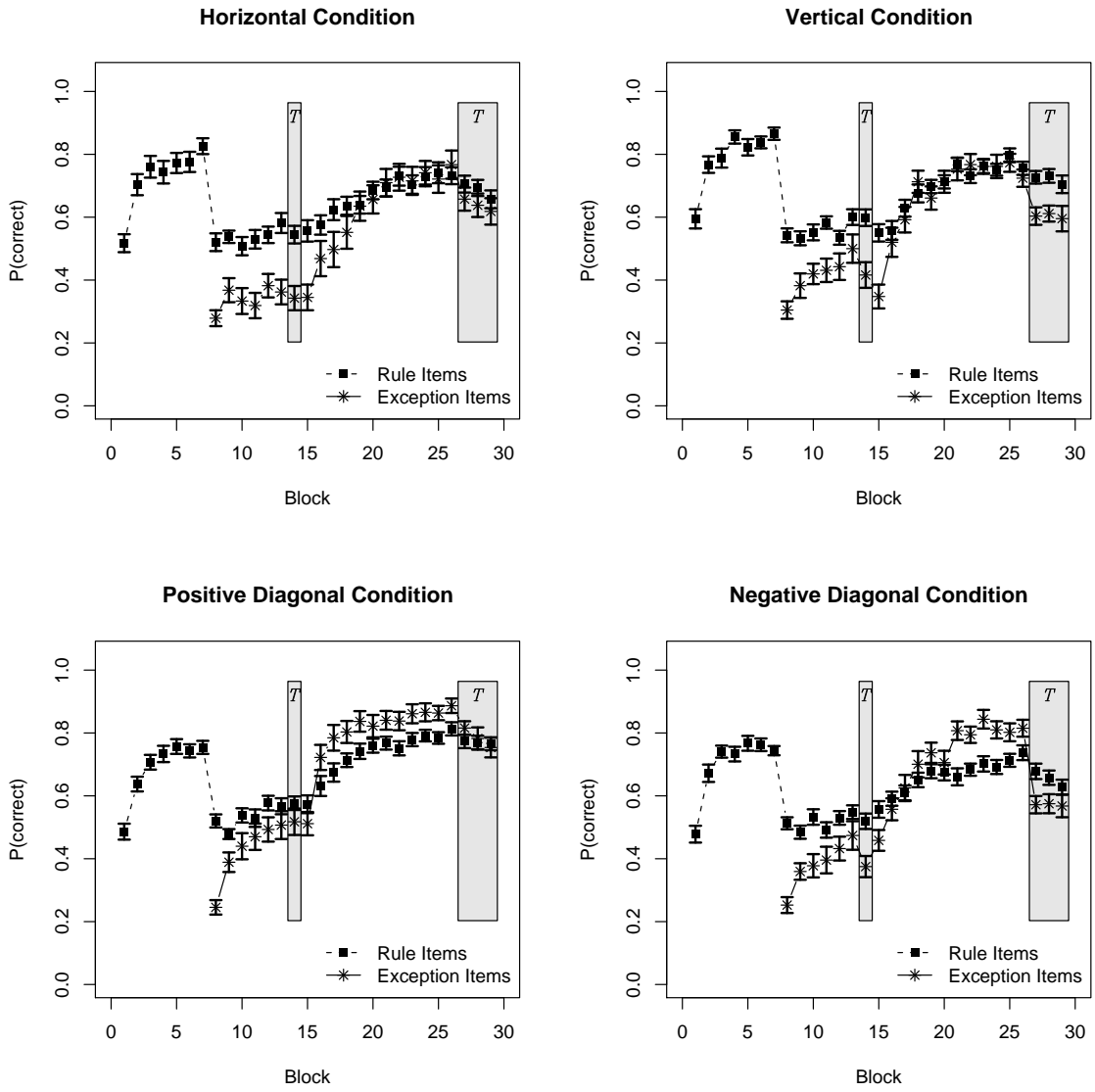


Figure 3.12: Learning curves for Experiment 2. The points are the mean participant accuracies per block by condition. The error bars are the standard errors of average participant accuracy for each block.

Transfer

Transfer and extreme transfer item performance in Experiment 2 can be seen in Figure 3.13. Participants in the horizontal condition did not have a significant difference between transfer and extreme transfer item performance. This pattern of performance is consistent with mixtures of categorization behaviors, which includes rule-based categorization. Participants in the vertical condition were found to have significantly more rule responses to transfer items than exception responses, $t(28) = 2.19, p = .037$. This pattern of was also present in the extreme transfer items, but was not significant, $t(28) = 0.98, p = .33$. This pattern of results is consistent with rule-based categorization and supports one-dimension rule use. They also support the use of rules in the one-dimensional conditions for separable dimensions, as has previously been found.

Participants in the negative diagonal condition were not found to have differences in transfer item performance, but were found to have significantly more exception response than rule responses on extreme transfer items, $t(31) = 2.13, p = .04$. Likewise participants in the positive diagonal condition were found to have significantly higher exception response rates on both transfer items, $t(35) = 5.90, p < .0001$, and extreme transfer items, $t(35) = 2.96, p = .005$. This pattern of behavior is consistent with exemplar-based categorization. The results of the transfer analysis for categories using separable and noncommensurate dimensions provide support for one-dimensional rules and no support for two-dimensional rules.

3.3 Experiment 3: Integral and Commensurate Dimensions

Experiment 3 was identical to Experiment 2, except the stimuli were composed of integral and commensurate dimensions.

3.3.1 Methods

Participants

The participants were 133 students from the University of California, Riverside enrolled in an introductory psychology class who participated to fulfill a class requirement. Applying the criteria from Experiment 1 resulted in the exclusion of data from 9 participants and retaining data from 124 participants (See Figure 3.10). The proportion of participants lost from each condition were not significantly different, $\chi^2(df = 3, N = 133) = 0.388, p = .94$ (see Table 3.3).

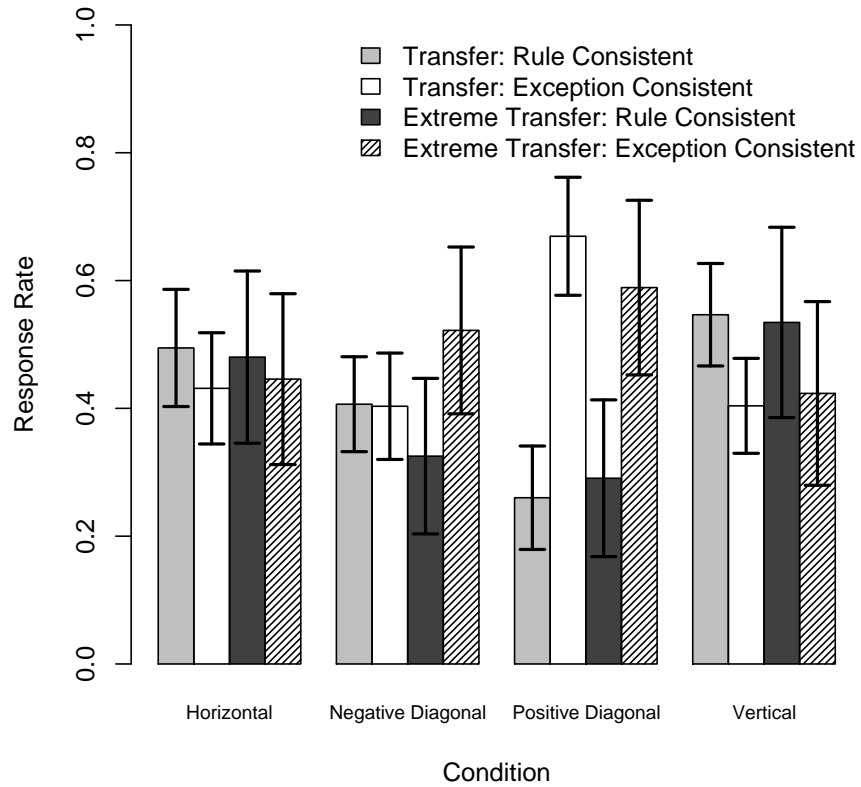


Figure 3.13: Responses to transfer items in Experiment 2 by condition. The extreme transfer items are the two stimuli on the outside corners of the category structure. The transfer items are the remaining 14 non-extreme transfer items. The error bars are the standard errors of average participant response rates.

Table 3.3: The Breakdown of Participants by Condition and Learning Criterion for Experiment 3.

Learning Criterion	Horizontal	Vertical	Negative Diagonal	Positive Diagonal	Total
Achieved	34	25	34	31	124
Failed	1	5	0	3	9
Total	35	30	34	34	133

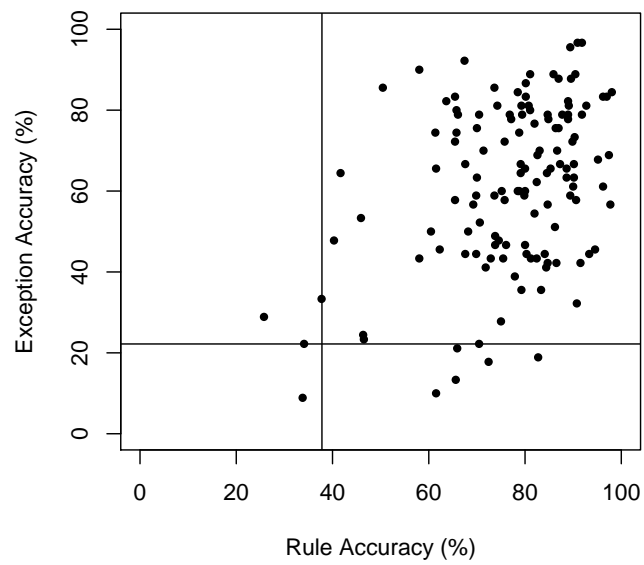


Figure 3.14: Participants from Experiment 3 are plotted by their accuracy on rule and exception items during the last transfer phase. Transfer items were not included in these calculations. The horizontal line is the rule accuracy criterion of 38%. The vertical line is the exception item accuracy criterion of 22%.

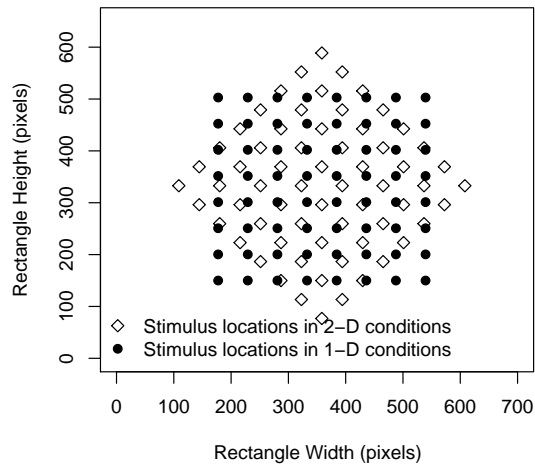


Figure 3.15: The physical dimensions of the stimuli in Experiment 3.

Stimuli

The stimuli were rectangles that ranged on the integral and commensurate dimensions of width and height (see examples in Figure 3.2). Rectangle width and height have been found to be integral dimensions (Dunn, 1983). Additionally, because the dimensions are both measured in units of distance they are commensurate. Each dimension had eight possible values, resulting in 64 possible stimuli, of which each participant only saw 40 (see Figure 3.4). The mapping of stimuli to height and width is shown in Figure 3.15.

For the one-dimensional conditions, rectangle width ranged from 177 pixels to 540 pixels in steps of 52 pixels, and the rectangle height ranged from 150 pixels to 502 pixels in steps of 50 pixels. The positions of the two-dimensional condition stimuli were created by rotating the positions of the one-dimensional stimulus by 45 or -45 degrees around the center of the category structure. For the two-dimensional conditions, the rectangle width ranged from 102 pixels to 614 pixels in steps of 73 degrees, and the rectangle height ranged from 77 pixels to 592 pixels in steps of 73 pixels.

3.3.2 Results and Discussion

Learning

Averaged participant learning curves per condition can be seen in Figure 3.16. All conditions show declines in performance in blocks 8, 14, and 27. In block 8, the exception items were first introduced, which increased the number of valid category labels to four. Blocks 14 and 27 were the start of the transfer phases, which added trials with unfamiliar transfer stimuli that were not followed by correct category labels nor accuracy feedback.

On the last block of training, participants averaged 79.5% rule accuracy. A one-way ANOVA revealed that final training block rule accuracy was not equal across conditions, $F(3, 120) = 7.09, p = .0002$. Participants in the positive diagonal condition had the highest accuracies on rules items (87.7%), followed by the negative diagonal condition participants (78.1%), the vertical condition participants (76.5%), and the horizontal condition participants (75.8%). Tukey's HSD tests found differences between participants in the positive diagonal condition (87.7%) and the horizontal condition (75.8%), $p = .0004$, between participants in the positive diagonal condition (87.7%) and the negative diagonal condition (78.1%), $p = .006$, and between participants in the positive diagonal condition (87.7%) and the vertical condition (76.5%). No differences were found between participants in the negative diagonal, vertical, or horizontal conditions. Additionally, a one-way ANOVA found that participants in the one-dimensional conditions were less accurate on rule items (76.1%) than participants in the two-dimensional conditions (82.7%), $F(1, 122) = 9.24, p = .003$.

On the last block of training, participants averaged 86.8% exception accuracy. A one-way ANOVA revealed significant differences between the conditions, $F(3, 120) = 2.90, p = .04$. Participants in the horizontal condition had the highest accuracies on exception items (91.4%), followed by the positive diagonal condition (87.9%), the negative diagonal condition (84.3%), and the vertical condition (82.7%). Tukey's HSD tests found differences between participants in the horizontal condition (91.4%) and in the vertical condition (82.7%), $p = .001$. No other differences approached significance. Additionally, a one-way ANOVA found that participants in the one-dimensional conditions (87.7%) did not have higher exception accuracy than participants in the two-dimensional conditions (86.0%), $F(1, 122) = 0.52, p = .472$.

To evaluate differences between rule and exception accuracy during the last training block by participant, a rule-advantage score was calculated by subtracting average exception accuracy from average rule accuracy. Overall, participants averaged a -7.3% rule-advantage score, participants tended to perform better on exception items than rule items. A one-way ANOVA revealed

significant difference among the conditions, $F(3, 120) = 8.42, p < .0001$. No condition showed an average positive rule-advantage score, all conditions performed better on exception items. Participants in the positive diagonal condition had the greatest rule-advantage scores (-0.2%), followed by the vertical condition (-6.1%), the negative diagonal condition (-6.2%), and the horizontal condition (-15.6%). Tukey's HSD tests revealed lower rule-advantage scores for participants in the horizontal condition (-15.6%) than in the vertical condition (-6.1%), $p < .0001$, the negative diagonal condition (-6.2%), $p < .013$, and the positive diagonal condition (-0.2%), $p < .0001$. Additionally, a one-way ANOVA found that participants in the one-dimensional conditions had lower rule-advantage scores (-11.6%) than participants in the two-dimensional conditions (-3.1%), $F(1, 122) = 12.27, p = .0006$.

Lastly, t-tests were performed on the rule-advantage scores to determine if they were different than 0. Participants in the horizontal condition (-15.6%) were found to have an average rule-advantage score less than 0, $t(33) = 7.78, p < .0001$, as well as participants in the vertical condition (-6.1%), $t(24) = 2.19, p = .04$, and the negative diagonal condition (-6.2%), $t(33) = 2.60, p = .014$. This suggests these participants were more accurate on exception items than rule items.

Transfer

Transfer and extreme transfer item performance in Experiment 3 can be seen in Figure 3.17. Participants in the horizontal and negative diagonal conditions were not found to have significant differences in transfer and extreme transfer item performance supporting mixtures of participants using rule-based and exception-based strategies. However, the trend in the horizontal condition of more exception-consistent responses than consistent rule responses supports exemplar-based categorization. In contrast, the trend in the negative diagonal condition of more rule responses than exception responses supports rule-based categorization. Thus there is tentative support for one-dimensional and two-dimensional rule use.

Participants in the vertical condition were found to have higher rule consistent response rates than exception-consistent response rates on transfer items, $t(24) = 4.04, p = .0005$, but no differences in extreme transfer performance. This pattern supports one-dimensional rule use. Additionally, participants in the positive diagonal condition were found to have more rule response than exception responses to both transfer items, $t(30) = 9.14, p < .0001$, and extreme transfer items, $t(24) = 5.32, p < .0001$. This supports provides support for two-dimensional rule use. The results

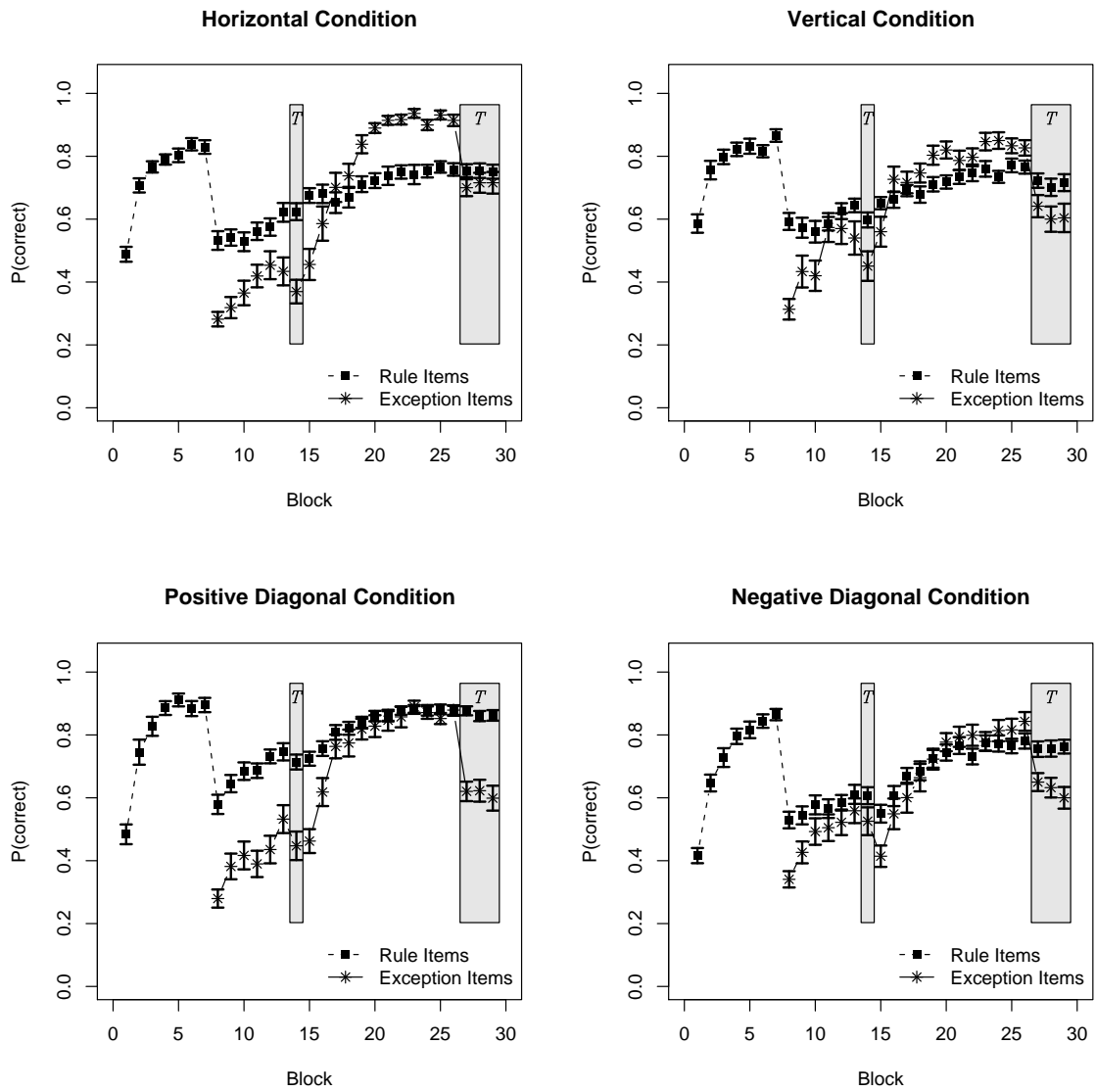


Figure 3.16: Learning curves for Experiment 3. The points are the mean participant accuracies per block by condition. The error bars are the standard deviations of average participant accuracy for each block.

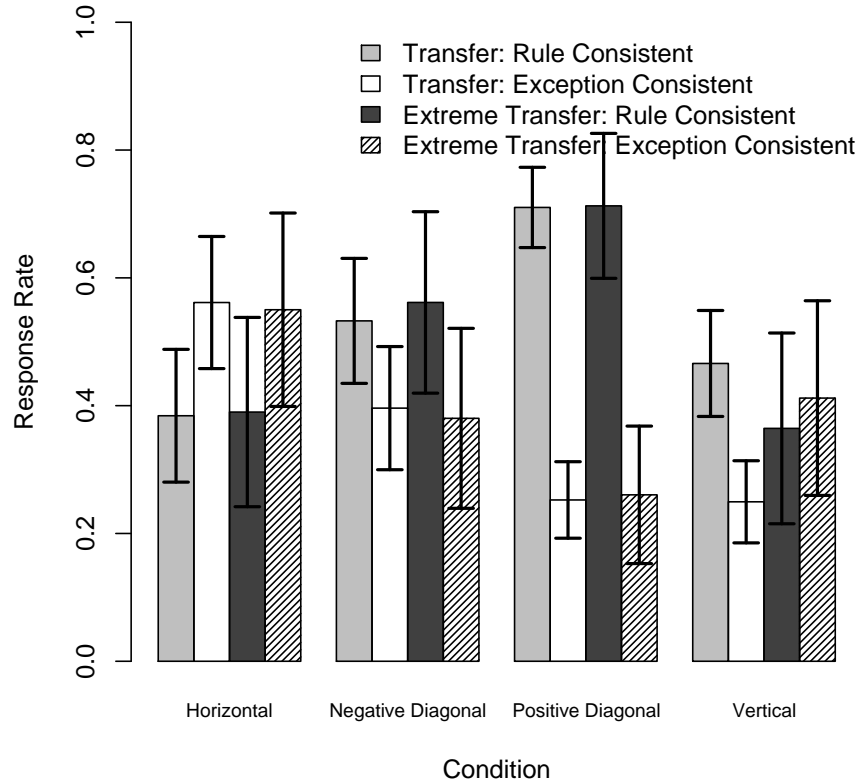


Figure 3.17: Responses to transfer items in Experiment 3 by condition. The extreme transfer items are the two stimuli on the outside corners of the category structure. The transfer items are the remaining 14 non-extreme transfer items. The error bars are the standard errors of average participant response rates.

of the transfer analysis for categories using integral and commensurate dimensions provide support for one-dimensional rules and stronger support for two-dimensional rules.

3.4 Experiment 4: Integral and Noncommensurate Dimensions

Experiment 4 was identical to Experiment 2 and 3, except a change in the stimuli to Fourier descriptors, stimuli containing integral and noncommensurate dimensions.

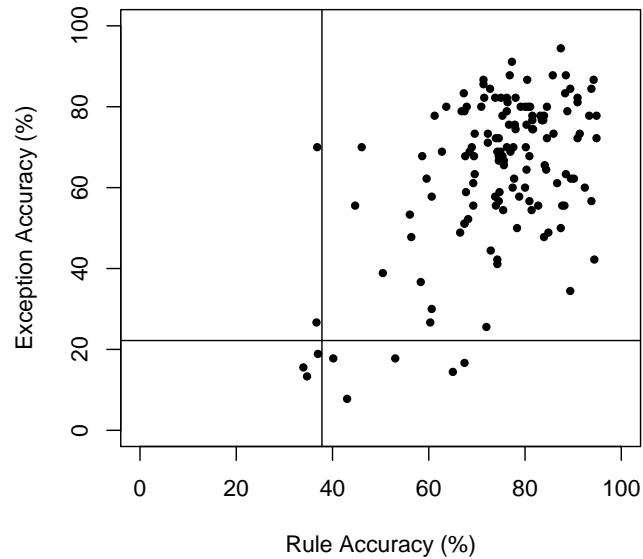


Figure 3.18: Participants from Experiment 4 are plotted by their accuracy on rule and exception items during the last transfer phase. Transfer items were not included in these calculations. The horizontal line is the rule accuracy criterion of 38%. The vertical line is the exception item accuracy criterion of 22%.

3.4.1 Methods

Participants

The participants were 139 students from the University of California, Riverside enrolled in an introductory psychology class who participated to fulfill a class requirement. Applying the criteria from Experiment 1 resulted in the exclusion of data from 10 participants and retaining data from 129 participants (see Figure 3.18). The proportion of participants lost from each condition were not significantly different, $\chi^2(df = 3, N = 139) = 0.125, p = .99$ (see Table 3.4).

Stimuli

The stimuli were constructed using Fourier descriptors. The Fourier descriptors varied on the integral and noncommensurate dimensions of amplitude and phase (see examples in Figure 3.2). Amplitude and phase, when used to construct Fourier descriptors, have been found to be integral dimensions (Op de Beeck, Wagemans, & Vogels, 2003; Cortese & Dyre, 1996). While

Table 3.4: The Breakdown of Participants by Condition and Learning Criterion for Experiment 4.

Learning Criterion	Horizontal	Vertical	Negative Diagonal	Positive Diagonal	Total
Achieved	33	32	32	32	129
Failed	1	5	2	2	10
Total	34	37	34	34	139

both dimensions are measures of rotation, they manifest themselves differently in visual stimuli. Since the dimensions are not measured in the same units, angle of rotation (phase) and complexity (amplitude), they are not commensurate. The Fourier descriptors were composed of three different sine waves of frequencies 2, 4, and 8 cycles per perimeter. The sine waves of frequency 2 and 4 cycles per perimeter had amplitudes of 0.5 radians and phases of 0 degrees. The frequency 8 cycle per perimeter sine wave varied along the dimensions of amplitude and phase. Each dimension possessed eight possible values, resulting in 64 possible stimuli, of which each participant only saw 40 (see Figure 3.4). Figure 3.19 contains the stimuli used in Experiment 4 arranged by condition. The mapping of stimuli to the amplitude and phase of variable sine wave is shown in Figure 3.20.

For the one-dimensional conditions, amplitude ranged from 0.5 radians to 1.55 radians in steps of 0.15 radians and the phase ranged from 0 degrees to 280 degrees in steps of 40 degrees. The positions of the two dimensional condition stimuli were created by rotating the positions of the one-dimensional stimulus by 45 or -45 degrees around the center of the category structure. For the two-dimensional conditions, amplitude ranged from 0.5 radians to 1.55 radians in steps of 0.075 radians and the phase ranged from 0 degrees to 280 degrees in steps of 20 degrees.

3.4.2 Results and Discussion

Learning

Averaged participant learning curves per condition can be seen in Figure 3.21. All conditions show declines in performance in blocks 8, 14, and 27. In block 8, the exception items were first introduced, which increased the number of valid category labels to four. Blocks 14 and 27 were the start of the transfer phases, which added trials with unfamiliar transfer stimuli that were not followed by correct category labels nor accuracy feedback.

On the last block of training, participants averaged 78.2% rule accuracy. A one-way ANOVA revealed that final training block rule accuracy was not equal across conditions, $F(3, 125) = 16.34, p < .0001$. Participants in the horizontal condition had the highest accuracies on rules items

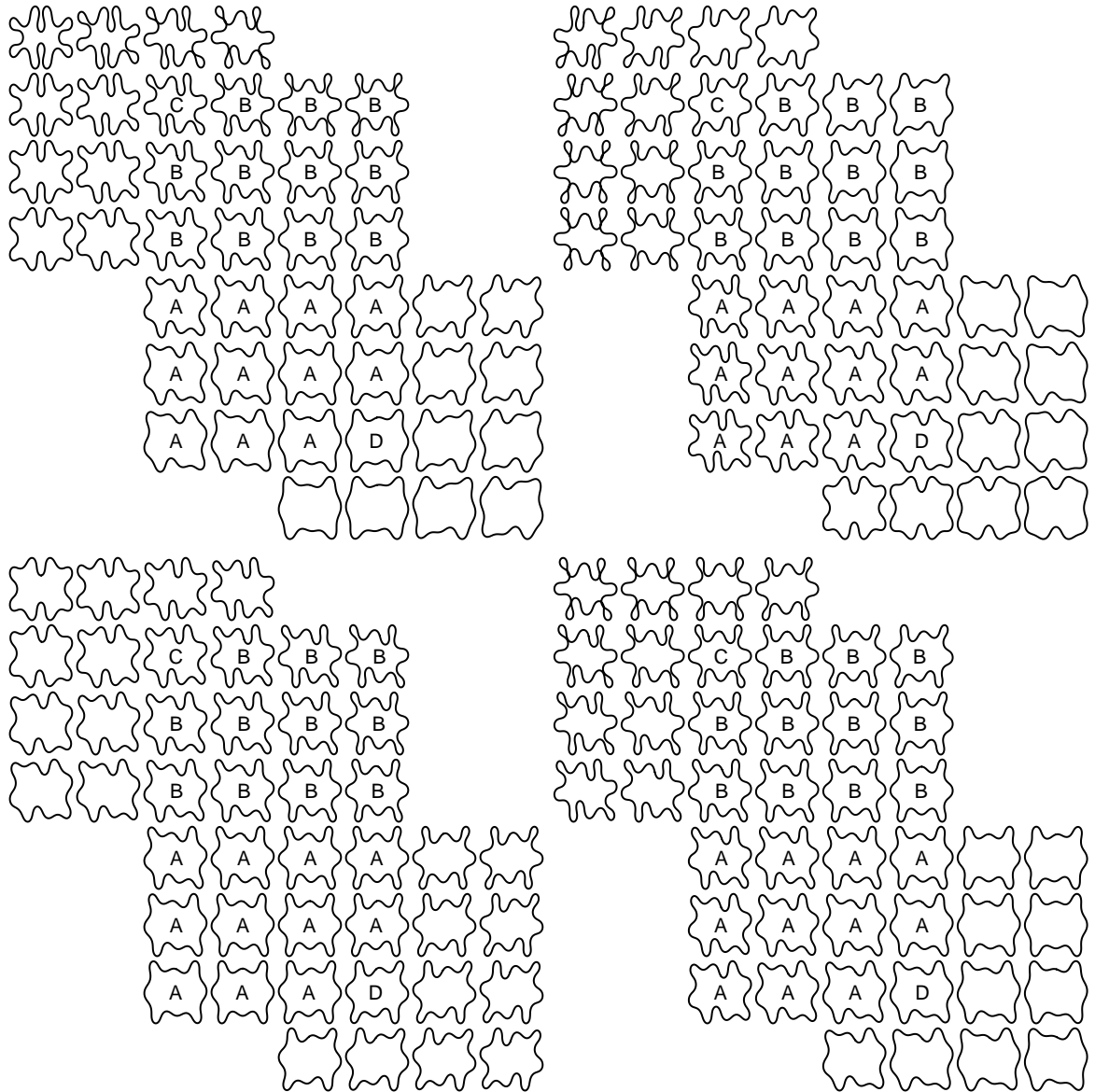


Figure 3.19: The Fourier descriptors used in Experiment 4. The category structures are rotated so that the rule-bounds are parallel to the x-axis. Upper left: Horizontal bounded category structure. Upper right: Vertical bounded category structure. Lower left: Negative diagonal bounded category structure. Lower right: Positive diagonal bounded category structure. Stimuli are labeled according to their category. The transfer stimuli are unlabeled.

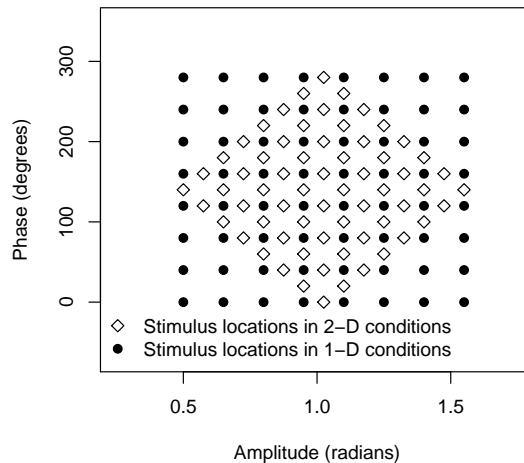


Figure 3.20: The physical dimensions of the stimuli in Experiment 4.

(87.6%), followed by the negative diagonal condition participants (78.1%), the vertical condition participants (77.1%), and the positive diagonal condition participants (70.0%). Tukey's HSD tests found differences between participants in the horizontal condition (87.6%) and the positive diagonal condition (70.0%), $p < .0001$, between participants in the horizontal condition (87.6%) and the vertical condition (77.1%), $p = .0003$, and between participants in the horizontal condition (87.6%) and the negative diagonal condition (78.1%). Participants in the negative diagonal condition (78.1%) were more accurate than participants in the positive diagonal condition (70.0%), $p = .010$, and participants in the vertical condition (77.1%) were also more accurate than participants in the positive diagonal condition (70.0%), $p = .030$. Additionally, participants in the one-dimensional conditions were found to be more accurate on rule items (82.4%) than participants in the two-dimensional conditions (74.1%), $F(1, 127) = 18.15, p < .0001$.

On the last block of training, participants averaged 83.7% exception accuracy. A one-way ANOVA did not reveal any significant difference between the conditions, $F(3, 125) = 2.19, p = .09$. Participants in the horizontal condition had the highest accuracies on exception items (87.1%), followed by the negative diagonal condition (85.4%), the vertical condition (82.8%), and the positive diagonal condition (82.7%). Additionally, a one-way ANOVA found that participants in the one-dimensional conditions (85.0%) did not have lower exception accuracy than participants in the two-dimensional conditions (82.3%), $F(1, 127) = 1.31, p = .255$.

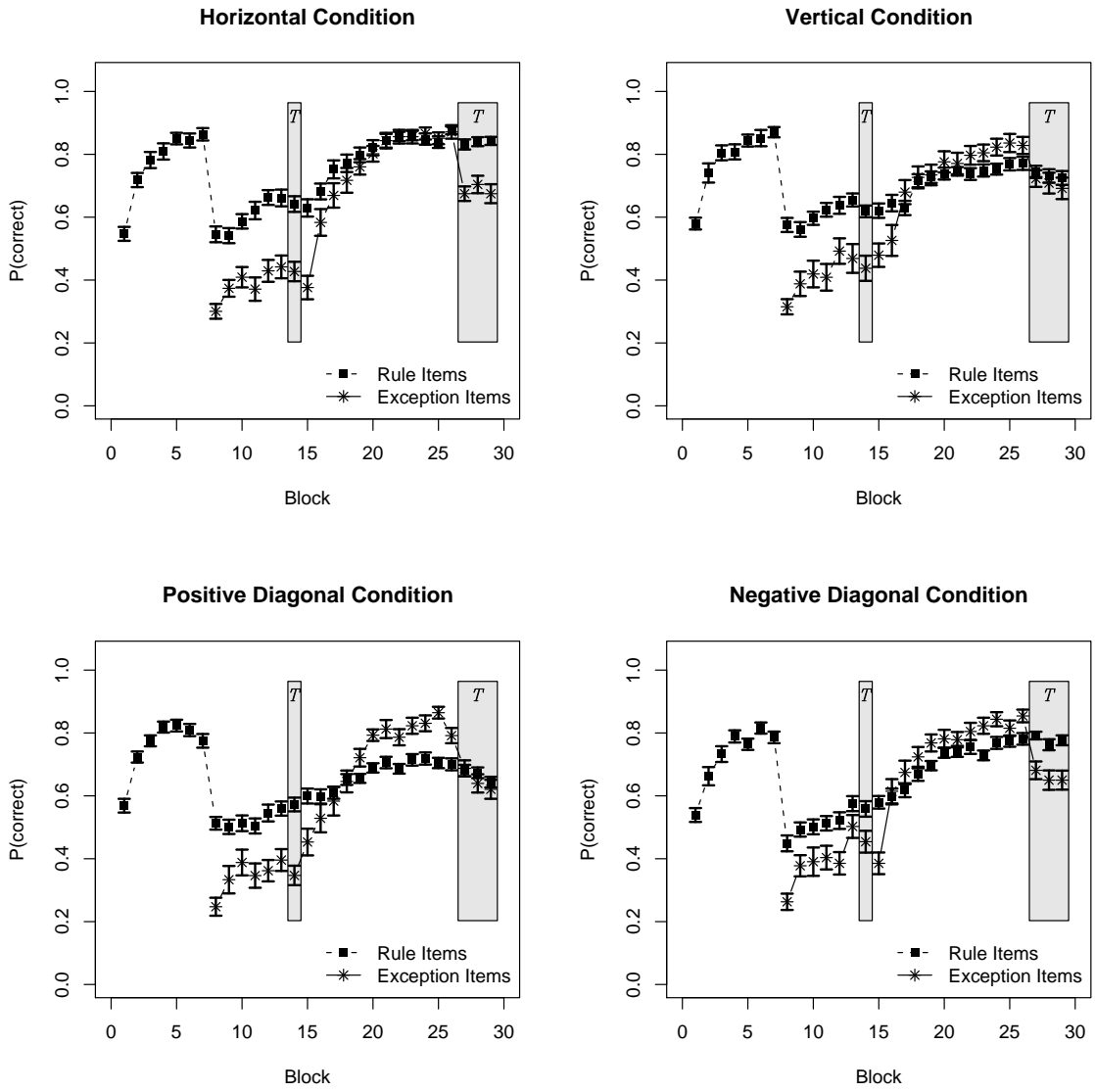


Figure 3.21: Learning curves for Experiment 4. The points are the mean participant accuracies per block by condition. The error bars are the standard deviations of average participant accuracy for each block.

To evaluate differences between rule and exception accuracy during the last training block by participant, a rule-advantage score was calculated by subtracting average exception accuracy from average rule accuracy. Overall, participants averaged a -5.4% rule-advantage score, participants tended to perform better on exception items than rule items. A one-way ANOVA did not reveal any significant differences between the conditions, $F(3, 125) = 2.41, p = .07$. Participants in the horizontal condition had the highest rule-advantage scores (0.4%), followed by the vertical condition (-5.7%), the negative diagonal condition (-7.3%), and the positive diagonal condition (-9.1%). Additionally, a one-way ANOVA found that participants in the one-dimensional conditions had higher rule-advantage scores (-2.5%) than participants in the two-dimensional conditions (-8.2%), $F(1, 127) = 4.36, p = .039$.

Lastly, t-tests were performed on the rule-advantage scores to determine if they were different than 0. Participants in the negative diagonal condition (-7.3%) were found to have an average rule-advantage score less than 0, $t(31) = 3.26, p < .003$, as were participants in the positive diagonal condition (-9.1%), $t(31) = 3.05, p = .005$. This suggests that these participants were more accurate on exception items than rule items.

Transfer

Transfer and extreme transfer item performance in Experiment 3 can be seen in Figure 3.13. Participants in the horizontal condition were found to have more rule-consistent responses than consistent exception responses for transfer items, $t(32) = 2.39, p = .023$. There were no differences in responses for extreme transfer items. This supports a mixture of participants using rule-based and rule- and exception-based categorization strategies. Participants in the vertical condition were found to have more rule responses than exception responses for both transfer items, $t(31) = 5.36, p < .0001$, and for extreme transfer items, $t(31) = 4.26, p = .0002$. This supports rule-based categorization.

Participants in the positive diagonal condition did not show significant differences in response rates. However, there was a trend toward more exception-consistent responses. This pattern of behavior supports a mixture of participants using exception-based and rule- and exception-based categorization strategies. Participants in the negative diagonal condition showed strong exception response rates for transfer items, $t(31) = 8.18, p < .0001$, and for extreme transfer items, $t(31) = 5.22, p < .0001$. This supports exception-based categorization. The results of the transfer analysis for categories using integral and noncommensurate dimensions provide support for one-dimensional rules and weak or no support for two-dimensional rules.

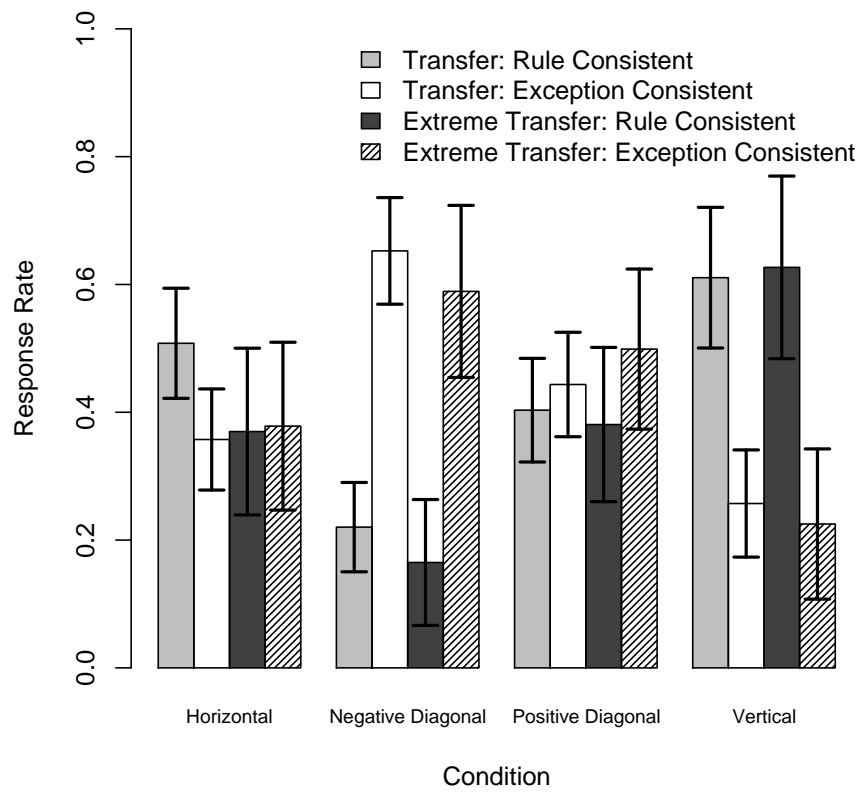


Figure 3.22: Responses to transfer items in Experiment 4 by condition. The extreme transfer items are the two stimuli on the outside corners of the category structure. The transfer items are the remaining 14 non-extreme transfer items. The error bars are the standard errors of average participant response rates.

3.5 Evaluation of Procedure Changes

After Experiment 1, adjustments were made to the procedure to enhance participant learning. The following experiments did have a large proportion of participants reaching the learning criterion than Experiment 1. In Experiment 1, 80% of the participants reached the learning criterion. This increased to 87.5% in Experiment 2, 93% in Experiment 3, and 92% in Experiment 4.

3.6 Summary

Four experiments were performed to test for two-dimensional rule use. The four experiments were designed to test one of the unique factorial combinations of the two factors of dimensionality (separable versus integral) and verbalizability (noncommensurate versus commensurate). Evidence of one-dimensional rule use was found across all experiments. However, evidence of two-dimensional rule use was only found in Experiments 1 and 3. These are the two experiments that used stimuli composed of commensurate stimulus dimensions. This is consistent with participants only using two-dimensional rules when stimuli are composed of commensurate dimensions.

Chapter 4

Cluster Analysis

When given a categorization task to perform, participants often use different strategies and develop different representations (Erickson & Kruschke, 1998; Ashby et al., 1998). When analyzing data it is necessary to select a level of analysis that can both discriminate between strategies and also retain sufficient information to distinguish consistent behaviors from random error. By combining across too much data, individual differences in strategy and representation are lost. By combining across too few data, consistent behaviors are distorted by random error. When choosing to analyze a data set, the questions that are being asked about the data set, determine the appropriate level of analysis.

In this paper, participants were analyzed at the level of strategy. Following Lee and Webb (2005), cluster analyses were used to group participants according to displayed strategies. These groups were then used for model fitting. This procedure allows for the advantages of both *individual* and *group* level of analysis. By fitting *groups* of participants, the influence of random factors on participant performance was reduced by the larger amount of stable nonrandom performance. Likewise, by fitting *individuals* grouped by similar categorization behaviors, unique strategies are preserved. Consequently, the technique of analyzing participants grouped by strategy allows both the noise-reduction advantages of group fitting and the retention of unique strategies of individual fitting.

To group participants using similar strategies a series of cluster analyses were performed upon participant transfer item performance. The participants were first clustered by experiment and then clustered across experiments. The individual experiment cluster analyses allow for the characterization of participant behaviors within a common set of conditions and stimuli. The combined experiment cluster analyses allow for the determination of participant categorization behaviors that

were consistent across experiments. In the following, individual cluster analyses by experiment are discussed and then followed by the combined experiment cluster analysis.

4.1 Cluster Analysis Method

Participant data from the transfer phase was prepared for the cluster analysis by converting each participant's performance into an abstracted form. Individual participant performance was characterized by the frequency at which a participant made each type of response to each of the 40 stimuli (see Figure 3.4). The four types of responses were: rule-consistent, rule inconsistent, exception consistent, and exception-inconsistent (see previous chapter for further clarification). Therefore, each participant was described by a vector of 40 stimuli by 4 types of responses, or 160 numbers.

These vectors were transformed into item profile plots for easier description and understanding of the strategies that produced each type of behavior. For simplicity, the horizontal condition category structure was chosen as the representative abstract structure. All subsequent figures and descriptions are based on this orientation of the category structure (see Figure 4.1). Each participant or group of participants is represented by a row of four profile plots. The plots in each of the four columns represents one type of response. Each of the cells represent one unique item, oriented according to its location in the category structure. The darkness in shading of each of the cells in each plot represents the frequency of that response. The darker the cell the greater frequency of that response. The training items are the items outlined by the thicker black box and the transfer items are the items in the upper left and lower right corners (for more detail refer back to Figure 3.4).

Figure 4.1 contains three different patterns that could be generated by a participant who was 100% accurate on training items. High performance on the training items requires both high proportions of rule-consistent responses to rule training items and high proportions of exception-consistent responses to exception training items. A high proportion of rule-consistent responses to rule training items is displayed in the rule-consistent response column (the first column) of Figure 4.1. A high proportion of exception-consistent responses to exception training items is displayed in the exception-consistent response column (the third column) of Figure 4.1.

The difference between the three patterns in Figure 4.1 reflects different types of performance on the transfer items. Transfer item performance can be used to determine the type of strategy used by the participant (Erickson & Kruschke, 1998). The top pattern, the *rule generalization* pattern, corresponds to a participant classifying transfer items according to the nearest

rule item. This type of pattern is generated by participants using rule- and exception-based categorization. The middle pattern, the *exception generalization* pattern, corresponds to a participant classifying transfer items according to the nearest exception item. This type of pattern is generated by participants using exemplar-based categorization. The bottom pattern, the *guessing generalization* pattern, corresponds to a participant who was equally likely to respond to a transfer item with each of the four possible category labels. This may be a result of a participant simply guessing on each of the transfer items. This type of pattern may also emerge when combining across participants using different strategies, whose averaged performance resembles that of a single participant using a guessing strategy. Regardless, this type of pattern does not provide evidence as to what kind of strategy was used to learn the categories.

These patterns are used as guidelines to interpret the patterns displayed by participants. If participants in the two-dimensional category conditions are found to perform in a manner similar to the *rule generalization* pattern, this provides support for the use of two-dimensional rules. Participants displaying other types of transfer item generalization may still be using two-dimensional rules, however they do not provide direct evidence of such use.

A series of cluster analyses using the Partitioning Around Medoids (PAM: Kaufman & Rousseeuw, 2005) method was performed upon these data. PAM partitions observations into clusters by maximizing the similarity of observations to their assigned cluster and maximizing the dissimilarity of the clusters to each other. To accomplish this, PAM searches a data set until it finds a series of representative objects, or *medoids*. Observations are then assigned to the cluster that possesses the closest medoid. PAM attempts to find the best k clusters, where k is provided by the user. The final selection of medoids and clusters minimizes the summed distance of all cluster members to their cluster medoid.

For this data, a series of different PAM cluster solutions using a Euclidean distance metric were considered as possible participant groupings. The final clustering solution was determined by a combination of the interpretability of the clusters, the number of clusters suggested by an agglomerative cluster solution using Ward's method, and the average silhouette width. Kaufman and Rousseeuw (2005) proposed silhouette width as a means of evaluating clustering solutions and the average silhouette width as a means of evaluating the fitness of the entire clustering solution. Silhouette plots are a graphical representation of the silhouette widths of all observations in a clustering solution, and are used to explain the findings. Figure 4.2 displays an example of a silhouette plot.

In a silhouette plot, silhouette width is calculated for each observation in the data and then arranged by cluster. The silhouette width is a measure of the goodness of fit that each observation

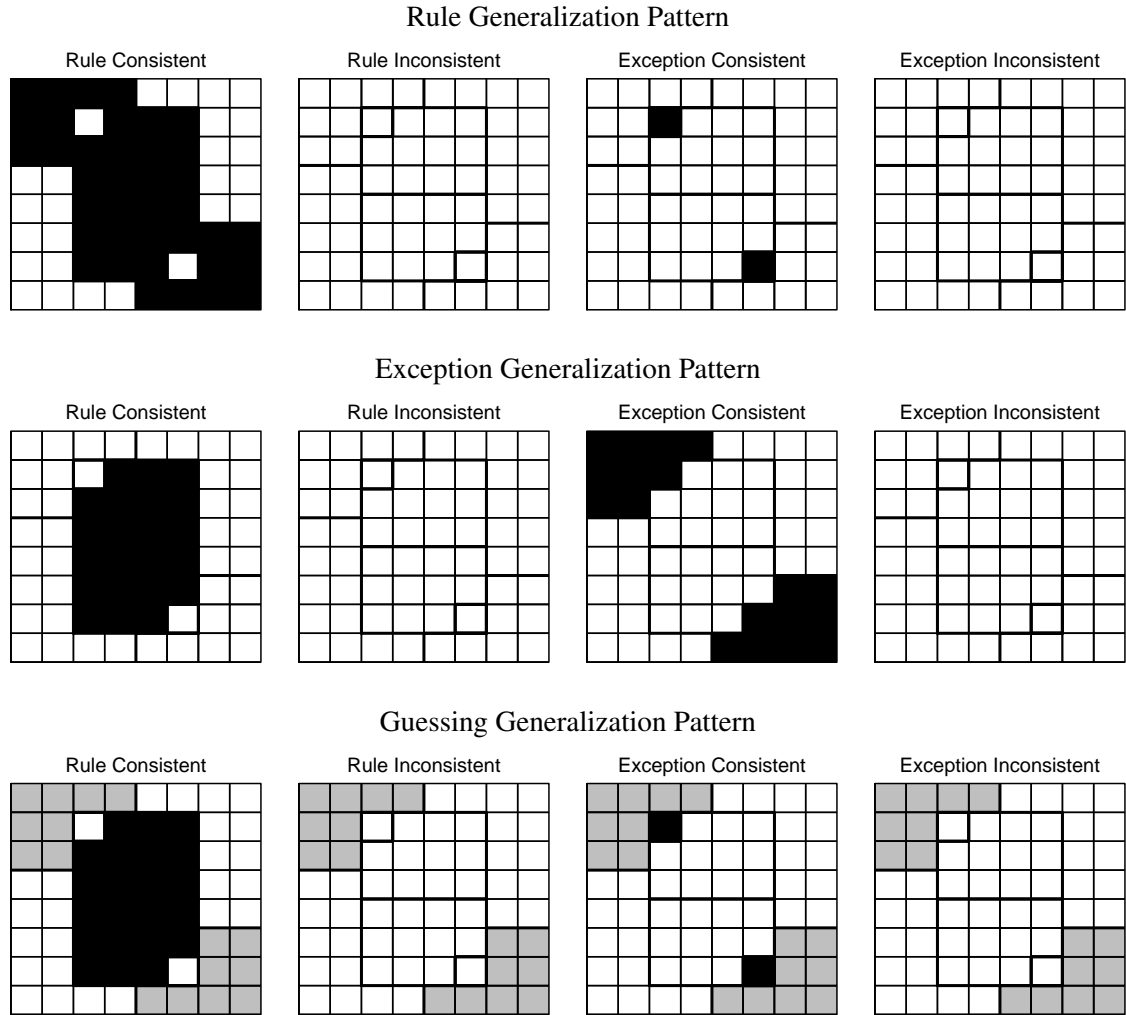


Figure 4.1: Examples of individual participant transfer performance.

has to its cluster. Silhouette widths range from -1 to 1, with positive values being good fits of an observation to a cluster, zero being an observation that is between two clusters, and negative values being an observation that may be more appropriately placed in another cluster.

The silhouette width for observation i , s_i , is defined as:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)}, \quad (4.1)$$

where a_i is the average dissimilarity of observation i to all other members of its assigned cluster, and b_i is the average dissimilarity of observation i to all members of its nearest neighboring cluster.

Table 4.1: Division of Participants into Clusters in Experiment 1

Cluster	1D	2D
1	16	15
2	5	16
3	14	0
4	2	5
5	3	0
6	0	4

The silhouette width is a measure of clustering fitness. It is a function of the difference between an observation and its cluster and the nearest neighboring cluster, weighted by the larger of the two differences. It is based on the average distance of each observation to the members of a cluster. It is formally equivalent to the distance between an observation and the mean of a cluster. When observation i approaches the average (or center) of its cluster, a_i is minimized, and the silhouette width is reduced to $\frac{b_i}{b_i}$ or 1. When the observation is midway between two clusters, a_i and b_i are approximately equal, and the silhouette width is reduced to $\frac{0}{\max(a_i, b_i)}$ or 0. When the observation approaches the average (or center) of its neighboring cluster, b_i is minimized, and the silhouette width is reduced to $\frac{-a_i}{a_i}$ or -1 .

In some cases silhouette plots show PAM producing clusters with negative silhouette widths. This is because PAM creates clusters by minimizing the distance between observations and the medoids of each cluster, whereas silhouette width is calculated using the distance between observations and the average of each cluster. Clustering solutions with many negative or near zero silhouette widths suggest that the clusters are poorly formed. This may be the result of large amounts of variance in the data leading to poorly defined clusters or choosing a clustering solution that does not reflect the underlying groups in the data.

4.2 Experiment 1: Separable and Commensurate Dimensions

Figure 4.2 contains the silhouette plot of the six cluster solution for Experiment 1 and the dendrogram created by the agglomerative clustering. Figure 4.3 shows the average participant performance for the six cluster solution for Experiment 1. Table 4.1 shows the breakdown of participants into clusters based on condition. This solution contains three large and consistent clusters

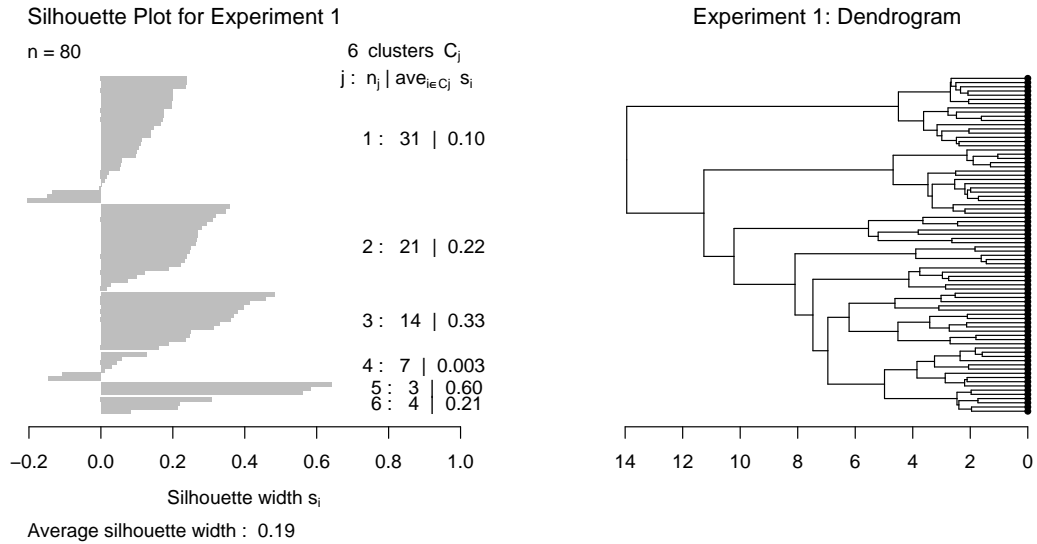


Figure 4.2: The left panel contains the silhouette plot for Experiment 1. The right panel contains the dendrogram for Experiment 1.

(Clusters 1, 2, and 3), two small and consistent clusters (Clusters 5 and 6), and one small and inconsistent cluster (Cluster 4).

Cluster 1 contains participants who classified transfer items as members of the rule-consistent category. This suggests that these participants used rule- and exemplar-based categorization. This cluster contains the most participants and has a positive average silhouette width. Cluster 1 also contains a mixture of participants from both the one-dimensional and two-dimensional conditions. Therefore, Cluster 1 provides evidence for rule use in one-dimensional conditions, as previously found with separable and commensurate dimensions, and may also provides evidence for two-dimensional rule use.

Cluster 2 contains participants who classified transfer items as members of the exception-consistent category. This suggests that these participants used exemplar-based categorization. This cluster contains the second most participants and has a positive average silhouette width. Cluster 2 contains one third as many participants from the one-dimensional conditions as Cluster 1. This suggests that participants were more likely to use rule generalization in the one-dimensional conditions than exception generalization. Cluster 2 contains approximately the same number of participants from the two-dimensional conditions as Cluster 1. This suggests that participants are equally likely to use rule generalization as exception generalization in the two-dimensional conditions. Although Cluster 2 does not suggest rule-use, it does contain a large portion of participants from the two-

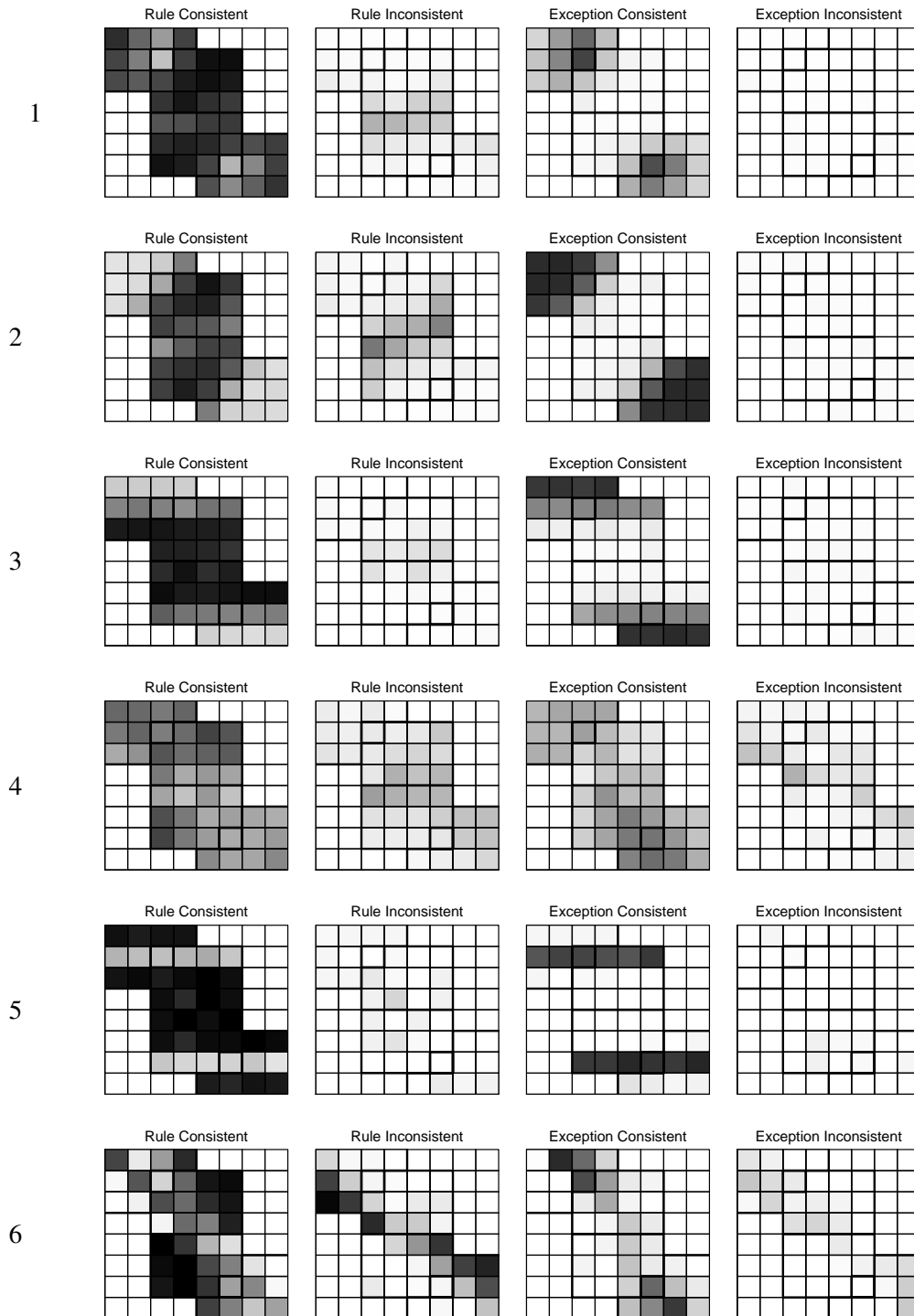


Figure 4.3: The proportion of responses for each stimulus during the transfer phase for the average participant per cluster of Experiment 1. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of response. The cluster number is displayed to the left of each row.

dimensional conditions and is a well-formed cluster. Thus Cluster 2 may still provide evidence that may lead to a better understanding of when two-dimensional rules may be used.

Cluster 3 contains participants who classified transfer items as members of the exception-consistent category, this suggests exemplar-based categorization. Cluster 3 contains only half the number of participants as Cluster 1, but has a high positive average silhouette width. Furthermore, Cluster 3 contains only participants from the one-dimensional conditions. While the participants did generalize to the exception-consistent category, participant strategies might not be exemplar-based.

Participants in Cluster 3 were found to have high rates of consistent exception responses to stimuli that shared the same feature (or a more extreme value) on the primary dimension as the exception training items. This strategy can be described by the theoretical rules of *equal to or more extreme*. The term *theoretical rule* is used here as a description of the strategy used by the participants and does not imply participants used a rule. The use of this strategy creates bands of exception responses that can even cross into the training items, as shown in Figure 4.3.

These patterns of response bands can be explained in different ways. One explanation is that participants did use exemplar-based categorization to learn the structure, as suggested by the high frequency of exception-consistent responses to the transfer items. Another possibility, is that participants solved the structure using four one-dimensional rules. These rules would correspond to two rules used to classify the rule items (i.e, category A if it is tall or category B if it is short) and the two theoretical rules described above (one for each group of exceptions). Regardless, Cluster 3 does not contain any participants from the two-dimensional conditions, and therefore does not provide any evidence for two-dimensional rule use.

Cluster 4 contains only 7 participants, and has an average silhouette width near 0. The averaged participant profile plot provides no strong indication of a meaningful strategy. Individual participant performance plots supported this finding. They suggested that these participants had not mastered the categorization task. The lack of consistent behaviors between participants and the lack of a meaningful strategy prevents Cluster 4 from providing any evidence for two-dimensional rule use.

Cluster 5 is the smallest cluster, containing only three participants, but also has the highest average silhouette width. These three participants performed similarly and came from the one-dimensional conditions. These participants' transfer phase performance was similar to participants in Clusters 3. However, whereas participants in Cluster 3 seem to have generalized according to theoretical rule of *equal to or more extreme*, Cluster 5 participant performance can be described as using the theoretical rule of *equal to*. This type of behavior suggests the use of multiple rules along

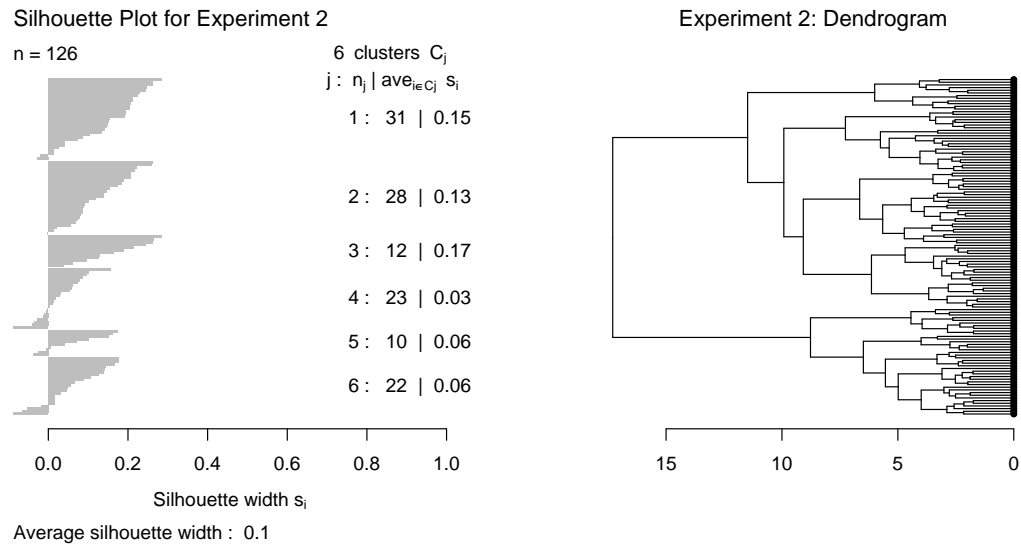


Figure 4.4: The left panel contains the silhouette plot for Experiment 2. The right panel contains the dendrogram for Experiment 2.

a primary dimension. Regardless, all participants were from the one-dimensional condition, and are therefore not informative on the use of two-dimensional rules.

Cluster 6 contains four participants from the two-dimensional conditions. Their strategy can be best described as attempting to use one-dimensional rules upon a two-dimensional structure. Although this cluster is not informative on the use of two-dimensional rules it does provide evidence that participants use one-dimensional rules, even when such strategies are not optimal.

In conclusion, the cluster analysis for Experiment 1 provides evidence for rule use in a task using stimuli with separable and commensurate dimensions. Additionally, with 37% of the participants in the two-dimensional conditions, Cluster 1 provides evidence for two-dimensional rule use.

4.3 Experiment 2: Separable and Noncommensurate Dimensions

Figure 4.4 contains the silhouette plot of the six cluster solution for Experiment 2 and the dendrogram produced by the agglomerative clustering. Figure 4.5 shows the averaged participant performance for the six cluster solution for Experiment 2. Table 4.2 shows the breakdown of participants into clusters based on condition. This solution contains two large and consistent clusters

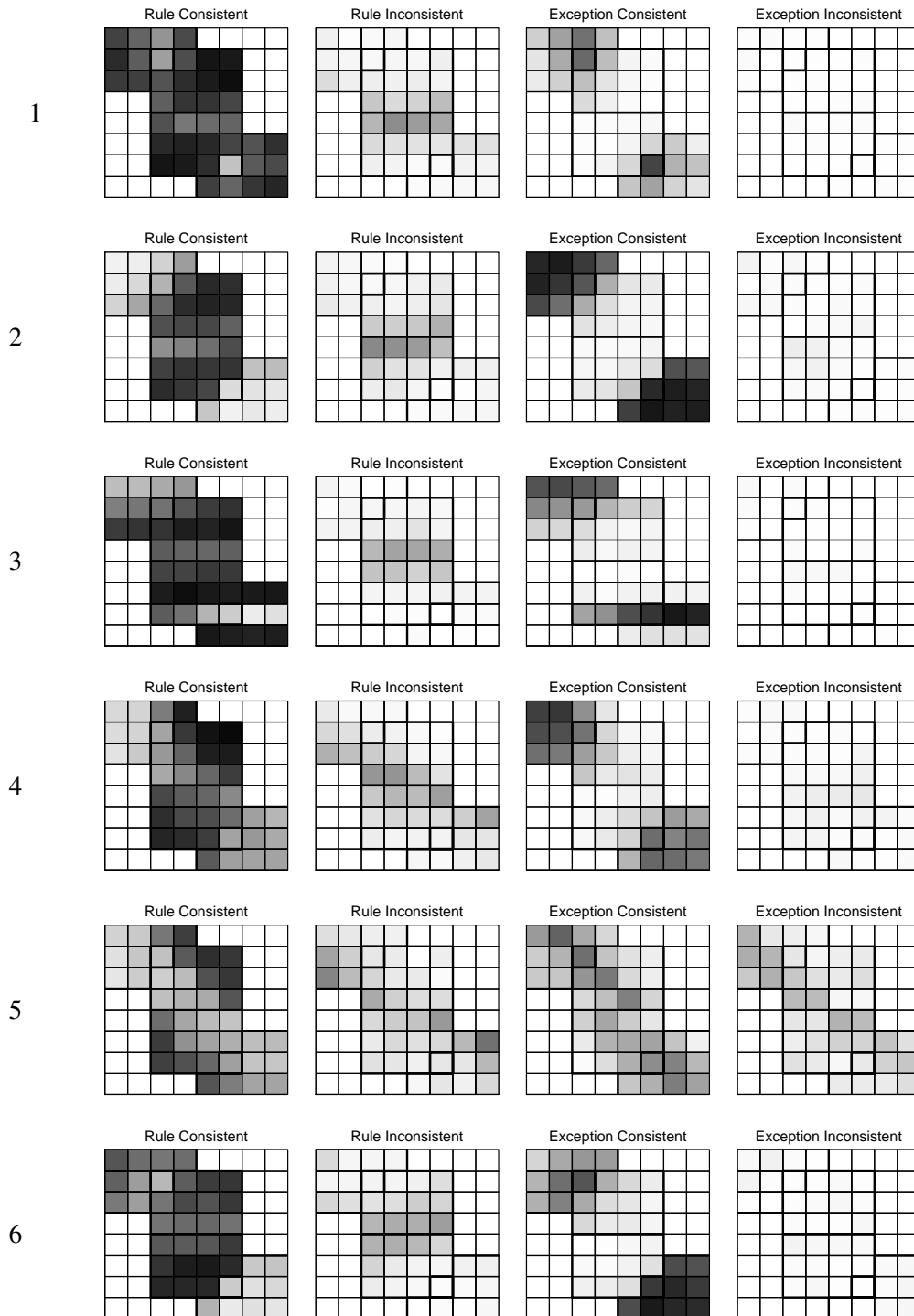


Figure 4.5: The proportion of responses for each stimulus during the transfer phase for the average participant per cluster of Experiment 2. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of response. The cluster number is displayed to the left of each row.

Table 4.2: Division of Participants into Clusters in Experiment 2

Cluster	1D	2D
1	20	11
2	11	17
3	12	0
4	6	17
5	3	7
6	6	16

(Clusters 1 and 2), two large and inconsistent clusters (Clusters 4 and 6), one small and consistent cluster (Cluster 3), and one small and inconsistent cluster (Cluster 5).

Cluster 1 contains participants who classified transfer items as members of the rule-consistent category. This suggests that these participants used both rule- and exemplar-based categorization. This cluster contains the most participants and has a positive average silhouette width. Cluster 1 also contains a mixture of participants from both the one-dimensional and two-dimensional conditions. Additionally, there were twice as many participants from the one-dimensional condition than from the two-dimensional condition which again suggests that two-dimensional rules may be more difficult to implement than one-dimensional rules. The participants using rule-based generalization in the one-dimensional conditions in Cluster 1 provide evidence for rule use in one-dimensional conditions. Likewise, the participants using rule-based generalization in Cluster 1 from the two-dimensional conditions also provide evidence for for two-dimensional rule use in the two-dimensional conditions.

Cluster 2 contains participants who classified transfer items as members of the exception-consistent category. This suggests that these participants used exemplar-based categorization. This cluster also contains the second most participants and has a positive average silhouette width. Cluster 2 also contains half as many participants from the one-dimensional conditions as Cluster 1. This suggests that participants were more likely to use rule-generalization in the one-dimensional conditions than exception generalization. Cluster 2 contains approximately 1.5 times as many participants in the two-dimensional conditions as Cluster 1. This suggests that participants were more likely to use exception generalization than rule generalization in the two-dimensional conditions. Although cluster 2 does not suggest rule-use, it does contain a large portion of participants from the two-dimensional conditions and is a well-formed cluster. Thus Cluster 2 may still provide evidence

leading to a better understanding of when two-dimensional rules may be used.

Cluster 3 contains participants who classified transfer items as members of the exception-consistent category. This suggests that these participants used exemplar-based categorization. Cluster 3 contains only half as many participants as Cluster 1, but has a high positive average silhouette width. Furthermore, Cluster 3 only contains participants from the one-dimensional conditions. While the participants did generalize to the exception-consistent category participant strategies might not be exemplar-based. These participants performed in a manner similar to the participants in Cluster 3 from Experiment 1. Thus these participants may have also used multiple one-dimensional rules to learn the category structure. Regardless of the strategy used by the participants, Cluster 3 contains no participants from the two-dimensional clusters. Therefore, Cluster 3 does not provide evidence for two-dimensional rule use.

Cluster 4 contains 23 participants and has an average silhouette width of near 0. This indicates a cluster of participants with a mixture of different generalization patterns. Containing three times as many participants from the two-dimensional conditions than one-dimensional conditions, this cluster has a large number of participants who used a one-dimensional rule to learn a two-dimensional category. Inspection of the individual participant profile plots support this interpretation. The high exception-consistent response rates to the transfer items also suggest the use of exemplar-based strategies. While these participants may seem to have performed poorly in the task, examination of the training items show that they learned the category. It is only when considering their performance on the transfer items do these participants seem to have done poorly. This cluster provides evidence for participants using different strategies of category learning, in this case using one-dimensional boundaries when two-dimensional boundaries are optimal. However, it does not provide support for two-dimensional rule use.

Cluster 5 is the smallest cluster, containing only 10 participants, and also has a near zero silhouette width. The transfer patterns of participants in Cluster 5 are similar to those of Cluster 4. Additionally, like Cluster 4, Cluster 5 is made up primarily of participants from the two-dimensional conditions. The main difference between the clusters is how participants generalized to the extreme transfer items. Participants from Cluster 4 tended to give exception consistent responses, while participants from Cluster 5 tended to give more rule inconsistent and exception-inconsistent responses. This pattern of results is consistent with using multiple rules along a single dimension. Like Cluster 4, the participants in this cluster do not provide support for two-dimensional rule use.

Cluster 6 contains 22 participants. Cluster 6 has a low average silhouette width and consists of a large distribution of participant response patterns. Furthermore, 13 participants were from

Table 4.3: Division of Participants into Clusters in Experiment 3

Cluster	1D	2D
1	20	50
2	18	2
3	21	13

the positive diagonal condition which means that the behavior in this cluster was primarily driven by these participants. The strategy used by the participants in the cluster can be best described as rule generalizing to some transfer items but exception generalizing to other transfer items. This may suggest that participants were using two different strategies based upon the location of the stimuli. When the participants were presented large circles with radial line segments possessing small angles, participants were more likely to generalize according to rule-consistent category. Whereas when the participants were presented with small circles with radial line segments possessing large angles, participants were more likely to generalize to the exception-consistent category. While this cluster is informative upon how participants may learn categories with stimuli having separable and noncommensurate dimension, it does not provide support for two-dimensional rule use.

In conclusion, the cluster analysis for Experiment 2 provides evidence for rule use in a task using stimuli with separable and noncommensurate dimensions. Additionally, with 11 of the participants in the two-dimensional conditions, Cluster 1 provides evidence of two-dimensional rule use. In contrast to the clustering solution of Experiment 1, only half of the participants in the two-dimensional conditions were in clusters that may provide meaningful evidence for two-dimensional rule use. The other participants from the two-dimensional conditions were members of clusters that were best explained by using one-dimensional rules. This suggests that participants may prefer to use one-dimensional rules when learning category structures with two-dimensional boundaries with stimuli that possess separable and noncommensurate dimensions. This contrasts the behaviors of participants in the two-dimensional conditions in Experiment 1, who were more likely to display two-dimensional rule-like behaviors.

4.4 Experiment 3: Integral and Commensurate Dimensions

Figure 4.6 contains the silhouette plot of the three cluster solution for Experiment 3 and the dendrogram produced by the agglomerative clustering. Figure 4.7 shows the averaged participant performance of the three cluster solution for Experiment 3. Table 4.3 shows the breakdown of

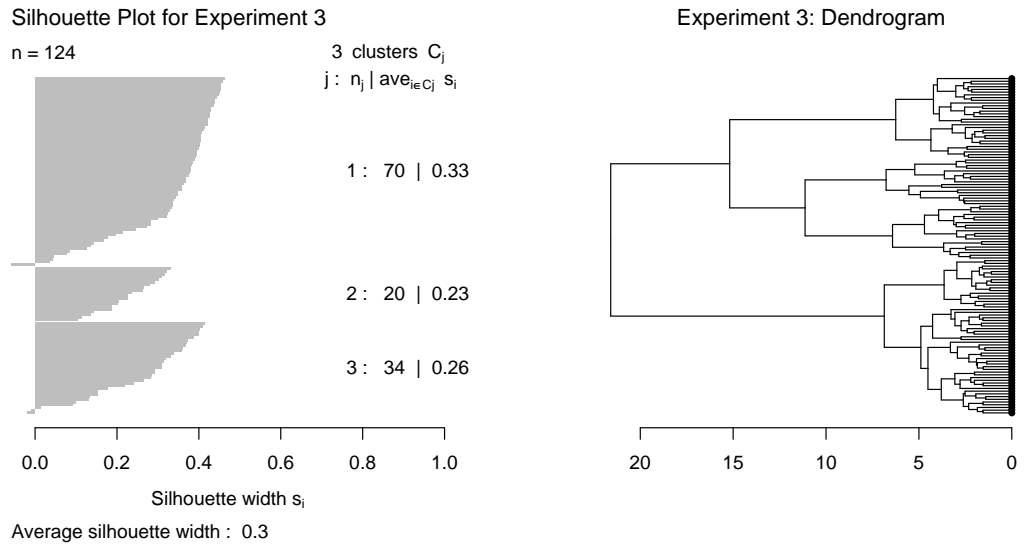


Figure 4.6: The left panel contains the silhouette plot for Experiment 3. The right panel contains the dendrogram for Experiment 3.

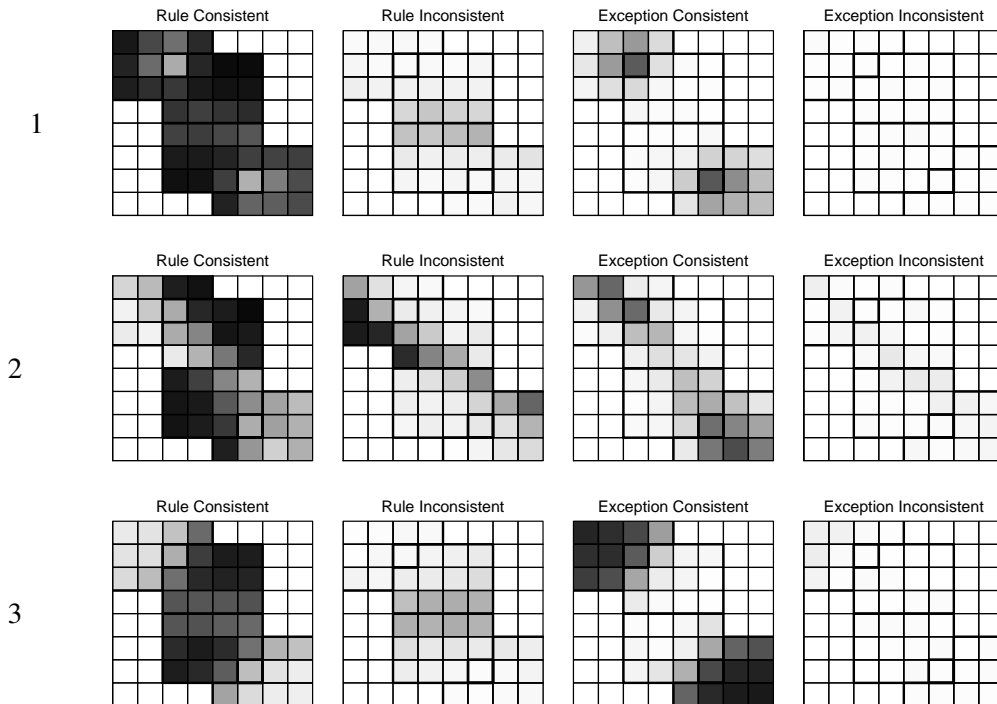


Figure 4.7: The proportion of responses for each stimulus during the transfer phase for the average participant per cluster of Experiment 3. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of response. The cluster number is displayed to the left of each row.

participants into clusters based on condition. This solution contains three large and consistent clusters. The dendrogram in Figure 4.6 suggests that a four cluster solution might be more appropriate, however upon adding a fourth cluster, the first cluster was broken up into two groups that did not have clear meaningful differences.

Cluster 1 contains participants who classified transfer items as members of the rule-consistent category. This suggests that these participants used both rule- and exemplar-based categorization. This cluster contains the most participants and has a positive average silhouette width. Cluster 1 also contains a mixture of participants from both the one-dimensional and two-dimensional conditions. Additionally, there were twice as many participants from the two-dimensional condition than from the one-dimensional condition. Indeed 77% of the participants in the two-dimensional conditions were placed into Cluster 1. This provides evidence for one-dimensional and two-dimensional rules, and suggests that with stimuli possessing integral and commensurate dimensions, two-dimensional rule use is a dominant strategy.

Cluster 2 contains participants who classified transfer item as members of the exception-consistent, rule-consistent and rule inconsistent categories. This cluster is predominately composed of participants from the one-dimensional conditions. This type of transfer pattern suggests that participants were using a two-dimensional rule bound to learn a one-dimensional category structure.

Previous research (e.g., Lewandowsky et al., 2006; Monahan & Lockhead, 1977) has found that in some cases integral dimensions, such as the dimensions used in this experiment (rectangles that vary in height and width), are perceived not as different dimensions, but instead as a single dimension. This may account for the behaviors of some participants. They may have viewed these stimuli as varying on area or the ratio of height to width rather than varying in height and width. If these participants applied one-dimensional rules, their resulting behavior would resemble two-dimensional rule use. These participants may be better classified as one-dimensional rule users rather than two-dimensional rule users. This suggests that the participants in Cluster 2, who followed what appears to be a two-dimensional rule in a one-dimensional category, may instead be using one-dimensional rules based on the dimension of area. Thus, while this cluster is informative upon how participants may learn categories with integral stimuli, it does not provide support for two-dimensional rule use. Additionally, the other participants in the two-dimensional conditions may have perceived the stimuli in similar ways. This would result in them performing as if they used two-dimensional rules yet actually using one-dimensional rules.

Cluster 3 contains participants who classified transfer items as members of the exception-consistent category. Their behavior is consistent with an exemplar-based strategy. This cluster is

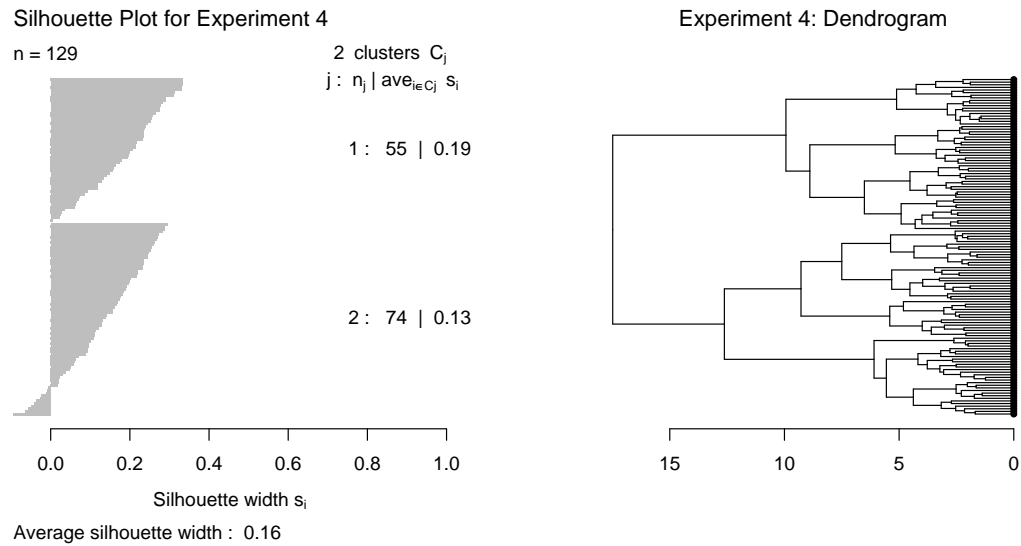


Figure 4.8: The left panel contains the silhouette plot for Experiment 4. The right panel contains the dendrogram for Experiment 4.

primarily composed of participants from the one-dimensional conditions, but also contained some participants from the two-dimensional conditions. The majority of participants in the two-dimension conditions were classified into Cluster 1. This suggests that participants were much more likely to demonstrate rule generalization to the transfer items than to the exception items.

In conclusion, the cluster analysis for Experiment 3 provides evidence for rule use in a task using stimuli with integral and commensurate dimensions. The clustering patterns indicated that not only were participants in the two-dimensional conditions using two-dimensional rule bounds, but that this was the most prevalent type of transfer pattern. However, the given the behaviors of the participants in Cluster 2, it may be the case that many participants viewed stimuli as being unidimensional. If this is so, then the two-dimensional rule-like behaviors displayed by the participants in Cluster 1, may actually be one-dimensional rule-like behaviors.

4.5 Experiment 4: Integral and Noncommensurate Dimensions

Figure 4.8 contains the silhouette plot of the two cluster solution for Experiment 4 and the dendrogram produced by the agglomerative clustering. Figure 4.9 shows the averaged participant performance of the two cluster solution for Experiment 4. Table 4.4 shows the breakdown of participants into clusters based on condition. This solution contains only two clusters. Both clusters are

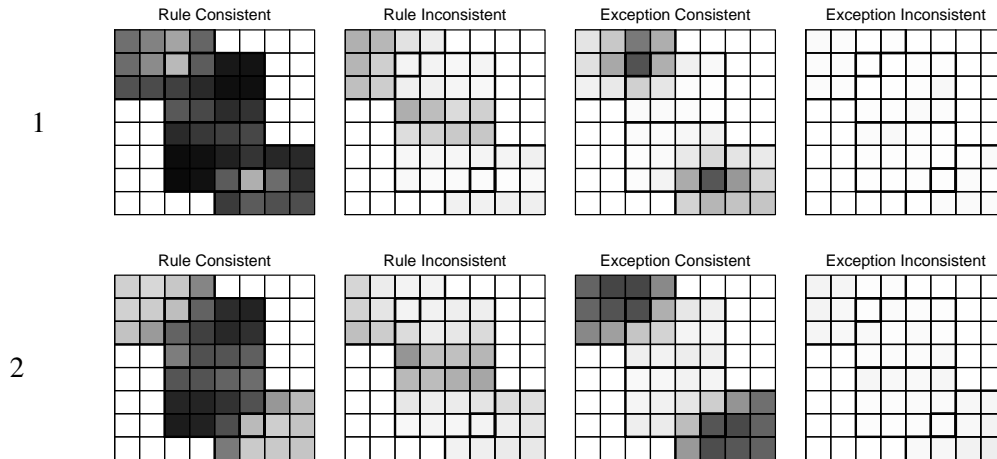


Figure 4.9: The proportion of responses for each stimulus during the transfer phase for the average participant per cluster of Experiment 4. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of response. The cluster number is displayed to the left of each row.

Cluster	1D	2D
1	45	10
2	20	54

Table 4.4: Division of Participants into Clusters in Experiment 4

relatively large and have positive average silhouette widths. The dendrogram in Figure 4.8 suggests that a three or four cluster solution might be more appropriate, however upon adding more clusters, the second cluster was broken up into multiple groups that were not meaningfully different.

Cluster 1 contains participants who classified transfer items as members of the rule-consistent category. This suggests that these participants used both rule- and exemplar-based categorization. Cluster 1 contains the majority of the participants from the one-dimensional conditions, along with a small number of participants from the two-dimensional conditions. This cluster provides evidence for one- and two-dimensional rule use with stimuli possessing integral and noncommensurate dimensions, although two-dimensional rule use is a minority strategy.

Cluster 2 contains participants who classified transfer items as members of the exception-consistent category. This cluster contains the majority of participants from the two-dimensional conditions, along with a small number of participants from the one-dimensional conditions. This suggests that participants in the two-dimensional conditions were most likely to make exemplar-based generalizations, and participants in the one-dimensional conditions were most likely to make rule-based generalizations.

In conclusion, the cluster analysis for Experiment 4 provides evidence for both rule- and exception-based strategies in a task using stimuli with integral and commensurate dimensions. However, the clustering solution found sparse support for the use of two-dimensional rules with most participants in the two-dimensional conditions using an exemplar-based generalization strategy.

4.6 Overall Experiment Cluster Analysis

The final series of cluster analyses were performed on the combined data from all experiments. By combining across all experiments larger, more consistent clusters may emerge, clusters with few members may gain enough sufficient members to separate themselves from other clusters, and it allows for more straightforward comparisons between experiments. Figure 4.8 contains the silhouette plot and the dendrogram, Figure 4.9 shows the averaged participant performance, and Table 4.4 shows the breakdown of participants into clusters based on condition. The overall experiment cluster analysis contains six clusters, two clusters are large and consistent (Clusters 1 and 2), one cluster is large and somewhat consistent (Cluster 5), one cluster is small and consistent (Cluster 3), and two clusters are small and inconsistent (Clusters 4 and 6).

Cluster 1 contains participants who classified the transfer items as members of the rule-consistent category. This suggests that these participants used both rule- and exemplar-based cat-

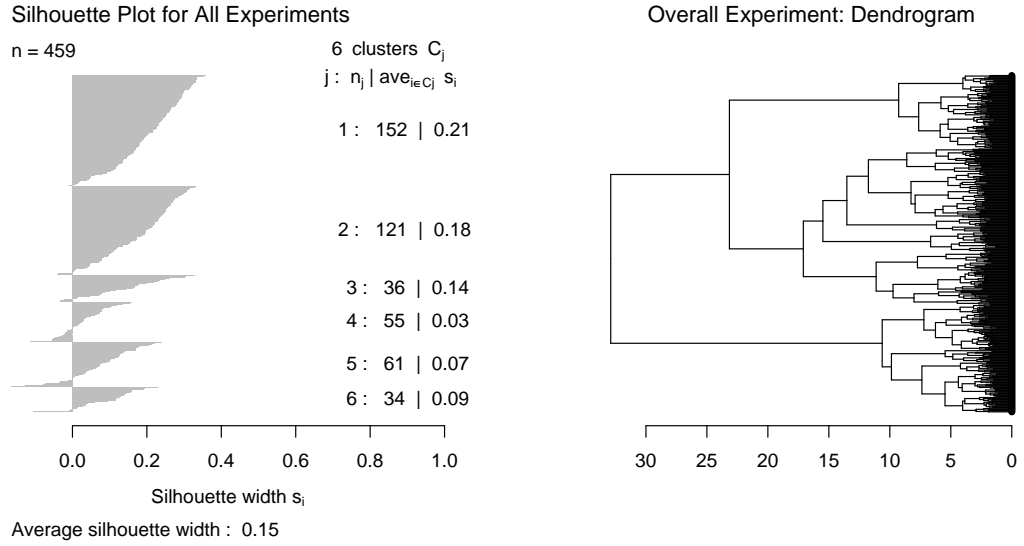


Figure 4.10: The left panel contains the silhouette plot for all participants combined across all experiments. The right panel contains the dendrogram for all participants combined across all experiments.

Cluster	Experiment									
	1		2		3		4		Total	
	1-D	2-D	1-D	2-D	1-D	2-D	1-D	2-D	1-D	2-D
1	14	10	22	10	16	43	32	5	84	68
2	5	15	12	27	16	11	6	29	39	82
3	16	0	9	0	0	1	10	0	35	1
4	4	5	5	17	6	5	5	8	20	35
5	0	5	0	10	20	3	9	14	29	32
6	1	5	10	4	1	2	3	8	15	19

Table 4.5: The division of participants into clusters based on condition across all experiments

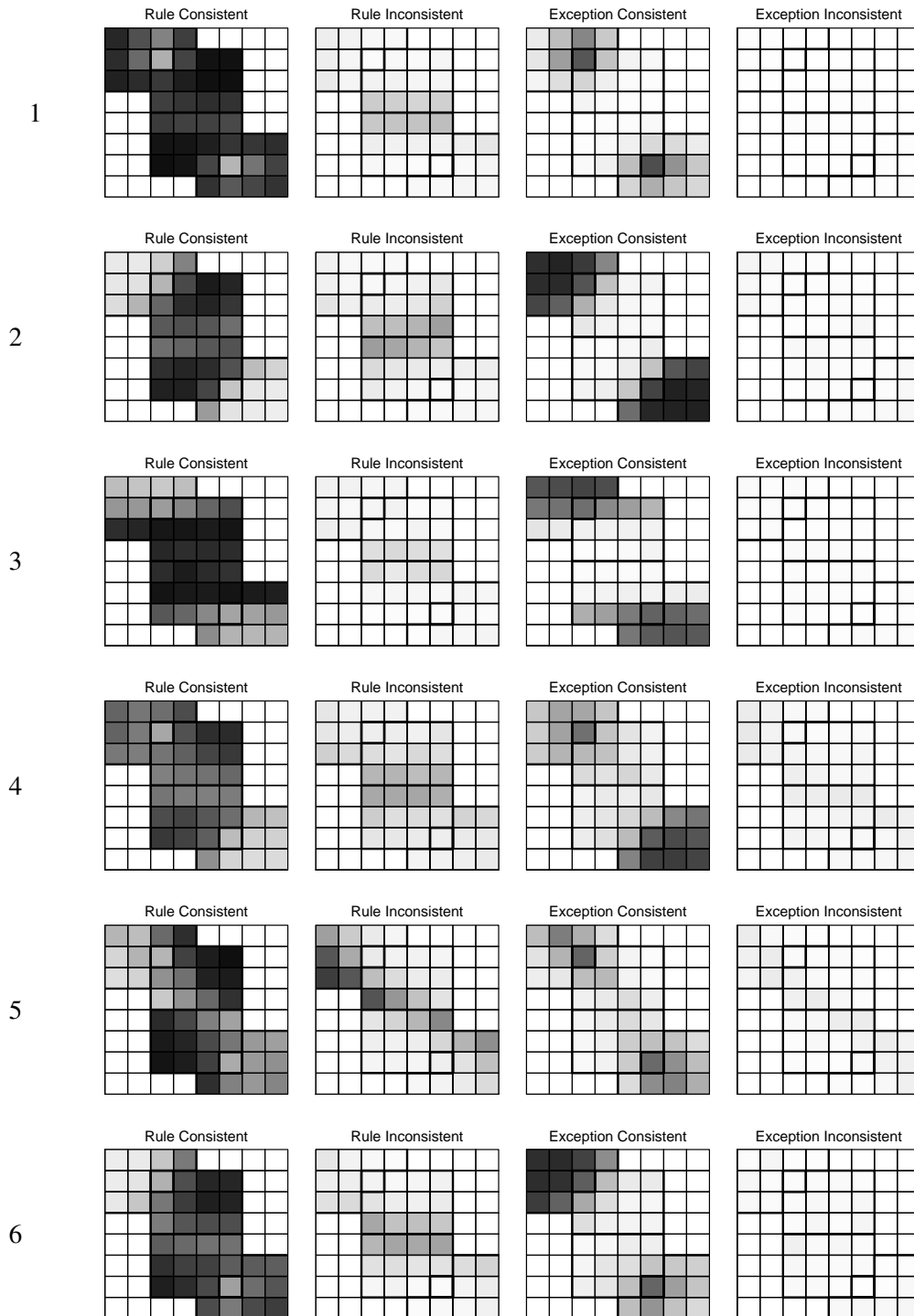


Figure 4.11: The proportion of responses for each stimulus during the transfer phase for the average participant per cluster for the combined experiment clustering solution. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of response. The cluster number is displayed to the left of each row.

egorization. This cluster contains the most participants and the highest positive average silhouette width. This is the cluster that provides the most support for the use of two-dimensional rules.

Cluster 2 contains participants who classified the transfer items as members of the exception-consistent category. This suggests that these participants used an exemplar-based strategy. This cluster contains over twice as many participants from the two-dimensional conditions as the one-dimensional conditions. This suggests that participants in the two-dimensional conditions were more likely to use exception generalization than rule-generalization.

Cluster 3 contains participants who classified the transfer items as members of the exception-consistent category. This suggests that these participants used an exemplar-based strategy. However, this pattern is also consistent with participants who classified stimuli according to the theoretical rule of *equal to or more extreme* as previously found in Experiments 1 and 2. Cluster 3 consists of a majority of participants from the one-dimensional conditions, there was only one participant from the two-dimensional conditions. This cluster contains low numbers of participants from Experiments 1, 2, and 4, and only one participant from Experiment 3.

Clusters 4, 5 and 6 are related clusters containing participants who classified transfer items as members of the rule-consistent, rule inconsistent, and exception-consistent categories. These clusters were a combination of four different strategies. Clusters 4 and 5 were composed of participants using a one-dimensional bounds on a two-dimensional category structure or a two-dimensional bound on a one-dimensional category structure. Cluster 4 contains participants that utilized a bound counter-clockwise to the optimal bound, while Cluster 5 contains participants that utilized a bound clockwise of the optimal bound.

Clusters 4 and 6 contain participants who classified transfer items differently, depending on where the transfer items were located. Cluster 4 contains participants who classified transfer items in the upper left as members of the rule-consistent category, whereas they classified transfer items in the lower right as members of the exception-consistent category. Cluster 6 contains participants who reversed this pattern. All three of these clusters contain low numbers of participants from all four experiments. None of these clusters provide evidence for two-dimensional rule use.

4.7 Summary

The cluster analysis revealed that all experiments contained a cluster of rule-consistent generalizers that support the use of one-dimensional and two-dimensional rules. Support for one-

dimensional rule use was more prevalent than support for two-dimensional rule use. However, evidence for two-dimensional rule use was still present in all experiments.

Experiments 1 (separable and commensurate dimensions) and 3 (integral and commensurate dimensions) contained many participants from the two-dimensional conditions that were rule generalizers. This supports the use of rules in commensurate conditions. Experiments 2 (separable and noncommensurate dimensions) and 4 (integral and noncommensurate dimensions) contained fewer participants from the two-dimensional conditions that were rule generalizers. This provides support for the use of two-dimensional rules but at a lower rate than in the other experiments. This suggests that two-dimensional rules are more difficult to use on stimuli with noncommensurate dimensions.

Experiment 1 had roughly equal amounts of participants from the one-dimensional and two-dimensional conditions in the rule generalizing cluster. This suggests that a two-dimensional rule strategy was frequently adopted by participants. Experiment 3 had roughly twice as many participants from the two-dimensional than one-dimensional conditions in the rule-generalizing cluster. This suggests that most participants adopted a two-dimensional rule.

Experiment 2 had half as many participants from the two-dimensional conditions as the one-dimensional conditions. This suggests that adopting a two-dimensional rule was possible, but generally more difficult. Experiment 4 had very few participants from the two-dimensional conditions in the rule-generalizing cluster. This is suggestive of a very low adoption rate of a two-dimensional rule.

Chapter 5

Mathematical Modeling

To determine how participants most likely approached the classification problems in the experiments, participant behaviors were fit by a series of mathematical models. The first two groups formed by the cluster analysis were selected for modeling. These clusters were the most likely to provide evidence for two-dimensional rule use, contained the majority of the participants, and were the most consistent of the clusters. The first cluster contained participants, who in the transfer phase, made a high proportion of rule-consistent responses to the transfer items. These participants generalized in a manner consistent with the rule items. Thus this cluster is referred to as the rule-generalization cluster. In contrast, the second cluster contained participants, who in the transfer phase, made a high proportion of exception-consistent responses to the transfer items. These participants generalized in a manner consistent with the exception items. Thus this cluster is referred to as the exception-generalization cluster.

Three models were fit to the data, ALCOVE (Kruschke, 1992), ATRIUM (Erickson & Kruschke, 1998), and a modified version of ATRIUM that utilizes two-dimensional rules, ATRIUM-DR. The fitness of the models were evaluated using the Bayesian information criterion statistic (BIC; Schwarz, 1978). The BIC is a fitness statistic that penalizes models for the number of free parameters it has and allows for comparisons to be made between different unnested models. The penalty applied by the BIC for free parameters is a function of the number of free parameters and sample size. So, as the number of trials increases, so does the penalty for for each free parameter. The BIC is then composed of a measure of how well the model matches the data (log-likelihood) and the penalty for free parameters and the amount of data. Models that accurately match the data have low BIC scores, whereas models that are inaccurate or contain more free parameters have higher BIC scores. Therefore, in the following fits, the model with the smallest BIC best describes

the data.

Within each cluster, each participant's data was independently fit by each model. Each model was fit to the exact series of trials presented to each participant. Model fitness was calculated based on performance in the final transfer phase. The resulting fitness values are the sum of all individual fitness values for a particular set of parameters. Thus, the resulting best fitting parameters were the parameters that most accurately matched the behaviors of the entire cluster of participants, rather than any one particular participant. In addition to the BIC, R^2 and Root Mean Square Deviation (RMSD) were calculated for each fit. These additional fit statistics are included in the figures with BIC. They can be used to evaluate how well the models fit each set of data. They will not be discussed further.

5.1 ALCOVE

ALCOVE is a single system exemplar-based connectionist model of category learning (Kruschke, 1992). The version of ALCOVE used in this paper has five free parameters that control model performance. Table 5.1 summarizes ALCOVE's parameters. The following is a short description of ALCOVE, for more details see Kruschke (1992).

In ALCOVE, exemplars are represented by nodes that are positioned within a category space according to the features of the exemplar. Each stimulus feature is represented along a continuous dimension. The contribution of each stimulus feature to a category response is weighted by the amount of attention allocated to each dimension and the relative salience of the dimensions. The amount of attention allocated to a dimension affects how the participants perceive the similarity of features along that dimension. The more attention allocated to a dimension, the more distinct the different features on that dimension. Likewise, the less attention allocated to a dimension, the more similar the different features on that dimension. As the model learns the category structure, it learns to pay more attention to relevant dimensions and less attention to irrelevant dimensions.

Similar to attention, the *relative salience* parameter, s , also represents a differential weighting of the relevance of a dimension. However, unlike attention, which is learned by the model, the relative salience of dimensions does not change during learning. Instead, the relative salience represents a comparison of the salience of one dimension to another dimension, or the perceptual prominence of one dimension over another. A dimension that easily captures attention, such as the brightness of a flashing red light is more salient than a dimension that is more subtle, such as the speed of a ceiling fan. The relative salience is the ratio of the salience of one dimension to another

dimension. When s is close to 1, the dimensions have similar saliences. As s differs from 1, the relative salience of the dimensions increases, with one dimension being more salient than another. So, in the previous example of a flashing red light and a ceiling fan, the relative salience of this pair of dimensions might be 5, or $1/5$ if the ceiling fan is the first dimension and the red light is the second.

In ALCOVE, the activation of an exemplar node is a function of the similarity of a presented stimulus and the exemplar n . The distance is weighted by salience and attention settings, and further modified by the *specificity* parameter, c . The *specificity* parameter magnifies the distance between the exemplar node and presented stimulus. Larger values of c result in increased distances between exemplar nodes and less generalization, whereas smaller values of c result in decreased distances between exemplar nodes and more generalization.

Each exemplar node is connected to all output nodes. The output nodes correspond to the model's possible category choices. The amount of activation that each output node receives from each exemplar node is a function of the activation of the exemplar node and learned connection weights between the exemplar node and the output node. ALCOVE produces response probabilities based upon the activation of the output nodes. The *choice probability scaling constant* parameter, ϕ , determines the extent to which the model's output is deterministic or probabilistic. Large values of ϕ result in more deterministic responses, while small values of ϕ result in more probabilistic responses.

The last two free parameters in ALCOVE are learning rate parameters (*lambda*); the attentional learning rate parameter λ_a and the exemplar learning rate parameter λ_w . These parameters control how quickly the model changes its distribution of attention across the input dimensions and the connection weights between the exemplar and output nodes. Larger values of the λ parameters result in faster changes, while smaller values result in slower changes. The model changes its attentional distribution and connection weights between exemplar nodes and output nodes as a function of the error produced by the model's output. The larger error, or the difference between the correct response and the model's output, the greater the change in the model.

Additionally, larger values of λ may cause the model to over-react by over-learning. The model is forced to compensate for this by making another (large) change. Smaller values of λ_r are less likely to cause over-reacting, and are better able to follow learning gradients. An analogy of this is trying to avoid a dog while driving. If a person turns the steering wheel by a large amount to avoid a dog, they must then quickly turn the steering wheel the other way to avoid going off the road. This second turn also runs the risk of causing another over-reaction, and the process continues

Table 5.1: Summary of ALCOVE Parameters

Parameter	Description
c	Specificity
λ_a	Attention learning rate
λ_w	Exemplar learning weight
ϕ	Choice probability scaling constant
s	Relative salience

until the car is stable again (or crashed into the side of a hill). Whereas if a smaller adjustment is made to avoid the dog, there is less difficulty and less risk in reversing this change after the dog is avoided.

5.2 ATRIUM

ATRIUM is a hybrid rule- and exemplar-based connectionist model of category learning (Erickson & Kruschke, 1998, 2002). The version of ATRIUM used in this paper has 11 free parameters that influence module performance. Table 5.2 summarizes ATRIUM’s parameters. The following is a short description of ATRIUM, for a more details, see Erickson and Kruschke (1998).

ATRIUM is a hybrid model that combines an exemplar module, a version of ALCOVE, with a number of rule modules. The modules process each stimulus simultaneously and the output from all modules is combined to produce a category response. The number of rule modules is dependent upon the category structure being fit, with one rule module per rule dimension. In these experiments, two rule modules, one for each relevant dimension, were used. For example, in the one-dimensional conditions of Experiment 1, one module classified stimuli based upon a *height-rule*, the other classified stimuli based upon a *line-segment-position-rule*.

The contribution of each module to a category response is dependent upon a gating module. The gating module uses an exemplar representation to learn which modules are best suited to classify each stimulus and to weight the contributions of each module appropriately. The gating module also controls the amount of feedback each module receives, with the most appropriate module receiving the most feedback. The amount of feedback a module receives also determines the rate of learning for the module, with larger amounts of feedback allowing for larger changes in a module.

In addition to possessing all of the parameters used by ALCOVE, ATRIUM has its own unique parameters. These parameters control the gating and rule modules. In the presented simulations, ATRIUM's rule modules have four free parameters and ATRIUM's gating modules have two free parameters.

In ATRIUM, the activation of a rule module is a sigmoidal function of the stimulus being presented. The steepness, or *gain*, of the sigmoid is represented by the *rule gain* parameter, γ . High values of γ result in sharply defined rule boundaries between categories, whereas low values of γ result in gradually defined rule boundaries.

Each rule module also has a *rule bias* parameter, β . In the presented simulations there are two rule modules, so there are two rule bias parameters, β_1 and β_2 . The rule bias parameters influence the contribution of the rule modules to the categorization response. They control the initial propensity to use each of the modules. The higher the module bias parameter the larger the contribution of that module to a categorization response. The exemplar module has a similar bias parameter. However, the sum of the rule and the exemplar biases is constrained to equal one, therefore, the exemplar module bias is determined by the rule modules biases and does not vary freely.

The last rule parameter is the *rule learning weight* parameter, λ_r . Similar to the other learning parameters, this parameter controls how quickly the model adjusts which rule it uses for a particular stimulus. Large values of λ_r cause rapid changes and may cause over-reacting, while small values cause slow changes and are less likely to cause over-reacting.

The gating module uses two free parameters, the *gating module probability scaling constant* and the *gate learning weight*, to control the gate modules behavior. The gating module probability scaling constant, ϕ_g , determines how the model combines the response outputs of the modules. Larger values of ϕ_g cause a *winner take all* type of performance, where the module with the greatest response output contributes the most to the category decision. Smaller values of ϕ_g cause more equal contributions from the modules. The gate learning weight parameter, λ_g , controls how quickly the model adjusts the contributions of the rule and exemplar modules. Similar to the other learning parameters, large values of λ_g cause rapid changes and may cause over-reacting, while small values cause slow changes and are less likely to cause over-reacting.

Table 5.2: Summary of ATRIUM Parameters

Parameter	Description
c	Specificity
ϕ	Choice probability scaling constant
λ_a	Attention learning rate
λ_w	Exemplar learning weight
γ	Rule gain
ϕ_g	Gating module probability scaling constant
β_1	Rule 1 bias
β_2	Rule 2 bias
λ_r	Rule learning weight
λ_g	Gate learning weight
s	Relative salience

5.3 ATRIUM-DR

ATRIUM-DR is a modified version of ATRIUM and has the same free parameters. There are two main differences between the models. The first difference is that in ATRIUM-DR the one-dimensional rule modules are replaced by two-dimensional rule modules. For example, in the two-dimensional conditions of Experiment 1, these two-dimensional rule modules corresponded to linear combinations of these two dimensions, either height divided by width, a *positive diagonal-rule*, or the difference between height and the maximum stimulus height divided by width, a *negative diagonal-rule*.

The second difference, is how dimensional attention relates to the rule modules. In ATRIUM and ALCOVE, dimensional attention allows the models to shrink or expand dimensions allowing more accurate categorization (Erickson & Kruschke, 1998; Nosofsky, 1986). By expanding a dimension, the similarity between adjacent values along that dimension decreases, resulting in greater influence of that dimension on the categorization response (i.e., features become more distinct and important). By shrinking the dimension, the similarity between adjacent values along that dimension increases, resulting in less influence of that dimension on the categorization response (i.e., features become less distinct and less important). This expansion and shrinking occurs on a per dimension basis and can mimic the effects of a one-dimensional rule. In ATRIUM-DR, the rule modules do not use one-dimensional rules, so dimensional attention can no longer directly mimic

the effects of the rules.

5.4 Modeling Details

When fitting the clusters, participant data with different category structures were combined together. This required a few manipulations to the free parameters to make them consistently meaningful for each of the clusters. Additionally, there are a few other details about the modeling process will be discussed.

5.4.1 Relative Saliency

No scaling studies were performed with the stimuli used in the experiments, so relative saliency was used as a free parameter. Additionally, the reported relative saliency for all three models always refers to the same primary dimension. For Experiment 1, the primary dimension was line segment position, for Experiment 2, the primary dimension was angle of the radial line segment, for Experiment 3, the primary dimension was rectangle width, and for Experiment 4, the primary dimension was amplitude. As a parameter, relative saliency should be interpreted as the relative contribution to the categorization response of the primary dimension in proportion to one unit of the other dimension. For example, in Experiment 1, if s was 2.0, this would mean that the position of the line segment was weighted twice as much as the height of the rectangle when making a classification.

5.4.2 Rule Bias Parameters (β_1 & β_2)

In ATRIUM and ATRIUM-DR the rule bias parameters were arranged so that β_1 always referred to the rule consistent with the rule used to create the category structure, and β_2 always referred to a rule orthogonal to the rule used to create the category structure.

5.4.3 Search Methods

A combination of different search methods were used to find the optimal parameters for each of the models. These methods included using a genetic algorithm, a hill-climbing algorithm, and using an artificial averaged per subcluster subject in a grid search. The best fit from each of these initial methods was then further refined through a hill-climbing algorithm. The besting fitting

Table 5.3: ALCOVE Fits for Experiment 1

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	39.92	38.85	17.17	5.20
ϕ	3.37	2.90	0.88	1.51
λ_a	0.003	0.003	1.474	1.315
λ_w	0.004	0.001	0.021	0.011
s	2.16	1.19	0.95	1.22
BIC	7155	5842	3328	9545
R^2	.92	.88	.86	.87
RMSD	.09	.11	.12	.12

solution was then selected from all of these fits. Additionally, similarity and distance in the models were computed using a Minkowski r -metric with $r = 1$ (city block or Manhattan distance).

5.5 Experiment 1: Separable and Commensurate Dimensions

The stimuli for Experiment 1 possessed separable and commensurate dimensions. They were rectangles that varied in the position of an internal line segment and height. The best fitting parameters for Experiment 1 for ALCOVE are shown in Table 5.3, for ATRIUM in Table 5.4, and for ATRIUM-DR in Table 5.5. The average participant performance and the predictions of the models can be seen in Figures 5.1 through 5.4. Additionally, when presenting the fit data for ATRIUM and ATRIUM-DR in cases where they were unable to outperform ALCOVE (e.g., Table 5.4) the ATRIUM specific free parameters are not reported, as they did not affect the behavior of the model. Nevertheless, the BICs for ATRIUM and ATRIUM-DR were computed using 11 free parameters.

Experiment 1: Rule-Generalizers

The rule-generalizer cluster contains participants who categorized the transfer items in a rule-consistent manner. In the one-dimensional conditions, the horizontal and vertical conditions, both ALCOVE and ATRIUM were able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC: 6324) was better able to account for this cluster of participants

Table 5.4: ATRIUM Fits for Experiment 1

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	33.763	1.499	10.482	5.2
ϕ	2.171	0.568	1.151	1.51
λ_a	0.006	0.405	0.025	1.315
λ_w	0.013	0.044	2.249	0.011
γ	2.798	3.211	21.647	–
ϕ_g	3.12	4.526	0.17	–
β_1	-1.684	2.386	3.841	–
β_2	1.266	-3.49	-0.615	–
λ_r	0.372	0.446	0.101	–
λ_g	0.025	0.215	0.756	–
s	1.56	3.021	2.289	1.22
BIC	6324	5872	3226	9587
R^2	.98	.88	.89	.87
RMSD	.05	.11	.10	.12

Note. In cases where ATRIUM’s fit was worse than ALCOVE’s the ATRIUM-specific free parameters were not reported, as they did not affect the behavior of the model. Nevertheless, the BIC for ATRIUM was still computed using 11 free parameters.

Table 5.5: ATRIUM-DR Fits for Experiment 1

Parameter	Rule-Generalizers	Exception-Generalizers
c	24.718	5.27
ϕ	0.644	1.937
λ_a	0.04	0.375
λ_w	0.043	0.004
γ	3.768	0.687
ϕ_g	0.77	1.092
β_1	2.206	2.166
β_2	-1.214	-0.01
λ_r	0.234	0.191
λ_g	0.874	0.254
s	1.186	0.871
BIC	5515	8938
R^2	.94	.99
RMSD	.08	.04

than ALCOVE (BIC: 7155). The key items that ALCOVE failed to account for, but ATRIUM was able to account for, were the extreme transfer items, as shown in Figure 5.1. The extreme transfer items are the transfer items furthest from the training items.

To model this cluster, ALCOVE used a high specificity to correctly match participant performance on the transfer items. This resulted in relatively few consistent-exception responses to the exemplar training items, unlike the participants. ALCOVE, with only exemplar representation, is unable to classify both the exception training items with high levels of exception-consistent responses and classify the extreme transfer items with high levels of rule-consistent responses.

In order to produce a high level of exception-consistent responses to exception training items, ALCOVE requires either strong connection weights (due to learning) between the exception exemplars and the exception-consistent output nodes or large exemplar specificity. With strong connection weights between the exception training exemplar and the exception-consistent output nodes, the presentation an exception item produces a relatively high exception response activation compared to the amount of rule-consistent response activation. This results in high exception-consistent response rates. Likewise, with large exemplar specificity, the influence of the rule-consistent exem-

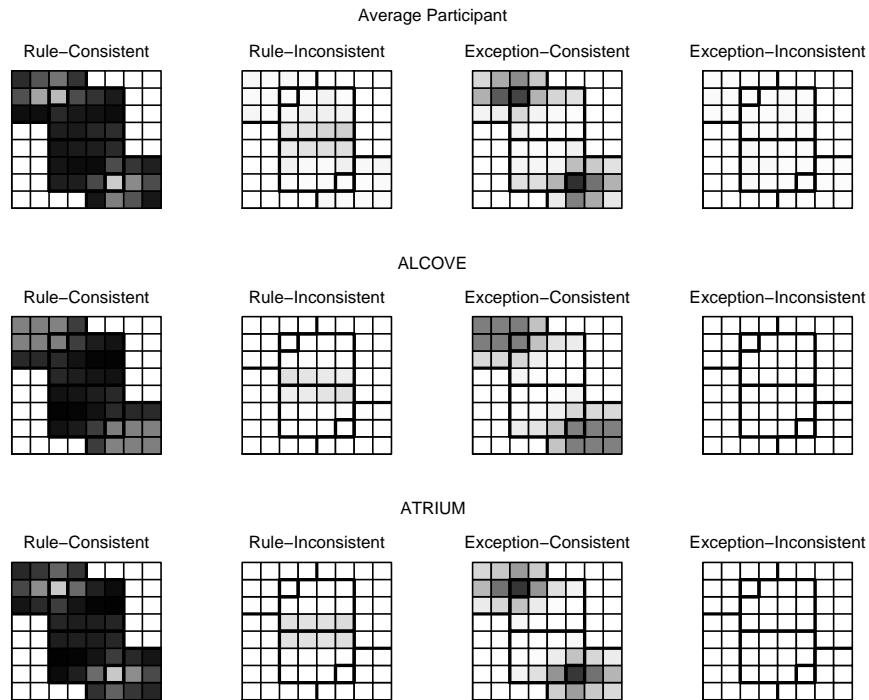


Figure 5.1: Proportion of responses for each stimulus during the transfer phase of Experiment 1 for rule-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

plars is relatively weak compared to the activation of exception-consistent exemplars, resulting in high exception-consistent response rates.

However, both ways of achieving high exception-consistent responses to the exception training item, high exemplar specificity or strong connections weights, result in classifying transfer items as exception-consistent items. This is unlike the rule generalization performance of the participants in this cluster. ALCOVE can either have high consistent-exception responses to exception training items or high consistent-rule responses to the transfer items, but not both.

One potential adaptation to ALCOVE that might allow ALCOVE to fit data like these is to allow specificity to vary between exemplars (Erickson & Kruschke, 1998). This would allow some exemplars to represent very specific stimuli, in this case, the exception training items, while other exemplar nodes would be allowed to represent broader ranges of stimuli, in this case the rule training items. Erickson and Kruschke (1998) tested a version of ATRIUM that included exemplar

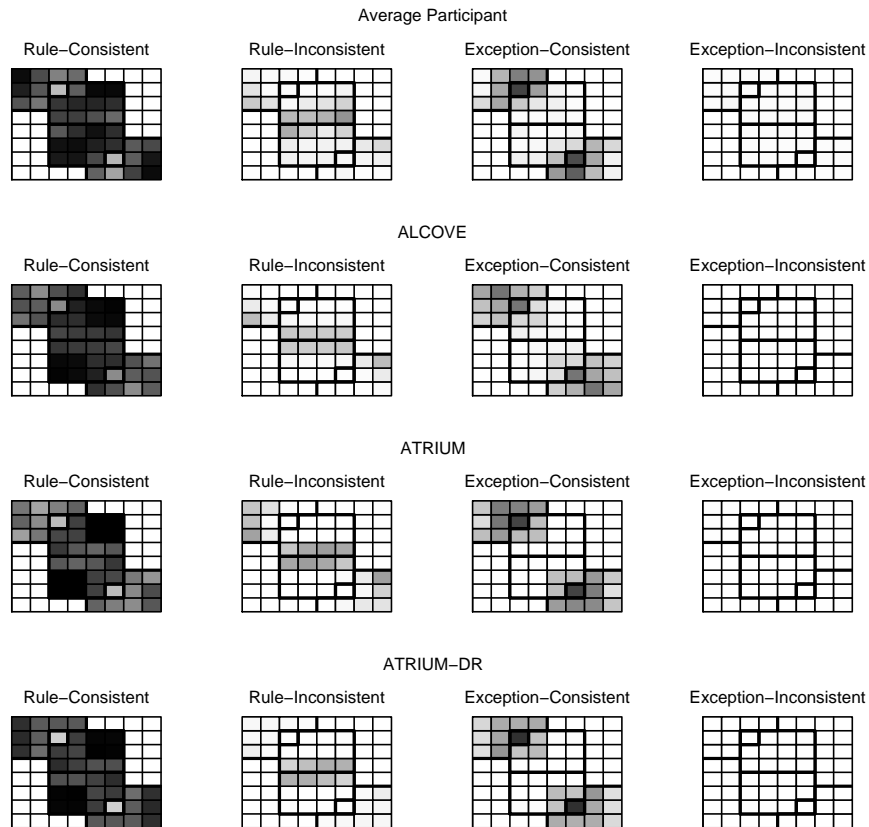


Figure 5.2: Proportion of responses for each stimulus during the transfer phase of Experiment 1 for Rule-Generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

specific attention. They found that the model was unable to learn adjust exemplar specificity to allow for both high levels of exemplar-consistent responses to exception training items and high levels of rule-consistent responses to transfer items. They concluded that such an enhancement was still insufficient to produce this type of generalization.

In contrast to ALCOVE, ATRIUM obtained a superior fit by using rules and exemplars. For ATRIUM the training items were primarily classified by the exemplar module, whereas the transfer items were primarily classified by the rule module. This allowed the model to classify the exception in an exception-consistent manner at high levels and to classify the transfer items in a rule-consistent manner. This suggests that participants are able to use one-dimensional rules to

learn category structures formed with a one-dimension rule boundary and composed of stimuli with separable and commensurate dimensions.

In the two-dimensional conditions, the positive and negative diagonal conditions, ALCOVE, ATRIUM, and ATRIUM-DR were all able to perform in a manner qualitatively similar to the participants. However, ATRIUM-DR (BIC = 5515) was better able to account for this cluster of participants than ALCOVE (BIC = 5842) or ATRIUM (BIC = 5872). Like in the one-dimensional conditions, the key items were the extreme transfer items as shown in Figure 5.2.

ALCOVE was again unable to account for the transfer performance of participants for the same reasons as in one-dimensional conditions. ATRIUM was likewise unable to match the transfer item performance, as its rules were not aligned with the diagonal rule used to create the category structure. However, ATRIUM-DR was able to match transfer item performance with its diagonal rules. This suggests that two-dimensional rule usage is possible for stimuli composed of separable and commensurate dimensions.

Experiment 1: Exception-Generalizers

The exception-Generalizer cluster contains participants who categorized the transfer items in an exception-consistent manner. In the one-dimensional conditions, the horizontal and vertical conditions, both ALCOVE and ATRIUM were able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 3226) was better able to account for this cluster of participants than ALCOVE (BIC = 3327). The two key sets of items that ALCOVE failed to account for, but ATRIUM was able to account for, were the extreme transfer items and the rule training items on opposite side of the category structure relative to the exception items as shown in Figure 5.3.

The first set of key items for comparing models and participants in one-dimensional conditions of exception-generalizers were the extreme corner transfer items. Participants made the most consistent exception responses to the extreme corner transfer items. The rate of consistent exception responses increased as a function of the distance from the training items. Participants' exception-consistent responses were lower for the exception training items than they were for the extreme transfer items. ALCOVE was unable to reproduce this pattern. ALCOVE's best attempt was to classify the transfer items with similar responses rates as the exception training item. ALCOVE was unable to duplicate the increase in exception-consistent responses as a function of the distance from the training items. While ATRIUM was unable to produce a response gradient in the transfer items as extreme as the participants, it was able to produce a gradient that was qualitatively similar.

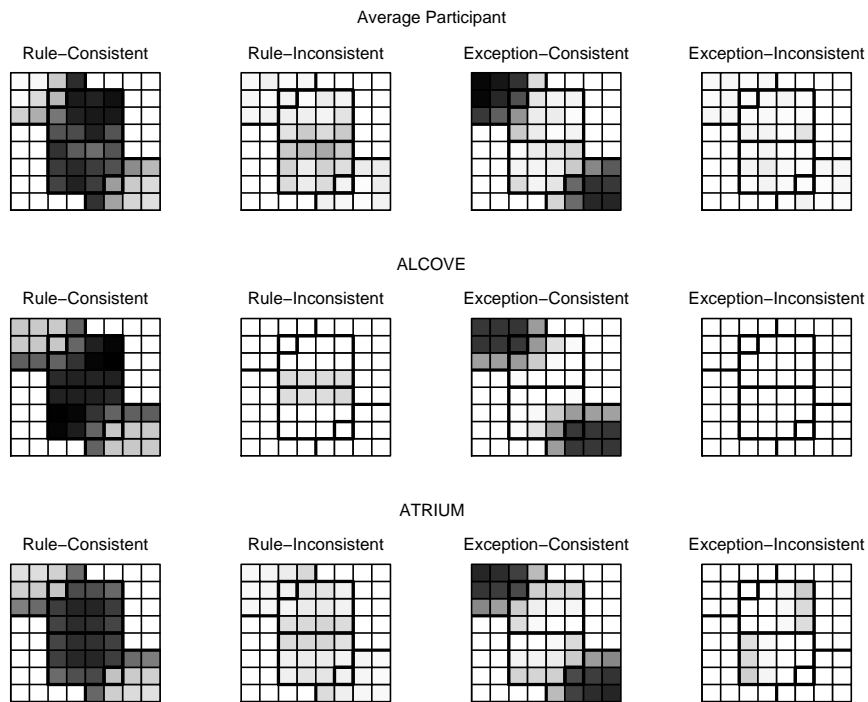


Figure 5.3: Proportion of responses for each stimulus during the transfer phase of Experiment 1 for exception-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

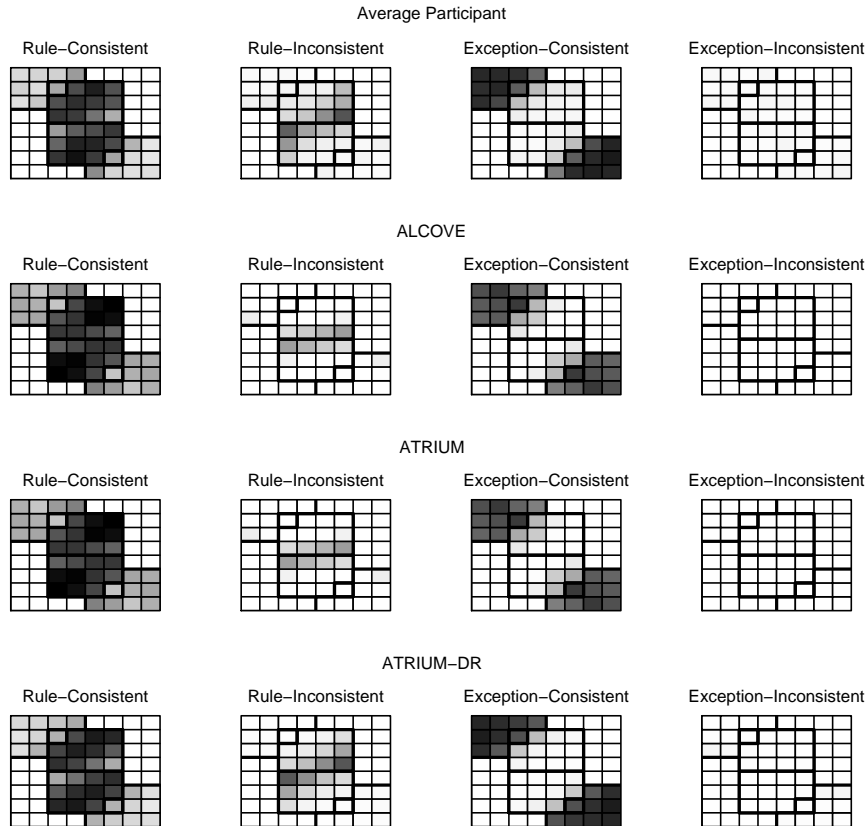


Figure 5.4: Proportion of responses for each stimulus during the transfer phase of Experiment 1 for exception-generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

The second set of key items for comparing models and participants in one-dimensional conditions of exception-generalizers were the rule training items. Participants in the one-dimensional conditions of exception-generalizers were less consistent in their responses than ALCOVE predicted. This is most noticeable in the rule training items on the opposite side of the category structure relative to the exception training items. Participants responded with high rates of rule-consistent responses, but also with noticeable rates of rule-inconsistent responses. ALCOVE was unable to match these types of responses. The high specificity of the best fitting solution forced these responses to be highly influenced only by their nearest neighbors. This resulted in ALCOVE having a high rule-consistent response rate for these items. This pattern was unlike participants, who were less consistent. In contrast, ATRIUM was able to produce less consistent behavior to these items, matching the behavior of the participants.

ATRIUM provided a superior fit to the data of the participants who generalized according to the exception items in one-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn categories structures formed with a one-dimension rule boundary and composed of stimuli with separable and commensurate dimensions.

In the two-dimensional conditions, the positive and negative diagonal conditions, ALCOVE, ATRIUM, and ATRIUM-DR were all able to perform in a manner qualitatively similar to the participants. However, ATRIUM-DR (BIC = 8938) was better able to account for this cluster of participants than ALCOVE (BIC = 9545) or ATRIUM (BIC = 9587). The key items for the models to fit were the exception items and the transfer items, as shown in Figure 5.4.

Participants in the two-dimensional conditions of exception-generalizers classified the exception items with high rates of exception-consistent responses, but also with a moderate rate of rule-consistent responses. Also, like in the one-dimensional conditions, the frequency of exception-consistent responses for the transfer items increased as the distance between the transfer items and the training items increased. ALCOVE was unable to produce this pattern of results. ALCOVE learned to classify the exception training items with a consistent exception response rate that was too high. Likewise, ALCOVE's responses to the transfer items did not match the gradient observed in the participants' data. ATRIUM was unable to find a way to improve upon ALCOVE's performance. Its best solution was to perform as ALCOVE.

ATRIUM-DR was more successful than the other models because its rule modules were aligned to the rule used to create the category structure. This allowed ATRIUM-DR to use its rule module in the transfer phase, resulting in a better fit to the data.

Another prominent feature of the participants' behavior was the boundary used to separate

the rule categories. Participants did not use an optimal boundary between the stimuli, but instead used one that was biased more toward one physical dimension of the stimuli. This is displayed in Figure 5.4 by the pattern of rule inconsistent responses. If the boundary was rotated clockwise the training items with high rule-inconsistent responses on edges of the category structure would then become rule-consistent responses. This consistent pattern of behavior is due to the large number of participants in the cluster who were in the same condition and biased towards the same dimension, the line segment position dimension. This pattern follows previous findings by Erickson and Kruschke (1998), who also found preference for line segment position over height in a similar stimulus set.

In fitting ATRIUM-DR, the model was forced to use rules that reflected the true boundary between the categories and was not allowed to vary them to fit the boundaries potentially used by a participant. However, if ATRIUM-DR was adapted to allow the model to change the slope and intercept of the boundaries in its rules, it would be better able to match participant performance. Such an adaptation would, however, go beyond the initial assumption that people can use two-dimensional rules. It would also require the assumption that people can construct two-dimensional rules in a variety of different combinations of the underlying perceptual dimensions. While this may be what people actually do, confirmation of two-dimensional rule usage is necessary before more extensive assumptions should be made. Regardless, ATRIUM-DR was better able to account for the cluster of participants than either ALCOVE or ATRIUM, using a combination of rule and exemplar modules.

ATRIUM-DR provided a superior fit to the data of the participants who generalized according to the exception items in two-dimensional conditions. This suggests that participants are able to use two-dimensional rules to learn category structures formed with a two-dimension rule boundary and composed of stimuli with separable and commensurate dimensions.

Experiment 1: Summary

The transfer behaviors of participants who were given category structures formed using one-dimensional boundaries composed of stimuli with separable and commensurate dimensions were better fit by ATRIUM than ALCOVE. ATRIUM produced superior fits for both clusters of participants, both rule-generalizing and exception-generalizing participants. This is consistent with previous literature supporting rule usage in category learning (Erickson & Kruschke, 1998, 2002).

The transfer behaviors of participants who were given category structures formed using two-dimensional boundaries composed of stimuli with separable and commensurate dimensions

Table 5.6: ALCOVE Fits for Experiment 2

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	36.48	29.50	58.53	7.15
ϕ	2.48	2.55	0.52	1.97
λ_a	0.005	0.001	0.001	0.119
λ_w	0.001	0.001	0.015	0.013
s	1.03	1.32	1.37	2.42
BIC	14208	6539	9292	15924
R^2	.92	.89	.87	.90
RMSD	.09	.10	.10	.10

were best fit by ATRIUM-DR. ATRIUM-DR produced superior fits for both clusters of participants. This suggests that people may use two-dimensional rules. Furthermore, the system constructing these rules may not form optimal rules. This may indicate a system with a more flexible rule construction method or one that is influenced by the perceptual characteristics of the stimuli.

5.6 Experiment 2: Separable and Noncommensurate Dimensions

The stimuli for Experiment 2 possessed separable and noncommensurate dimensions. They were circles that varied in the angle of radial line segment and height. The best fitting parameters for Experiment 2 for ALCOVE can be seen in Table 5.6, for ATRIUM in Table 5.7, and for ATRIUM-DR in Table 5.8. The average participant performance and the predictions of the models can be seen in Figures 5.5 through 5.8.

Experiment 2: Rule-Generalizers

The rule-generalizer cluster contains participants who categorized the transfer items in a rule-consistent manner. In the one-dimensional conditions, ALCOVE (BIC = 14208) was better able to account for this cluster of participants than ATRIUM (BIC = 14252), as seen in Figure 5.5.

Examining the gating module of ATRIUM revealed that ATRIUM relied upon its exemplar module to learn the category structure, and did not use its rule modules. However, ALCOVE

Table 5.7: ATRIUM Fits for Experiment 2

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	36.48	29.5	58.53	33.11
ϕ	2.48	2.55	0.52	1.05
λ_a	0.005	0.001	0.001	0.129
λ_w	0.001	0.001	0.015	0.063
γ	–	–	–	21.75
ϕ_g	–	–	–	0.92
β_1	–	–	–	1.62
β_2	–	–	–	-0.72
λ_r	–	–	–	0.05
λ_g	–	–	–	0.25
s	1.03	2.32	1.37	2.06
BIC	14252	6580	9332	15106
R^2	.92	.89	.87	.96
RMSD	.09	.10	.10	.06

Note. In cases where ATRIUM’s fit was worse than ALCOVE’s the ATRIUM-specific free parameters were not reported, as they did not affect the behavior of the model. Nevertheless, the BIC for ATRIUM was still computed using 11 free parameters.

Table 5.8: ATRIUM-DR Fits for Experiment 2

Parameter	Rule-Generalizers	Exception-Generalizers
c	29.50	2.38
ϕ	2.55	0.31
λ_a	0.005	0.188
λ_w	0.001	0.147
γ	–	0.001
ϕ_g	–	0.640
β_1	–	-0.958
β_2	–	-0.602
λ_r	–	0.039
λ_g	–	0.170
s	2.32	2.94
BIC	6580	18890
R^2	.92	.95
RMSD	.10	.07

Note. In cases where ATRIUM-DR's fit was worse than ALCOVE's the ATRIUM-specific free parameters were not reported, as they did not affect the behavior of the model. Nevertheless, the BIC for ATRIUM-DR was still computed using 11 free parameters.

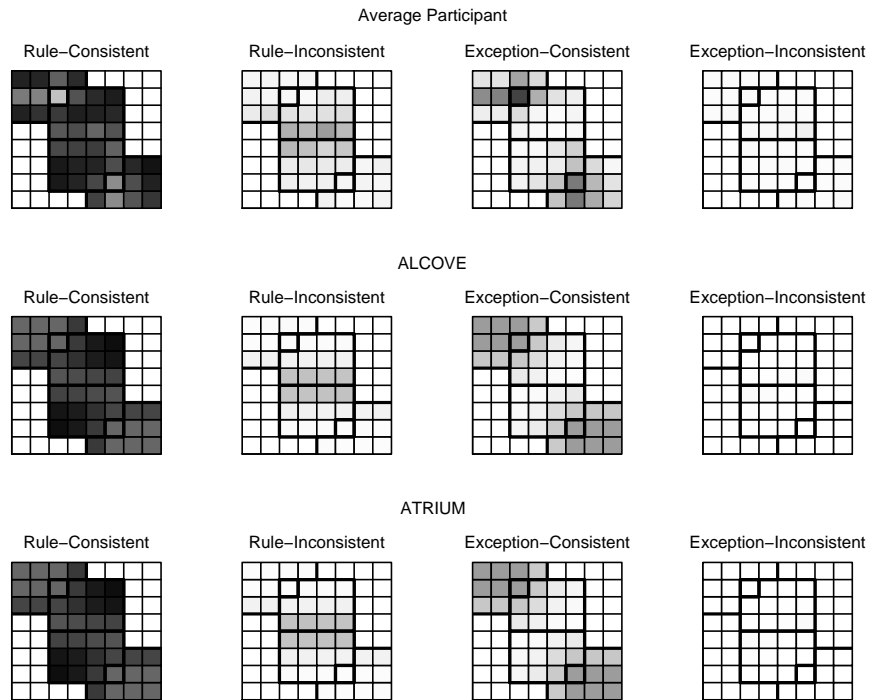


Figure 5.5: Proportion of responses for each stimulus during the transfer phase of Experiment 2 for rule-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

was able to capture the participants' high level of exception-consistent responses to training items nor match the participants' transfer item performance.

The superior fit by ALCOVE suggests that participants who generalized according to the rule items were not using one-dimensional rules to learn this category structure. Therefore, rules may not be necessary to learn category structures formed with a one-dimensional rule boundary and composed of stimuli with separable and noncommensurate dimensions.

In the two-dimensional conditions, ALCOVE (BIC = 6539) out performed both ATRIUM (BIC = 6580) and ATRIUM-DR (BIC = 6580). Comparing the performance of the participants with the models in Figure 5.6, the models managed to capture the rule generalization aspect of the participants' performance. However, like in the one-dimensional conditions, the models failed to respond with high exception training item accuracy and match transfer item behaviors. Also like in the one-dimensional conditions, examining the gating modules of ATRIUM and ATRIUM-DR

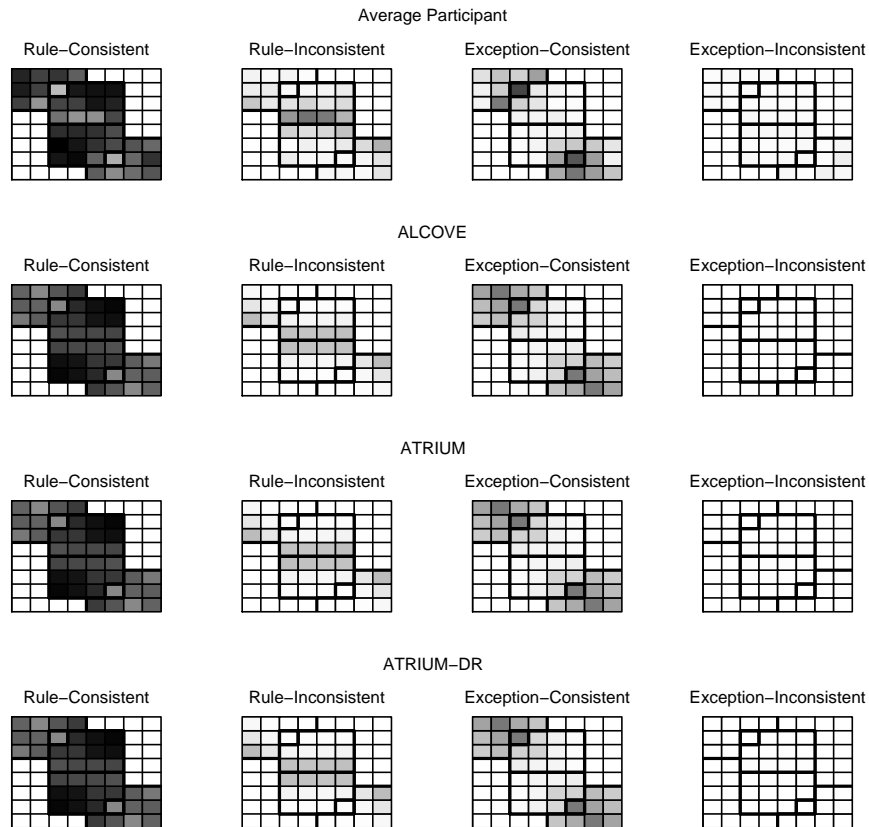


Figure 5.6: Proportion of responses for each stimulus during the transfer phase of Experiment 2 for rule-generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

revealed that both models exclusively relied upon their exemplar modules to learn the category structure.

In both the models' outputs and the participants' behaviors, there were very distinct transfer item performance behaviors. Looking at the rule-consistent and consistent exception responses for the averaged participant behaviors, there are distinct patterns in the transfer items. In the upper left transfer items, participants made higher exception responses to items that were to the upper right and the lower left of the exception training item. If this category structure was realigned with the actual stimulus dimensions of the stimuli (i.e., unrotating it), these would be the stimuli that share one particular feature nearly identically with the exception item. For example, given stimuli with a rule

boundary on size, participants were more likely to make exception-consistent responses to transfer items with the *same* size as the exception training item, and more likely to make rule-consistent responses to transfer items with *different* sizes.

In contrast to the single feature generalization occurring in the upper left transfer items, participants made different types of generalizations to the lower right transfer items. Instead of generalizing to a single feature, participants produced a pattern of behavior that was more consistent with generalization to both features of the exception item. Examination of the individual participant behaviors eliminates the possibility that this is an artifact of averaging over cluster members and reveals the same consistent pattern. Some participants display very different generalization patterns to symmetrical category structures.

The models failed to duplicate these two different patterns of behaviors. ALCOVE, the model that best fit the participants' behaviors, did so by generalizing in the opposite direction of the participants. However, given that the differences in behavior between the top left and bottom right transfer items were not evident during the training period of the experiment, it is not surprising that the models were unable to predict both of these outcomes. During the fitting process the models were unable to find a set of parameters that allowed the models to predict two different types of behaviors given a symmetrical category structure. ATRIUM and ATRIUM-DR could produce these sorts of patterns, however this would require direct manipulation of the models' modules and weights. Such artificially constructed versions of the models would not reflect the learning behaviors that occurred during the experiment, and would only produce the divergent strategies used by the participants during the transfer phase.

Experiment 2: Exception-Generalizers

The exception-generalizer cluster contains participants who categorized the transfer items in an exception-consistent manner. In the one-dimensional conditions, ALCOVE (BIC = 9291) was better able to account for this cluster of participants than ATRIUM (BIC = 9332). The behaviors of the models, as seen in Figure 5.7, show that ATRIUM was performing like ALCOVE.

Both models managed to capture the rule-generalization aspect of the participants' data. Yet, neither model was able to capture the participants' increases in exception-consistent responses to the transfer items. Examining the gating module of ATRIUM revealed that it was only using its exemplar module and was not using rules.

The superior fit by ALCOVE suggests that participants who generalized according to the exception items were not using one-dimensional rules to learn the category structure. Therefore,

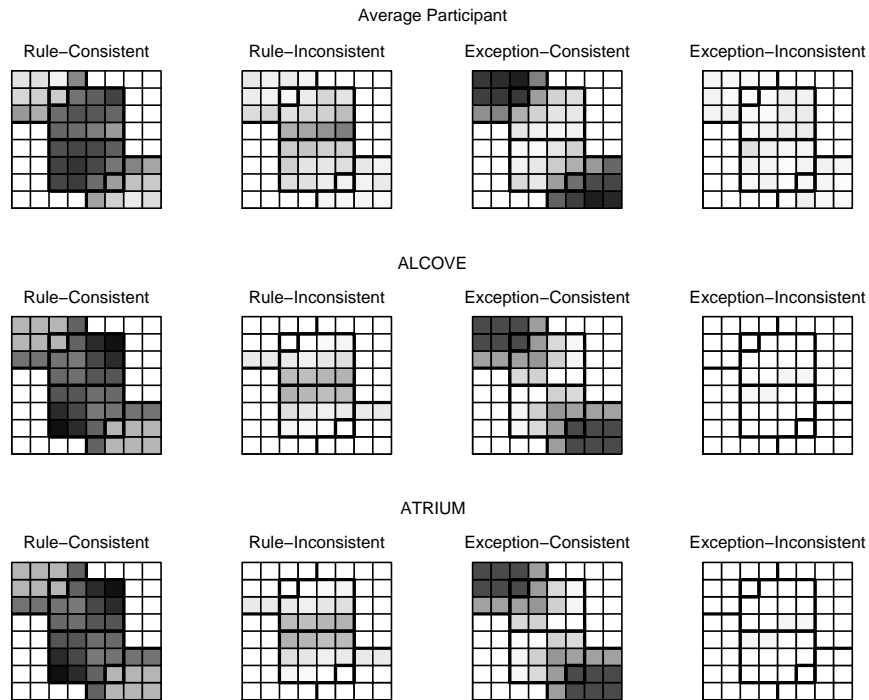


Figure 5.7: Proportion of responses for each stimulus during the transfer phase of Experiment 2 for exception-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

rules may not be necessary to learn category structures formed with a one-dimensional rule boundary and composed of stimuli with separable and noncommensurate dimensions.

In the two-dimensional conditions, ATRIUM (BIC = 15106) out performed both ALCOVE (BIC = 15924) and ATRIUM-DR (BIC = 15720). Comparing the performance of the participants with the models in Figure 5.8, the models managed to capture the exception generalization aspect of the participants' performance. The key items that distinguished between the behaviors of ALCOVE, ATRIUM, and ATRIUM-DR were the transfer items and the frequency of rule inconsistent responses close to the boundary between the rule items.

ALCOVE's transfer generalization pattern was influenced by its initial salience parameter and its attentional settings. This resulted in ALCOVE being more influenced by one dimension than the other dimension. This influence was manifested in ALCOVE's generalization patterns. When ALCOVE generalized from the exception training items during the transfer phase, it did so along

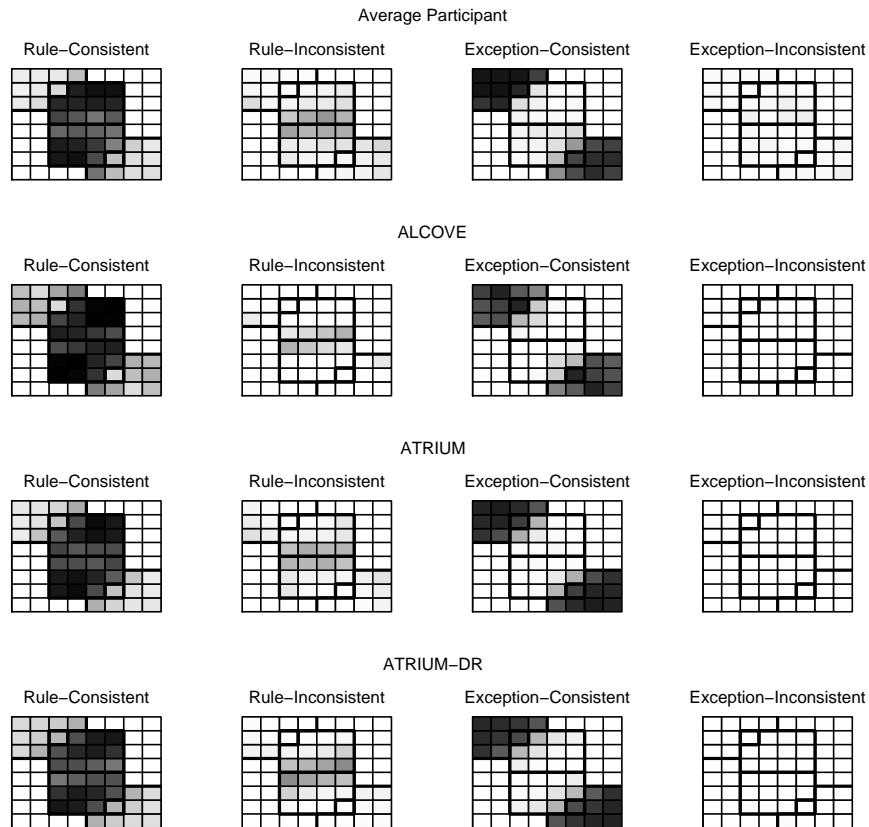


Figure 5.8: Proportion of responses for each stimulus during the transfer phase of Experiment 2 for exception-generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

one dimension. This resulted in the transfer items, to the top left and bottom right of the exception training items, having a large proportion of exception consistent responses. The participants did not have this type of generalization pattern. ATRIUM and ATRIUM-DR were not as heavily influenced by attentional settings and better matched participant transfer performance.

The participants' behavior at the boundary between the rule training items was characterized by a large proportion of rule-inconsistent responses. This behavior is characteristic of participants with an ill-defined boundary between the two categories. ATRIUM-DR attempted to match this performance using only its exemplar module, whereas ATRIUM used both a rule module and its exemplar module. Both models used a rule module to produce exception generalization to the

transfer items. However, because ATRIUM uses one dimensional rules, it was also able to use the same rule module to learn training item classifications. ATRIUM's use of two modules allowed it to better match participant performance than ATRIUM-DR's single module.

ATRIUM provided a superior fit to the data of the participants who generalized according to the exception items in two-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn category structures formed with a two-dimension rule boundary and composed of stimuli with separable and noncommensurate dimensions. It also suggests that participants who were exception-generalizers in the two-dimensional conditions were not required to use two-dimensional rules.

Experiment 2: Summary

The transfer behaviors of participants who were given category structures formed using one-dimensional boundaries and composed of stimuli with separable and noncommensurate dimensions were better fit by ALCOVE than ATRIUM. Likewise, ATRIUM-DR failed to provide a better fit than ALCOVE in the two-dimensional conditions. This experiment failed to find evidence of two-dimensional rule use. However, given the very different transfer item generalization patterns produced by participants to symmetrical category structures, caution is suggested when interpreting these results.

5.7 Experiment 3: Integral and Commensurate Dimensions

The stimuli for Experiment 3 possessed integral and commensurate dimensions. They were rectangles that varied in width and height. The best fitting parameters for Experiment 3 for ALCOVE can be seen in Table 5.9, for ATRIUM in Table 5.10, and for ATRIUM-DR in Table 5.11. The average participant performance and the predictions of the models can be seen in Figures 5.9 through 5.12.

Experiment 3: Rule-Generalizers

The rule-generalizer cluster contains participants who categorized the transfer items in a rule-consistent manner. In the 1-D conditions, the horizontal and vertical conditions, both ALCOVE and ATRIUM were able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 8584) was better able to account for this cluster of participants than ALCOVE

Table 5.9: ALCOVE Fits for Experiment 3

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	2.15	23.93	90.86	14.80
ϕ	2.43	3.73	0.94	1.95
λ_a	1.125	0.001	0.069	0.161
λ_w	0.002	0.003	0.018	0.010
s	1.41	0.85	0.93	0.81
BIC	9268	22952	9552	6288
R^2	.86	.92	.84	.90
RMSD	.13	.09	.13	.10

Table 5.10: ATRIUM Fits for Experiment 3

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	14.18	34.37	19.74	4.73
ϕ	0.58	4.37	1.72	1.87
λ_a	1.853	0.001	0.002	0.928
λ_w	0.103	0.001	0.001	0.005
γ	1.508	0.969	7.187	2.650
ϕ_g	1.05	1.791	0.743	1.772
β_1	-1.761	-2.876	3.762	1.722
β_2	2.060	0.250	0.050	-1.48
λ_r	0.738	0.010	0.036	0.474
λ_g	0.135	0.202	0.562	0.094
s	0.854	1.018	1.289	1.628
BIC	8584	22091	9032	6074
R^2	.93	.95	.95	.94
RMSD	.09	.08	.08	.08

Table 5.11: ATRIUM-DR Fits for Experiment 3

Parameter	Rule-Generalizers	Exception-Generalizers
c	33.750	16.816
ϕ	3.217	1.635
λ_a	0.001	0.018
λ_w	0.037	0.013
γ	1.238	12.811
ϕ_g	1.644	1.931
β_1	1.118	2.327
β_2	-1.347	4.111
λ_r	0.006	0.015
λ_g	0.051	0.066
s	0.978	0.906
BIC	21713	6243
R^2	.97	.93
RMSD	.06	.09

(BIC = 9268). The key items that ALCOVE failed to account for, but ATRIUM was able to account for, were the exception training items and the transfer items, as shown in Figure 5.9.

ALCOVE, with only an exemplar representation, was unable to capture both the high exception-consistent response rates to the exception training items and produce rule-consistent generalization to the transfer items. However, ATRIUM was able to produce these patterns of behavior. ATRIUM learned to categorize the training items with its exemplar module and used its rule module to learn to classify the transfer items.

ATRIUM provided a superior fit to the data of the participants who generalized according to the rule items in the one-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn category structures formed with a one-dimension rule boundary and composed of stimuli with integral and commensurate dimensions.

In the two-dimensional conditions, all models were able to qualitatively match the performance of the participants. However, ATRIUM-DR (BIC = 21713) was better able to account for this cluster of participants than ALCOVE (BIC = 22952) or ATRIUM (BIC = 22091). The key items distinguishing between the models were the transfer items.

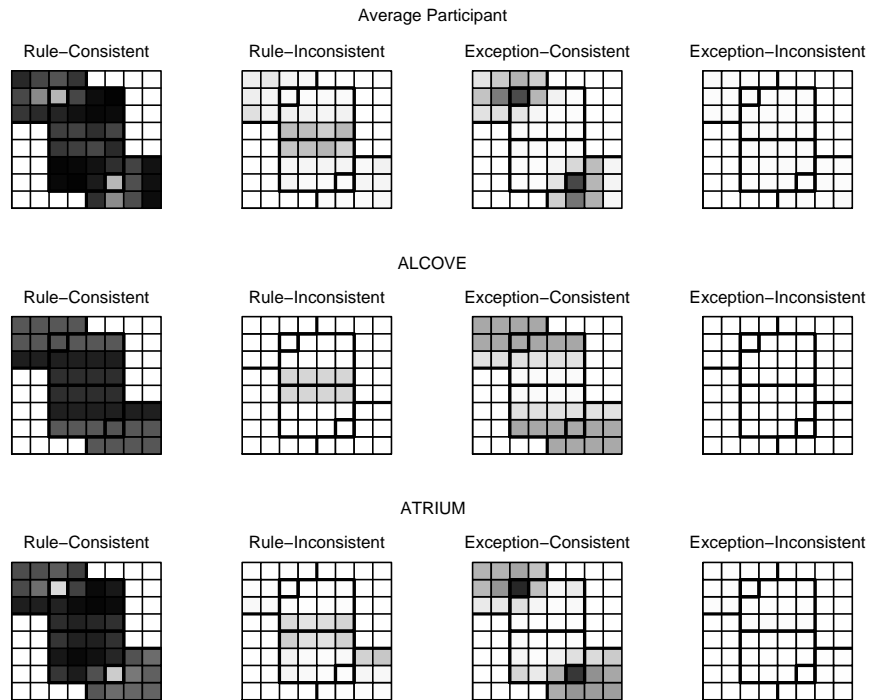


Figure 5.9: Proportion of responses for each stimulus during the transfer phase of Experiment 3 for rule-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

ALCOVE and ATRIUM were both heavily influenced by dimensional attention when making transfer item generalizations. This is demonstrated by the large proportion of exception-consistent responses to the transfer items that were aligned with the exception training item along a single dimension. These items were the transfer item to the upper left of the exception training item and the transfer item to the lower right of the exception training item. Examining ATRIUM's gating module, revealed that ATRIUM's responses were dominated by its exemplar module, resulting in very ALCOVE-like performance. However, ATRIUM-DR was more successful in accounting for participant transfer item performance. It was able to use both its rule and exemplar modules to account for the behavior of the participants.

ATRIUM-DR provided a superior fit to the data of the participants who generalized according to the rule items in the two-dimensional conditions. This suggests that participants are able

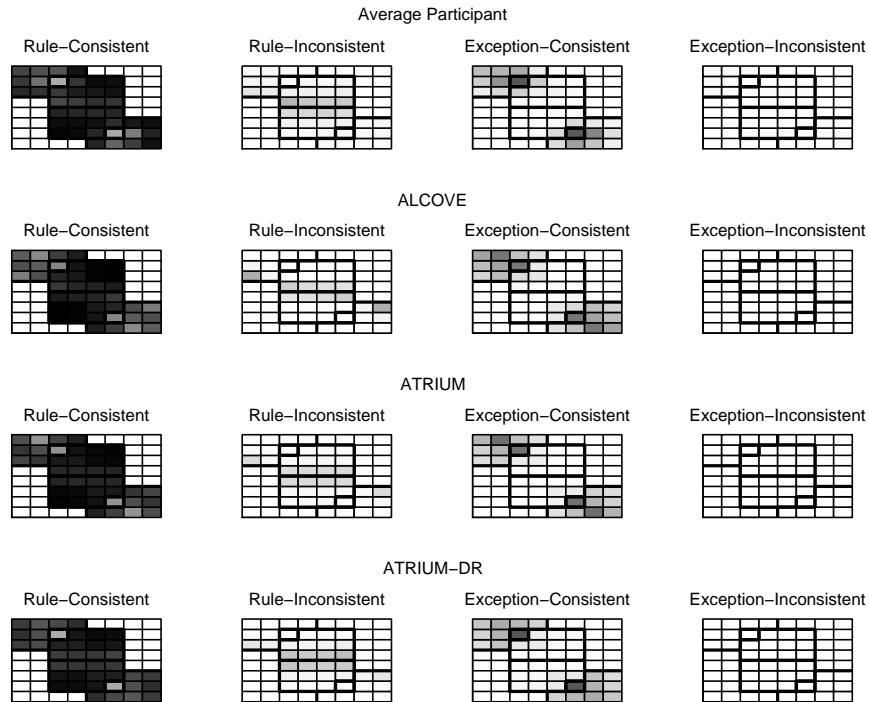


Figure 5.10: Proportion of responses for each stimulus during the transfer phase of Experiment 3 for rule-generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

to use two-dimensional rules to learn category structures formed with two-dimension rule boundaries and composed of stimuli with integral and commensurate dimensions.

Experiment 3: Exception-Generalizers

The exception-generalizer cluster contains participants who categorized the transfer items in an exception-consistent manner. In the one-dimensional conditions, the horizontal and vertical conditions, both ALCOVE and ATRIUM were able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 9032) was better able to account for this cluster of participants than ALCOVE (BIC = 9552). The transfer items were the key items that ALCOVE failed to account for, but ATRIUM was able to account for, as shown in Figure 5.3.

To account for the high exception-consistent response rate to the exception training items, ALCOVE was required to have high specificity and therefore low generalization between exemplars.

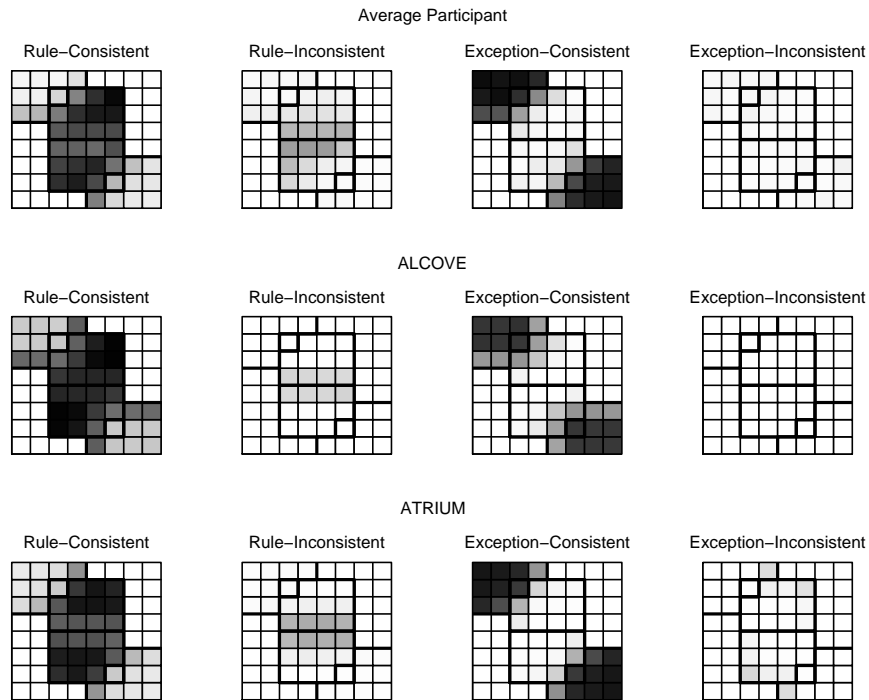


Figure 5.11: Proportion of responses for each stimulus during the transfer phase of Experiment 3 for exception-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

This resulted in the module having a high exception-consistent response rate to the transfer items in the corners, but a low exception-consistent rate to the transfer items on the edges. These transfer items were closer to the rule-consistent training items than the exception-consistent training items, and were therefore categorized with a higher proportion of rule-consistent responses. However, the participants did not display this pattern of generalization. Participants responded with high exception-consistent responses to all transfer items. In contrast to ALCOVE, ATRIUM was able to produce this pattern of results by using both its rule and exemplar modules.

ATRIUM provided a superior fit to the data of the participants who generalized according to the exception items in the one-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn category structures formed with a one-dimension rule boundary and composed of stimuli with integral and commensurate dimensions.

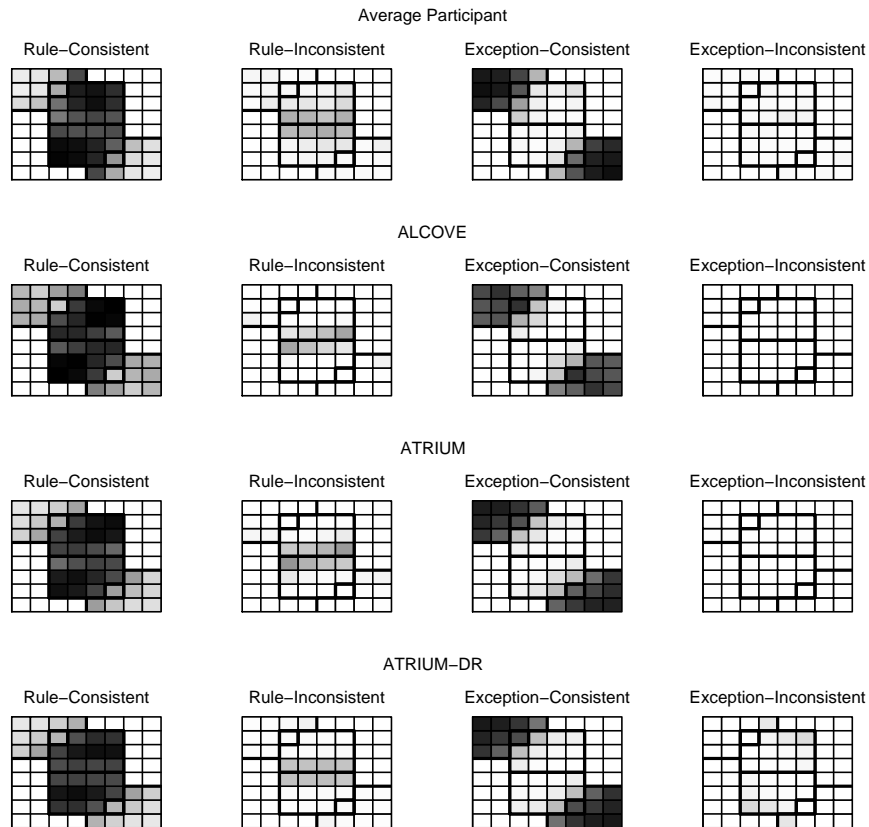


Figure 5.12: Proportion of responses for each stimulus during the transfer phase of Experiment 3 for exception-generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

In the two-dimensional conditions, the positive and negative diagonal conditions, ALCOVE, ATRIUM, and ATRIUM-DR were all able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 6074) was better able to account for this cluster of participants than ALCOVE (BIC = 6288) or ATRIUM-DR (BIC = 6243). The keys items for the models to fit were the transfer items and the training items near the boundary between the rule items, as shown in Figure 5.4.

ALCOVE was unable to match the transfer item generalization pattern displayed by participants. In learning the training items, ALCOVE allocated a large amount of dimensional attention to one dimension. This caused ALCOVE's transfer generalization to be driven by a single

dimension, unlike participants who generalized more equally across both dimensions. ATRIUM and ATRIUM-DR were more successful than ALCOVE in producing this pattern of behavior. ATRIUM and ATRIUM-DR were able to use different modules to classify training and transfer items. This allowed them to perform more like the participants.

ATRIUM was better able to match participant behaviors than ATRIUM-DR. Specifically, the models performed differently on the training items close to the boundary between the two rule categories. ATRIUM was able to use its exemplar module to learn the training items. The exemplar module is more flexible for learning training items because it has exemplars for each of the items. Each exemplar is able to be associated with an output response independently of other items. In contrast, ATRIUM-DR relied upon a rule module to classify the training items and was unable to produce the slightly skewed boundary between the two categories that the participants displayed. Rules are less flexible than exemplars because rules produce the same classification for all items that are equidistant from the rule.

In addition, because ATRIUM uses one-dimensional boundaries, it would still be able to produce a skewed boundary pattern if it was required to use its rule module to learn the training items. ATRIUM-DR's rule modules are aligned to the actual category structure. This forces it to fit too well to the category structure, unlike the participants who were more uncertain about the location of the boundary. Interestingly, ATRIUM-DR did not use its exemplar module to learn this category structure, it only used rule modules. If ATRIUM-DR was required to use its exemplar module, it may match ATRIUM's performance.

ATRIUM provided a superior fit to the data of the participants who generalized according to the exception items in two-dimensional conditions. This suggests that participants may not use two-dimensional rules in learning category structures formed with a two-dimension rule boundary and composed of stimuli with integral and commensurate dimensions.

Experiment 3: Summary

The transfer behaviors of participants who were given category structures formed using one-dimensional boundaries and composed of stimuli with integral and commensurate dimensions were better fit by ATRIUM than ALCOVE. This is consistent with previous literature supporting rule use in category learning (Erickson & Kruschke, 1998, 2002).

The transfer behaviors of participants who were given category structures formed using two-dimensional boundaries and composed of stimuli with integral and commensurate dimensions were best fit by ATRIUM-DR when transfer item generalization was rule consistent. However,

when participant transfer item generalization was exception-consistent, the data were best fit by ATRIUM. This provides limited evidence that people may use two-dimensional rules with integral and commensurate dimensions.

ATRIUM-DR's failure to account for the behaviors of participants who generalized according to the exception may be a result of the model having too many possible rules. The second rule module allowed for it to account for participants' transfer and training item performance, without requiring the exemplar module. ATRIUM's success over ATRIUM-DR may be due to its use of only one rule module and its exemplar module. A version of ATRIUM-DR with only one rule module may be able to produce a better fit. Currently, rules in ATRIUM and ATRIUM-DR are assigned by the experimenter. The models can learn when it is appropriate to use a particular rule, but they cannot recruit new rules or remove inappropriate rules. This is a design feature of the models that reflect the implicit nature of rule development and use.

5.8 Experiment 4: Integral and Noncommensurate Dimensions

The stimuli for Experiment 4 possessed integral and noncommensurate dimensions. They were Fourier descriptors that varied in initial phase and the amplitude of a sine wave component. The best fitting parameters for Experiment 4 for ALCOVE can be seen in Table 5.12, for ATRIUM in Table 5.13, and for ATRIUM-DR in Table 5.14. The average participant performance and the predictions of the models can be seen in Figures 5.13 through 5.16.

Experiment 4: Rule-Generalizers

The rule-generalizer cluster contains participants who categorized the transfer items in a rule-consistent manner. In the one-dimensional conditions both ALCOVE and ATRIUM were able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 19953) was better able to account for this cluster of participants than ALCOVE (BIC = 21532). Yet, neither model was able to match the exception-consistent response rates for the exception training items displayed by the participants, as shown in Figure 5.13. Examining the gating module of ATRIUM reveals that it was primarily using its exemplar module with only a small contribution from its rule modules. This small contribution was sufficient to allow ATRIUM to outperform ALCOVE.

While ATRIUM provided a superior fit to the data of the participants who generalized according to the rule items in one-dimensional conditions, it failed to utilize a strong rule module. The contributions of the rule modules were rather weak, suggesting that rules may not be responsible

Table 5.12: ALCOVE Fits for Experiment 4

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	40.43	94.60	13.65	27.30
ϕ	2.17	0.38	1.07	1.59
λ_a	1.592	0.003	0.362	0.139
λ_w	0.001	0.035	0.025	0.012
s	0.81	0.33	0.46	0.45
BIC	21532	3595	4507	19929
R^2	.86	.92	.84	.90
RMSD	.13	.09	.13	.10

Table 5.13: ATRIUM Fits for Experiment 4

Parameter	Rule- Generalizers		Exception- Generalizers	
	1-D	2-D	1-D	2-D
c	4.451	8.608	8.552	11.283
ϕ	2.546	1.983	1.108	1.315
λ_a	0.082	0.229	0.131	0.092
λ_w	0.010	0.005	0.131	0.018
γ	0.012	0.436	20.44	9.997
ϕ_g	1.192	0.943	1.020	1.095
β_1	0.042	0.353	0.870	1.562
β_2	0.92	-0.003	-0.002	-0.007
λ_r	0.002	0.001	0.112	0.179
λ_g	0.074	0.552	0.036	0.198
s	0.872	0.400	0.564	0.477
BIC	19952	3509	4287	18786
R^2	.93	.95	.95	.94
RMSD	.09	.08	.08	.08

Table 5.14: ATRIUM-DR Fits for Experiment 4

Parameter	Rule-Generalizers	Exception-Generalizers
c	26.52	14.186
ϕ	0.542	1.144
λ_a	0.044	0.098
λ_w	0.059	0.016
γ	6.098	8.929
ϕ_g	0.236	0.663
β_1	-1.125	2.452
β_2	-3.88	-0.456
λ_r	0.048	0.096
λ_g	0.018	0.855
s	0.186	0.899
BIC	3593	19265
R^2	.97	.93
RMSD	.06	.09

for ATRIUM's superior performance. This provides weak evidence that participants are able to use one-dimensional rules to learn category structures formed with a one-dimension rule boundary and composed of stimuli with integral and noncommensurate dimensions.

In the two-dimensional conditions ALCOVE, ATRIUM, and ATRIUM-DR were all able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 3509) was better able to account for this cluster of participants than ALCOVE (BIC = 3595) or ATRIUM-DR (BIC = 3593). The transfer items were the key items for distinguishing between the models, as shown in Figure 5.14.

ALCOVE and ATRIUM-DR's performance on the transfer stimuli was characterized by a strong preference for one dimension. Examination of ATRIUM-DR's gating module revealed that the model was dominated by its exemplar module, resulting in very ALCOVE-like performance. ATRIUM, however, was able to utilize both rule and exemplar modules. This allowed the model to match participant behaviors better than ALCOVE or ATRIUM-DR. However, ATRIUM still failed to match participant response patterns to the exception training items.

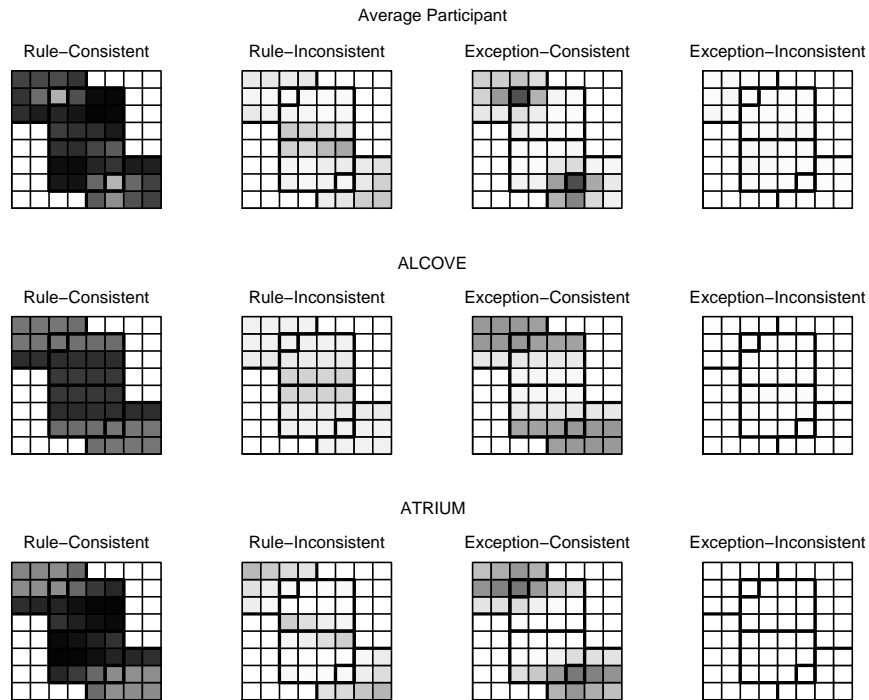


Figure 5.13: Proportion of responses for each stimulus during the transfer phase of Experiment 4 for rule-generalizer in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

ATRIUM provided a superior fit to the data of the participants who generalized according to the rule items in two-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn category structures formed with a two-dimension rule boundary and composed of stimuli with integral and noncommensurate dimensions. It also fails to provide evidence for two-dimensional rule use in these types of categories structures.

Experiment 4: Exception-Generalizers

The exception-generalizer cluster contains participants who categorized the transfer items in an exception-consistent manner. In the one-dimensional conditions both ALCOVE and ATRIUM were able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 4287) was better able to account for this cluster of participants than ALCOVE (BIC = 4507). There

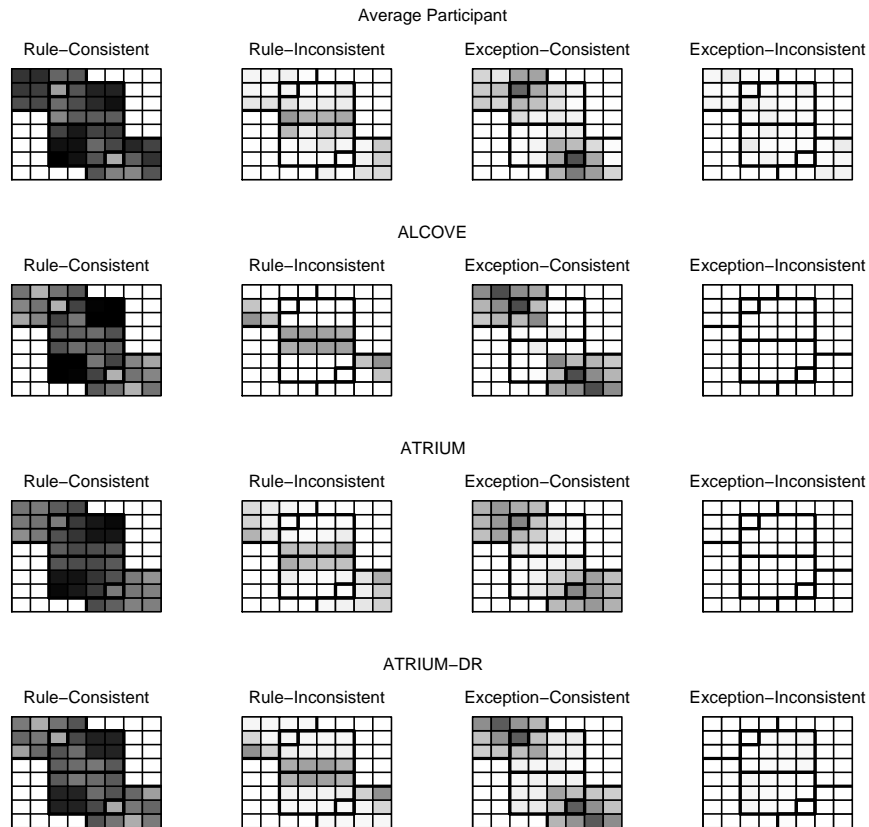


Figure 5.14: Proportion of responses for each stimulus during the transfer phase of Experiment 4 for rule-generalizer in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

are two key sets of items that ALCOVE failed to account for, but ATRIUM was able to account for, the transfer items and the number of rule inconsistent responses, as shown in Figure 5.15.

To learn to classify the exception training items with a high proportion of exception-consistent responses, ALCOVE used a high specificity value. This resulted in few exception-consistent responses to the transfer items closest to the rule-consistent training items. This is in contrast to participant behavior, which was characterized by a high proportion of exception-consistent responses to all transfer items. However, ATRIUM was able to produce this pattern of behavior by using its exemplar module to classify the training items and a combination of exemplar and rule modules to classify the transfer items.

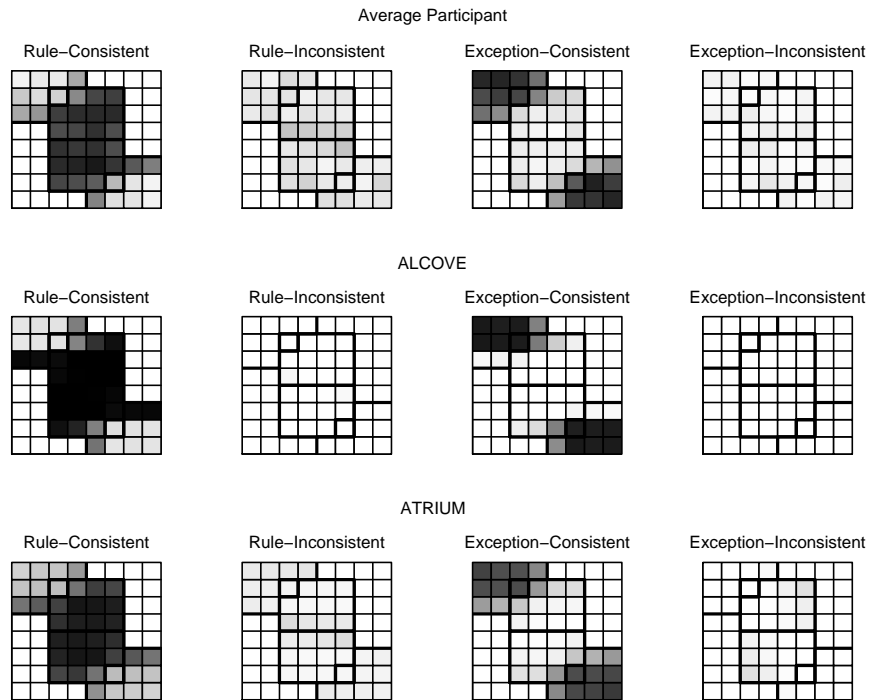


Figure 5.15: Proportion of responses for each stimulus during the transfer phase of Experiment 4 for exception-generalizers in the 1-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM.

The participants' transfer performance was also characterized by a large proportion of rule-inconsistent responses. This suggests that participants found the task difficult. ALCOVE was unable to produce this large error rate, however ATRIUM was able to do so.

ATRIUM provided a superior fit to the data of the participants who generalized according to the exception items in the one-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn category structures formed with a one-dimension rule boundary and composed of stimuli with integral and noncommensurate dimensions.

In the two-dimensional conditions, the positive and negative diagonal conditions, ALCOVE, ATRIUM, and ATRIUM-DR were all able to perform in a manner qualitatively similar to the participants. However, ATRIUM (BIC = 18786) was better able to account for this cluster of participants than ALCOVE (BIC = 19929) or ATRIUM-DR (BIC = 19265). The key items for the models to fit were the transfer items and the training items near the rule boundary, as shown in

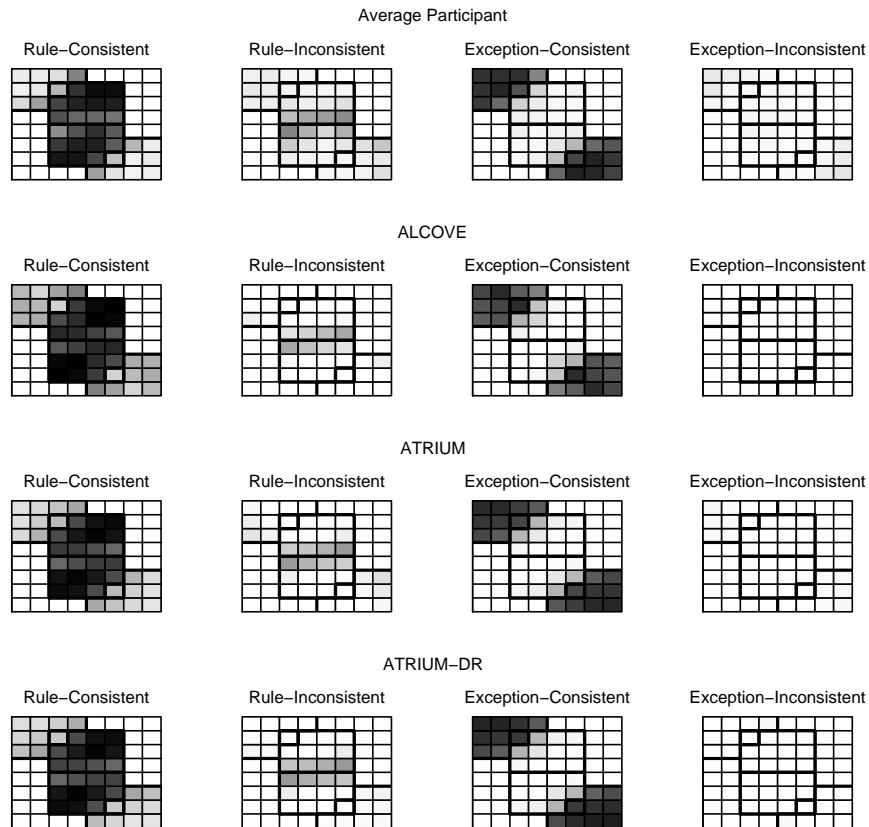


Figure 5.16: Proportion of responses for each stimulus during the transfer phase of Experiment 4 for exception-generalizers in the 2-D conditions. The shading in each cell corresponds to the proportion of responses for each response type. Dark cells indicated a high proportion of responses, light cells indicate a low proportion of responses. The top row displays the empirical data averaged over participant. The second row displays the predictions made by ALCOVE. The third row displays the predictions made by ATRIUM. The fourth row displays the predictions made by ATRIUM-DR.

Figure 5.16.

ALCOVE's transfer item performance was characterized by a strong preference for one dimension. This caused ALCOVE's transfer generalization to be driven by a single dimension, unlike participants who generalized more equally across both dimensions. ATRIUM and ATRIUM-DR were more successful than ALCOVE in producing this pattern of behavior. The models were able to use different modules to classify training and transfer items, which allowed them to perform more like the participants.

Like in the rule-generalizer cluster, participants made a large number of rule-inconsistent responses. This suggests that both clusters of participants found the task difficult. Additionally, the

frequency of rule-inconsistent responses on the outside edges of the rule boundary suggests participants were not using an optimal rule, but instead one biased toward one dimension. Examination of the gating modules revealed that ATRIUM and ATRIUM-DR used rule modules to classify transfer stimuli and exemplar modules to classify training stimuli. ATRIUM's rule modules are optimal for categorizing the transfer stimuli, allowing it to use one rule module for the transfer items and the other rule module to facilitate training item performance. ATRIUM-DR, however, used both rule modules to classify the transfer items, and only the exemplar module was used to classify training items.

ATRIUM provided a superior fit to the data of the participants who generalized according to the exception items in two-dimensional conditions. This suggests that participants are able to use one-dimensional rules to learn category structures formed with a two-dimension rule boundary and composed of stimuli with integral and noncommensurate dimensions. It also fails to provide evidence for two-dimensional rule use in these types of categories structures.

Experiment 4: Summary

The transfer behaviors of participants who were given a category structure formed using one-dimensional boundaries composed of stimuli with integral and noncommensurate dimensions were better fit by ATRIUM than ALCOVE. This is consistent with previous literature supporting rule use in category learning (Erickson & Kruschke, 1998, 2002).

The transfer behaviors of participants who were given category structures formed using two-dimensional boundaries composed of stimuli with integral and noncommensurate dimensions were best fit by ATRIUM. This fails to provide evidence for two-dimensional rules for categories using stimuli with integral and noncommensurate dimensions.

5.9 Summary

The summary of the modeling fitting can be seen in Table 5.15. Support for two-dimensional rule use was only found in Experiments 1 and 3. Experiments 1 and 3 both contained stimuli composed of commensurate dimensions in contrast to Experiments 2 and 4 which contained stimuli composed of noncommensurate dimensions.

In Experiment 1, both rule-generalizing and exception-generalizing clusters were found to be best described by ATRIUM-DR. This suggests that 2D rules were used by participants using both strategies. In Experiment 3, only those participants who generalized according to the rule items

Table 5.15: Summary of Model Fits

Experiment	Cluster 1: Rule-Generalizers		Cluster 2: Exception-Generalizers	
	1-D	2-D	1-D	2-D
1: Separable & Commensurate	ATRIUM	ATRIUM-DR	ATRIUM	ATRIUM-DR
2: Separable & Noncommensurate	ALCOVE	ALCOVE	ALCOVE	ATRIUM
3: Integral & Commensurate	ATRIUM	ATRIUM-DR	ATRIUM	ATRIUM
4: Integral & Noncommensurate	ATRIUM	ATRIUM	ATRIUM	ATRIUM

were found to be best described by ATRIUM-DR. Participants in Experiment 3 who generalized according to the exception items were found to be best described by ATRIUM.

Chapter 6

General Discussion

In this dissertation, I addressed the nature of rules in categorization. Specifically, I investigated the use of two-dimensional rules in category learning and the conditions that allow for two-dimensional rule use. Two factors were explored in four experiments designed to test two-dimensional rule use. These factors were the dimensionality (integrality or separability) and the verbalizability (commensurate or noncommensurate) of the stimulus dimensions.

The data from these experiments were organized into groups based upon the performance of the participants. These groups were then used for the fitting of three different models of category learning, ALCOVE, ATRIUM, and ATRIUM-DR. Evidence for two-dimensional rule use was provided by participant behaviors that were best described by ATRIUM-DR. The results of the model fitting were used to evaluate the claims and predictions made by the various rule-based accounts of category learning (refer back to Table 2.2).

The results of the model fitting (see Table 5.15) suggest that two-dimensional rules can be used by people when learning categories under certain conditions. Participants from Experiment 1 and 3 were found to show two-dimensional rule use, as they were best modeled by ATRIUM-DR. In Experiment 1, participants in the two-dimensional conditions that were classified as rule-generalizers or exception-generalizers were found to be best described by ATRIUM-DR. Likewise, in Experiment 3, participants in two-dimensional conditions that were classified as rule-generalizers were found to be best described by ATRIUM-DR. This suggests that people can use two-dimensional rules when learning two-dimensional category structures.

Experiments 1 and 3 both contained stimuli composed of verbalizable dimensions. This suggests that two-dimensional rules can only be used to learn categories with stimuli composed of verbalizable dimensions. This is not consistent with any of the rule accounts that were discussed in

Chapter 1 (refer back to Table 2.2 for the predictions of the accounts). The account that is closest to the current results is the account provided by Ashby et al. (1998).

Ashby et al. (1998) claim that two-dimensional rule use is only possible with stimuli composed of separable and verbalizable dimensions. The results of Experiment 1 are consistent with this claim. However, the results of Experiments 3 are not consistent with this claim. Experiment 3 used categories with stimuli composed of integral and verbalizable dimensions, finding two-dimensional rule use with these types of stimuli is not directly predicted by Ashby et al. (1998). Thus, Ashby et al.'s claim is not completely supported.

6.1 Two-Dimensional Rules or Predecisional Integration

There is a possible explanation for these data that would support Ashby et al.'s account. It is possible that participants were not using rules, but instead were performing predecisional integration that resulted in rule-like behaviors (Ashby et al., 1998; Ashby, Ell, & Waldron, 2003; Ashby & Gott, 1988; Shaw, 1982). The stimuli used in Experiment 3 possessed integral dimensions, they were rectangles that varied in height and width. Some participants have been found to perceive integral stimuli, not as combinations of the dimensions (i.e., of height and width), but instead along an emergent unitary dimension (i.e., of area, a ratio of height to width, or similar emergent composite dimensions). If some participants did perceive these stimuli as rectangles that varied along the dimension of area, then these participants would appear to use two-dimensional rules when learning a category with a two-dimensional boundary. However, they would actually be using a one-dimensional bound along the single dimension of area.

Support for this possibility comes from the distributions of participants in the clusters. In general, people prefer to use simple strategies (e.g., Nosofsky et al., 1994). This leads to a preference for using one-dimensional rules even when the category structures used in the task were created using two-dimensional boundaries (e.g., Erickson, 2008). Experiments 1, 2, and 4 are consistent with this idea. In these experiments, more participants in the one-dimensional conditions grouped into the rule-generalization cluster than the exception-generalization cluster. In contrast, more participants in the two-dimensional conditions were grouped into the exception-generalization cluster than into the rule-generalization cluster. This suggests that in these experiments, participants found rule-generalization to be harder in two-dimensional conditions than in one-dimensional conditions. This also supports the finding that two-dimensional rule use is more difficult than one-dimensional rule use.

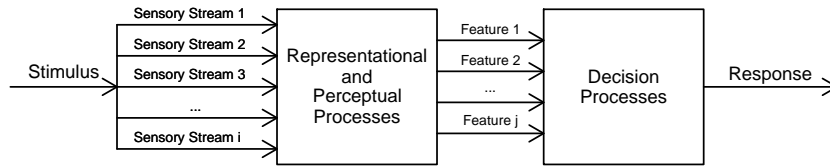


Figure 6.1: An information processing model displaying the processes that occur during categorization.

However, in Experiment 3, this pattern of behavior is reversed. In Experiment 3, participants in the one-dimensional conditions were *equally likely* to use rule-generalization as exception-generalization. Additionally, participants in the two-dimensional conditions were *more* likely to use rule-generalization than exception generalization. This suggests that participants in Experiment 3 found two-dimensional rule use easier than one-dimension rule use, in contrast to Experiments 1, 2, and 4. This reversal in difficulty may be a result of participants treating the stimuli as varying along the dimension of area. By representing the stimuli as varying along the dimension of area, participants are able to use a simpler one-dimensional rule to learn the two-dimensional category structure. Likewise, representing the stimuli as varying along the dimension of area would also result in some participants having to use a more difficult two-dimensional rule to learn the one-dimensional category structure. This would result in the conditions being reversed. Participants would use one-dimensional rules to learn categories with one-dimensional boundaries and use two-dimensional rules to learn categories with one-dimensional boundaries. This reversal would also affect difficulty, resulting in categories with two-dimensional boundaries being easier to learn than categories with one-dimensional boundaries.

In addition, Experiment 3 provided the largest number of participants classified into Cluster 5. This was the cluster that contained participants who were best described as using a two-dimensional rule on a one-dimensional category structure. This is consistent with the assumption that some participants treated the stimuli as varying along the dimension of area. For the participants using a one-dimensional area rule, the task of categorizing stimuli based on height or width would be like a two-dimensional category structure. If participants in Experiment 3 were displaying two-dimensional rule use, but were actually using a one-dimensional rule along a composite dimension, their behavior may not truly be two-dimensional rule use.

To better illustrate this point requires differentiating between perceptual and decision integration (Ashby & Townsend, 1986; Shaw, 1982). Figure 6.1 displays a basic information processing model containing the processes that occur during categorization. This model is simplified, it contains no references to the recurrent or continuous nature of human cognitive functions (McClelland, 1979), but it is sufficient to demonstrate the time-line of categorization. A stimulus has a set of sensory components that can be detected and processed by the human sensory system. These raw sensory components are processed pre-attentively to produce streams of sensory information. These streams contain information such as color, luminance, shape, and motion. These sensory streams are then combined using attention to form object representations (Treisman & Gelade, 1980). Next, these object representations are used in making category decisions. Finally, the output of the decision process is used to make a category response.

The processes underlying categorization can be divided into two types of processes, perceptual and decision processes (Ashby & Townsend, 1986; Shaw, 1982). Perceptual processes operate on the raw sensory information provided by the stimulus, transforming it into a psychological representation. Decision processes operate on the psychological representation created by perceptual processes and result in a category response. This method of modeling the processes that occur in categorization is relatively straightforward in the unidimensional case, however the addition of other relevant stimulus dimensions greatly increases the complexity of the task. When stimuli with multiple relevant dimensions are categorized, at some point the information from each dimension must be combined. This integration process can occur in the representational stages or the decisional stages (Ashby & Townsend, 1986; Massaro & Friedman, 1990; Shaw, 1982).

Representational integration refers to the integration of stimulus dimensions that occurs before or during the object representation of a stimulus. This type of integration is performed for the purpose of constructing object representations. It contains all types of integration that occur pre-attentively. This includes low level sensory integration of perceptual primitives such as luminance, shape, and motion. It also includes higher level integration such as the changes to object representations that are due to perceptual learning about members of a category (Foard & Kemler Nelson, 1984; J. D. Smith & Kemler Nelson, 1984). This type of integration is also contained within the term, *predecisional representation* (Ashby & Gott, 1988; Ashby et al., 1998). I use *representational integration* to emphasize that the integration does not occur at a basic sensory level, but instead when representing the stimulus.

An example of representational integration is color. The underlying dimensions of color, (i.e., hue, saturation, and brightness) are combined to form a single color percept. These dimensions

are classical examples of integral dimensions. The combination occurs pre-attentively, after which, the system loses access to the information used (e.g., hue, saturation, and brightness) to create the new feature.

Decisional integration refers to the integration of dimensions that occurs after a stimulus has been fully represented. It is during the decisional processes stage that rule-based categorization occurs. In the case of a unidimensional category, a stimulus is represented as a single feature along a dimension. This dimension is evaluated by a one-dimensional rule and the outcome is used to produce a response. In the case of a multidimensional category being classified using conjunction rules, the process is more complicated, but still occurs within the decisional processes stage. In this case, a stimulus is represented by two independent features. These features are evaluated separately. The outcome of each of these evaluations is then combined and a new evaluation is performed. The result of this third evaluation is used to produce a response.

An example of decisional integration is the category defined by the conjunction rule, *a black cat is a cat and is black*. When given the stimulus of a white cat, the decision process is as follows. The stimulus (a white cat) is represented by the dimensions of animal (cat) and color (white). The color dimension is white, this is evaluated by the rule, *is black*, and is found to be false. The animal dimension is cat, this is evaluated by the rule, *is a cat*, and is found to be true. The results of both of these comparisons are combined (*true* and *false*), and the result, *false* is used to produce a category response. In this case, the answer is *a white cat is not a member of the category black cat*.

Likewise, two-dimensional rule use also occurs in this decisional processes stage. In this case, unlike the one-dimensional comparisons, both dimensions are combined and evaluated in a single step. The result of this one evaluation is then used to produce a response (refer back to Chapter 2 for examples of two-dimensional rules).

Having defined *representation integration* and *decisional integration*, it is now more straightforward to analyze participants who perceived rectangles varying in height and width as possessing a single unitary dimension of area (or similar emergent dimension), rather than possessing the two dimensions of height and width. For these participants, the features of the stimuli are combined by representational processes and a single feature is then passed forward to the decisional processes. The decisional processes then evaluates this single dimension according to a one-dimensional rule. The output is used to produce a category response. For these participants, the integration of these dimensions occurs in the representation phase, thus these participants are performing representational integration. Although these participants are behaving as if they were

using two-dimensional rules, they are actually using a one-dimensional rule.

Interpreting the data provided by the participants in Experiment 3, those participants given integral stimuli, as being the result of representational integration supports Ashby et al.'s account. However, it is not straightforward to characterize all the participants as using representation integration or decisional integration. The composition of participant strategies in Experiment 3 is unknown.

The difficulty lies in establishing which participants are using an area representation with a one-dimensional rule and which are using a height and width representation with a two-dimensional rule. The participants in Cluster 5 of Experiment 3, those using two-dimensional boundaries to solve one-dimensional category structures, provide evidence that some participants are using an area representation. However, this is not sufficient to declare that all participants in Experiment 3 used representational integration. It may be the case that only the participants classified into Cluster 5 used representational integration.

Support for this counter claim comes from the participants classified into Cluster 1 in the one-dimensional conditions. These are the participants who were given one-dimensional categories to learn and classified transfer items as members of the rule consistent category. The transfer phase behaviors of these participants are similar to other participants classified into Cluster 1 from the other experiments. If all participants in Experiment 3 had used representational integration, then the participants in Cluster 1 from the one-dimensional conditions would have behaved more like those participants in either Clusters 2 (exception-generalizers) or 5 (those using a two-dimensional boundary in a one-dimensional structure). Since this was not the case, it is unlikely that all participants in Experiment 3 were using representational integration.

Regardless of the evidence provided by these indirect measures of participant strategies, there were no definitive tests applied in these experiments. It is therefore, inappropriate to make conclusions based upon any one particular group composition. The participants in Experiment 3 may have used either representational integration or decisional integration.

However, it is possible to separate these groups of participants. For example, one method to distinguish between participants using representational integration or true two-dimensional rule use is to evaluate their use of dimensional attention. If participants formed a representation along the dimension of area, they may have done so in the representational processing stage by combining the dimensions of height and width. If these dimensions are combined at this stage of processing, participants may no longer have direct access to the underlying dimensions. This is akin to how information about the saturation level of a stimulus is lost after the creation of a color percept. If

participants lack access to these dimensions, then they may be unable to directly apply dimensional attention to them.

This argument parallels the argument proposed by Nosofsky (1986), that participants are only allowed to make dimensional adjustments along single dimensions. Having lost access to these dimensions, the participants would be forced to apply dimensional attention only to their emergent area dimension. This would manifest as a strong influence to stimuli that are diagonally adjacent to stimuli in psychological space rather than stimuli that are horizontally or vertically adjacent in psychological space. Examples of this type of influence can be seen in the behaviors of ALCOVE and ATRIUM when these models were fit to the data of participants in the two-dimensional conditions (e.g., see Figure 5.4).

However, if participants do not lose direct access to the dimensions used to form new emergent dimensions, then this dimensional bias will not be displayed. In this case, participants may retain the ability to allocate attention to the underlying dimensions of height and width, even though they use a one-dimensional area rule. This suggests that participants have transformed a two-dimensional category (height and width) into a three-dimensional category (height, width, and area). However, it is unlikely that this is the case. If participants did have access to all three dimensions, then it would be more likely that participants would apply a one-dimensional rule to solve one-dimensional category structures instead of using a less accurate one-dimensional rule on their composite dimension, as displayed by the participants in Cluster 5 of Experiment 3.

Two possible methods for testing which representation participants were using are to directly test participant representations and to perform mathematical modeling of their categorization behaviors. To directly test participant representations, a filtration and condensation task may be used (Kruschke, 1993). These tasks can be used to distinguish between participants using decisional integration and those using representation integration. If participants are given a filtration and condensation after performing the categorization experiments used in this dissertation, the nature of their representation can be verified.

The second method for testing participant representations is to perform mathematical modeling of their categorization behaviors. This would be a further extension of the modeling performed in this paper. An additional set of models would be tested against those currently being tested. These models would consist of versions of each of the current models. The version of ALCOVE would have the dimensions and exemplars rotated to align with the representation integrated dimension of area. The version of ATRIUM would possess the rotated version of ALCOVE and one-dimensional rules aligned with the new dimensions. The version of ATRIUM-DR would like-

wise possess the rotated version of ALCOVE and two-dimensional rules aligned as one-dimensional rules with the original dimensions.

6.2 Dimensional Attention with Two-Dimensional Rules

One of the difficulties in modeling the data was how to interpret dimensional attention (i.e., the dimension weights) in ATRIUM-DR. In ATRIUM, rules are one-dimensional boundaries that are formed along the same dimensions as those used by the exemplar module. In ATRIUM-DR, rules are two-dimensional boundaries that are formed along a new dimension and is a combination of the underlying dimensions. This new dimension is not directly represented in the exemplar module. It may be argued that dimensional weights for attention should include this new composite, either in addition to the two underlying dimensions or as a replacement for those dimensions.

I would argue that this is inappropriate. In ATRIUM (and ATRIUM-DR) attention is initially set by the experimenter (in the simulations, the attention weights were set equal to each other). Any changes that occur in the attention weights are caused by the exemplar module. The rule modules do not directly affect attention weights. Likewise, the attention weights do not directly affect the rule modules. In ATRIUM, the attention weights affect the activation of the exemplar nodes, which in turn affect the activation of the gating nodes, which in turn affect the contributions of the rule modules to the output node activations. However, rule use does not directly influence dimensional attention and vice versa. If rules and dimensional attention do not directly affect each other, a dimension available only to the rule system should not affect the exemplar system. Following this reasoning, dimensional attention in ATRIUM-DR was set to be only available to the same dimensions used by the exemplar-modules.

In contrast to participants who used two-dimensional rules, participants performing representational integration *would* most likely represent stimuli along their new composite dimension. While discussed more thoroughly earlier in this chapter, the conclusion is that these participants are mostly likely representing stimuli along their new composite dimension and may or may not retain direct access to the underlying dimensions. However, these participants were not the focus of the modeling efforts, and were not specifically modeled for this dissertation.

6.3 The Continuum of Separability and Integrality

While dimensions have been previously discussed as being either separable or integral, this is a oversimplification. Previous research has found (e.g., Foard & Kemler Nelson, 1984; Garner, 1974; Lockhead, 1972; J. D. Smith & Kemler Nelson, 1984) that separability and integrality are better represented as continuum rather than a dichotomy. This issue was not discussed earlier because it was incidental to the questions raised about two-dimensional rule use. However, the stimuli were designed to account for this perspective. They were chosen because they exist on the extremes of the separable-integral continuum, and were therefore also representative of the dichotomous scale.

In life however, the distinctions between separability and integrality is less defined. Previous research has found that the likelihood of processing stimuli in an integral or separable fashion can be manipulated by a number of factors (e.g., J. D. Smith & Kemler Nelson, 1984; Foard & Kemler Nelson, 1984; Garner & Felfoldy, 1970). Relevant factors include stimulus factors, task factors, and subject factors.

Evidence that stimulus factors can influence the perception of separability and integrality comes from Foard and Kemler Nelson (1984). They found that when participants are given stimuli that vary in the magnitude of stimulus differences, performance shifts to reflect the amount of similarity. When a category contains stimuli that possess features that are close together along their dimensions, the dimensions are more likely to be treated as being integral. Whereas, when a category contains stimuli that possess features that are farther apart along their dimensions, the dimensions are more likely to be treated as being separable.

Foard and Kemler Nelson (1984) also found evidence for task factors influencing the perception of separability and integrality. When participants were given instructions that emphasized analytical processing, participants were more likely to treat dimensions as being separable. Likewise, when participants were instructed to make more holistic judgments, they were more likely to treat dimensions as being integral. Additionally, when participants are placed under time constraints, they are more likely to treat stimuli as having integral dimensions (e.g., J. D. Smith & Kemler Nelson, 1984).

Garner and Felfoldy (1970) also demonstrated that under some circumstances stimuli demonstrate asymmetrical dimensional relationships. Garner and Felfoldy (1970) presented participants with card sorting tasks testing the the effect of a secondary dimension on a one-dimensional sorting task. The stimuli were constructed from two dichotomous dimensions. The correct classifi-

cation of a card was determined by the primary dimension. The secondary dimension varied based on condition and was either a fixed value, correlated with the primary dimension, or orthogonal to the primary dimension. The type of secondary dimension and relationship of the primary dimension to the secondary dimension was found to help, hinder, or not affect the accuracy of the card sorting task.

With stimuli composed of circles with radial line segments that varied in the size of the circle and the angle of the radial line segment. Garner and Felfoldy (1970) found that when the primary dimension was size, participant behaviors were not affected by secondary dimension of the orientation of the line segment. This suggests that size was separable from line segment orientation and that the participants did not benefit from the redundancy when categorizing the second type of stimuli. However, when the primary dimension was the angle of orientation of the radial line segment, participant behaviors were affected by the secondary dimension of size. When participants were given the stimuli with a secondary dimension correlated to the primary dimension, they were more accurate than when they were given the stimuli that possessed a fixed secondary dimension or when the secondary dimension was orthogonal to the primary dimension. Garner and Felfoldy (1970) further found that they could manipulate this relationship by changing the relative salience of the dimensions. This reversal and the malleability of dimensional relationships demonstrates their asymmetrical properties.

Evidence for subject factors influencing the perception of dimensional relationships is supported by studies that examine how strategies develop over time and studies that compare children to adults. People have been found to change how they represent stimuli as they gain experience. When participants are allowed to practice a task, the dimensions may shift in their degree of separability versus integrality over time (Foard & Kemler Nelson, 1984). This shift tends to encourage analytical analysis and perception of the stimuli as possessing features along separable dimensions. Similar effects arise when comparing the categorization behaviors of children and adults (J. D. Smith & Kemler Nelson, 1984). Children are found to be guided by holistic or overall similarity more than they are by analytical processes. This results in them displaying behaviors consistent with interpreting dimensions as being integral. In contrast, adults show the opposite pattern, they tend to use analytical processes more than similarity based processes, which is consistent with interpreting dimensions as being separable. Further research has demonstrated that these findings may be driven by cultural attitudes. When comparing adults educated in Western European cultures versus those educated in East Asian cultures, East Asian educated adults are more likely to use intuition and less likely to perform analytically (Norenzayan, Smith, Kim, & Nisbett, 2002). Taken together, these

findings suggest that expertise and task difficulty influence how dimensions are represented.

The perspective that separability and integrality are better interpreted as a continuum dependent upon stimuli, task, and subject factors, is consistent with the findings of this dissertation. In the previous experiments, separability and integrality were not found to predict two-dimensional rule use. Participants displayed two-dimensional rule use when categorizing both integral and separable dimensions, hence to some degree, integral and separable dimensions were treated the same by participants.

This is not to say that dimensional relationships had no effect on participant behaviors. Dimensional relationships caused a dramatic behavioral difference in participants. This difference is demonstrated by the behaviors of participants in Cluster 5 of Experiment 3, those classified as using an emergent dimension, such as area, to classify the stimuli. The effect of separability and integrality may be best described as allowing some participants the opportunity to use different representations. By representing the stimuli differently, these participants were able to solve a different categorization problem, an easier one. To be more specific, because the stimuli were composed of integral dimensions, some participants were able to represent the stimuli along the area dimension. This allowed those participants to reduce the task of learning a two-dimensional rule to the task of learning a one-dimensional rule.

6.4 Two-Dimensional Rule Commensurability

Participants were found to use two-dimensional rules when given stimuli with commensurate dimensions, but not when given stimuli with noncommensurate dimensions. This suggests that participants require commensurate dimensions to use two-dimensional rules. It further suggests that rule use may be a function of hypothesis testing and thus requires executive functions. This supports the requirements for two-dimensional rule use provided by Ashby et al. (1998).

6.5 Extensions to the Current Research

In accounting for the data, the models performed well. However, there are extensions to the modeling and to the experimental design, that would provide additional evidence for the nature of two-dimensional rule use. As previously discussed, more in depth modeling of the participants' behaviors may be appropriate. Testing the data with models that can distinguish between two-dimensional rule use and representational integration would be especially useful. Likewise adding

filtration and condensation tasks after the primary categorization task would provide information towards the representations used by participants.

Another issue previously discussed is that ATRIUM (and ATRIUM-DR) may be able to account for behavior better with a simplified rule representation. These models represent rules as boundaries in psychological space. The version of ATRIUM used in this dissertation uses a double thermometer encoding in the rule modules. This type of rule-module is useful for learning categories structures where the different categories are organized into bands along a dimension (e.g., ABAB). ATRIUM has difficulty in learning not to use all of the modules and representations available to it. When participants perform in ways that are not optimal, as displayed in the previous experiments, ATRIUM has difficulty in matching these behaviors. Simpler versions of ATRIUM, ones that incorporate only a single rule module or do not possess double thermometer encoding rule modules, may be more successful in displaying the suboptimal behaviors displayed by participants.

An additional area to address is how participants were organized into groups by strategy. In this dissertation, groups were created using the PAM cluster analysis technique and based on participant transfer phase performance. Another possible method would be to fit participants individually with the models and then perform a cluster analysis on the free parameters used by the participants. This would group participants together based upon how the models would describe them. This type of clustering may better represent the strategies employed by the participants. This technique maximizes the information from individual participants, but strategies may be distorted due to the large amount of random error.

Another potential method of clustering, would be to develop a series of different prototypical strategy performance profiles. These performance profiles would represent the strategies potentially used by participants to do the task. These profiles would include strategies such as rule-generalization, exception-generalization, random-guessing, and multiple one-dimensional rule use. These performance profiles can then be used as the seeds to form a cluster analysis. This method would sort participants into groups based upon established strategies. However, it would also reduce the emergence of any unexpected strategies. If participants did display any strategies not contained within the original performance profiles, they might be evident from this type of analysis.

6.6 Two-Dimensional Rules and Knowledge Partitioning

Yang and Lewandowsky (2004) explored knowledge partitioning in category learning and found evidence for two-dimensional rule use. In testing categories formed with diagonal bound-

aries, they found that a version of ATRIUM with two-dimensional rules was better able to model the results than ALCOVE. This was followed by Lewandowsky et al. (2006) demonstrating that knowledge partitioning was possible with stimuli that varied in the type of relationship between the dimensions (integral or separable) and the commensurability of the dimensions. Their conditions were equivalent to the conditions used by the experiments in this dissertation. The results of my experiments provide three extensions to these works.

The first extension relates to when knowledge partitioning occurs. Lewandowsky et al. (2006) found knowledge partitioning in all conditions except the easiest condition, where stimuli were composed of rectangles that varied in height and width (i.e, the dimensions were integral and commensurate). This condition corresponds to the present Experiment 3, which contains participants displaying representational integration. My results suggest that this condition may have been the easiest because participants were using representational integration. This would allow participants to use a one-dimensional rule to learn Lewandowsky and Yang's (2004) category structure, whereas in the other conditions, participants still need to process a two-dimensional category boundary.

The second extension relates to how ATRIUM performs knowledge partitioning. Yang and Lewandowsky (2004) found that a version of ATRIUM using diagonal rules was better able to fit their data than ALCOVE. They suggested this was due to different rule modules being able to learn the different parts of the knowledge partitioning task, in contrast to ALCOVE which possesses only one exemplar module. However, they further allowed that it was ATRIUM's multiple module systems that allowed the model to succeed in this task, and that this success may not be due to the particular type of module (rule or exemplar).

The results from the current experiments support the idea of the different modules being necessary and not the specific type of module. In Experiment 1 (stimuli with commensurate and separable dimensions), participants were able to use two-dimensional rules. This suggests that participants may use different rules to knowledge partition in tasks with stimuli composed of commensurate and separable dimensions. In contrast, in Experiments 2 and 4 (stimuli with noncommensurate and either separable or integral dimensions), participants were not able to use two-dimensional rules. This suggests that participants may use different exemplar modules to knowledge partition in tasks with stimuli composed of noncommensurate of dimensions. Thus evidence was found to support knowledge partitioning relying upon different modules and not a specific type of module.

The third extensions relates to the differences in the modeling. When modeling their results, Yang and Lewandowsky (2004) used a version of ATRIUM with one rule per module, whereas

the simulations in this dissertation used a version of ATRIUM with several rules per module. While previously discussed, it may have been the case that ATRIUM attempted to use the multiple rules in a way that caused the model to perform less like the participants. Thus, modeling the data from my experiments with a version of ATRIUM using rule-modules with a single rule may demonstrate two-dimensional rule use. Likewise, modeling the data from Yang and Lewandowsky's experiments with a version of ATRIUM using multiple rules per module may demonstrate different ways to account for knowledge partitioning.

6.7 Summary

This dissertation found evidence for two-dimensional rule use in category learning. Participants were able to use two-dimensional rules when learning categories that contained stimuli with commensurate dimensions. However, evidence was not found for two-dimensional rule use when the categories contained stimuli with noncommensurate dimensions. These findings support the requirements for two-dimension rule use proposed by Ashby et al. (1998). These findings also challenge current theories of rule-based categorization which state that rule-based category representations operate on single features. The experiments within this dissertation found evidence for rules operating on multiple features that create a two-dimensional boundary. Furthermore, when modeling categories with two-dimensional boundaries with mathematical models such as ATRIUM, the inclusion of two-dimensional rules in the models should be considered.

References

- Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General*, *120*, 3–19.
- Ashby, F. G., Alfonso-Reese, L., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442–481.
- Ashby, F. G., Ell, S., & Waldron, E. (2003). Procedural learning in perceptual categorization. *Memory and Cognition*, *31*(7), 1114–1125.
- Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Ashby, F. G., & Maddox, T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 598–612.
- Ashby, F. G., & Maddox, W. T. (1998). Measurement, judgement, and decision making. handbook of perception and cognition (2nd ed.). In M. H. Birnbaum (Ed.), (pp. 251–301). San Diego, CA, US: Academic Press.
- Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.
- Ashby, F. G., Noble, S., Vincent, F. J., Waldron, E. M., & Ell, S. W. (2003). Category learning deficits in parkinson's disease. *Neuropsychology*, *17*, 115–124.
- Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.
- Cortese, J. M., & Dyre, B. P. (1996). Perceptual similarity of shapes generated from fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 133–143.

- Dunn, J. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 9(2), 242–257.
- Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, 36(4), 749–761.
- Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, 127, 107–140.
- Erickson, M. A., & Kruschke, J. K. (2002). Rule-based extrapolation in perceptual categorization. *Psychonomic Bulletin and Review*, 9, 160–168.
- Foard, C., & Kemler Nelson, D. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General*, 113(1), 94–111.
- Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*.
- Garner, W. R. (1974). *The processing of information and structure*. Hillsdale, NJ: Erlbaum.
- Garner, W. R. (1977). The effect of absolute size on the separability of the dimensions of size and brightness. *Bulletin of the Psychonomic Society*, 9(5), 380–382.
- Garner, W. R., & Felfoldy, G. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1, 225–241.
- Janowsky, J. S., Shimamura, A. P., Kritchevsky, M., & Squire, L. R. (1989). Cognitive impairment following frontal lobe damage and its relevance to human amnesia. *Behavioral Neuroscience*, 103, 548–560.
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Hoboken, NJ: Wiley-Interscience.
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44.
- Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science*, 5, 3–36.
- Lee, M., & Webb, M. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin and Review*, 12(4), 605–621.
- Leng, N. R., & Parkin, A. J. (1988). Double dissociation of frontal dysfunction in organic amnesia. *British Journal of Clinical Psychology*, 27, 359–362.
- Lewandowsky, S., Roberts, L., & Yang, L. (2006). Knowledge partitioning in categorization: Boundary conditions. *Cognition*.
- Lockhead, G. (1972). Processing dimensional stimuli: A note. *Psychology Review*, 79(5), 410–419.

- Maddox, W. T. (1992). Multidimensional models of perception and cognition. In F. G. Ashby (Ed.), (pp. 147–180). Hillsdale, NJ: Erlbaum.
- Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 650–662.
- Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225–252.
- McClelland, J. L. (1979). On the time-relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.
- Monahan, J., & Lockhead, G. (1977). Identification of integral stimuli. *Journal of Experimental Psychology: General*, *106*, 94–110.
- Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive Science*, *26*, 653–684.
- Nosofsky, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.
- Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 282–304.
- Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin and Review*, *3*, 222–226.
- Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.
- Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin and Review*, *5*, 345–369.
- Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, *101*, 53–79.
- Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General*, *132*, 491–511.
- Posner, M. I., & Keele, S. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, *83*(2), 304–308.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*(77), 353–363.

- Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Shaw, M. L. (1982). Attending to multiple sources of information: I. The integration of information in decision making. *Cognitive Psychology*, 14, 353–409.
- 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. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, 1, 54–87.
- Shepard, R. N., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology*, 65, 94–102.
- Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75(13). (Whole No. 517)
- Shwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464.
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- Smith, J. D., & Kemler Nelson, D. (1984). Overall similarity in adult's classification: The child in all of us. *Journal of Experimental Psychology: General*, 113(1), 137–159.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin and Review*, 8(1), 168–176.
- Yang, L., & Lewandowsky, S. (2004). Knowledge partitioning in categorization: Constraints on exemplar models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1045–1064.