A STUDY OF RULE-BASED CATEGORIZATION WITH REDUNDANCY

by

Farzin Shamloo

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Psychological Sciences

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Sebastien Hélie, Chair
Department of Psychological Sciences
Dr. Richard Schweickert
Department of Psychological Sciences
Dr. Gregory S. Francis
Department of Psychological Sciences
Dr. F. Gregory Ashby
Department of Psychological & Brain Sciences,
University of California, Santa Barbara

Approved by:

Dr. David Rollock
Head of the Graduate Program

Dedicated to my grandparents

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Shamloo, Farzin. PhD
Institution: Purdue University
Degree Received: May 2019
Title: A Study of Rule-Based Categorization With Redundancy
Committee Chair: Sebastien Hélie

In tasks with more than one path to succeed, it is possible that participants' strategies vary and therefore, participants should not be analyzed as a homogeneous group. This thesis investigates individual differences in a two-dimensional categorization task with redundancy (i.e., a task where any of the two dimensions by itself suffices for perfect performance). Individual differences in learned knowledge and used knowledge are considered and studied. Participants first performed a categorization task with redundancy (training phase), and afterward were asked to do categorizations in which the previously redundant knowledge becomes decisive (testing phase). Using the data from the testing phase, dimension(s) *learned* by each participant were determined and the response patterns of each participant in the training phase was used to determine which dimension(s) were *used*. The used knowledge was assessed using two separate analyses, both of which look at accuracy and response time patterns, but in different ways. Analysis 1 uses iterative decision bound modeling and RT-distance hypothesis and Analysis 2 uses the stochastic version of general recognition theory. In Analysis 1, more errors and slower response times close to a decision bound perpendicular to a dimension indicate that a participant is using that dimension. Analysis 2 goes a step further and in addition to determining which dimension(s) are used, specifies in what way they were used (i.e., identifying the strategy of each participant). Possible strategies are described heuristically (unidimensional, time efficient and conservative) and then each heuristic is translated into a drift diffusion model by the unique way that strategy is assumed to affect trial-by-trial difficulty of the task. Finally, a model selection criterion is used to pick the strategy that is used by each participant.

INTRODUCTION

We are constantly categorizing things that we encounter in everyday life. For example, we categorize people we see in the street (stranger, friend), food we eat (healthy, unhealthy), etc. Categorization can be described as the process of assessing features of an encountered stimulus and labeling the stimulus based on previous observations (Kruschke, 2005; Murphy & Medin, 1985; Wisniewski & Medin, 1991). A feature can be defined as the perceived (or measured) value of a specific dimension in the dimensional space of the encountered stimulus and the dimensional space can be defined as an organizing principle that structures perception in a consistent way (Burns & Shepp, 1988). The encountered stimuli are comprised of numerous dimensions but in many cases, there is redundancy of information and therefore, learning the differences between a subset of dimensions suffice to successfully categorize objects. An example is categorizing an apple as a Gala or Fuji when grocery shopping. Even though the color by itself is enough to categorize them, there are individuals who can tell subtler differences in shape and texture between the two apples. The efficient strategy from a computational point of view might be to selectively pay attention only to the color of the apple and not 'waste' attentional resources on other dimensions that are not going to increase the performance accuracy. On the other hand, acquiring knowledge on other dimensions might become useful in future tasks. For example, in distinguishing between a Gala and a Jonagold apple, those individuals who acquired knowledge on shape of the Gala apple in the previous task will have the upper hand in this new task. Therefore, both strategies have their advantages and disadvantages: The first strategy is more efficient for the

current task, whereas the second strategy is better prepared for new situations where knowledge of previously redundant information becomes decisive.

While some previous categorization studies have focused on the effect of task instructions and learning process on knowledge acquisition (e.g., Ell, Smith, Peralta, & Hélie, 2017; Hélie, Shamloo & Ell, 2017; Levering & Kurtz, 2015), in this thesis the focus is on the individual differences (IDs).

Why Studying Individual Differences Matters?

In an interview given in 1974 (Skinner, 2014), Jacques Lacan said, “Let’s get rid of this average Joe, who does not exist. He is a statistical fiction. There are individuals, and that is all.” Even though Lacan was talking about psychoanalysis and the uniqueness of the anxieties of each individual, this issue is relevant to cognitive psychology as well. In fact, Levinson (2012) makes a very similar point about cognitive psychology: “The cognitive science revolution was based on a fundamental idealization, the myth of “the human mind””. Levinson (2012) argues that the injunction to find universal cognitive characteristics has resulted in underestimating the importance of IDs, which can prevent researchers from understanding the underlying mechanisms of cognitive abilities. For example, Kidd, Donnelly & Christiansen (2017) showed that focusing on IDs highlights how experience can affect language acquisition and language architectures, which can be used to assess psycholinguistic theories. There are various ways to improve the understanding of cognitive processes by studying IDs, some examples in the categorization literature are discussed in the next section.

Individual Differences in the Categorization Literature

There are numerous studies that have attempted to find the relation between domain general cognitive constructs such as working memory capacity (WMC) and categorization performance (e.g., DeCaro, Thomas, & Beilock, 2008; Erickson, 2008; Tharp & Pickering 2009). Due to the contradictory results of such studies, Lewandowski (2011) studied the relation between WMC and categorization in six categorization tasks (type I - type VI) developed by Shepard et al. (1961). The stimuli used in the study by Lewandowski (2011) were shapes (square or circle) that varied in color (unfilled or red) and size (small or large). Participants were asked to learn to associate each stimulus to one of two categories using trial-and-error learning. The six different conditions (i.e., six different ways of labeling the stimuli) corresponded to some of the most important categorization tasks (rule-based, information integration and unstructured; Ashby & O'Brien, 2005). The results in Lewandowski (2011) showed a strong relation between WMC and categorization accuracy in all of these tasks. Moreover, the study looked at how the IDs were manifested in a computational model of categorization (ALCOVE; Kruschke, 1992) and concluded that the IDs in all the different categorization tasks were linked to a single parameter of the model representing learning speed. Variations in this parameter were shown to be captured by a single latent variable, which was associated with WMC. As a result, Lewandowsky concluded that working memory mediates various types of category learning.

In a subsequent study by Craig & Lewandowski (2012), the relation between WMC and categorization was studied from a different perspective. Craig & Lewandowski (2012) studied IDs in tasks where different strategies could be used for

successful categorization, and whether WMC could predict the strategy selected by the participants. The study included two experiments: the 5-4 task (Medin & Schaffer, 1978; Smith & Minda, 2000) and the correlated cues task (Medin et al., 1982). In both of the experiments, the stimuli were four-dimensional with binary features and therefore, there were 16 unique stimuli. Each experiment was divided into training and transfer phases. During the training phase, participants were shown a subset of stimuli and received feedback. In the transfer phase, participants were shown the remaining stimuli (those that were not shown in the training phase) and were asked to categorize them without receiving feedback. In the 5-4 task, relative success in the training phase was possible using any of these strategies: a unidimensional rule on dimension 1, a unidimensional rule on dimension3, or an exemplar-based strategy (i.e., categorizing based on the overall similarity of the stimulus to the previously encountered stimuli in training phase). The data from the transfer phase made it possible to identify the strategy of each participant in training. Similarly, in the correlated cues task, relative success was possible by using three different rule-based strategies, two of them were unidimensional rules and the other was a correlated cue rule (two categories could be perfectly separated by observing the correlation between the third and fourth dimensions). The results show that even though WMC predicted the categorization performance (similar to the result from Lewandowski, 2011), it could not predict the strategy selected by the participants. However, one limitation of this experiment is that the unidimensional rules in both experiments were suboptimal compared to the exemplar and correlated cue strategies. This could explain some of the obtained results.

McDaniel et al. (2014) compared IDs in the context of a function learning paradigm (DeLosh, Busemeyer, & McDaniel, 1997). Unlike Craig and Lewandowski (2012), performance in the training phase of this study did not depend on the choice of strategy. During the training phase (interpolation trials), participants learned the association between two variables (stimulus and response) with a V shaped relation. In the test phase (extrapolation trials), participants were asked to respond to stimuli with values outside of the range of the training trials. Participants who abstracted rules extrapolated based on either a V shaped or a sinusoidal function (VVV shaped). In contrast, exemplar based learners extrapolation was close to the responses associated with the extreme values of the training phase ($\overline{\overline{\text{V}}}$ shaped). The results showed that WMC associates with a tendency to use a rule-based approach. Whether or not WMC predicts a tendency toward using rule-based strategies can be helpful in assessing the assumption of computational models of categorization. For example, COVIS (Ashby, Alfonso-Reese, & Waldron, 1998) is a multiple system model of categorization learning that postulates that there are at least two systems for category learning: A hypothesis-testing system that can learn rule-based tasks and a procedural-learning system that can learn categories that are not easily verbalizable. Since COVIS assumes that WMC plays an important role in hypothesis-testing but not in procedural learning, the conclusions from Craig & Lewandowski (2012) seem to be different from COVIS' predictions but the conclusions from McDaniel et al. (2014) are in line with predictions from COVIS.

Minda, Desroches and Church (2008) studied another aspect of COVIS by looking at IDs in category learning between children and adults. COVIS assumes that the hypothesis-testing system includes prefrontal cortex while the procedural learning system

relies on subcortical structures (Hélie, Roeder, & Ashby, 2010; Waldschmidt & Ashby, 2011). Since prefrontal cortex's development occurs later than other brain regions (Bunge & Zelazo, 2006), COVIS predicts that adults should have a better performance in rule-based categorization compared to children. More specifically, children should perform like adults in tasks that rely on procedural learning, but might have difficulty in rule-based tasks, especially in cases where the rules are complex. Minda et al. (2008) tested this prediction using the category structures introduced by Shepard et al. (1961). They showed that in accordance with the assumptions of COVIS, young children performed similarly to adults in tasks that relied on procedural learning and in task with unidimensional rules, but did worse than adults did when the categories were separated by a disjunctive rule.

Individual Differences in This Thesis

Similar to Craig and Lewandowski (2012), the goal of this thesis is to study individual differences of strategies. However, the experiment of this thesis is not designed to distinguish between rule-based and exemplar-based strategies and it is not an attempt to relate an external factor (such as WMC) to the strategies used by the participants. What it aims at is to link the response patterns of participants to their strategy using computational modeling. IDs are studied in a categorization task with two-dimensional stimuli, where both dimensions are diagnostic and knowledge on any one of them suffices for perfect accuracy. Since there is redundancy of information and using any of the two dimensions by itself results in perfect accuracy, there are three general possibilities for the strategy of each participant: Using Dimension 1, using Dimension 2 and using both dimensions. Similar to Craig and Lewandowski (2012), the experiment in

this study has a test phase. Performance of a participant in the test phase shows whether s/he learned Dimension 1, Dimension 2, or both dimensions. Knowing which dimension(s) were learned by each participant allows to partially validate the methods used to identify the strategy of participants when there was redundancy in the task (more details are discussed in the methods section where the experiment is explained).

Determining which dimensions were used by each participant is done in two different ways (Analysis 1 and Analysis 2). The independent variables in both analyses are accuracy and response time patterns of participants, which are used in different ways. Below, there is a short description of each of them, but more details will be provided in the subsequent sections.

Analysis 1: The static version of general recognition theory (Ashby & Townsend, 1986) and RT-distance hypothesis (Ashby & Maddox, 1991; Ashby & Maddox, 1994) are used to look at accuracy and RT patterns respectively. Accuracy patterns are taken into account by looking at the location of errors (i.e., where in the category space is the participant making errors) and RT patterns are taken into account (broadly speaking) by looking at the locations in the category space where the response times are slower.

Analysis 2: The dynamic version of the general recognition theory (Ashby, 2000) is used to identify the dimension(s) used by each participant. Dynamic GRT is a drift diffusion model (DDM) applied to categorization tasks of the type that are traditionally analyzed by static GRT and RT-distance hypothesis. Accuracy and RT are taken into account with a single structure that imposes certain RT distributions on correct and incorrect responses.

In the remainder of the introduction, relevance and diagnosticity of a dimension are defined, distinction between learned and used knowledge is discussed, and some of the theoretical tools used to study the two set goals of the thesis are overviewed: General recognition theory (GRT), RT-distance hypothesis, and dynamic GRT.

Relevance and Diagnosticity

In a specific categorization structure, a dimension can be either relevant or irrelevant. Figure 1a shows an example of a categorization structure with four categories, where only dimension 1 is relevant and Figure 1b and 1c are two examples of categorization structures where both dimensions are relevant. In other words, in a category space a dimension is relevant when knowledge about its value for a stimulus adds to the information about the label of that stimulus. To put it more formally, in a multidimensional space with features $[x_1, x_2, \ldots., x_p]$, and
K categories, the i^{th} feature is irrelevant if and only if:

$$P(stimulus \in category\ k \mid known\ x_i)\ is\ equal\ for\ all\ k \in K$$

It is possible to distinguish between relevance of a dimension in a category space and its diagnosticity in a specific categorization task. Often in categorization tasks it is not required to choose a label for an observed stimulus among all the possible labels. For example, consider the categorization structure shown in Figure 1b. It is clear that both dimensions are relevant. However, consider a trial where the participant is asked to distinguish between “A” and “B”. In this specific trial, knowledge on dimension1 is not needed to perform the task on hand and therefore, it is not diagnostic. To put it more

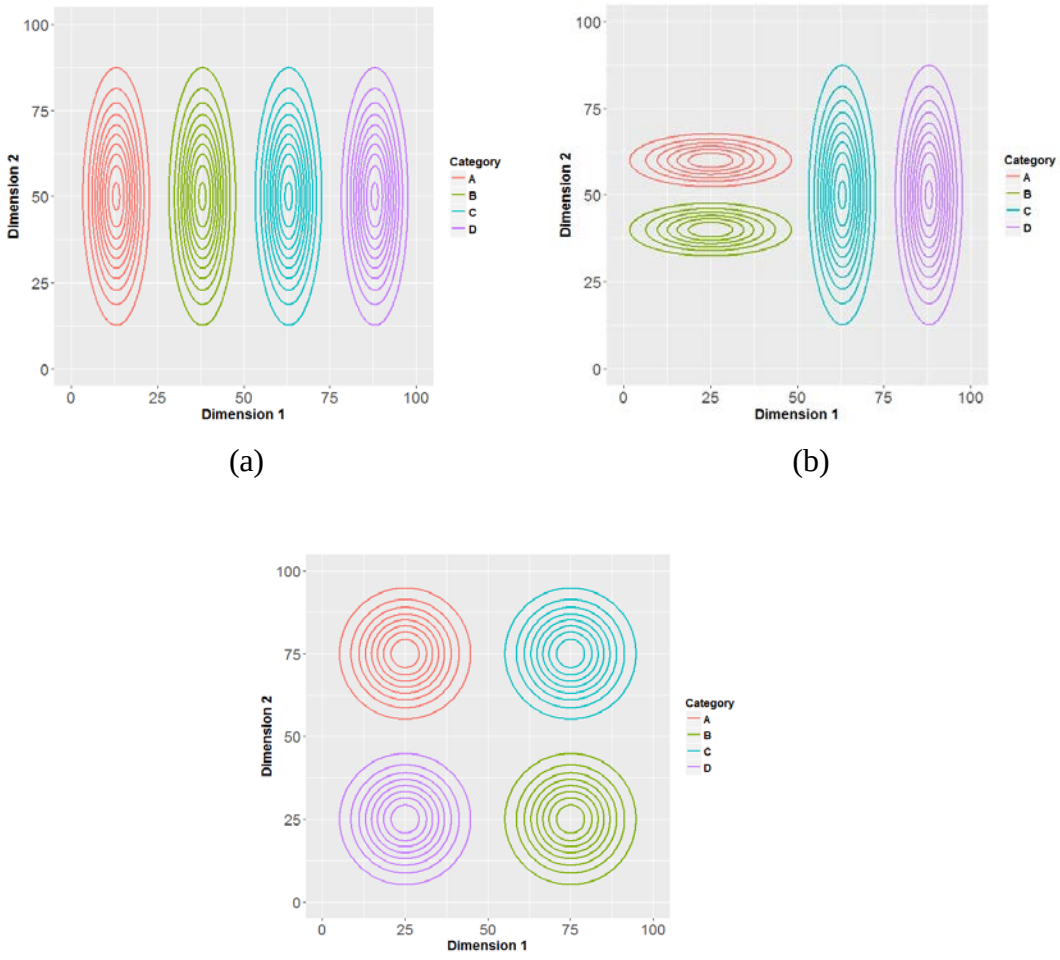


Figure 1. Likelihood contour of four categories where (a) dimension 1 is relevant, (b) dimension 2 is irrelevant, and (c) both dimensions are relevant.

formally, in a multidimensional space with features $[x_1, x_2, \ldots, x_p]$, and K categories, the i^{th} feature is non-diagnostic in a categorization task where an ideal participant is asked to choose a label from $H \subset K$ set if and only if:

$$P(stimulus \in category\ h \mid known\ x_i)\ is\ equal\ for\ all\ h \in H$$

One way to study knowledge acquisition on diagnostic and non-diagnostic dimensions is to separate the experiment into two different phases: A training phase where participants learn the categories by receiving feedback and a test phase where they do not receive any feedback. This structure (known as generalization criterion method; Busemeyer & Wang, 2000) opens up the possibility to study knowledge acquisition under different conditions. In this thesis, the experiment is a rule-based categorization task with redundancy in the training phase and a test phase that determines which dimensions each participant learned.

Figures 2 shows the category space corresponding to this task. The experiment is a two choice classification task: In each trial, a stimulus and a question appear on the screen and the question asks the participant to assign the stimulus to one of the two categories (e.g., “A or B?”). In Figure 2, the arrows indicate the categories that are compared against each other. As shown by the arrows in Figure 2, participants perform only two comparisons during the training phase (“A or B?” and “C or D?”) and for both comparisons knowledge of any of the two dimensions suffices for perfect accuracy. In the test phase, all possible pairs are compared with each other and there are trials that only Dimension1 is diagnostic (e.g., “A or C?” trials) and there are trials that only Dimension2 is diagnostic (e.g., “A or D?” trials). Therefore, it is possible to find out whether a

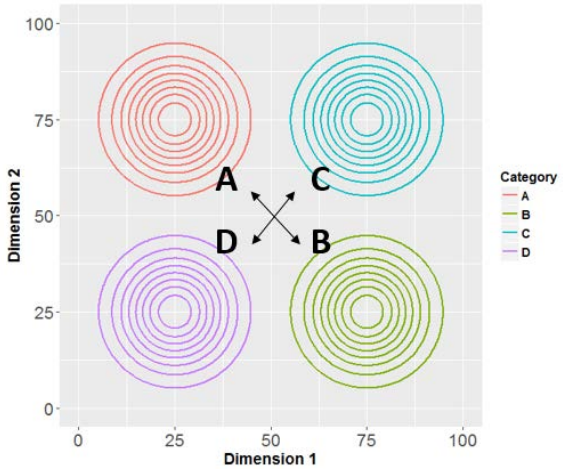


Figure 2. Category structure with two relevant dimensions. Black arrows show categorization tasks that participants have to do in the training phase. In test phase, participants have to perform categorization of all possible pairs of categories.

participant learned only one dimension or both dimensions. In real life categorization scenarios where objects are comprised of multiple dimensions, there is usually more than one diagnostic dimension, and therefore, understanding categorization mechanisms under such conditions is the focus of this thesis.

Learned Knowledge vs. Used Knowledge

A main goal of this thesis is to divide participants instead of analyzing their data as if they are part of a homogenous group. This division is going to be based on two indicators: the learned knowledge and the used knowledge. In the experiment shown in Figure 2, it is possible to learn and use only one dimension, use only one dimension but learn both, and learn and use both. Potential possibilities are shown in Table 1. Possibilities 1 and 2 are unidimensional strategies, possibilities 3 and 4 are also unidimensional, but the dimension that is not being used is being latently learned, and possibility 5 is a two-dimensional strategy.

Table 1

Possibilities for a Successful Participant in a Categorization Task That Corresponds to the Training Phase of Figure 2

	Learned Knowledge	User Knowledge
Possibility 1	Only Dimension 1	Only Dimension 1
Possibility 2	Only Dimension 2	Only Dimension 2
Possibility 3	Both dimensions	Only Dimension 1
Possibility 4	Both dimensions	Only Dimension 2
Possibility 5	Both dimensions	Both dimensions

The training/test structure mentioned in the previous section makes it possible to identify the dimension(s) that are learned, simply by looking at mean accuracy in the test phase (note that in the test phase, participants categorize all the possible pairs, not just "A or B?" and "C or D?"). More specifically, a high mean accuracy on "A or C?" and "B or D?" trials indicate that the participant learned Dimension 1, and a high mean accuracy on "A or D?" and "B or C?" trials indicate that the participant learned Dimension 2. The main interest however, is to find out what they use when there is redundancy (i.e., the training phase). Since all of the possibilities listed in Table 1 result in high accuracy in the training phase, looking at the mean accuracy does not suffice and more sophisticated tools are needed to identify which dimension(s) were used in the training phase (A brief description of the tools used in this thesis is provided in the next section).

Both Analysis 1 and Analysis 2 aim at identifying which dimensions were used by each participant in the training phase of an experiment with a structure similar to what is shown in Figure 2. Analysis 2 goes a step further and for participants that use both dimensions distinguishes between different two-dimensional strategies.

Methodological Tools

Some of the classic methods of the categorization literature are used to study the goals of this thesis. Decision bound models and RT-distance hypothesis look separately at accuracy patterns and response time respectively. These two methods are used in Analysis 1 and the stochastic version of the general recognition theory is used in Analysis 2.

Decision Bound Models

General recognition theory (Ashby & Townsend, 1986) is an extension of signal detection theory in multidimensional spaces that has been used in numerous contexts in the past 30 years (e.g., Ashby & Gott, 1988; Ashby & Perrin, 1988; Maddox, Ashby & Waldron, 2002). When GRT is used to model participants' response patterns in a categorization task, it is often called a decision bound model (Ashby & Soto, 2015). Decision bound models (DBM) assume that participants divide perceptual space using bounds and use these bounds to perform the categorization task.

For example in a categorization task with two categories (A and B), DBM models the percept of a p dimensional stimulus as $X = \left[x_1, x_2, \ldots., x_p\right]^T$ and postulates that a participant partitions the p dimensional space into two regions corresponding to the A and B categories. The set of all points that separate the two regions is called the decision bound and the probability of responding A or B is equal on the decision bound. The probability changes in favor of one of the two categories as the stimulus moves further away from the decision bound. Whether the probability changes in favor of A or B depends on the direction of the move. In order to formulate this characteristic, a discriminant function $y = h(X)$ (Ashby, 2000) is defined where:

$$y = h(X) > 0 \; for \; X \; that \; fall \; on \; the \; region \; corresponding \; to \; category \; A$$
$$y = h(X) = 0 \; for \; X \; that \; fall \; exactly \; on \; the \; decision \; boundary$$
$$y = h(X) < 0 \; for \; X \; that \; fall \; on \; the \; region \; corresponding \; to \; category \; B$$

Two sources of variability affect y: The variability in perception and the criterial noise. The percept (X) is often assumed to have a multivariate normal distribution with

mean μ and covariance structure of $\sum$ and criterial noise (ε) is modeled as a zero mean normal with variance σ^2. Assuming that the participant is using a linear bound (Ashby, 2000):

$$y = h(X) = bX + C + \varepsilon$$

Where b is a $1 \times p$ vector and C is a constant

Therefore:

$$P(responding\ A) = P(y > 0) = P\ (h(X) > 0) =$$
$$P(bX + C + \varepsilon > 0) = P(bX + \varepsilon > -C)$$

$$Assuming\ the\ percept\ (X)\ and\ criterial\ noise\ (\varepsilon)\ are\ independent:$$

$$bX + \varepsilon \sim N(b\mu\,, b\sum b' + \sigma^2)$$

$$P(bX + \varepsilon > -C) = P\left(\frac{bX + \varepsilon - b\mu}{\sqrt{b\sum b' + \sigma^2}} > \frac{-C - b\mu}{\sqrt{b\sum b' + \sigma^2}}\right)$$

$$= P\left(Z > \frac{-C - b\mu}{\sqrt{b\sum b' + \sigma^2}}\right)$$

Where Z is a standard normal random variable.

Therefore:

$$P(responding\ A) = 1 - P\left(Z < \frac{-C - b\mu}{\sqrt{b\sum b' + \sigma^2}}\right) = 1 - \Phi\left(\frac{-C - b\mu}{\sqrt{b\sum b' + \sigma^2}}\right)$$

$$= P(responding\ A) = \Phi\left(\frac{C + b\mu}{\sqrt{b\sum b' + \sigma^2}}\right)$$

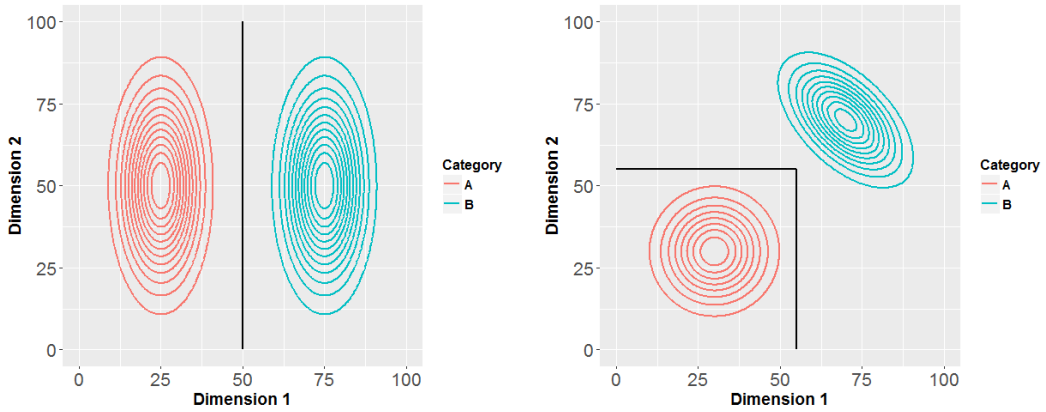
And:

$$P(responding\ B) = 1 - P(responding\ A) = \Phi\left(\frac{-C - b\mu}{\sqrt{b\sum b' + \sigma^2}}\right)$$

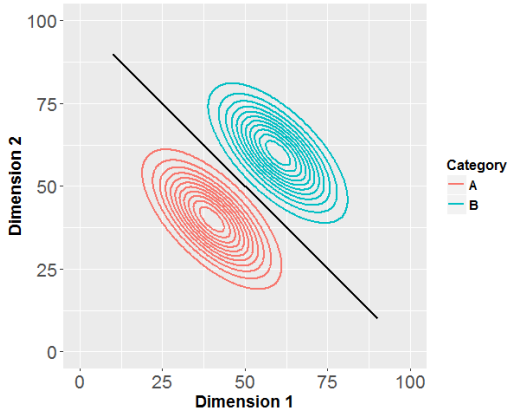
Where Φ is the CDF of a standard normal distribution.

Decision bounds can be classified into two types, those that can be described verbally and those that cannot be described verbally. Any decision bound that is comprised of a set of bounds perpendicular to the two dimensions can be described by a verbal rule. Learning that results in a verbally describable decision bound is called "explicit reasoning learning". Figure 3a and Figure 3b show two examples of such decision bounds. In contrast, Figure 3c shows an example of a decision bound that cannot be described verbally in a meaningful way (assuming dimension1 and dimension2 are non-commensurable). For example, if the two dimensions are length and orientation of a line, then an attempt to translate the bound would be something like "If length attribute of the line is bigger than its orientation attribute, it belongs to A, otherwise it belongs to B". This is meaningless because it is impossible to compare a length attribute to an orientation attribute. Therefore, the two dimensions have to be integrated before making a decision. This is true for any decision bound where at least part of it is not perpendicular to any of the two dimensions. Learning that results in a verbally indescribable decision bound is called "procedural learning" (Maddox & Ashby, 2004).

Note that the procedural learning strategy subsumes the explicit reasoning strategies (i.e., a bound perpendicular to one of the axes is a special case of a bound that can have any orientation). Hence, using any of the two learning mechanisms can potentially lead to perfect accuracy in a rule-based task and looking only at participants' assigned labels, it may be impossible to know which learning mechanism was used to categorize the stimuli. However, it is assumed that participants start by testing explicit reasoning strategies (Ashby et al., 1998), and if an explicit reasoning strategy (similar to the task shown in Figure 2) can result in perfect accuracy, it is reasonable to assume that



(a) (b)



(c)

Figure 3. Three examples of decision bounds: (a) unidimensional rule-based, (b) conjunctive rule-based, and (c) verbally indescribable decision bound.

nearly all participants use explicit reasoning strategies, either a unidimensional rule or a conjunctive rule. Iterative decision bound modeling (iDBM; Hélie, Turner, Crossley, Ell, & Ashby, 2017) is used to identify the bound that is used by participants in each trial to classify the stimulus and identify any possible changes in the bound that is being used by a participant.

Iterative Decision Bound Modeling

iDBM is a tool developed by Hélie et al. (2017b) to identify participants' strategy in a categorization experiment based on the way they label stimuli. The original version of iDBM considers three possible models: guessing models, explicit reasoning models and procedural learning models. Guessing models assume that participants randomly assign stimuli to categories, explicit reasoning models assume that the decision bound(s) being used are perpendicular to one of the dimensions and procedural learning models assume that participants use a bound that is not perpendicular to any of the dimensions, but it limits them to be linear. In each iteration, iDBM identifies the best fit for each of the mentioned models (guessing, explicit reasoning and linear procedural) using maximum likelihood (Ashby, 1992) and then compares them using Bayesian Information Criterion (BIC; Schwarz, 1978) and outputs the strategy used by each participant in each trial of the experiment (details can be found in Hélie et al., 2017b).

RT Distance Hypothesis

Decision bound models only take the accuracy patterns into account and response times are excluded from these models. In order to incorporate RT into decision bound models, Ashby and Maddox (1991, 1994) introduced the RT-distance hypothesis which is based on the previous empirical findings (Bornstein & Monroe, 1980; Cartwright, 1941)

that mean RT decreases as the distance between stimulus and the decision bound increases. RT-distance hypothesis states that the function relating RT and distance to bound is monotonically decreasing but it does not specify the exact relation between them. There are specific versions of RT-distance hypothesis that assume a special form for the decreasing function relating distance to bound and RT (Murdock 1985; Shepard, 1981), but the specific shape of the monotonically decreasing function is not the focus of this thesis.

Stochastic GRT

Drift diffusion models (DDM; Ratcliff, 1978; Ratcliff & McKoon, 2008) have been used to model two alternative forced choice tasks in many different research areas (e.g., Aging: Ratcliff, Thapar, Gomez, & McKoon, 2003; Aphasia: Ratcliff, Perea, Colangelo, & Buchanan, 2004). Ashby (2000) used DDM to study accuracy and response times in categorization tasks and introduced a dynamic version of the decision bound models called stochastic GRT. A DDM has a noisy evidence accumulator and two decision boundaries. Percept and accumulated evidence are stochastic processes that are usually modeled as a discrete random walk process. Figure 4 shows an accumulator with visualization of the process for three trials of a two choice task.

In order to prevent confusion between the decision bounds in a perceptual space and the decision bounds in a DDM, following Ashby (2000) we call the later ‘absorbing barriers’. Some of the most important parameters of a DDM that are shown in Figure 4 are:

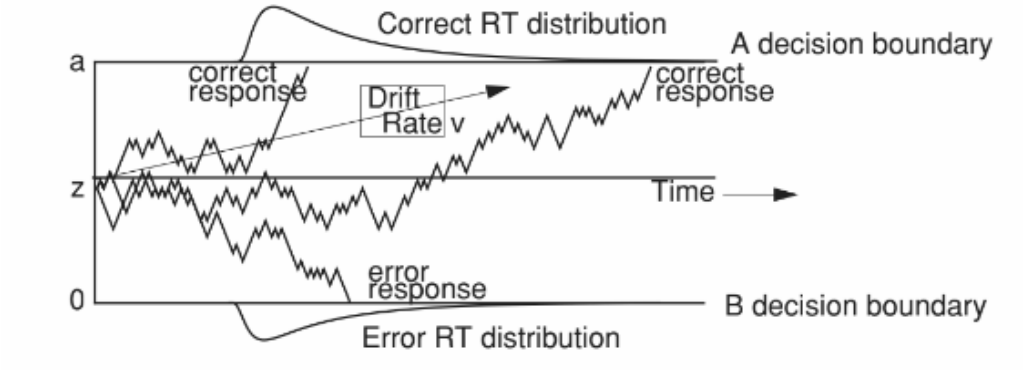


Figure 4. Three instances of a drift diffusion process. Figure is from Ratcliff and McKoon (2008).

1. Drift rate (v): Reflects the quality of evidence (i.e., difficulty of a trial). The positive values mean that the drift is toward choice A and negative values mean that the drift is toward choice B. The drift rate is zero on the decision bound used by the participant.
2. Starting point (z): Determines whether there is any bias towards one of the choices.
3. Boundary separation (a): Reflects speed accuracy trade off: bigger values model higher accuracy and slower RT and vice versa.
4. Non-decision time (t0): Assuming that RT is sum of stimulus encoding, decision making and response execution, t0 models sum of stimulus encoding and response execution components of RT.

The advantage of stochastic GRT over its static version is that it makes predictions for accuracy and response time in one single structure. Another advantage of stochastic GRT is that it may be more biologically plausible because evidence accumulation in DDM is reminiscent of firing pattern of neurons (Smith & Ratcliff, 2004). A more detailed comparison between the two analyses is provided in the discussion section.

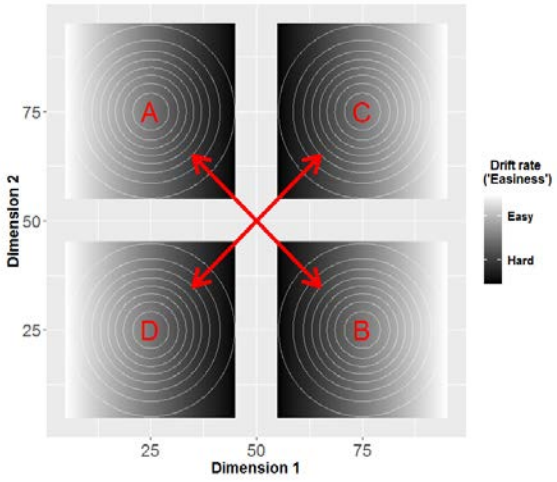
Hypothesis

In a two-dimensional categorization task where both dimensions are diagnostic, there are participants who learn and use both dimensions, participants that learn both dimensions but only use one of them, and participants that learn and use only one of the dimensions. We hypothesize that whether a dimension was used or not by a participant manifests itself in the error and RT patterns. Analysis 1 and Analysis 2 are two ways to

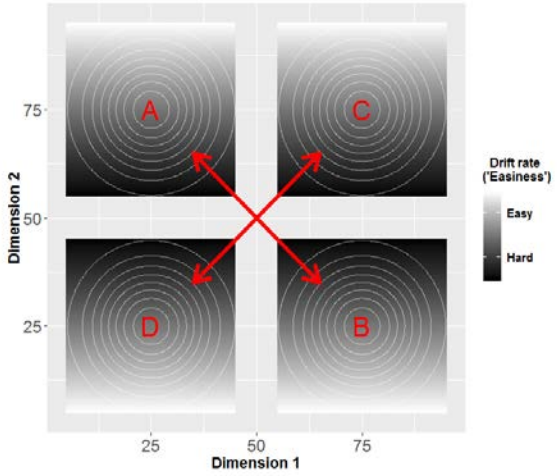
identify the dimension(s) used by each participant and there are hypotheses and assumptions inscribed in each analysis about different ways by which using a dimension should affect the error and RT patterns.

More errors and slow response times close to a bound perpendicular to a dimension indicate that a participant is using that dimension to categorize stimuli (Ashby & Soto, 2015, Ashby & Maddox, 1994). In Analysis 1, existence of any of these two characteristics (more errors and slower RTs around a decision bound perpendicular to a dimension), is considered as evidence for that dimension being used. iDBM and RT-distance hypothesis are used to directly assess whether there is evidence for any of the two mentioned characteristics.

Analysis 2 uses drift diffusion models to identify which dimension(s) each participant used. A desirable characteristic of DDMs is that the main parameters of the model are selectively affected by different kinds of manipulations (Voss, Rothermund & Voss, 2004). More specifically, drift rate, starting point, boundary separation, and non-decision time are selectively affected by difficulty of trial, payoff structure, speed-accuracy instruction, and ease of executing the motor response (respectively). The assumption of Analysis 2 is that the difficulty of trial (reflected in the drift rate of DDM) depends on the location of stimulus in the two-dimensional category space, and the relation between location of the stimulus on category space and drift rate depends on the strategy of participant. To make it more clear, possible strategies with their corresponding drift rates are shown in Figure 5 and Figure 6. Figure 5 shows the unidimensional strategies: Figure 5a shows the expected dependency of the stimulus difficulty on its location in the category space for participants that use only Dimension1. Figure 5a



(a)



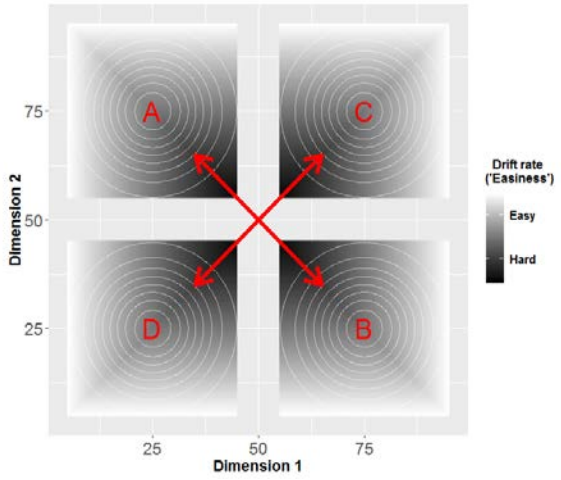
(b)

Figure 5. Relation between location of the stimulus on category space and its relative difficulty in unidimensional strategies. Red arrows show that only two of the possible comparisons ("A or B?" and "C or D?") were asked of participants. (a) Categorization based on only Dimension 1. (b) Categorization based on only Dimension 2.

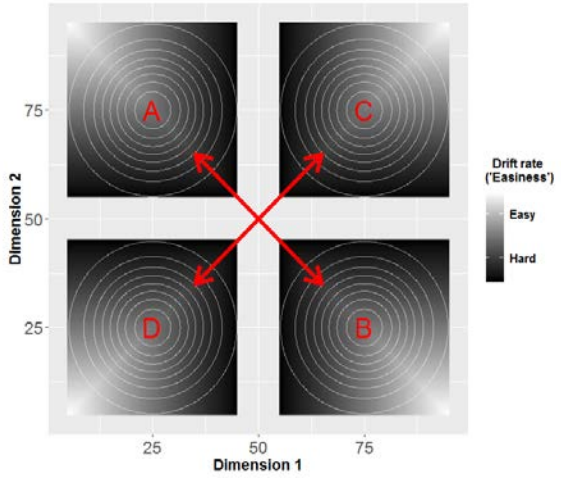
reflects two different groups of participants: (1) Participants that use Dimension1, and Dimension 2 does not even enter their attentional frame (i.e., Possibility 1 in Table 1), and (2) participants that use only Dimension1 but are latently learning the differences in Dimension 2 (i.e., Possibility 3 in Table 1). Figure 5b shows the other unidimensional strategy that corresponds to participants that use only Dimension 2. Similar to Figure 5a, Figure 5b reflects two group of participants, namely participants that used Dimension2 and did not learn Dimnesion1 (i.e., Possibility 2 in Table 1), and participants that used only Dimension2 but latently learned Dimension1 (i.e., Possibility 4 in Table 1).

Figure 6 shows two of the possible ways that location of a stimulus can affect trial difficulty when a participant uses both dimensions. Figure 6a shows a 'Time efficient' strategy. 'Time efficient' means that a stimulus is perceived to be relatively easy as long as the attribute of any one of its dimensions are far from the corresponding attributes of the members of the other category. Figure 6b shows a 'Conservative' strategy. 'Conservative' means that a stimulus is perceived to be relatively easy only if both dimension attributes are far from the corresponding attributes of the other category's members.

Theoretically, there are more two-dimensional strategies than just the two cases shown in Figure 6, but Figure 6a and 6b depict two extremes, and therefore the 'in between' strategies are likely to be captured by one of the two: Relatively time efficient strategies by Figure 6a and relatively conservative strategies by Figure 6b. In addition to the four strategies depicted in Figures 5 and 6, a fifth strategy that assumes perceived difficulty of a stimulus is independent of its location in category space is considered. This strategy reflects response pattern of a participant for whom all stimuli are perceived



(a)



(b)

Figure 6. Relation between location of the stimulus on category space and its relative difficulty in two-dimensional strategies. Red arrows show that only two of the possible comparisons ("A or B?" and "C or D?") were asked of participants. (a) A "time efficient" strategy. (b) A "conservative" strategy.

equally easy, irrespective of their location in category space (Hélie, Waldschmidt, & Ashby, 2010). In Analysis 2, five DDMs corresponding to the five mentioned strategies are fitted to each participant's data and the strategy corresponding to the best fitting model is considered to be the strategy of the participant.

To summarize, we hypothesize that in a categorization task with redundancy, there are individual differences and participants choose (happen) to learn and/or use different dimensions. The differences in strategy of participants must be reflected in the accuracy and RT patterns, and the main goal of this thesis is to identify and quantify the differences using two separate analyses.

THE EXPERIMENT

The experiment studied categorization learning in a two-dimensional space where perfect accuracy can be reached by using any one of the two dimensions. Whether a participant acquired knowledge on one of the dimensions or both dimensions is assessed using a test phase. The stated goals of the thesis are studied using accuracy and RT patterns of participants.

Method

Participants

One hundred seventy Purdue University undergraduate students participated in the study and received credit to fulfill a course requirement.

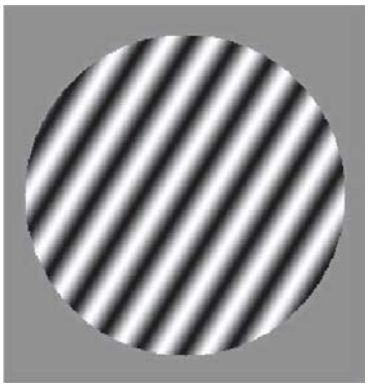
Material

The stimuli were sine-wave gratings of constant contrast and size that differed in frequency (ranging from 1.65 to 2.21 cycles per degree) and orientation (ranging from 0.82 to 1.44 radians). Stimulus presentation and response recording was done using the Psychophysics Toolbox in MATLAB (Brainard, 1997). There were four categories (arbitrary labeled “A”, “B”, “C” and “D”) and in each trial participants were shown a stimulus and asked to choose between two of the categories. Stimuli were generated using bivariate normal distributions with the following parameters: $\mu_A = (1.736,1.322), \mu_B = (2.096,0.945), \mu_C = (2.096,1.322), \mu_D = (1.736,0.945), \sum_A = \sum_B = \sum_C = \sum_D = \begin{pmatrix} 0.0022 & 0 \\ 0 & 0.0024 \end{pmatrix}$. Ninety-six stimuli were generated (twenty-four from each category) which were shuffled in the beginning of each of the training blocks. One hundred forty-four stimuli (thirty-six of each category) were generated for the test

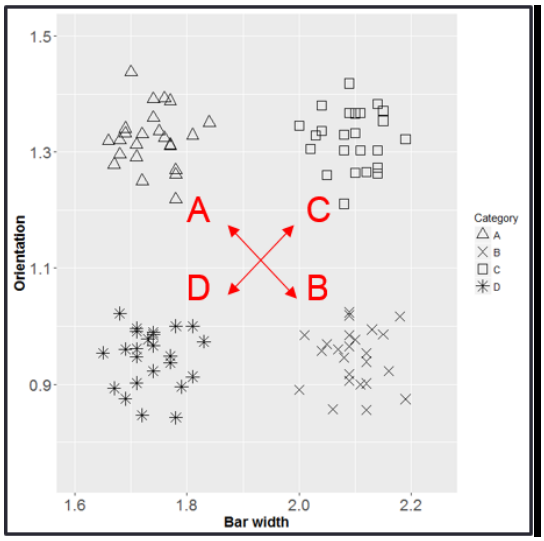
phase. Figure 7a shows a sample stimulus and Figure 7b and 7c show the generated stimuli in the training and test phase, respectively. Participants only performed two types of categorization trials in the training phase. In each trial the question shown on the screen was either “A or B?” or “C or D?”, and note that in both of “A or B?” and “C or D?” knowledge on any one of the two dimensions is enough to distinguish between the two categories. In the test phase, participants performed all possible two choice categorizations (“A or B?”, “A or C?”, “A or D?”, “B or C?”, “B or D?”, “C or D?”). Red arrows in Figure 7b and Figure 7c show the categories that were compared together in each phase of the experiment. Participants responded using a standard keyboard and in all of the trials, ‘d’, ‘k’, ‘x’ and ‘m’ keys were used to choose categories ‘A’, ‘B’, ‘C’ and ‘D’ respectively.

Procedure

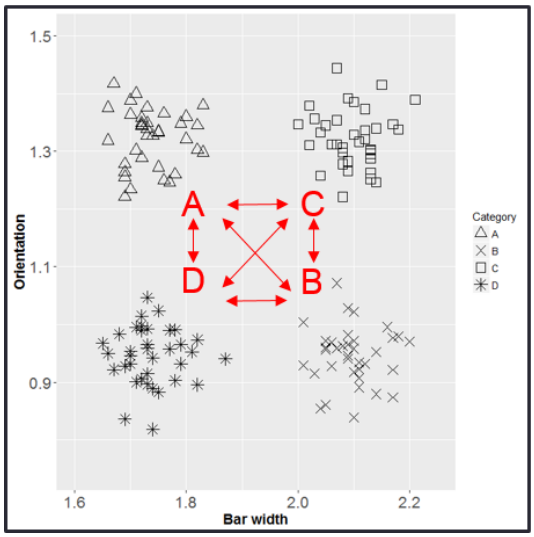
Participants were told that they would be participating in a two choice categorization task where stimuli are sine-wave gratings that differed in bar width and orientation. They were told that there are four categories, and in each trial, a question on top of the screen asks them to choose between two of the four categories. The experiment was divided into six blocks and participants categorized ninety-six stimuli in each of the first five blocks and 144 stimuli in the sixth block. Participants were told that during the first five blocks, they will receive feedback but no feedback will be given in the last block. Each training trial started with a fixation cross that was presented at the center of the screen for 1500 ms. Then the fixation cross was replaced by the stimulus and categorization question. As soon as the participant responded, the stimulus and question were replaced by feedback (green “Correct” for correct responses and red “Incorrect” for



(a)



(b)



(c)

Figure 7. Red arrows indicate the comparisons participants were asked to do in each phase of the experiment. (a) An example stimulus. (b) The stimuli used in the training phase. (c) The stimuli used in the test phase.

incorrect responses) which stayed on the screen for 750 ms. In trials where the participant did not respond within five seconds, a blank screen with a “Too Slow!” text in the center replaced the stimulus and question. The timed-out trials were counted as error. Test trials followed the same sequence except that no feedback was given to participants. Figure 8 shows the display sequence for a trial in the training phase.

Results

The results are divided into two sections: ‘Learned knowledge’ and ‘Used knowledge’ that correspond to the analysis of test and training phase respectively. As stated in the introduction, the main goal of this thesis is to understand the strategy of participants when there is redundancy of information, which corresponds to the training phase of the experiment. However, analyzing the test phase and determining the dimension(s) that each participant learned matters because it partially validates the methods that are used to identify the strategy of participants: Assume a participant acquired knowledge only on one of the dimensions. Logically, her strategy must be a unidimensional strategy. Therefore, if the method that is used to identify the strategy of participants identify a two-dimensional strategy, the method is faulty. Assessing the learned knowledge is straightforward (using the test block) and after determining the learned knowledge of each participant, the used knowledge (during training) will be studied using two separate approaches (Analysis 1 and Analysis 2).

Learned Knowledge

During the training phase, the category structure allowed participants to reach perfect accuracy using only one of the dimensions. Therefore, participants could acquire knowledge on bar width, orientation, or both. The last block of the experiment (the test

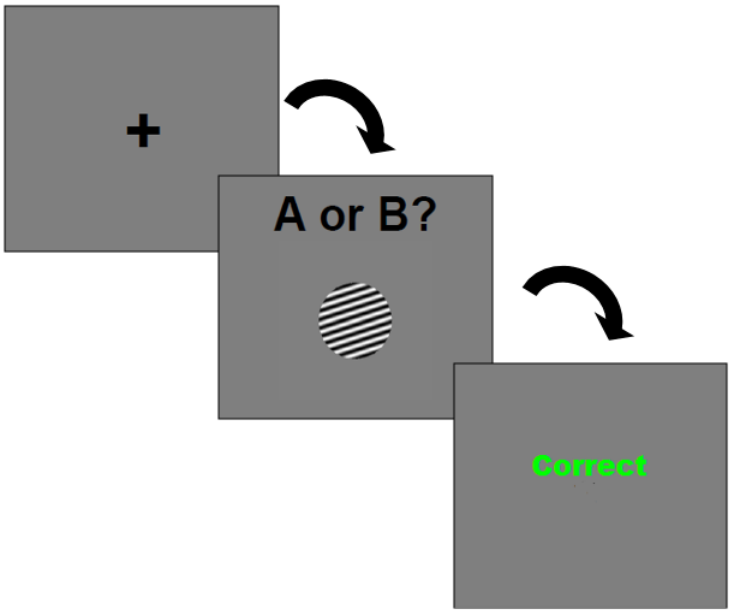
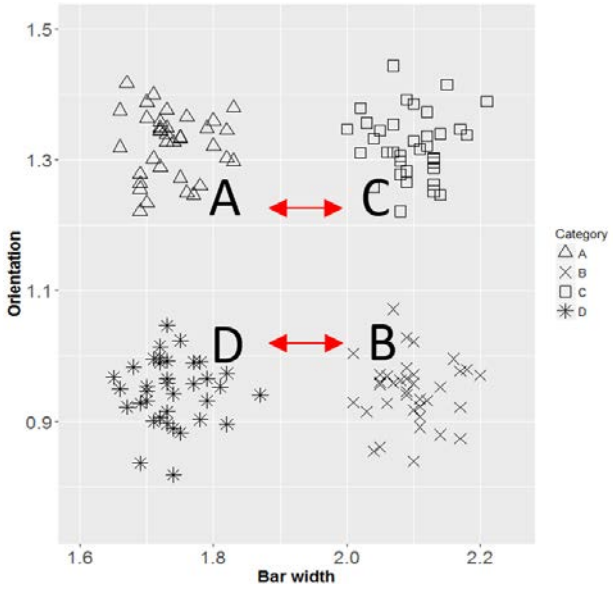


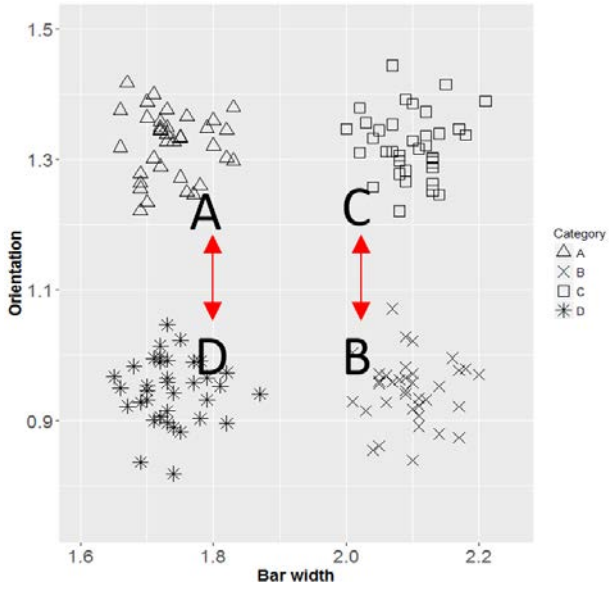
Figure 8. An example of a trial sequence in the training phase. Test phase trials were similar, only no feedback was given to participants.

phase), was added in order to find out the dimension(s) learned by each of the participants. Figure 9 shows two types of categorization tasks participants performed in the test phase. Figure 9a shows trials in which knowledge on bar width was decisive (from now, called BW trials) and Figure 9b shows trials in which knowledge on orientation was decisive (from now, called OR trials).

Participants received no feedback in the test phase and therefore, it is safe to assume that success on BW trials meant that knowledge on bar width was acquired during training and, similarly, success on OR trials meant that knowledge on orientation was acquired during the training. Participants were divided into four groups based on their learned knowledge. The four groups correspond to participants that acquired knowledge on both dimensions (‘Learned_Both’), participants that acquired knowledge only on bar width (‘Learned_BW’), participants that acquired knowledge only on orientation (‘Learned_OR’) and participants that learned none (‘Non-Learner’). There were forty-eight BW trials and forty-eight OR trials in the test phase, and using two separate one tailed binomial tests ($p = 0.5$, $N = 48$) on each of BW and OR trials, participants’ learned knowledge label were assigned. A minimum of thirty-one correct responses (out of fourty-eight total, corresponding to p-value < 0.05) was considered evidence that a participant’s accuracy was better than chance. Figure 10 summarizes the test performance and assigned learned knowledge. Each circle represents a participant (the number inside the circle is the participant number) and the x-axis and y-axis are participant’s accuracy in BW and OR trials respectively. Note that as previously stated, no feedback was given in the test phase, and therefore, the knowledge learned by each participant was acquired during the training phase. Figure 10 confirms that there are IDs



(a)



(b)

Figure 9. Four types of categorizations that participants did in test phase. (a) BW trials: "A or C?" and "B or D?" trials, where knowledge on bar width differences is necessary. (b) OR trials: "A or D?" and "B or C?" trials, where knowledge on orientation differences is necessary.

*Figure 1*0. Test performance of participants. The x-axis is the mean accuracy on trials where knowledge on bar width was necessary to categorize the stimulus and the y-axis is the mean accuracy on trials where knowledge on orientation was necessary to categorize the stimulus. Each circle is a participant and color of each participant shows whether they learned both dimensions, only bar width, only orientation or none.

in the learned knowledge when there is redundancy. Some participants acquire knowledge on both dimensions (northeast of the plot) while some participants only acquire knowledge on one of the dimensions (located either on the northwest of the plot or on the southeast). The 'Non-Learner' participants were not included in the remaining analyses.

Now that the dimension(s) learned by each participant is established, it is helpful to restate the goal of the thesis more specifically. In a categorization task with redundancy of information, knowing that a participant 'learned' only bar width means that she also 'used' only bar width. Are there tools good enough to detect patterns in accuracy and RT showing that the only used dimension was bar width? What about participants that learned both dimensions? Did they use both dimensions, or did they use only one dimension while latently learning the other? If both dimensions were used, in what specific way? These questions are pursued using two separate analyses (Analysis 1 and Analysis 2), both of which look at the training data of each participant and determine the used knowledge.

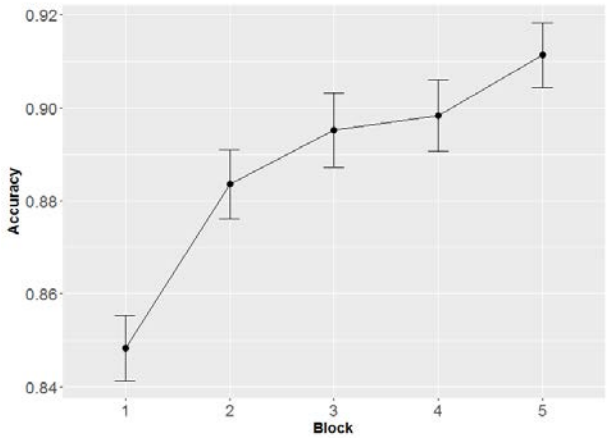
Used Knowledge

In a two-dimensional categorization task with redundancy of information, participants differ in terms of which dimension(s) they learn and use. A test phase was added to the experiment, which made it possible to know which dimension(s) each participant *learned* and the rest of the thesis is an attempt to identify the dimension(s) *used* by each participant by looking at the data from the training phase (the first five blocks). As it was shown in Table 1, there is not a one-to-one mapping between learned knowledge and used knowledge of participants (due to the possibility of latent learning).

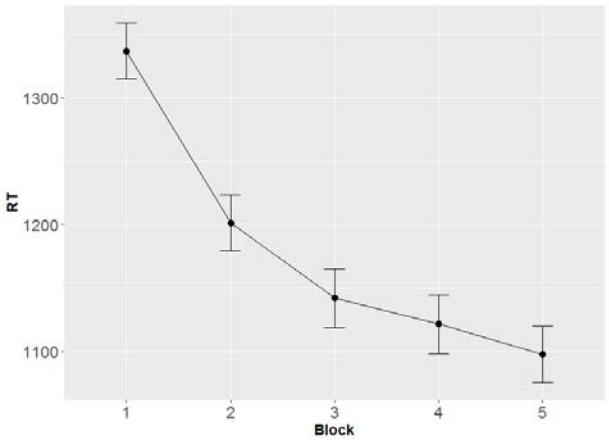
This means that even though we know which dimension(s) are learned by each participant, the 'true' used knowledge is not known and therefore: 1) Identifying which dimension(s) each participant used is not going to be as easy as extracting some features and assessing the predictive power of those features by fitting a supervised classifier. 2) It is impossible to validate the methods used to identify participants' strategies because there is no 'true' used knowledge. The first point means that in order to identify participants' strategies there needs to be a theoretical framework that asserts a relation between participants' strategies and response patterns; Analyses 1 and 2 provide two such frameworks. The second point (the impossibility of validating the methods) is very problematic. However, note that there is a one to one mapping between learned and used knowledge of participants that belong to Learned_BW and Learned_OR. Therefore, it is possible to partially validate the theories and their implementations by making sure that Learned_BW participants are identified as using only bar width and Learned_OR participants are identified as using only orientation.

Analysis 1 and Analysis 2 look at the data in its totality without summarizing the response patterns into measures such as mean accuracy and mean RT but it is worth looking at the mean accuracy and mean RT of participants before starting Analyses 1 and 2. Figure 11 shows the average performance of participants in the first five blocks (training phase). Training accuracy started at around 85% and by the end of training reached around 91% (Figure 11a) and response time started at around 1340 ms and reached around 1100 ms by the end of training (Figure 11b).

A central emphasis of this thesis is to distinguish between what people learn, and what they use, and the assumption that used knowledge manifests itself in the response



(a)

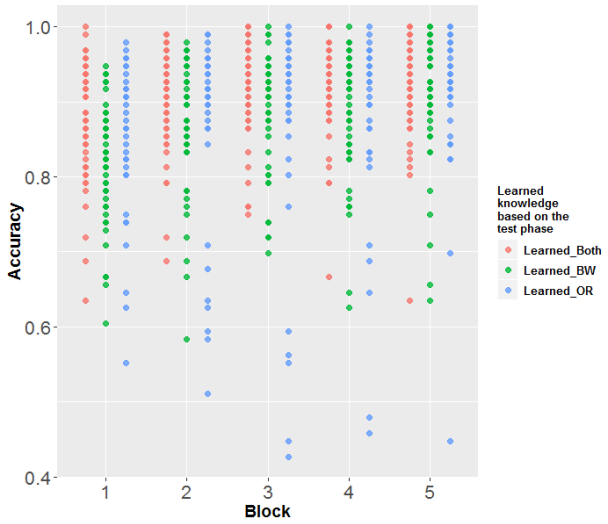


(b)

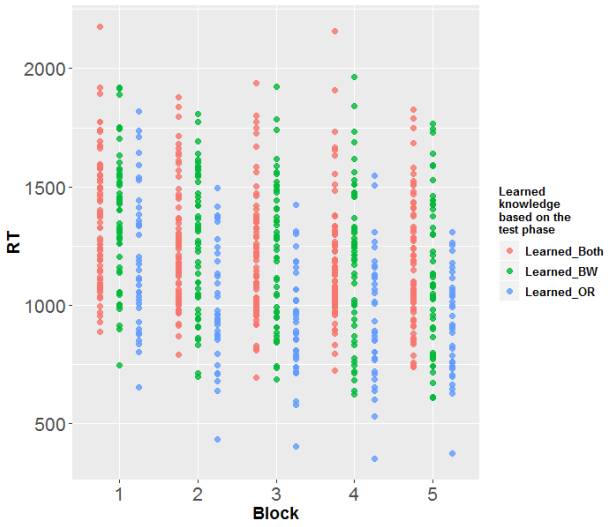
Figure 11. Performance of participants during the first five blocks (training phase): (a) average accuracy and (b) average response time. Error bars represent one standard error.

patterns, whereas learned knowledge might not. Therefore, since some of the Learned_Both participants do not use both dimensions, any measure that uses the features extracted from data to predict the learned knowledge will fail. Nevertheless, it is worth visualizing the differences between participants based on their learned knowledge. Figure 12 shows mean accuracy and mean RT for Learned_BW, Learned_OR and Learned_Both participants. The difference between the three groups are hard to detect, which suggests that mean accuracy and mean RT would not be enough to predict the learned knowledge of participants. In order to show this formally, a multinomial regression was fit to data. The dependent variable was the learned knowledge (with three levels: Learned_Both, Learned_BW and Learned_OR) and the independent variables were mean accuracy and mean RT of blocks 1 to 5 (10 variables). Using MLR package of R (Bischl et al., 2016), the cross validation accuracy of the model (3-fold cross validation with stratified sampling, repeated 500 times and averaged to get a robust estimate) was computed to be 49.31%. Note that there are seventy-two Learned_Both, fifty-one Learned_BW and thirty-six Learned_OR participants and by simply assigning all to Learned_Both, the accuracy is 45.28% and therefore, 49.31% accuracy using ten variables is extremely low.

This result is not surprising: the Learned_Both participants are not a uniform group, some might be using only bar width, some might be using only orientation, and some use both dimensions and among participants that use both dimensions, there might be differences in the specific ways the two dimensions are used. Therefore, even if mean accuracy and mean RT were enough to identify used knowledge, they would fail in identifying the learned knowledge. However, as Table 1 shows, there is a one to one mapping between learned and used knowledge of Learned_BW and Learned_OR



(a)



(b)

Figure 12. Mean accuracy and mean RT, grouped based on the dimension(s) learned. Each dot represents a participant. (a) Average accuracy. (b) Average response time.

participants. The problematic Learned_Both participants were excluded and a logistic regression was used to assess the predictive power of mean accuracy and mean RT of blocks 1 to 5 (10 variables) with dependent variable being the learned knowledge (with two levels: Learned_BW and Learned_OR. Using MLR package of R, the cross validation accuracy of the model (3-fold cross validation with stratified sampling, repeated 500 times and averaged to get a robust estimate) was computed, to be 68.98%. There are fifty-one Learned_BW and thirty-six Learned_OR participants and by simply assigning all to Learned_BW, the accuracy is 58.62% and therefore adding mean accuracy and mean RT improved the predictive power slightly. Looking at the p-values of the fitted logistic model shows that two features reach statistical significance: accuracy of block 1 (p-value =0.00504) and RT of block 3 (p-value=0.00815). Note that it is also possible to look at the effect of learned knowledge on accuracy and RT, but the focus of this study is to predict used knowledge of each individual. The group differences are studied later in the thesis (Appendix) after dividing participants based on their strategy in Analysis 2.

The point of the first analysis (multinomial regression) was to show that it is not possible to use a supervised classifier to predict the learned knowledge, which may be because of the potential IDs in the strategies of Learned_Both participants. The second analysis (logistic regression) showed that using mean accuracy and mean RT is not good enough: even when Learned_Both participants were excluded, the model was not able to distinguish between Learned_BW and Learned_OR participants.

Analysis 1 and Analysis 2 take an entirely different approach: Both analyses look at the accuracy and RT of all the trials instead of the mere block averages and

additionally, the learned knowledge of participants (determined by the test phase performance) does not play any role in building the models that determine participants' strategies. In other words, Analysis 1 and Analysis 2 rely on theories that tell what response pattern of a participant should look like depending on which dimension(s) s/he is using. Validity of theories that are used in Analyses 1 and 2 and their implementation will be tested by looking at Learned_BW and Learned_OR participants, which must be identified as using only bar width and using only orientation respectively.

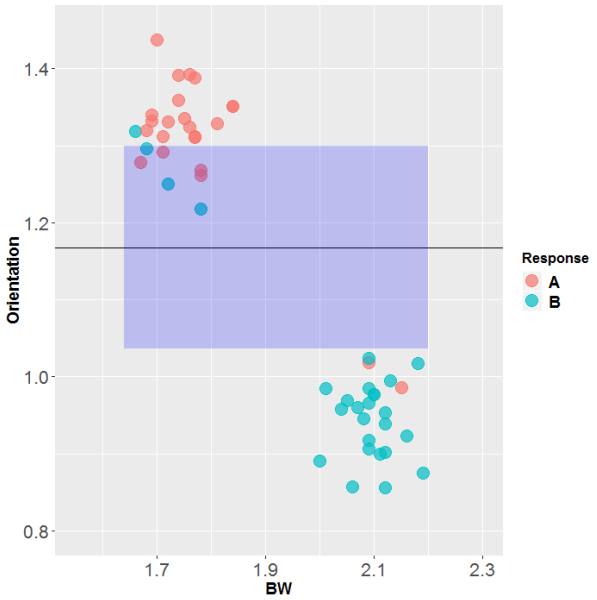
ANALYSIS 1

Methods

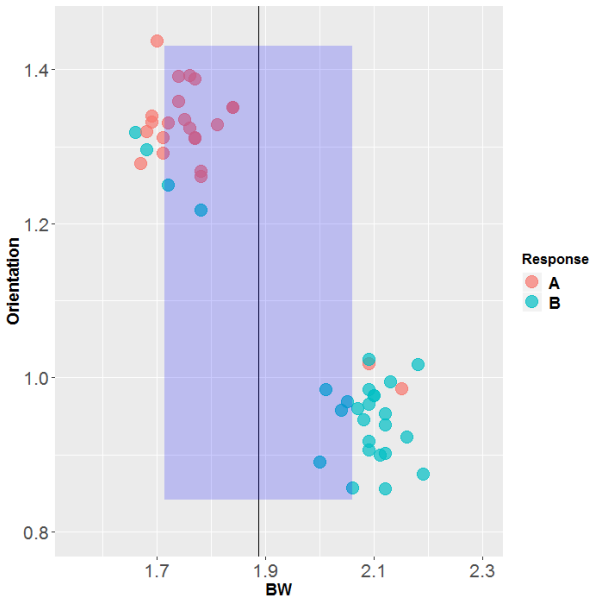
There are two types of evidence indicating a dimension is being used: More errors and slower RT close to a decision bound. Analysis 1 looks at each of them separately and uses iDBM and RT-distance hypothesis to detect if either of the two pieces of evidence exist.

iDBM: Looking at Error Patterns

The COVIS model of categorization (Ashby et al., 1998) postulates that participants usually start by testing unidimensional explicit rules. In the training phase of the experiment, a unidimensional rule on any of the two relevant dimensions is sufficient for perfect accuracy. Therefore, if a participant starts by testing any of the two bounds, there is no need (in the sense that any of the two by itself suffices for perfect accuracy) to use a different bound. However, it is possible that a participant notices differences in the other dimension after a while and starts responding based on the other dimension. In this experiment, a rule-based strategy suffices and based on the assumptions of COVIS, the participants will not use a non-verbal strategy. Therefore, the procedural strategies (i.e., diagonal bounds) were excluded from the list of models that iDBM fits and considers. “A or B?” and “C or D?” trials are considered two different tasks and iDBM was fit separately on these two tasks. Figure 13 is an illustration of what iDBM does on its first iteration. The best fitting bound on each dimension is fitted to trials 1-100 of one of the participants and the maximum likelihood values are compared. In the instance shown in Figure 13, the participant seems to be using the orientation dimension, since the errors are close to the bound on orientation, which is reflected in the likelihood values: -14.65 for



(a)



(b)

Figure 13. A visualization of how iDBM works. The bounds are fitted to trials 1-100 of participant 109. (a) The bound fitted on bar width. (b) The bound fitted on orientation.

the bound on orientation and -20.76 for the bound fitted on the bar width dimension (More details at Hélie et al., 2017).

There are three general possibilities for each participant:

1. Participant starts by guessing and at some point switches to a unidimensional rule on one of the dimensions and then never switches again.
2. Participant starts by guessing and switches to different unidimensional rules for “A or B?” and “C or D?” tasks, therefore, both dimensions have entered the attentional frame.
3. Participant starts by guessing and switches to a unidimensional rule on one of the dimensions for both “A or B?” and “C or D?” tasks and later on switches to a unidimensional rule on the other dimension for one or both of “A or B?” and “C or D?” set of trials.

One of the dimensions was never used by the participants in the first described case (based on the iDBM at least), but it is possible for them to have passively obtained information on a dimension without using it (therefore its effect did not appear in their response patterns). It is possible to look at the test performance of these participants and verify whether they were successful in both OR and BW trials. On the other hand, participants in cases 2 and 3 used both dimensions to perform categorization at some point of the training phase. Therefore, these participants are expected to be at least partially successful in both BW and OR trials of the test phase.

RT-Distance Hypothesis

If a participant uses a rule on a dimension to perform the categorization task at training, then based on the RT-distance hypothesis there should be a negative relation between response time and distance to the bound used by the participant. In order to assess whether this negative relation exists, for each dimension the following procedure was applied to the last three blocks of the training phase (where accuracy seems to be stable) of each participant. The distance between each stimulus and the ideal bound was calculated. Then the distances were normalized and rounded to two digits and the median response time of all stimuli located on the same distance from the bound were selected. Finally, the Spearman correlation between response time and distance to bounds was calculated. Spearman's rank correlation is used since RT-distance hypothesis posits the relation between distance to bound and RT to be monotonically decreasing and assumes nothing about the shape of the relation. Figure 14 shows the distance to bound graph of a participant that shows longer response time close to BW bound and belongs to Learned_BW group. Each dot is a stimulus that is specified by its distance to an optimal bound in the BW dimension (x-axis) and the RT associated with it (y-axis). Left panel shows response time for all stimuli, and right panel shows the median response time of those in the same distance from the bound. The correlation is computed based on the data corresponding to the right panel. In order to be brief, from now on the correlation between RT and distance to a bound is referred to as distance to bound effect (D2B) of that bound (e.g., D2B of BW bound).

Bayesian Spearman correlations are calculated and participants are divided into five groups based on the value of Bayes factor (BF):

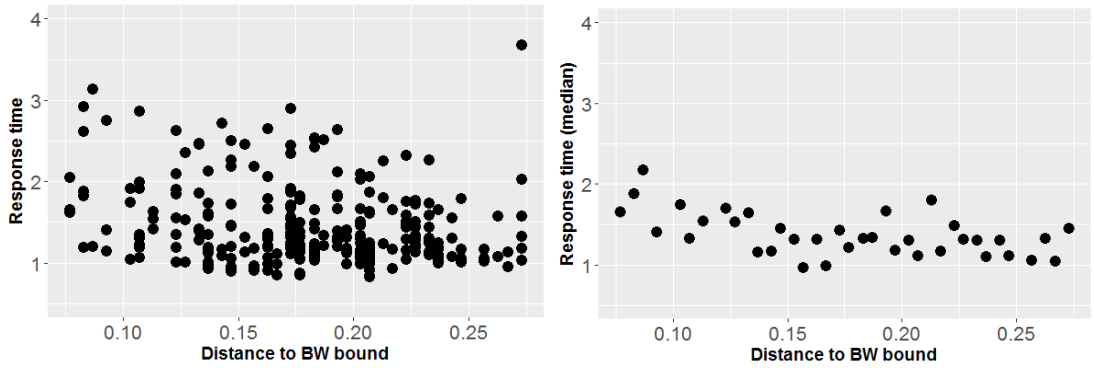


Figure 14. An example of how distance to bound measure was calculated. Left panel shows all trials and in the right panel, the median RT of stimuli in the same distance from the BW bound is shown.

1. No D2B: There is good evidence that these participants do not show D2B effect for any of the bounds.
2. Uncertain: There is not enough evidence to claim anything about existence or nonexistence of a D2B for any of the bounds.
3. Used bar width: There is evidence for the existence of D2B of bar width, but no evidence for D2B of orientation.
4. Used orientation: There is evidence for the existence of D2B of orientation, but no evidence for D2B of bar width.
5. Used both: There is evidence for the existence of both D2B of bar width and D2B of orientation.

Table 2 summarizes the labeling procedure.

Results

Results of iDBM and RT-distance hypothesis are first presented separately and combined in the end to conclude Analysis 1. The validity of each analysis is tested by looking at participants that learned only one of the bounds, which must be best fit by the models that correspond to the unidimensional strategies.

iDBM Results

Figure 15 shows the result of iDBM for Learned_BW participants and Learned_OR participants. Each circle represents a participant, the location of the circle codes the learned knowledge and color of each circle is the identified used knowledge. Based on the procedure described for iDBM in the methods section, there are three possible outputs for used knowledge, labeled as Used Bar width, Used Orientation and

Table 2

The way Each Participant is Labeled Based on the Bayes Factor of Spearman Correlations Between RT and Distance to Decision Bounds on Bar Width and Orientation

BF for the D2B of Bar Width	BF for the D2B of Orientation	Assigned Used Knowledge
BF < 0.3	BF < 0.3	No D2B
0.3 < BF < 3	0.3 < BF < 3	Uncertain
BF > 3	BF < 0.3	Used Bar width
BF > 3	0.3 < BF < 3	Used Bar width
BF < 0.3	BF > 3	Used Orientation
0.3 < BF < 3	BF > 3	Used Orientation
BF > 3	BF > 3	Used Both

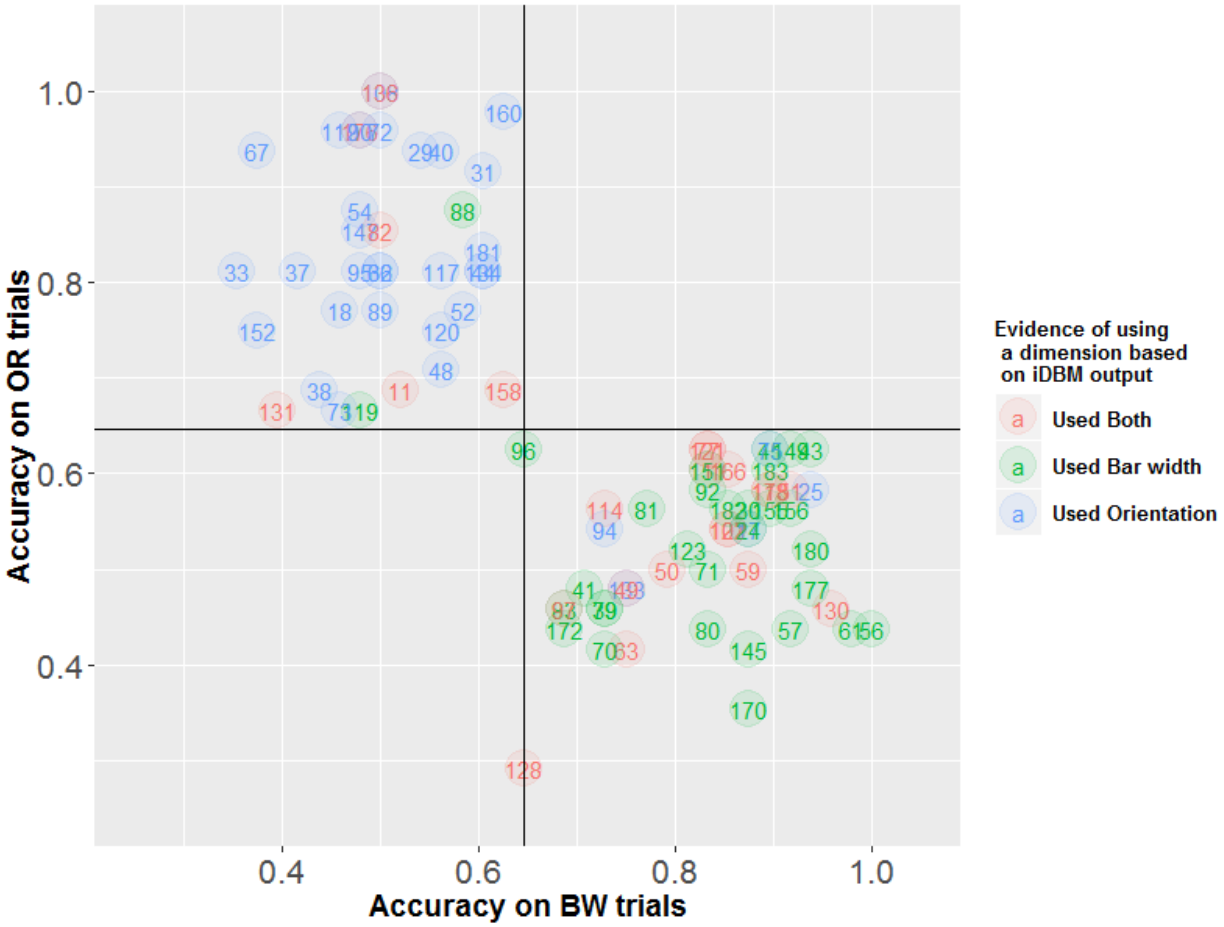


Figure 15. The identified strategies of Learned_BW and Learned_OR participants. Each circle represents a participant. Color of a circle shows the used knowledge (based on iDBM) and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

Used Both. Learned_BW and Learned_OR participants should be assigned to Used Bar width and Used Orientation respectively. As Figure 15 shows, the majority of Learned_BW participants (bottom right corner of the figure) are green (i.e., assigned to Used Bar width) and the majority of Learned_OR participants (top left corner of the figure) are blue (i.e., assigned to Used Orientation).

Table 3 summarizes the correspondence between labels assigned by iDBM and the learned knowledge for participants that learned only one dimension. Out of eighty-seven participants that learned only one dimension, iDBM correctly identifies the strategy of fifty-seven of them (i.e., 65% accuracy compared to 33% random assignment). Note that the errors are mostly due to mistakenly identifying a unidimensional strategy as a two-dimensional strategy. In other words, if a participant uses a dimension, iDBM successfully detects the patterns caused by using that dimension, but there are cases that iDBM detects signs of a dimension being used even though in fact that dimension was not learned.

Now that the validity of iDBM labels is demonstrated, the strategy of participants that learned both dimensions is shown in Figure 16. Theoretically, Learned_Both participants can belong to any of the considered strategies, it is possible that they used only one of the dimensions while latently learning the other, or it is also possible that they used both dimensions.

Table 4 shows the identified strategies of the Learned_Both participants. An interesting pattern to note is that among participants that learned both dimensions many used only one and latently learned the other dimension. This pattern can be seen by the large number of blue and green dots in the northeast corner of Figure 16.

Table 3

The Confusion Table for the Relation Between Identified Used Knowledge (Based on iDBM) and Learned Knowledge for Participants That Learned Only One of the Dimensions

	Learned Knowledge	
Used Knowledge Based on iDBM	Learned_BW	Learned_OR
Used Bar width	29	2
Used Orientation	5	28
Used Both	17	6

Table 4

Identified Used Knowledge (Based on iDBM) for Participants That Learned Both Dimensions

	Learned Knowledge
Used Knowledge Based on iDBM	Learned_Both
Used Bar width	17
Used Orientation	25
Used Both	30

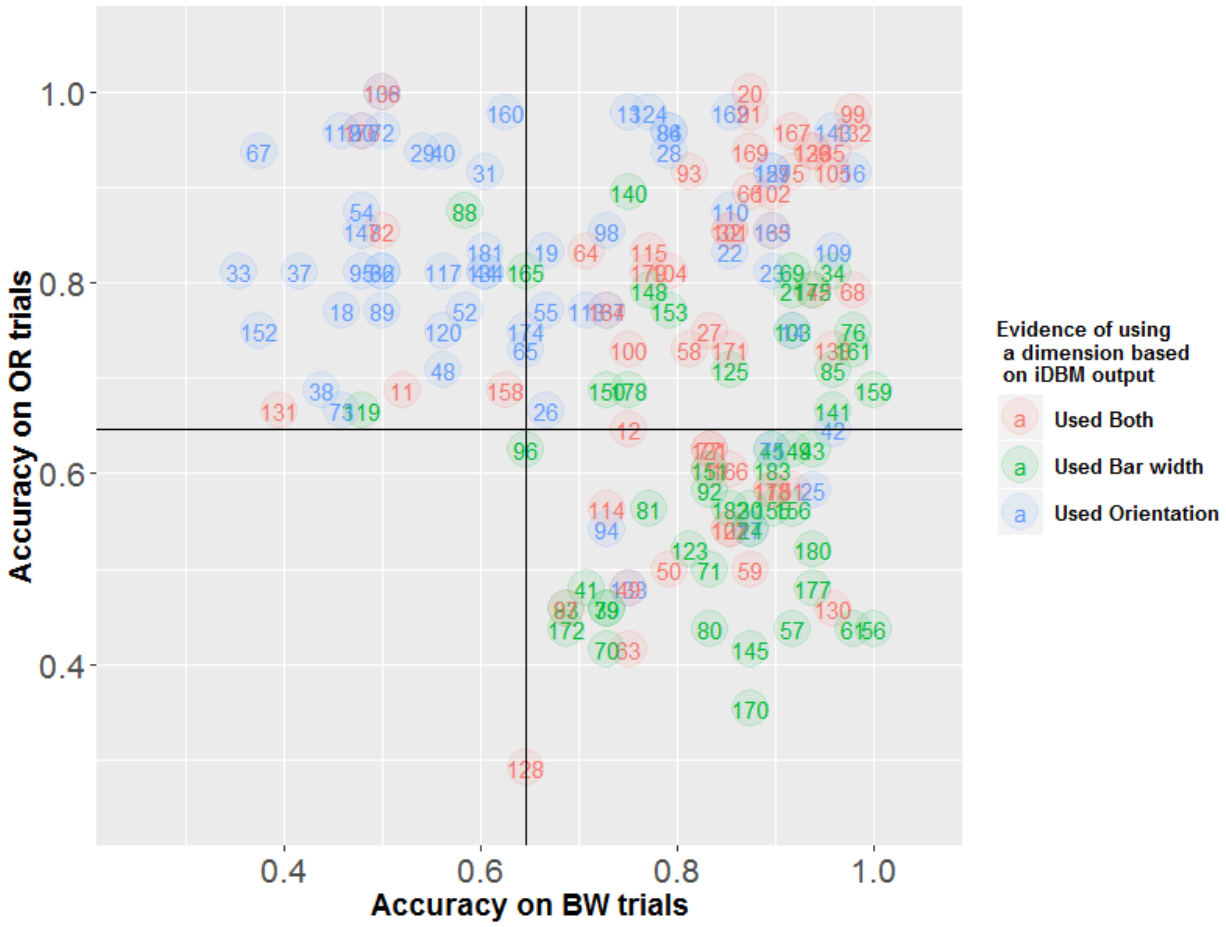


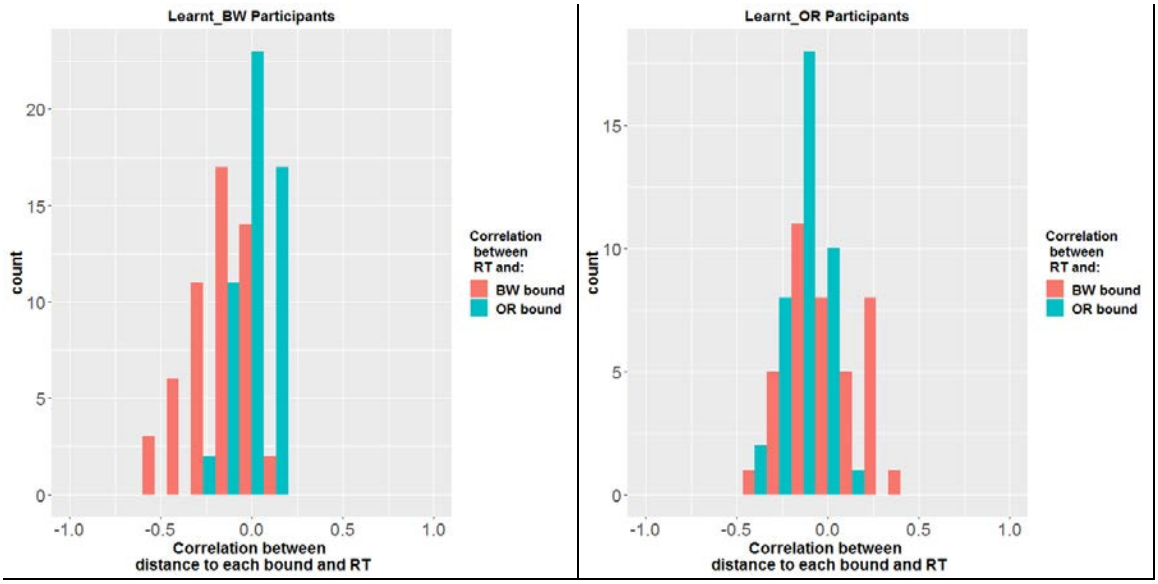
Figure 16. Each circle represents a participant. Color of a circle shows the used knowledge (based on iDBM) and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

RT-Distance Hypothesis Results

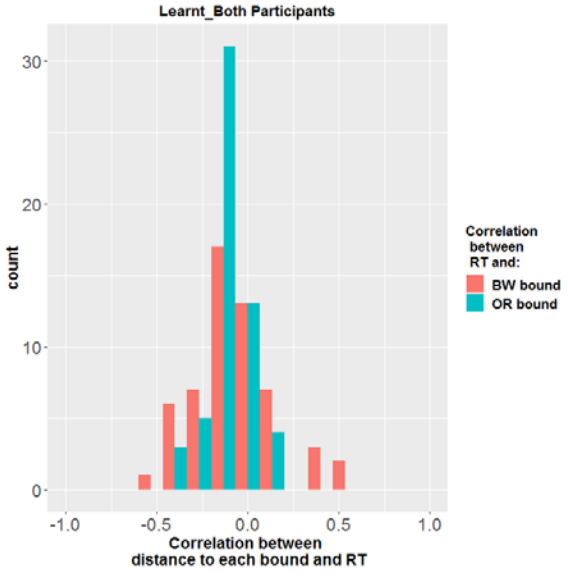
The correlations between response time and distance to each of the bounds was calculated as explained in the methods section. Figure 17 shows the histogram of correlation for Learned_BW, Learned_OR and Learned_Both participants separately. As expected, the histogram of participants in Learned_BW and Learned_OR groups are approximately centered on zero for the bound in the dimension that they have not acquired knowledge on and for the bound which they are using, the correlation between RT and distance to bound is negative for most of the participants. In the Learned_Both group, most of the correlations are negative for both bounds.

Following the procedure described in the methods section, participants were divided into the five groups mentioned in Table 2. Labels assigned to participants that learned only one of the dimensions is shown in Figure 18. Similar to previous sections each circle represents a participant. Location and color of circles code the learned knowledge and used knowledge respectively. As the figure shows, the data were not clean enough to assert existence or non-existence of D2B for most of the participants (i.e., most Bayes factors are between 0.3 and 3). This suggests that the current implementation of RT-distance hypothesis needs improvement. Some possible solutions are considered in the discussion section.

Table 5 summarizes the correspondence between labels based on D2B and learned knowledge for Learned_BW and Learned_OR participants. Note that even though the D2B measure failed to identify the used knowledge for most participants, among the very few participants that it did, it was successful. Four Learned_BW



(a) (b)



(c)

Figure 17. The histogram of correlation between RT and distance to the ideal BW and OR bounds: (a) Learned_BW participants, (b) Learned_OR participants, (c) Learned_Both participants.

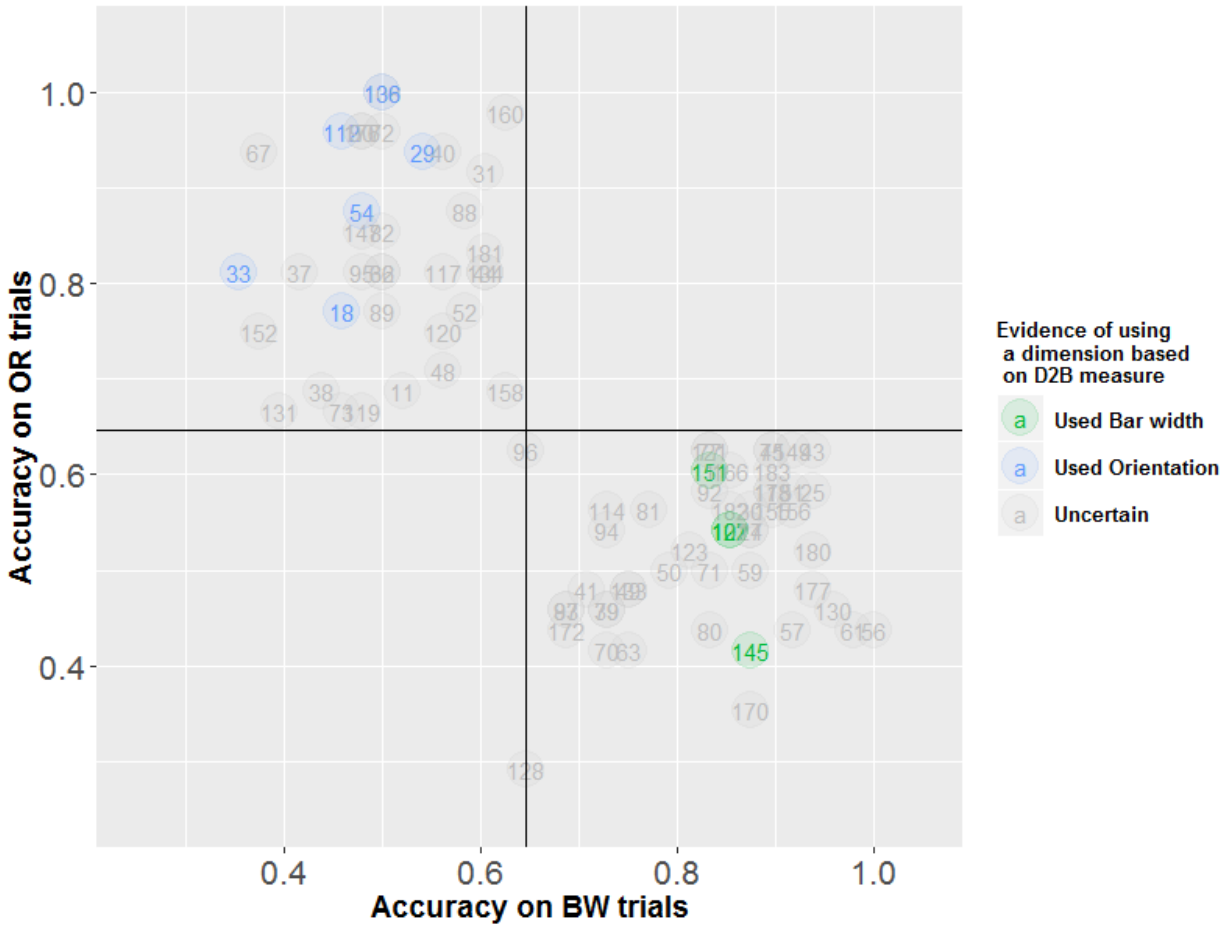


Figure 18. The identified strategies of Learned_BW and Learned_OR participants based on D2B. Each circle represents a participant. Color of a circle shows the used knowledge and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

Table 5

The Confusion Table for the Relation Between Identified Used Knowledge (Based on D2B) and Learned Knowledge for Participants That Learned Only One of the Dimensions

	Learned Knowledge	
Used Knowledge Based on D2B	Learned_BW	Learned_OR
Used Bar width	4	0
Used Orientation	0	6
Used Both	0	0
Uncertain	47	30
No D2B	0	0

Even though D2B is not able to identify almost any of the participants, just to be complete, the labels assigned to Learned_Both participants are shown in Figure 19. Again, D2B fails to identify most Learned_Both participants.

Table 6 shows the identified used knowledge of the Learned_Both participants based on D2B measure.

Combining Evidence From iDBM and D2B

The goal of this study is to identify the strategy of participants in a categorization task with redundancy. Accuracy and RT patterns were analyzed separately to identify the dimension(s) that were used by each participant. More errors close to a decision bound was considered evidence for a bound being used, quantified using iDBM. Slower RTs

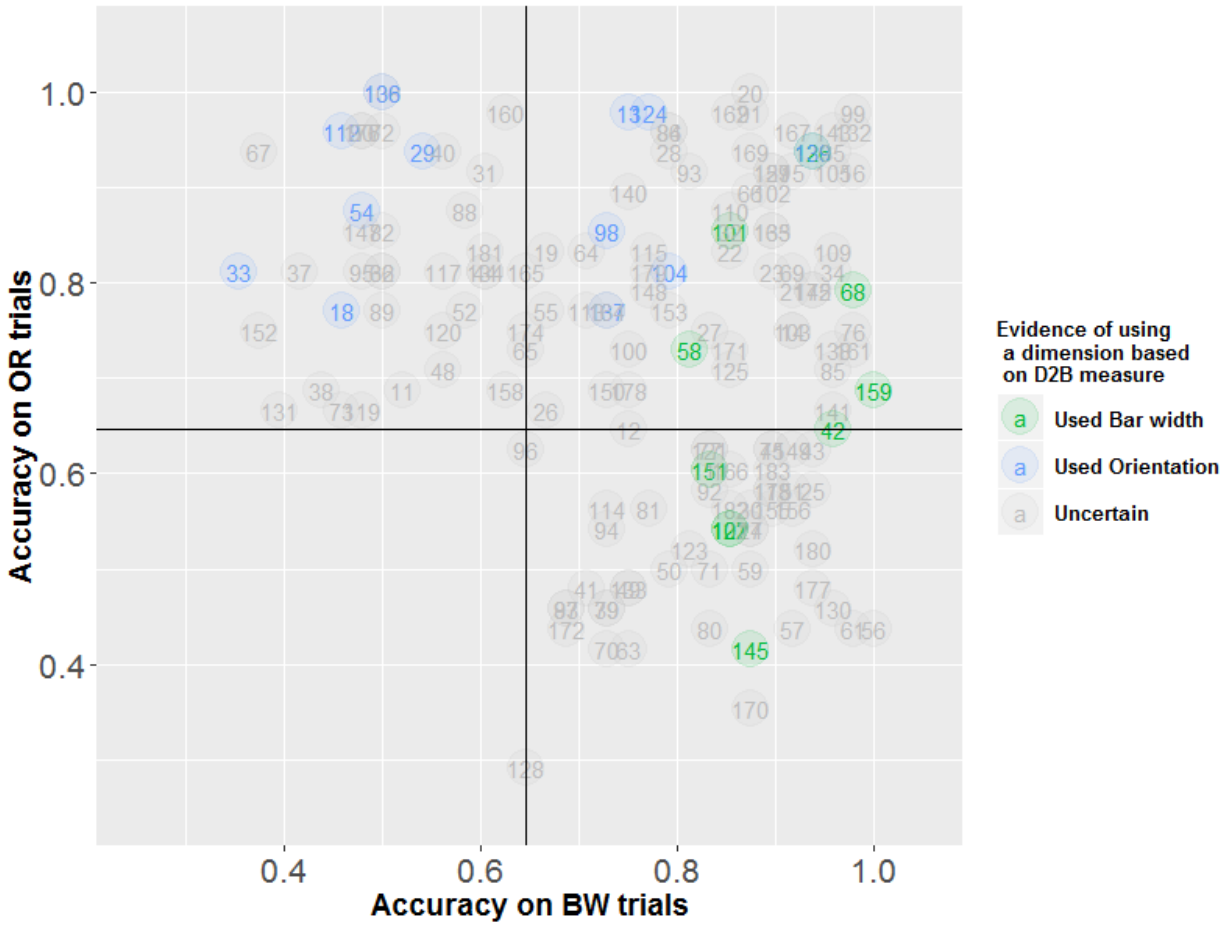


Figure 19. Each circle represents a participant. Color of a circle shows the used knowledge (based on D2B) and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

Table 6

Identified Used Knowledge (Based on D2B) for Participants That Learned Both Dimensions

Used Knowledge Based on D2B	Learned Knowledge Learned_Both
Used Bar width	6
Used Orientation	6
Used Both	0
Uncertain	60
No D2B	0

close to a bound was also considered evidence of a bound being used, which was quantified by computing rank correlations between RT and distance to each of the bounds. In this section, we combine the two in the following way: If any of the two mentioned characteristics exist for a bound, we conclude that the dimension was used. Since iDBM was more sensitive than D2B effect in detecting evidence, the used knowledge based on both iDBM and D2B effect is almost the same as the used knowledge based on only iDBM. Combined evidence map (shown in Figure 20) is very similar to Figure 16, where strategies were labeled based on only iDBM's output. Therefore, the assessment of the combined evidence is similar to what was discussed in the iDBM section.

Table 7 summarizes the correspondence between used knowledge (based on iDBM and D2B) and learned knowledge.

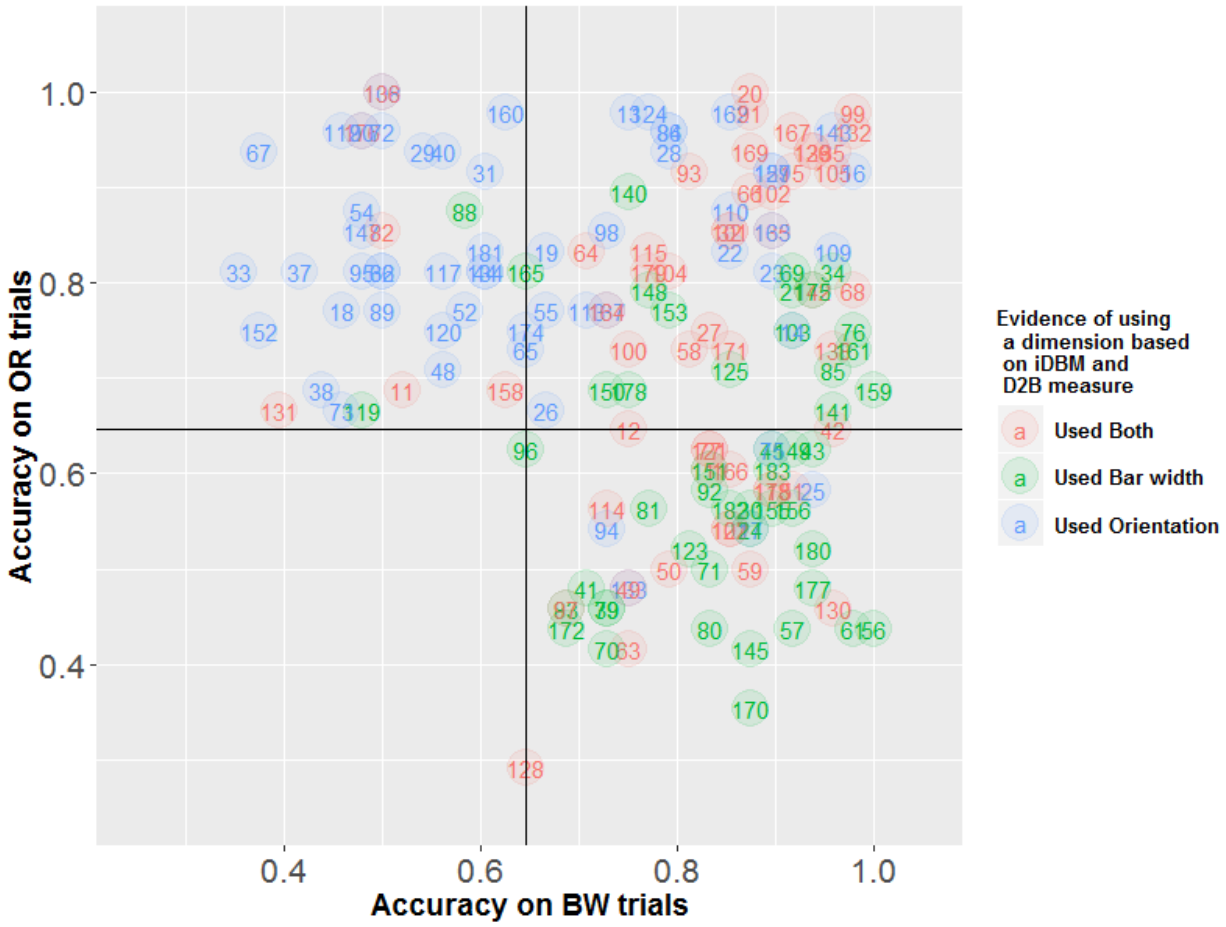


Figure 20. Each circle represents a participant. Color of a circle shows the used knowledge (based on iDBM and D2B) and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

Table 7

The Confusion Table for the Relation Between Used Knowledge (Based on iDBM and D2B Measure) and Learned Knowledge

	Learned Knowledge		
Used Knowledge on iDBM and D2B	Learned_Both	Learned_BW	Learned_OR
Used Both	31	20	6
Used BW	17	28	2
Used OR	24	3	28

ANALYSIS 2

The test phase of the experiment confirmed that in a two-dimensional categorization task with redundancy there are IDs in terms of which dimension(s) are learned: Some participants learned only bar width (Learned_BW participants), some learned only orientation (Learned_OR participants) and some learned both (Learned_Both participants). Analysis 1 showed that among participants that learned both dimensions, some used both, and some used only one of the dimensions (and learned the other dimension latently). Analysis 2 repeats Analysis 1 using different tools and goes a step further by dividing participants that used both dimensions into two different groups based on the specific way that each of them used the two dimensions. Consider an "A or B?" trial where one of the stimuli circled in Figure 21 is shown on the screen. The circled stimuli are far from a decision bound on the bar width dimension but are close to a decision bound on orientation, and therefore, the trial should be relatively easy for participants that are using only bar width and relatively difficult for participants using only orientation. However, it is possible to imagine that IDs exist among participants that use both dimensions, and these IDs might be reflected in the perceived difficulty of the circled trials.

Different two-dimensional strategies are distinguished by different difficulty maps in the category space (as discussed in the Hypothesis section). In the Methods section, models corresponding to the considered strategies are described and the strategy selection process is explained. Similar to Analysis 1, the credibility of the methods used in Analysis 2 is evaluated by checking whether Learned_BW and Learned_OR participants are identified as using unidimensional strategies and after making sure that the model

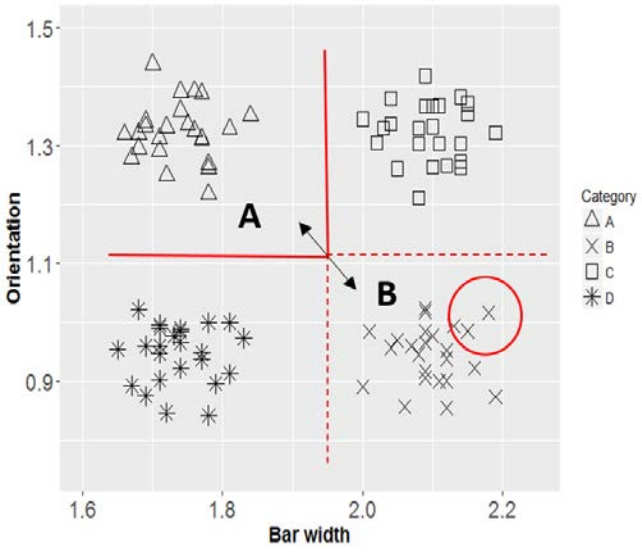


Figure 21. Analysis 2 distinguishes between different two-dimensional strategies. In an "A or B?" trial, the circled stimuli can be perceived as easy or difficult depending on participant's strategy.

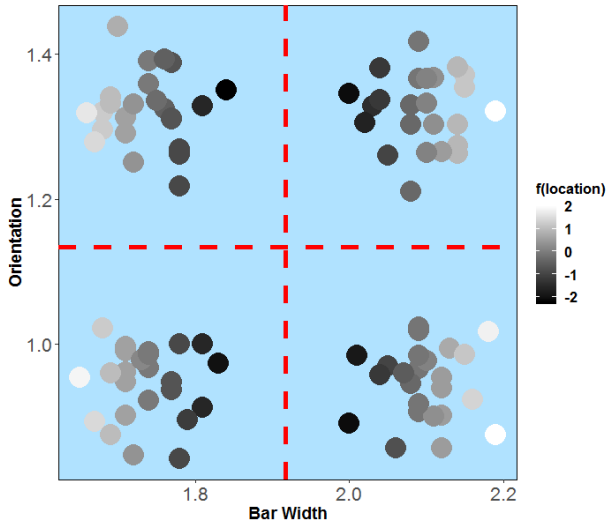
selection process is relatively successful, differences between two-dimensional strategies are explored.

Methods

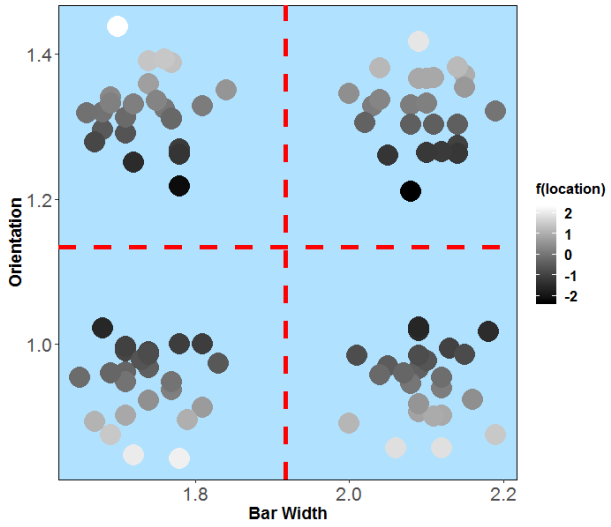
DDM was introduced over forty years ago (Ratcliff, 1978) but there is still ongoing research on parameter estimation methods (for a review: Ratcliff & Childers, 2015). In our analysis we used HDDM (Wiecki, Sofer, & Frank, 2013), a Python-based package that uses hierarchical Bayesian modeling. HDDM allows adding covariates to DDM parameters, which makes it possible to assess the effect of trial-by-trial variability on the parameters. The trial-by-trial variability that we consider is going to be a function of the location of a stimulus on the category space. We hypothesize that perceived difficulty of a stimulus depends on the strategy of a participant. Since trial difficulty is captured by the drift rate (v), the trial-by-trial variability measure (determined by the location of the stimulus) will be regressed on the drift rate:

$$v = v_0 + \beta \times f(location\ of\ stimulus)$$

The first step is to determine $f(location\ of\ stimulus)$ (i.e., the covariate), which is going to be different for each strategy. Figure 22 shows two covariate maps corresponding to the unidimensional strategies. Figure 22a corresponds to a unidimensional strategy on bar width and Figure 22b corresponds to a unidimensional strategy on orientation. In Figure 22a, the covariate is smallest close to an ideal bound on bar width dimension, and it increases further away from the bound. Similarly, in Figure 22b, the smallest covariate values are those close to an ideal bound on orientation dimension and it increases further away from the bound. The exact values are



(a)



(b)

Figure 22. The covariate maps expected to fit best to participants with a unidimensional strategy. (a) Unidimensional strategy on bar width. (b) Unidimensional strategy on orientation.

standardized distances to an optimal bound on each of the dimensions. The models corresponding to Figure 22a and Figure 22b will be referred to as BW model and OR model.

Figure 23 shows the covariate maps of two-dimensional strategies. Figure 23a shows the 'Time efficient' strategy and Figure 23b shows the 'Conservative' strategy (more explanation was provided in the hypothesis section). The covariates of the two-dimensional strategies are calculated in the following way:

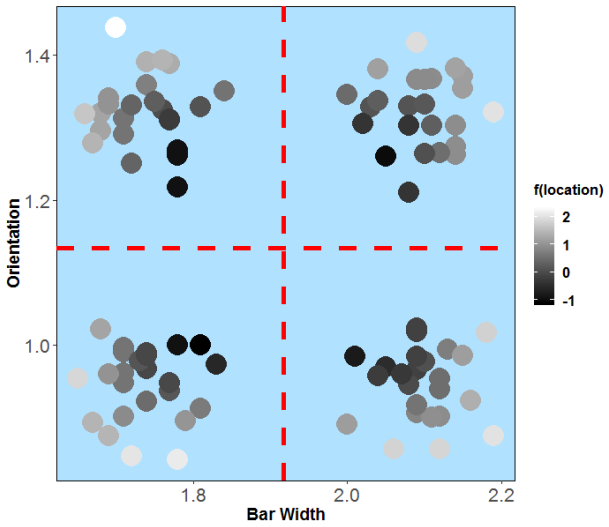
$$f_{Time\ efficient}(location\ of\ stimulus) =$$
$$Max\left(f_{BW}(location\ of\ stimulus)\ , f_{OR}(location\ of\ stimulus)\right)$$
$$and$$
$$f_{Conservative}(location\ of\ stimulus) =$$
$$Min(f_{BW}(location\ of\ stimulus)\ , f_{OR}(location\ of\ stimulus))$$

Where $f_{BW}(location\ of\ stimulus)$ and $f_{OR}(location\ of\ stimulus)$ correspond to covariates of the unidimensional strategies on bar width and orientation respectively.

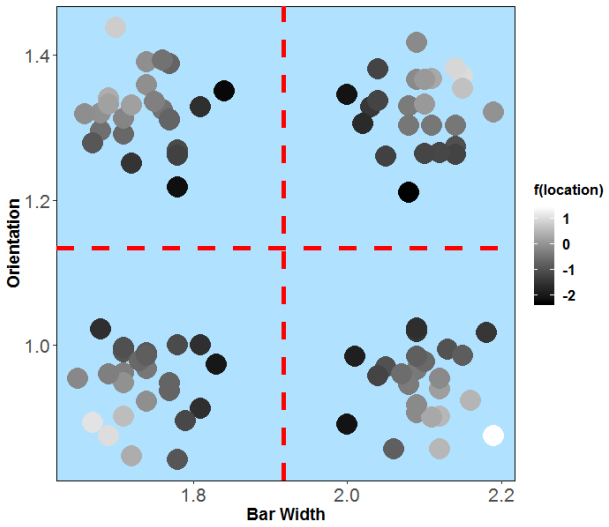
In addition to the four covariate maps, a DDM with no covariate (i.e., $\beta = 0$) was also fitted, which is expected to fit best to participants for whom the relative difficulty of a trial does not depend on its location in the category space. The five DDMs were fitted to the last three blocks of training of each participant separately and using a model selection criterion (described below), the best fitting model(s) were identified.

Model Selection Process

Two measures were used to assess the best fitting model(s), the DIC score and percentage of posterior samples of β that are bigger than zero. DIC is a measure similar to AIC, which is used when the model fitting is done with Bayesian methods and posterior



(a)



(b)

Figure 23. The covariate maps expected to fit best to participants that used both dimensions. (a) Time efficient strategy. (b) Conservative strategy.

samples of parameters are available (Spiegelhalter et al., 2002). Similar to AIC, a DIC score is a goodness of fit measure that penalizes the number of parameters but unlike AIC, it is not possible to translate the scores to relative probabilities (i.e., the probability that a model provides the best description for the data among the considered models). However, it seems that the rule of thumb used in AIC works for DIC scores as well (Spiegelhalter et al., 2002), which is to consider all models that are within 1 point of the best model (i.e., the model with lowest DIC) to be relatively good. For example, if the BW model has the lowest DIC, and $DIC_{Time\ efficient}$ - $DIC_{BW} < 1$, then both models are selected. In addition to DIC scores, the posterior samples of β parameter for each model was tested and a model is considered only if 99% of its β samples are greater than zero.

To summarize, a model is picked if (a) it has the smallest DIC score or its score is within 1 point of the minimum and (b) if its β parameter is positive with a probability of 99% or higher. If none of the four models with covariates satisfy $p(\beta > 0) > 0.99$ and the model with no covariate does not have a small enough DIC (i.e., $DIC_{No\ covariate} - DIC_{min} > 1$), then Analysis 2 does not assign a strategy to that participant.

Results

Similar to Analysis 1, before looking at all participants, validity of the analysis is tested by looking at participants that learned only one of the bounds, which must be best fit by the models that correspond to the unidimensional strategies, or to the model with no covariate. Figure 24 shows the result of the model selection. As before, each circle represents a participant, the location of the circle codes the learned knowledge and the color of each circle is the identified strategy. Most Learned_BW participants (bottom right corner of the figure) are either green (i.e., best fit by the model with distance to an

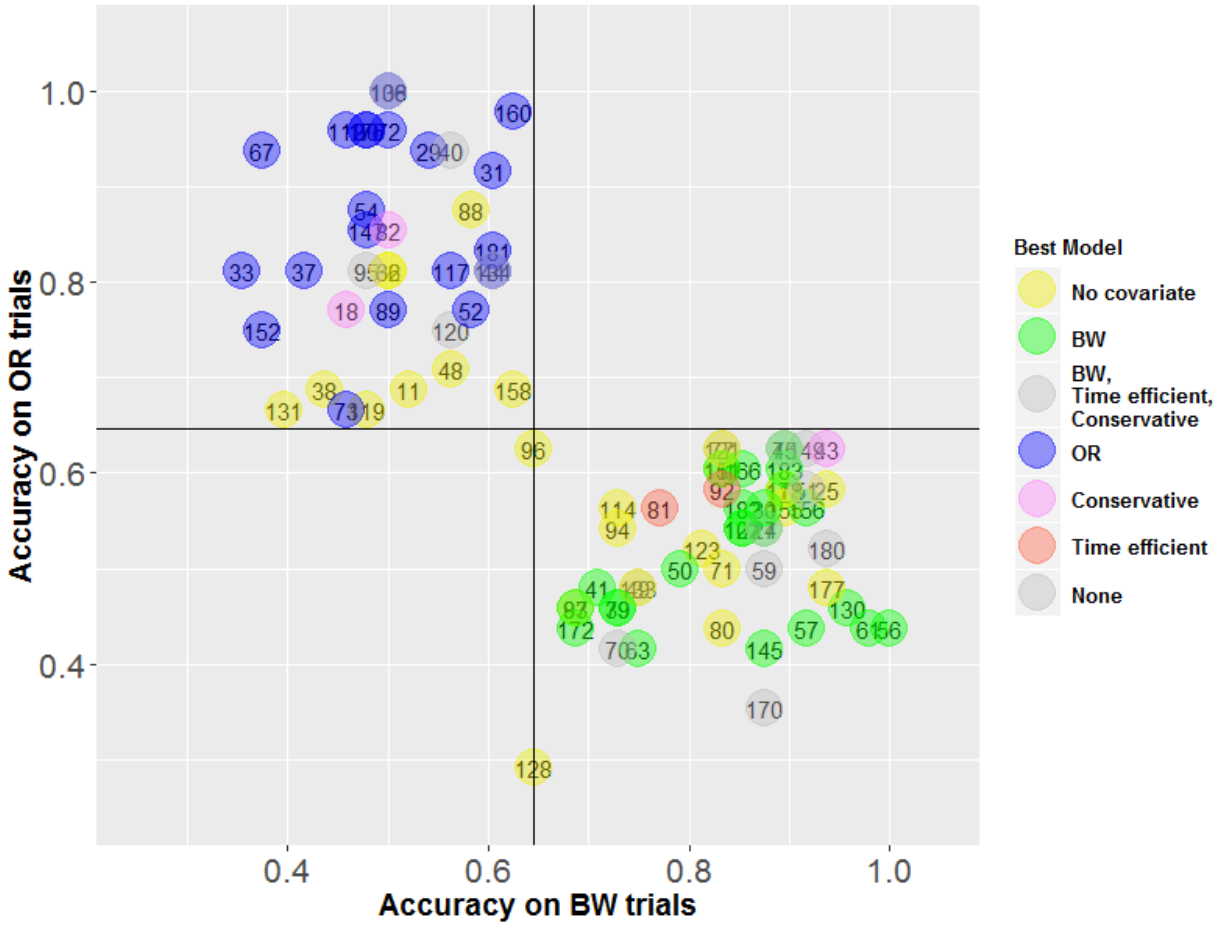


Figure 24. The identified strategies of Learned_BW and Learned_OR participants based on Analysis 2. Each circle represents a participant. Color of a circle shows the strategy (based on Analysis 2) and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

ideal bound on bar width as covariate) or gold (i.e., best fit by the model with no covariate). Similarly, most Learned_OR participants (top left corner of the figure) are either blue (i.e., best fit by the model with distance to an ideal bound on orientation as covariate) or gold (i.e., best fit by the model with no covariate).

Table 8 summarizes the correspondence between labels based on DDM models and the learned knowledge for participants that learned only one dimension. Excluding participants that were not fit to any of the models (the 'None' strategy), there are only five participants (out of 72) that are best fit to the model that corresponds to a wrong strategy, which shows that the strategy identification process works well. There was one participant for whom there were three well-fitting strategies (BW, Time efficient and Conservative). This participant was colored gray (same as the 'None' participants) in Figure 22, because if the model selection process says three out of five models are describing the data well, then in some sense it is not different than saying none of them fits the data well.

Now that the validity of model selection process is demonstrated, the strategy of participants that learned both dimensions is shown in Figure 25. Theoretically, Learned_Both participants can belong to any of the considered strategies, it is possible that they used only one of the dimensions while latently learning the other, or it is also possible that they used both dimensions. In order to make the figure easier to read, in cases where two models are selected and one of them is Time efficient, the color is red, same as cases where there is only one best fitting model and it is Time efficient. Similarly, in cases where two models are selected and one of them is Conservative, the color is pink, same as cases where there is only one best fitting model and it is

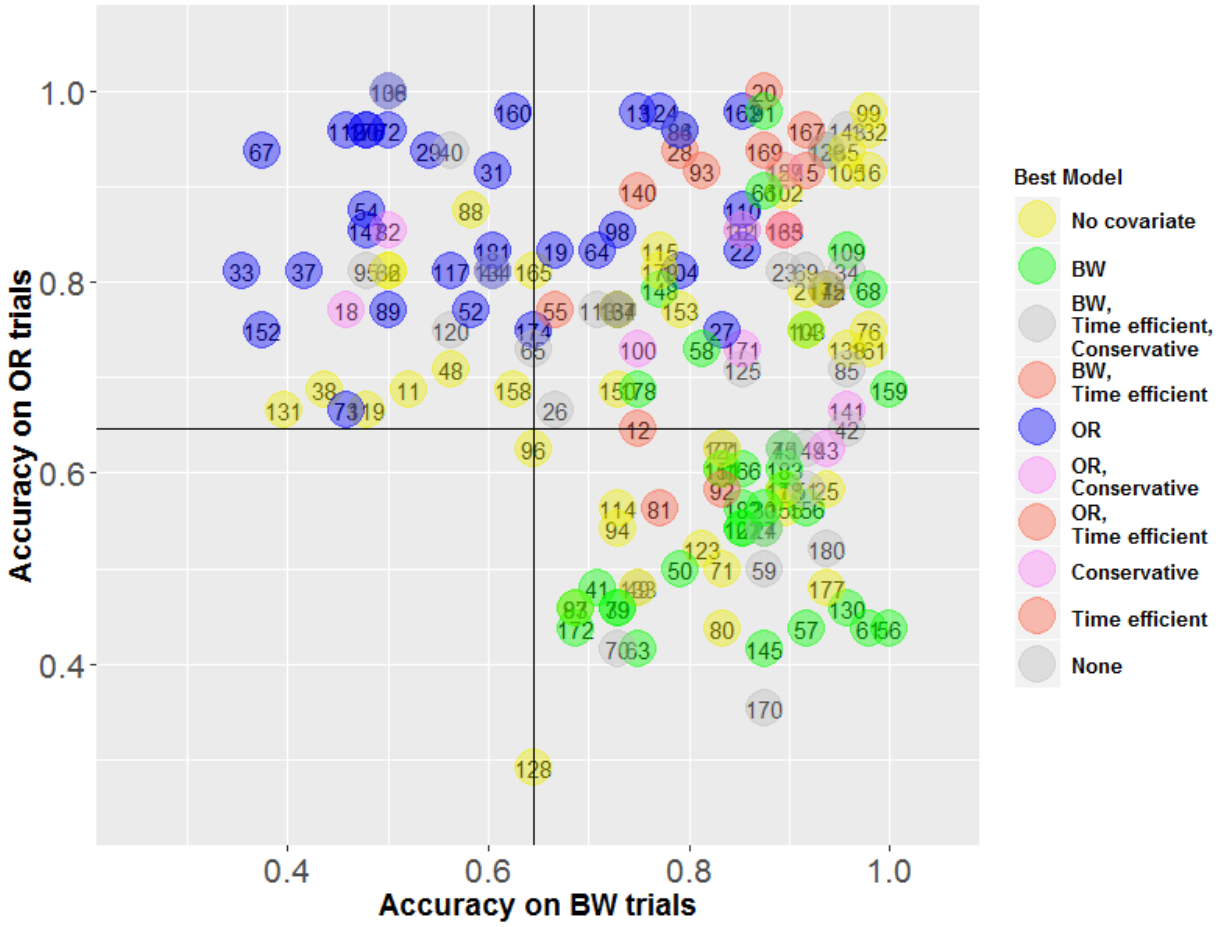


Figure 25. The identified strategies participants based on Analysis 2. Each circle represents a participant. Color of a circle shows the strategy (based on Analysis 2) and its location shows the learned knowledge (x-axis is test accuracy on BW trials and y-axis is test accuracy on OR trials).

Table 8

The Confusion Table for the Relation Between Identified Strategy (Based on Analysis 2) and Learned Knowledge for Participants That Learned Only One of the Dimensions

	Learned Knowledge	
Strategy	Learned_BW	Learned_OR
Unidimensional, BW	23	0
Unidimensional, OR	0	20
No Covariate	15	9
Two-Dimensional Strategies	3	2
None	10	5

Conservative. Participants best fit by the No covariate model were hypothesized to be the 'elite' participants that perceive stimuli close to any of the two boundaries no harder than other stimuli. However, Figure 25 shows that based on the learned knowledge of each participant (coded by the location of each participant in the figure), it seems that not all of the No covariate participants (i.e., gold circles) belong to the 'elite' group. It is true that there are four or five gold circles in the top right corner of the figure, but there are also gold circles in the regions that learned knowledge level is not particularly high on neither of the dimensions (e.g., participants 131, 48, 96 and 128). A possible explanation is that the less engaged participants that may have noisy data do not benefit from the extra parameter (β) and therefore are best fit by the No covariate model simply because this

model has fewer parameters compared to the other four models. Therefore, the No covariate can be thought of as a generic model that fits best to two type of participants: The elite participants that perceive all stimuli equally easy irrespective of the location of a stimulus and participants whose strategy is not well-described by the other models (BW, OR, Time efficient and Conservative).

Table 9 shows the identified strategies of the Learned_Both participants. In cases where there were two best fitting models and one of them was Time efficient, the participant was counted as Time efficient and similarly, in cases where there were two best fitting models and one of them was Conservative, the participant was counted as Conservative.

Table 9

Identified Strategy (Based on Analysis 2) for Participants That Learned Both Dimensions

	Learned Knowledge
Strategy	Learned_Both
Unidimensional, BW	11
Unidimensional, OR	13
No covariate	18
Time efficient	11
Conservative	7
None	12

GENERAL DISCUSSION

The thesis focused on individual differences in a rule-based categorization task with redundancy. A central emphasis was to show that participants might achieve success in different ways and therefore, they have to be divided properly before doing any group analysis. The reason is that different types of participants may have different response patterns, which is ignored if all participants are pooled together. Participants were divided based on their learned knowledge and their used knowledge. The learned knowledge of each participant was determined by adding a test phase to the experiment. Identifying the learned knowledge was relatively straightforward and the main challenge was to identify the used knowledge, which was done in two different ways (Analysis 1 and Analysis 2). In this section, different aspects of the two analyses are compared and there is a discussion on individual differences that are ignored by both analyses.

Comparing the Analyses

The two analyses are compared from two different perspectives. First, a simple comparison is made based on how well each of them identified used knowledge/strategy of participants. Then the details of the method implementations are compared and the way the details of implementation has affected the results are discussed.

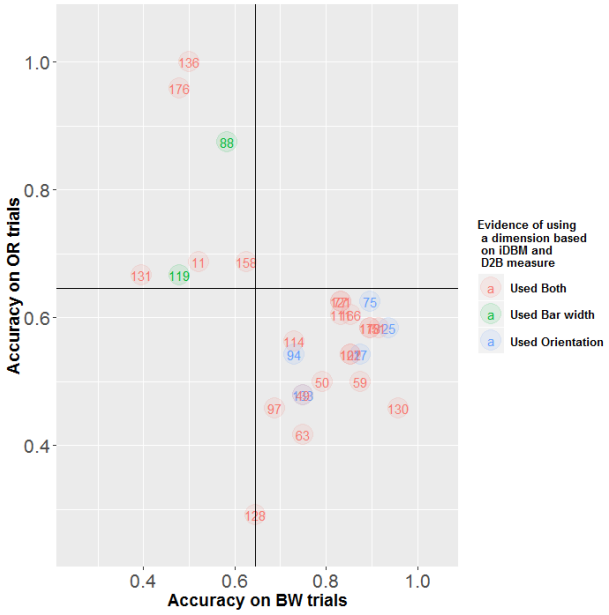
Comparing the Results

Analyses 1 and 2 were concerned with identifying the used knowledge of participants. The test phase of the experiment established the learned knowledge of each participant, but it is not possible to assess the validity of identified used knowledge based on the learned knowledge, because the participants that learned both dimensions could use any of the two dimensions. However, participants that learned only one dimension

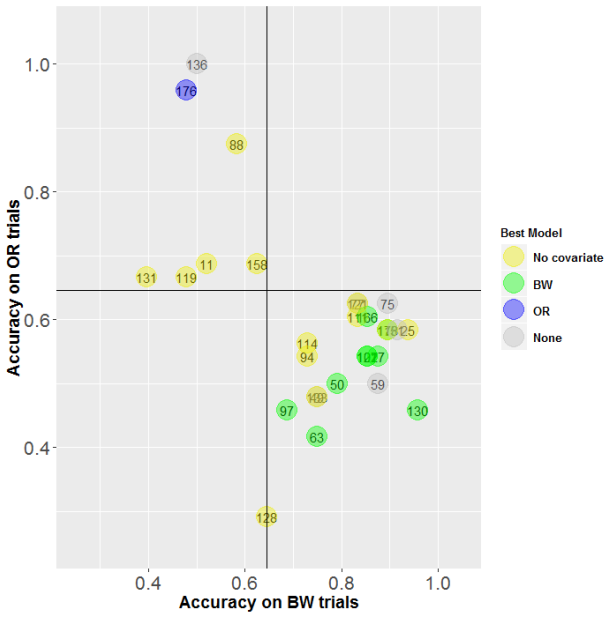
must have used only that dimension, and for this reason, Learned_BW and Learned_OR participants were used to partially validate the results of Analysis 1 and 2. There were eighty-seven unidimensional (Learned_BW and Learned_OR) participants, and fifty-three of them were identified correctly by both analyses. There was also one participant that was misidentified by both analyses. In order to compare differences in the errors made by each analysis, the focus is on participants that were misidentified by only one of the analyses. Participants that were identified by Analysis 2 as 'None' or 'No covariate' are not considered misidentification. The reason is that 'No covariate' can be viewed as a generic model that theoretically can represent any strategy and 'None' simply shows that there is no good fit.

Figure 26 shows participants that are misidentified by Analysis 1 but not by Analysis 2. Figure 26a shows the labels assigned by Analysis 1 and Figure 26b shows the same participants and the labels assigned by Analysis 2. There are twenty-nine participants that are misidentified by Analysis 1 and Analysis 2 assigns nineteen out of the twenty-nine to the more obscure models ('None' and 'No Covariate') and the remaining ten to the correct models (BW and OR). This shows that Analysis 2 avoids misidentification of some participants because it can assign them to no model at all (i.e., 'None' label) or to a generic model ('No covariate').

Figure 27 shows the participants that are misidentified by Analysis 2 but not by Analysis 1 and as the figure shows, there are only four such participants. Figure 27a shows the labels assigned by Analysis 1 and Figure 27b shows the same participants and the labels assigned by Analysis 2.

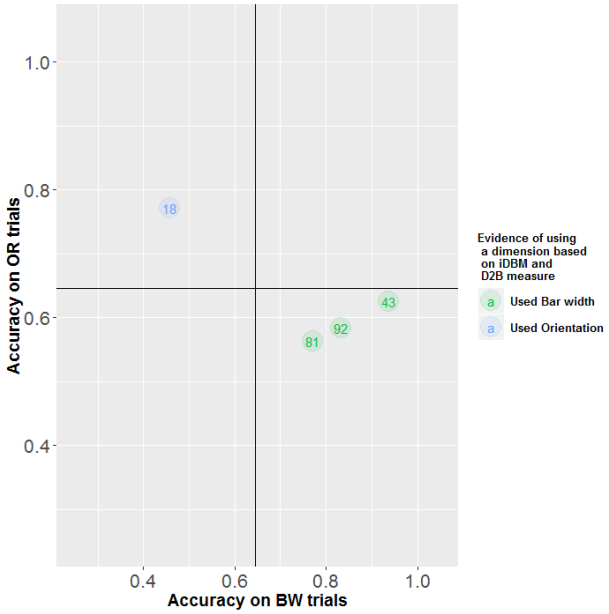


(a)

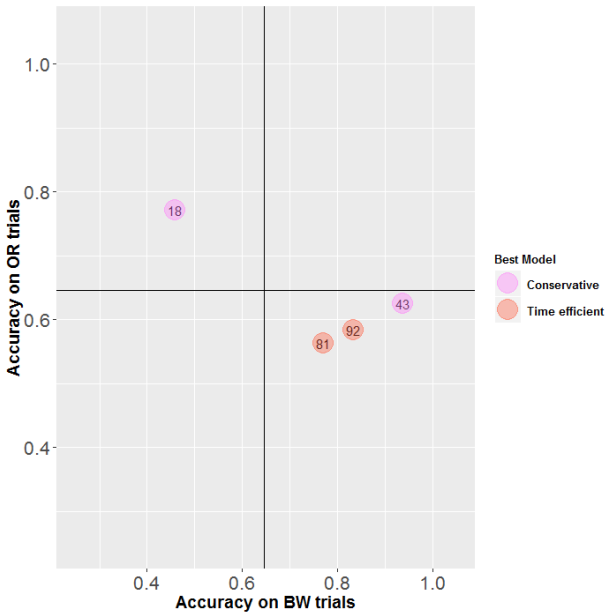


(b)

Figure 26. Participants that were misidentified by Analysis 1. (a) Labels based on Analysis 1. (b) Labels based on Analysis 2.



(a)



(b)

Figure 27. Participants that were misidentified by Analysis 2. (a) Labels based on Analysis 1. (b) Labels based on Analysis 2.

Comparing the misidentified participants showed that Analysis 2 made fewer mistakes and it is partly because of the possibility of assigning participants that do not fit any of the unidimensional and two-dimensional strategies to a generic model (No covariate) or to no model at all (None). In subsequent sections, there is a discussion on why having models similar to 'No covariate' and 'None' is beneficial to any model selection scheme.

Comparing the Implementation

This section compares the differences in the implementation between the two analyses and discusses the advantages and disadvantages of each. Analysis 1 used two tools (iDBM and D2B) that were implemented separately and therefore will be discussed separately.

The input data and trial order. There are five training blocks (each block is ninety-six trials) and iDBM uses all of them and considers the order of trials, while D2B measure and Analysis 2 use the last three blocks and ignore the order of trials. The reason for this difference is that the goal of iDBM is to detect switches in participant's response pattern, from guessing to using a rule and possible subsequent switches. Therefore, iDBM needs to start from the beginning and has to consider the order of trials. On the other hand, D2B measure and Analysis 2 aspire to detect a quality about a participant's response pattern as a whole, without considering the changes in strategy throughout the experiment. For this reason, D2B measure and Analysis 2 used the last three blocks in order to exclude the messy data from the early stages of the experiment when the participant does not have a stable strategy yet. The advantage of the approach used by iDBM is that it does not make the additional assumption that all participants settle on a

strategy, and the advantage of the approach used in D2B measure and Analysis 2 is its computational simplicity (no need to fit models iteratively). The D2B measure assigned most participants to the 'Uncertain' category, which might have been due to ignoring the order of trials. It has been shown that as the task becomes lest effortful, the D2B effect diminishes (Hélie, Waldschmidt, & Ashby, 2010) and therefore, ignoring the trial order is possibly a reason for the low detection power of the current implementation of RT-distance hypothesis. Imagine a participant that shows D2B effect in blocks three and four but as the task becomes easier for her, the negative correlation between RT and distance to bound diminishes and by block five there is no D2B effect. The lack of D2B effect at block five reduces the magnitude of overall D2B and therefore, the probability of detecting it is reduced. Note that even though all the mentioned problems exist for Analysis 2, it did well, and arguably, Analysis 2 was more successful than Analysis 1 (based on the previous section). One reason is that while D2B effect used a simple rank correlation coefficient, Analysis 2 uses a more sophisticated method with lots of parameters, and while D2B effect used only RT, Analysis 2 used both accuracy and RT. However, note that even in Analysis 2 there are many participants that the model selection process was not able to assign to any strategy, which might change if an iterative version of Analysis 2 were to be implemented.

The model space. Each analysis considered a set of models (model space) and assigned each participant to one of the models. Table 10 shows the model space of each analysis. As Table 10 shows, there are two differences between Analysis 1 and 2. The first difference is that in Analysis 2, there are two models that correspond to a two-dimensional strategy (Time efficient and Conservative), while in Analysis 1 there is only

one such model (Used both). In other words, Analysis 2 is a little more specific compared to Analysis 1 and distinguishes between different two-dimensional strategies. The second difference is that Analysis 2 has two models that are labeled as 'Other' in Table 10, which include 'No covariate' and 'None'. The 'Other' category covers a) participants who are using a strategy that cannot be formulated by BW, OR, Time efficient or Conservative and b) participants that for various reasons (e.g., not being attentive enough to the experiment) have noisy data.

Table 10

Model Space of Each Analysis

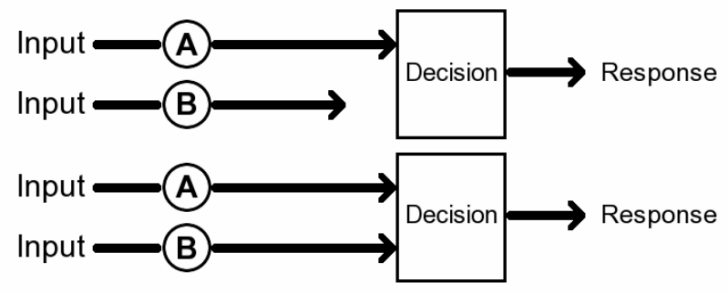
	Analysis 1	Analysis 2
Unidimensional Strategies	Used Bar Width	BW
	Used Orientation	OR
Two-Dimensional Strategies	Used Both	Time Efficient
		Conservative
Other	—	No Covariate
		None

The possibility of assigning participants to a generic model (i.e., 'No covariate') or no model at all (i.e., 'None') reduces misidentifications. Comparing the results of the two analyses confirmed this claim by showing that Analysis 1 makes more mistakes than Analysis 2 and the majority of participants misidentified by Analysis 1 are identified as either 'No covariate' or 'None' by Analysis 2. In general, for any model selection

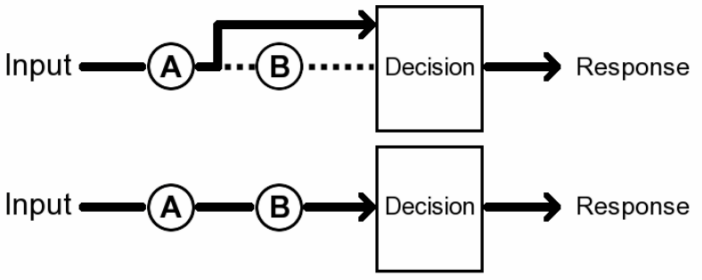
scheme having models similar to ‘No covariate’ and ‘None’ can be beneficial. Such models cover possibilities that are not conceptually considered and when most cases are assigned to none of the defined models, it may be an indication of shortcomings in the implementation or may simply suggest that the data is too noisy.

More IDs?

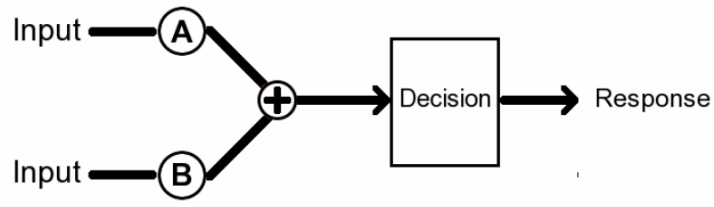
The study assumed two type of IDs in a categorization task with redundancy: learned knowledge and used knowledge. The test phase of the experiment showed that there are IDs in the learned knowledge and Analysis 1 and 2 attempted to identify the IDs in used knowledge. There seems to be no conceptual ambiguity about dividing participants into three groups based on the learned knowledge (Learned_BW, Learned_OR and Learned_Both). However, used knowledge is not a specific enough term and does not determine how the two dimensions are combined. Systems Factorial Technology (SFT) is a framework that is formulated to study how information from different sources are combined (Townsend & Nozawa, 1995). Analysis 1 and 2 are reassessed using SFT in order to evaluate whether participants were properly divided or not. SFT uses four characteristics to describe a two-dimensional process: architecture (serial or parallel), stopping rule (AND or OR), stochastic dependence (dependence or independence) and workload capacity (limited, unlimited or super capacity). Discussing the details of each characteristic and their mathematical formulation is not the intention of this section and the goal is to use SFT to examine what is lacking in Analysis 1 and 2. Figure 28 shows the schematic of five different SFTs. The two dimensions can be processed parallel (Figure 28a and 28c) or serial (Figure 28b) and a response can be made as soon as a target is detected (OR processing; top panel of Figure 28a and 28b) or after



(a)



(b)



(c)

Figure 28. a) Top panel: Parallel, OR processing. Bottom panel: Parallel, AND processing. b) Top panel: Serial, OR processing. Bottom panel: Serial, AND processing. c) Coactive model (special case of parallel architecture). Figure is taken from Houpt, Blaha, McIntire, Havig, & Townsend (2014).

processing all available information (AND processing; bottom panel of Figure 28a and 28b).

It is not possible to situate Analysis 1 within the SFT framework, because Analysis 1 determines whether both dimensions were used or not and does not imply anything about the way two dimensions are used. On the other hand, Analysis 2 implicitly assumes a coactive architecture (i.e., Figure 28c) in its two-dimensional models by fitting only one evidence accumulator. Figure 29 shows the architecture assumed by Analysis 2. Time efficient and Conservative strategies differed in the way two dimensions were combined.

A coactive architecture is not the only possible way to implement Time efficient and Conservative strategies. Two alternative architectures are shown in Figure 30. Figure 30a shows a parallel OR processing model and Figure 30b shows a parallel AND processing model, which can be viewed as alternative implementations of Time efficient and Conservative strategies respectively. Another possibility (not visualized) is serial OR and serial AND models.

The importance of having well-defined strategies and properly dividing participants into them is that a valid group analysis depends on it. This section showed a possible improvement using a framework (SFT) that is specifically designed to model two-dimensional processes. Even though the current implementation of Analysis 2 lacks a specification of the architecture and implicitly assumes a coactive architecture while ignoring other possibilities, it is still possible to perform some group analysis. Three group analyses are done in Appendix comparing difficulty of two dimensions, and testing the effect of latently learning a dimension.

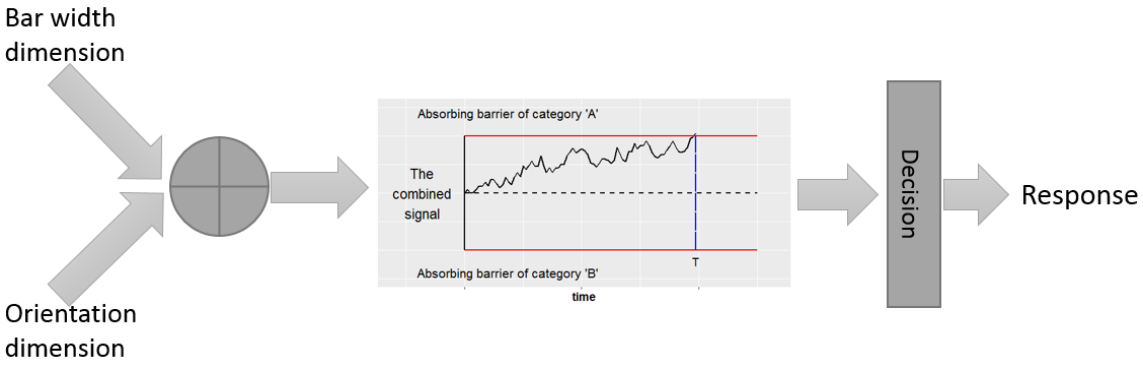
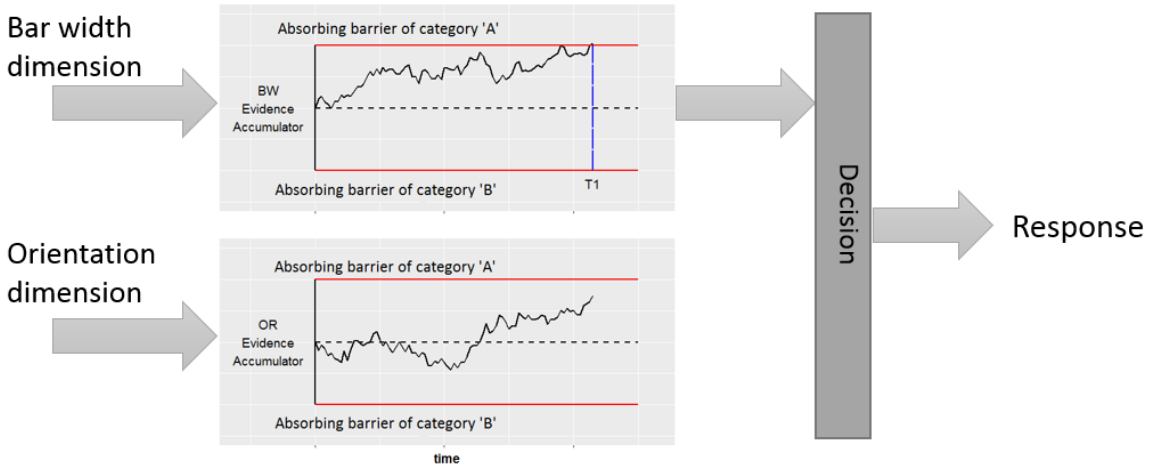
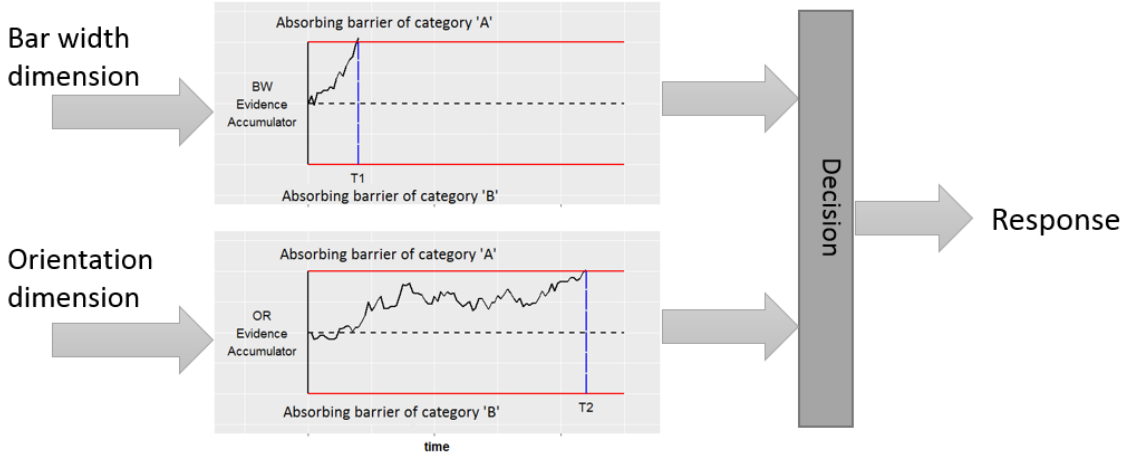


Figure 29. Visualization of two-dimensional model implemented in Analysis 2.



(a)



(b)

Figure 30. Alternative architectures to model two-dimensional strategies in Analysis 2. a) An alternative model for Time efficient strategy. b) An alternative model for conservative strategy.

Future Work

A possible direction is to improve implementation of current methods. iDBM will benefit from having generic models such as those defined for D2B and Analysis 2, in order to avoid misidentifications. On the other hand, D2B and Analysis 2 will benefit from the iterative characteristic of iDBM. Ignoring the initial phase of category learning and ignoring the order of trials results in missing changes in participants' decision-making behavior and it was probably the reason for D2B's failure and the abundance of participants that could not be described by any of Analysis 2's models. The next step beyond improving current methods would be to consider different architectures for two-dimensional strategies using SFT framework. The benefit would be having strategies that are defined more accurately and the possibility to use the insights of the SFT framework.

A non-methodological topic that can be investigated is the effect of time: Given enough time, will all participants learn both dimensions? If yes, will all eventually use the same strategy? Or will there be differences in the final strategy that participants settle on?

This thesis was concerned with formulating IDs and analyzing *how* participants are different, but did not investigate *why* these differences exist. One possible reason could be that participants that learned only one dimension are better at inhibiting task irrelevant information in general and therefore, after starting to test one of the dimensions and finding out that it works, filtered the other dimension completely and never noticed the differences between categories in the other dimension. It is also possible that IDs are not linked to any general cognitive characteristic of participants. Finally, it is worth mentioning that this study was limited to two-alternative forced choice task and studying

IDs in a categorization task with redundancy can be done in different settings such as Yes/No tasks or forced choice tasks with more than two alternatives.

LIST OF REFERENCES

Ashby, F. G. (1992). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1034). Hillsdale, NJ: Erlbaum.

Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44*(2), 310-329.

Ashby, F. G., Alfonso-Reese, L. A., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442-481.

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*(1), 33-53.

Ashby, F. G., & Maddox, W. T. (1991). A response time theory of perceptual independence. In *Mathematical psychology* (pp. 389-413). New York, NY: Springer.

Ashby, F. G., & Maddox, W. T. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38*(4), 423-466.

Ashby, F. G., & O'Brien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Sciences, 9*(2), 83-89.

Ashby, F. G., & Perrin, N. A. (1988). Toward a unified theory of similarity and recognition. *Psychological Review, 95*(1), 124-150.

Ashby, F. G., & Soto, F. A. (2015). Multidimensional signal detection theory. In J. R. Busemeyer, Z. Wang, J. T. Townsend, & A. Eidels (Eds.), *Oxford library of psychology. Oxford handbook of computational and mathematical psychology* (pp. 13-34). New York, NY: Oxford University Press.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*(2), 154-179.

Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., . . . Jones, Z. M. (2016). mlr: Machine Learning in R. *The Journal of Machine Learning Research, 17*(1), 5938-5942.

Bornstein, M. H., & Monroe, M. D. (1980). Chromatic information processing: Rate depends on stimulus location in the category and psychological complexity. *Psychological Research, 42*(3), 213-225.

Brainard, D. H., & Vision, S. (1997). The psychophysics toolbox. *Spatial Vision, 10*, 433-436.

Broadbent, D. E. (2013). *Perception and communication*. Amsterdam, The Netherlands: Elsevier.

Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science, 15*(3), 118-121.

Burns, B., & Shepp, B. E. (1988). Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception & Psychophysics, 43*(5), 494-507.

Busemeyer, J. R., & Wang, Y. M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, *44*(1), 171-189.

Cartwright, D. (1941). Relation of decision-time to the categories of response. *The American Journal of Psychology*, *54*(2), 174-196.

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*(11), 1462-1467.

Craig, S., & Lewandowsky, S. (2012). Whichever way you choose to categorize, working memory helps you learn. *The Quarterly Journal of Experimental Psychology*, *65*(3), 439-464.

DeCaro, M. S., Thomas, R. D., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition*, *107*(1), 284-294.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 968-986.

Ell, S. W., Smith, D. B., Peralta, G., & Hélie, S. (2017). The impact of category structure and training methodology on learning and generalizing within-category representations. *Attention, Perception, & Psychophysics*, *79*(6), 1777-1794.

Erickson, M. A. (2008). Executive attention and task switching in category learning: Evidence for stimulus-dependent representation. *Memory & Cognition*, *36*(4), 749-761.

Garner, W. R. (2014). *The processing of information and structure*. Hove, East Sussex: Psychology Press.

Hélie, S., Shamloo, F., & Ell, S. W. (2017). The effect of training methodology on knowledge representation in categorization. *PloS one*, *12*(8), e0183904.

Hélie, S., Turner, B. O., Crossley, M. J., Ell, S., & Ashby, F. G. (2017). Trial-by-trial identification of categorization strategy using iterative decision bound modeling. *Behavior Research Methods, 49*, 1146-1162.

Hélie, S., Waldschmidt, J. G., & Ashby, F. G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics*, *72*(4), 1013-1031.

Hélie, S., Roeder, J. L., & Ashby, F. G. (2010). Evidence for cortical automaticity in rule-based categorization. *Journal of Neuroscience, 30*(42), 14225-14234.

Houpt, J. W., Blaha, L. M., McIntire, J. P., Havig, P. R., & Townsend, J. T. (2014). Systems factorial technology with R. *Behavior Research Methods, 46*(2), 307-330.

Kidd, E., Donnelly, S., & Christiansen, M. H. (2017). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences, 22*(1), 154-169.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22-44.

Kruschke, J. K. (2005). Category learning. In K. Lamberts & R. L. Goldstone (Eds.), *The handbook of cognition* (pp. 183-201). London, England: Sage.

Levering, K. R., & Kurtz, K. J. (2015). Observation versus classification in supervised category learning. *Memory & Cognition, 43*(2), 266-282.

Levinson, S. C. (2012). The original sin of cognitive science. *Topics in Cognitive Science, 4*(3), 396-403.

Lewandowsky, S. (2011). Working memory capacity and categorization: Individual differences and modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(3), 720-738.

Maddox, W. T., Ashby, F. G., & Waldron, E. M. (2002). Multiple attention systems in perceptual categorization. *Memory & Cognition, 30*(3), 325-339.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes, 66*(3), 309-332.

Markman, A. B. (2013). *Knowledge representation*. Hove, East Sussex:Psychology Press.

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. *Journal of Experimental Psychology: General, 143*(2), 668-693.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*(1), 37-50.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207-238.

Minda, J. P., Desroches, A. S., & Church, B. A. (2008). Learning rule-described and non-rule-described categories: A comparison of children and adults. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1518-1533.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8-14.

Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory & Cognition*, *13*(6), 511-521.

Murphy, G. L., & Medin, D. L. (1985). The roles of theories in conceptual coherence. *Psychological Review*, *92*, 289-316.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59-108.

Ratcliff, R., & Childers, R. (2015). Individual differences and fitting methods for the two-choice diffusion model of decision making. *Decision*, *2*(4), 237-279.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873-922.

Ratcliff, R., Perea, M., Colangelo, A., & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain and Cognition*, *55*(2), 374-382.

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2003). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, *18*(3), 415-429.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of Statistics*, *6*(2), 461-464.

Shepard, R. N. (1981). *Discrimination and classification: A search for psychological laws*. Presidential address to the Division of Experimental Psychology of the American Psychological Association, Los Angeles, CA.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied, 75*(13), 1-42.

Singmann, H., Brown, S., Gretton, M., & Heathcote, A. (2016). *rtdists: Response Time Distributions*. R package version 0.6-6.

Skinner, J. (2014, July 22). *There can be no crisis of psychoanalysis*. Retrieved from http://www.versobooks.com

Smith, D. J., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 3-27.

Smith, P. L., & Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences, 27*(3), 161-168.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639.

Tavris, C., & Wade, C. (1995). *Psychology in perspective*. New York, NY: HarperCollins College Publishers.

Tharp, I. J., & Pickering, A. D. (2009). A note on DeCaro, Thomas, and Beilock (2008): Further data demonstrate complexities in the assessment of information–integration category learning. *Cognition, 111*(3), 410-414.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189-208.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial, and coactive theories. *Journal of Mathematical Psychology*, *39*(4), 321-359.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, *32*(7), 1206-1220.

Waldschmidt, J. G., & Ashby, F. G. (2011). Cortical and striatal contributions to automaticity in information-integration categorization. *Neuroimage*, *56*(3), 1791-1802.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in Neuroinformatics*, *7*, 14.

Wisniewski, E. J., & Medin, D. L. (1991). The fiction and non-fiction of features. In R. S. Michalski & G. T. Tecuci (Eds.), *Machine learning: A multi-strategy approach* (pp. 63-84). San Francisco, CA: Morgan Kaufmann.

APPENDIX A

The appendix aims at exploring group differences, which was avoided due to IDs that existed in two domains: Learned knowledge and used knowledge. The discussion section pointed to some shortcomings in the conceptualization of the used knowledge (lack of an explicit formulation of architecture), but this shortcoming does not prohibit all group comparisons. The lack of explicit formulation of architecture affects two-dimensional strategies, and even considering the shortcomings of implementation and conceptualization, it seems valid to compare unidimensional strategies. In this section three group analyses are done. The first comparison is between participants that learned and used only bar width and participants that learned and used only orientation. The goal of this analysis is to test whether the two dimensions were equally easy or whether there is an asymmetry in the difficulty levels. The second and third analyses compare participants that used only one dimension but learned both. The goal of these comparisons are to check the effect of latently learning a dimension without using it on performance. Participants are labeled in the following format: BW-BW are participants that learned and used bar width, OR-OR are participants that learned and used orientation, Both-BW are participants that learned both dimensions but used only bar width and Both-OR are participants that learned both dimensions but used only orientation. The used knowledge labels are based on Analysis 2. In each comparison, the accuracy and RTs are visualized and compared using a Bayesian linear mixed model. Finally, a DDM is fit to the four groups where separate drift rate (v), boundary separation (a) and non-decision time (t0) are estimated for each group and posterior samples of each parameter are compared. The goal of group analysis using DDMs is to decompose the

differences in accuracy and RT into interpretable parameters. This allows distinguishing between a slower RT due to task difficulty and a slower RT due to an increase in non-decision component of RT. One thing to note is that the comparisons cannot be used to infer any causal relation because the groups are defined based on observation and participants are not randomly assigned to the different groups. Additionally, number of participants in Both-BW and Both-OR groups are relatively low and overall, this section is for the most part speculative.

BW-BW vs. OR-OR

Figure 31 shows participants that are going to be compared. The green circles are BW-BW participants (i.e., learned bar width and used bar width) and blue circles are OR-OR participants (i.e., learned orientation and used orientation).

Figure 32a and 32b show mean accuracy and mean RT of each group during the five blocks of training respectively. Two separate Bayesian linear mixed models with block (1, 2, …, 5) and group (BW-BW vs OR-OR) as fixed effects and participant as random effect were fitted, one with accuracy as dependent variable and one with RT as dependent variable. Model's estimate for the effect of BW-BW on accuracy is 0.85 and the estimate for OR-OR is 0.86 (difference of only 1% in accuracy). Figure 32c shows the posterior samples of group factor's coefficient in the model with accuracy as dependent variable. On the other hand, model's estimate for the effect of BW-BW on RT is 1393.42 and the estimate for OR-OR is 1136.21 (difference of 257 ms in RT). Additionally, more than 99% of the posterior samples of the OR-OR coefficient are smaller than BW-BW coefficients. Figure 32d shows the posterior samples of group factor's coefficient in the model with RT as dependent variable. Overall, Figure 32

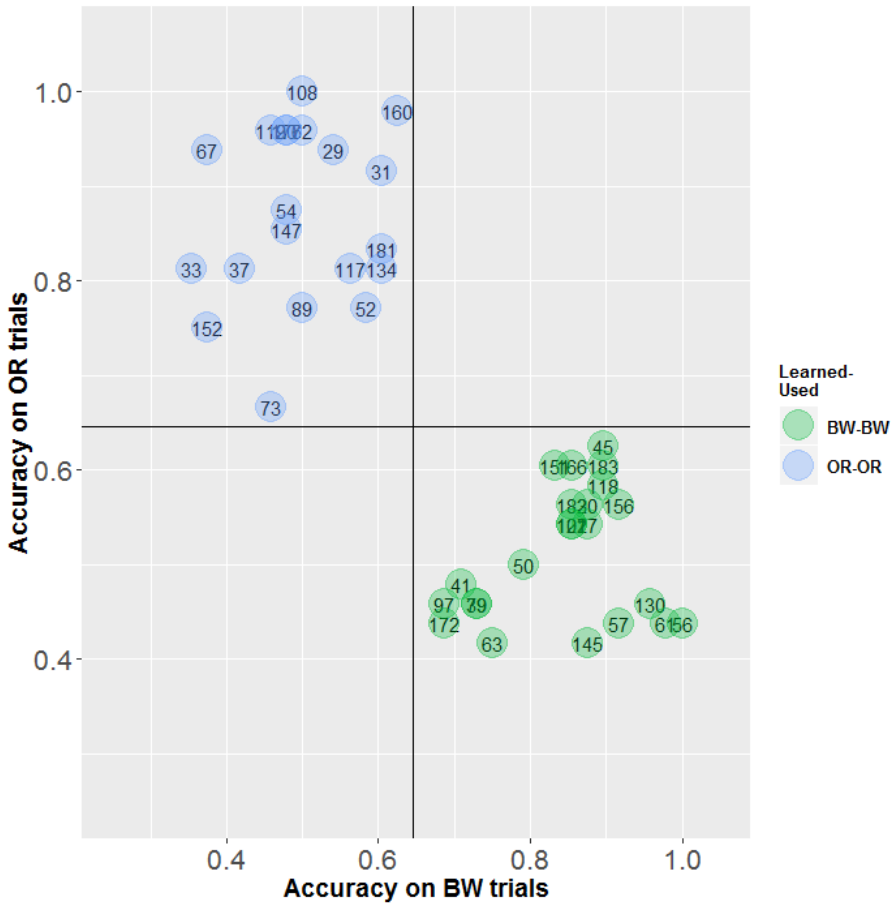
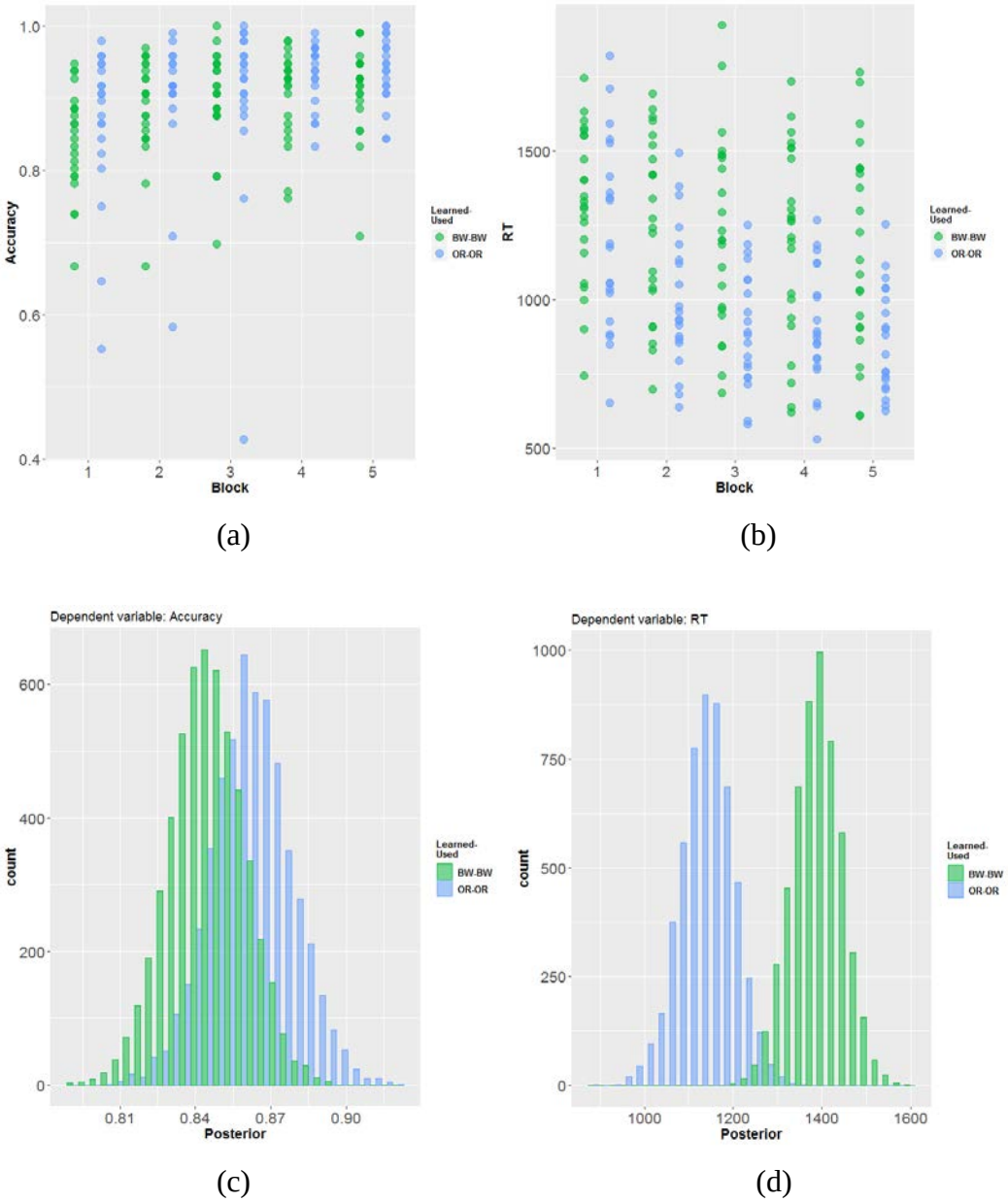


Figure 31. Comparing participants that learned and used bar width (green circles) and participants that learned and used orientation (blue circles).



(a) (b)

(c) (d)

Figure 32. Comparing BW-BW and OR-OR participants. a) Mean accuracy of participants in each block (each circle represents a participant). b) Mean RT of participants in each block (each circle represents a participant). c) Posterior samples of each group, when dependent variable is mean accuracy. d) Posterior samples of each group, when dependent variable is mean RT.

suggests that OR-OR participants are faster than BW-BW participants are, but not more accurate.

The differences between the two groups are explored using a hierarchical DDM. The difference in RTs can be translated into differences in drift rate or non-decision time. There is no reason to assume that the differences should be reflected in the non-decision time parameter, and the most intuitive explanation is that orientation attribute is more salient than bar width and categorizing based on orientation is easier compared to categorizing based on bar width. Figure 33 shows the posterior samples of three parameters of the DDM. As shown by the figure, the only difference seems to be between the drift rates (over 99% of OR-OR posterior samples were bigger than the BW-BW posterior samples), which suggests that the differences in RT between BW-BW and OR-OR was due to differences in difficulty of categorization based on bar width and orientation. Additionally, around 90% of boundary separation posterior samples from BW-BW are bigger than OR-OR.

There was no reason to expect that the differences in RT between the two groups may manifest itself in the non-decision component (i.e., stimulus encoding and response execution components of RT), therefore, the result of drift diffusion models seem to be understandable. Overall, there seems to be enough evidence to conclude that categorizing based on orientation was easier. However, since it is an observational study, the differences might be caused by ‘good’ participants choosing orientation dimension to categorize and ‘bad’ participants choosing bar width dimension.

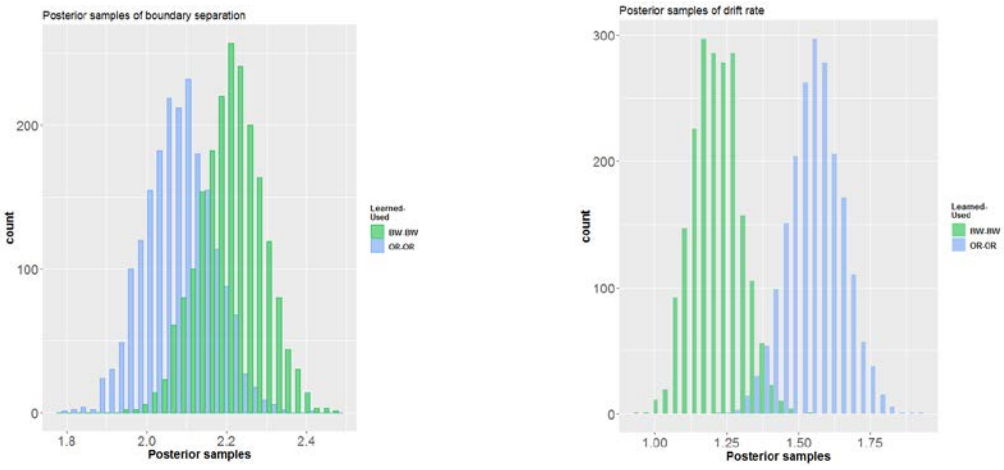


Figure 33. Posterior samples of DDMs for BW-BW and OR-OR participants. a) Posterior samples of boundary separation. b) Posterior samples of drift rate. c) Posterior samples of non-decision time.

BW-BW vs. Both-BW

Figure 34 shows participants that are going to be compared. The green circles are BW-BW participants (i.e., learned bar width and used bar width) and red circles are Both-BW participants (i.e., learned both dimensions and used only bar width).

Figure 35a and 35b show mean accuracy and RT of each group during the five blocks of training respectively. Two separate Bayesian linear mixed models with block (1, 2, …, 5) and group (BW-BW vs Both-BW) as fixed effects and participant as random effect were fitted, one with accuracy as dependent variable and one with RT as dependent variable. Model's estimate for the effect of BW-BW on accuracy is 0.84 and the estimate for Both-BW is 0.88 (difference of 4% in accuracy). Figure 35c shows the posterior samples of group factor's coefficient in the model with accuracy as dependent variable. More than 96% of the posterior samples of the Both-BW coefficient are bigger than BW-BW coefficients. On the other hand, model's estimate for the effect of BW-BW on RT is 1354.78 and the estimate for Both-BW is 1434.07 (difference of 79 ms in RT). Figure 35d shows the posterior samples of group factor's coefficient in the model with RT as dependent variable. Overall, Figure 35 suggests that Both_BW participants are more accurate compared to BW-BW, but not faster.

The differences between the two groups are explored using a hierarchical DDM. Figure 36 shows the posterior samples of three parameters of the DDM. As shown by the figure, the main difference seems to be between the non-decision time components (Figure 36c; over 98% of Both-BW posterior samples were bigger than the BW-BW posterior samples). Additionally, around 89% of boundary separation posterior samples

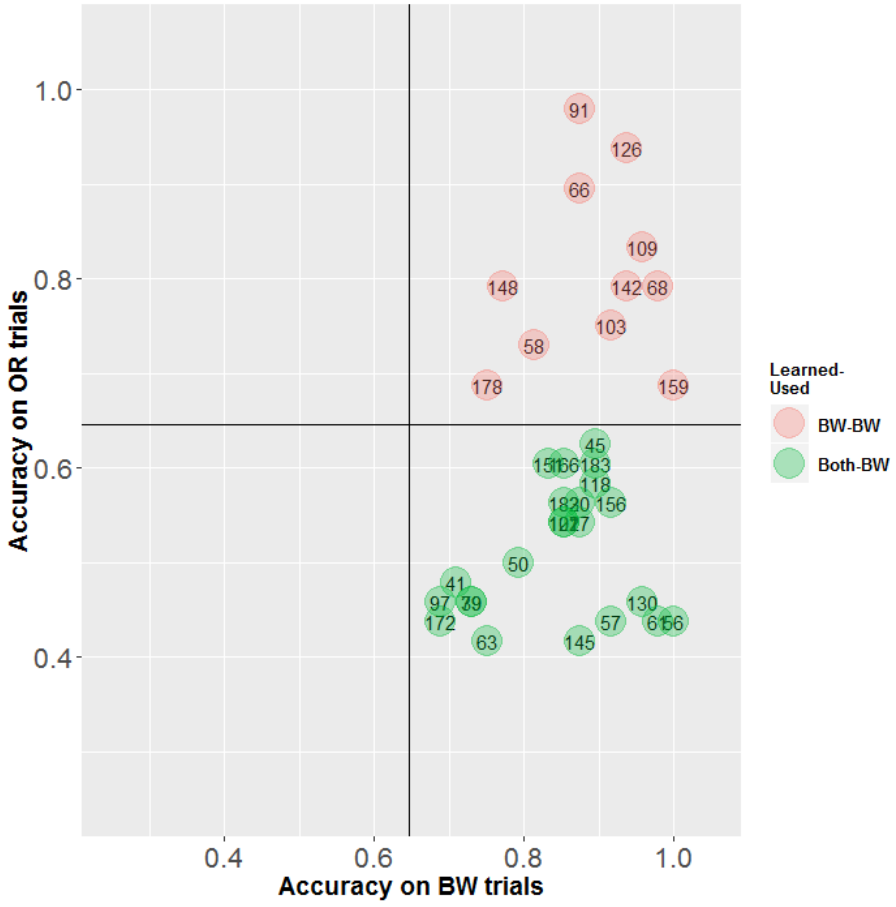


Figure 34. Comparing participants that learned and used bar width (green circles) and participants that learned both dimensions but used only bar width (red circles).

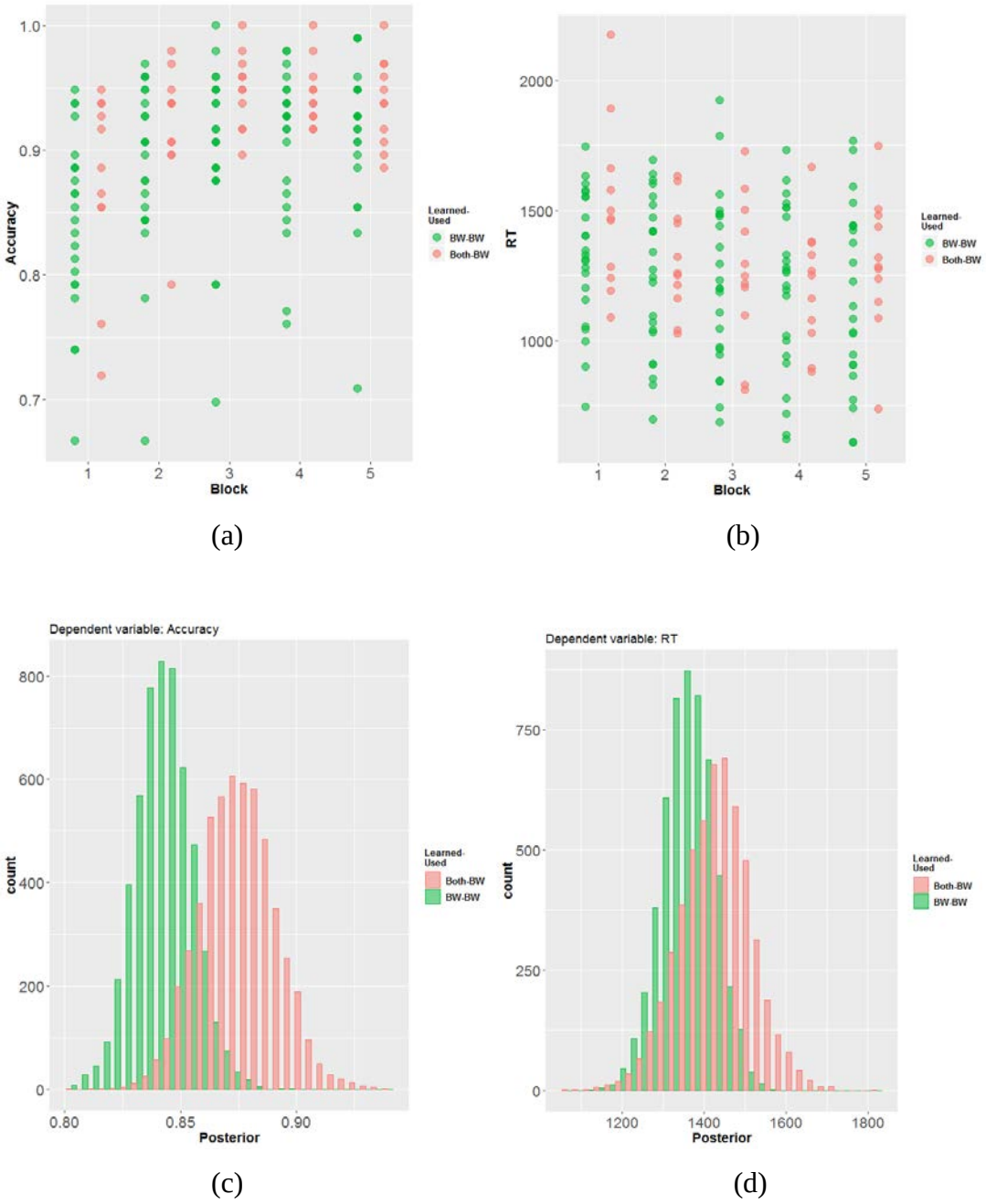


Figure 35. Comparing BW-BW and Both-BW participants. a) Mean accuracy of participants in each block (each circle represents a participant). b) Mean RT of participants in each block (each circle represents a participant). c) Posterior samples of each group, when dependent variable is mean accuracy. d) Posterior samples of each group, when dependent variable is mean RT.

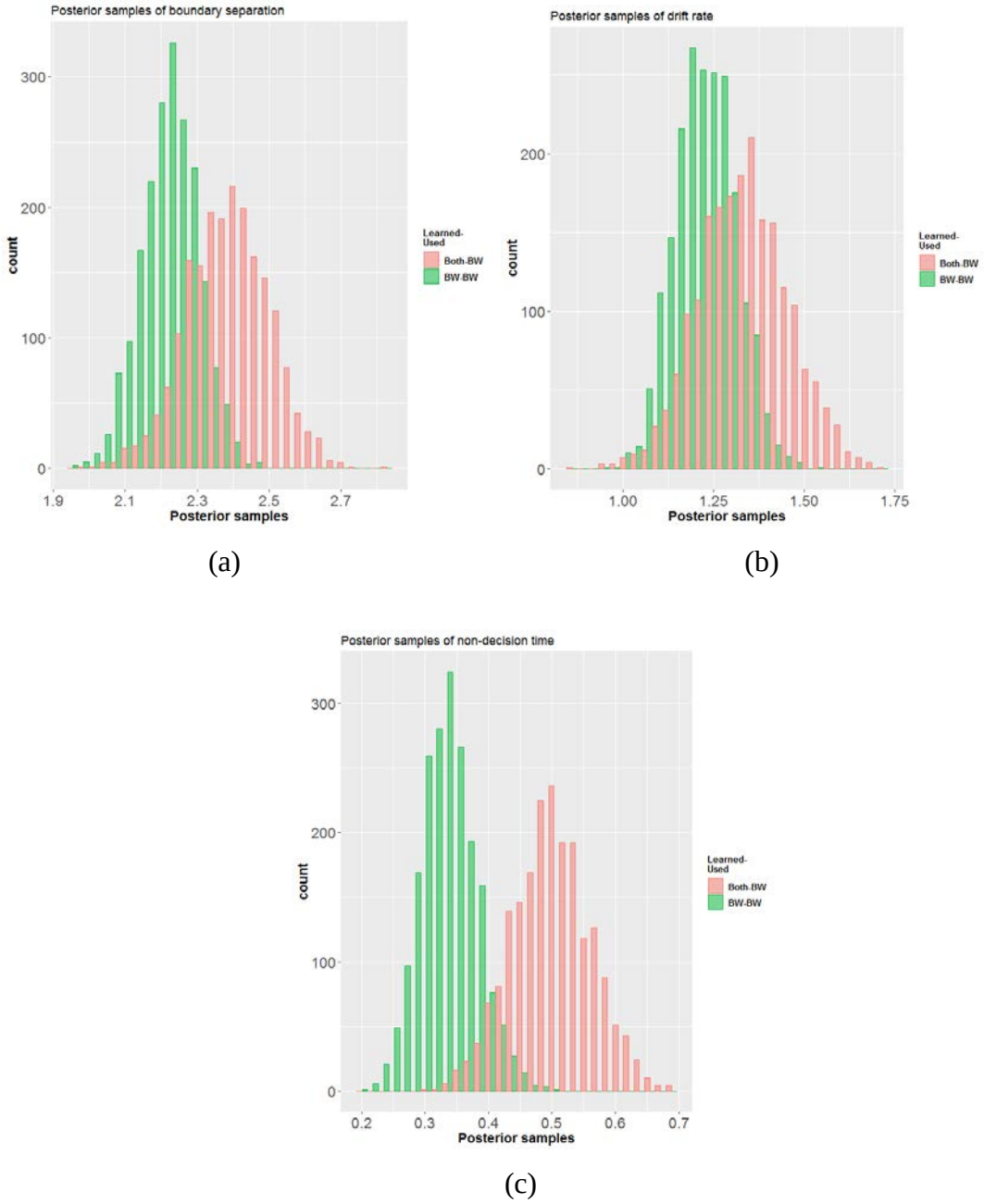


Figure 36. Posterior samples of DDMs for BW-BW and Both-BW participants. a) Posterior samples of boundary separation. b) Posterior samples of drift rate. c) Posterior samples of non-decision time.

from Both-BW are bigger than BW-BW (Figure 36a) and around 78% of drift rate posterior samples from Both-BW are bigger than BW-BW (Figure 36b).

A possible explanation for Both-BW participants having a higher non-decision time component can be given using the concept of workload capacity (a SFT terminology). SFT assumes three possibilities for the effect of adding sources of information: degradation of performance (limited capacity), no change in performance (unlimited capacity) and better performance (super capacity). The reason that Both-BW participants have higher non-decision time component compared to BW-BW participants might be caused by additional workload of latently learning the orientation dimension.

OR-OR vs. Both-OR

Figure 37 shows participants that are going to be compared. The blue circles are OR-OR participants (i.e., learned orientation width and used orientation) and red circles are Both-OR participants (i.e., learned both dimensions and used only orientation).

Figure 38a and 38b show mean accuracy and RT of each group during the five blocks of training respectively. Two separate Bayesian linear mixed models with block (1, 2, …, 5) and group (OR-OR vs Both-OR) as fixed effects and participant as random effect were fitted, one with accuracy as dependent variable and one with RT as dependent variable. Model's estimate for the effect of OR-OR on accuracy is 0.88 and the estimate for Both-OR is also 0.88 (no difference in accuracy). Figure 38c shows the posterior samples of group factor's coefficient in the model with accuracy as dependent variable. On the other hand, model's estimate for the effect of OR-OR on RT is 1152.07 and the estimate for Both-OR is 1337.64 (difference of 185 ms in RT). More than 99% of the

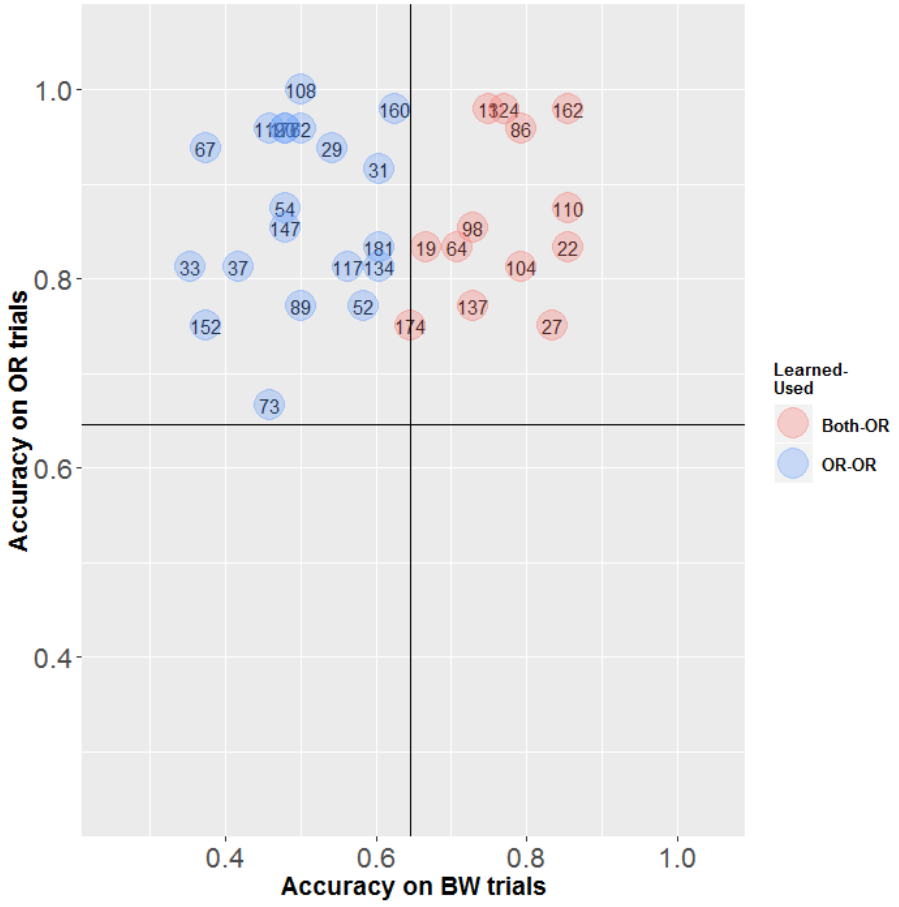


Figure 37. Comparing participants that learned and used orientation (blue circles) and participants that learned both dimensions but used only orientation (red circles).

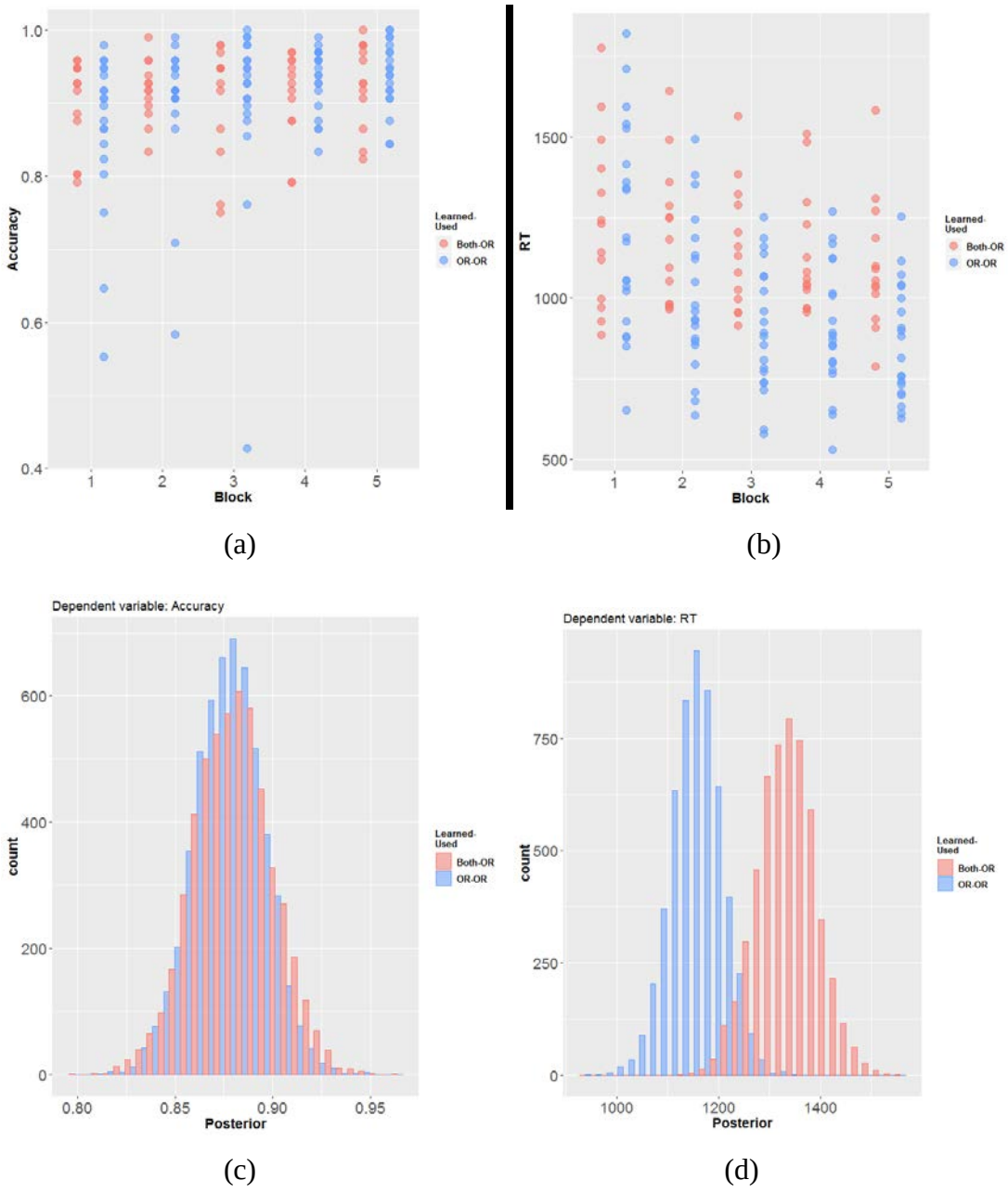
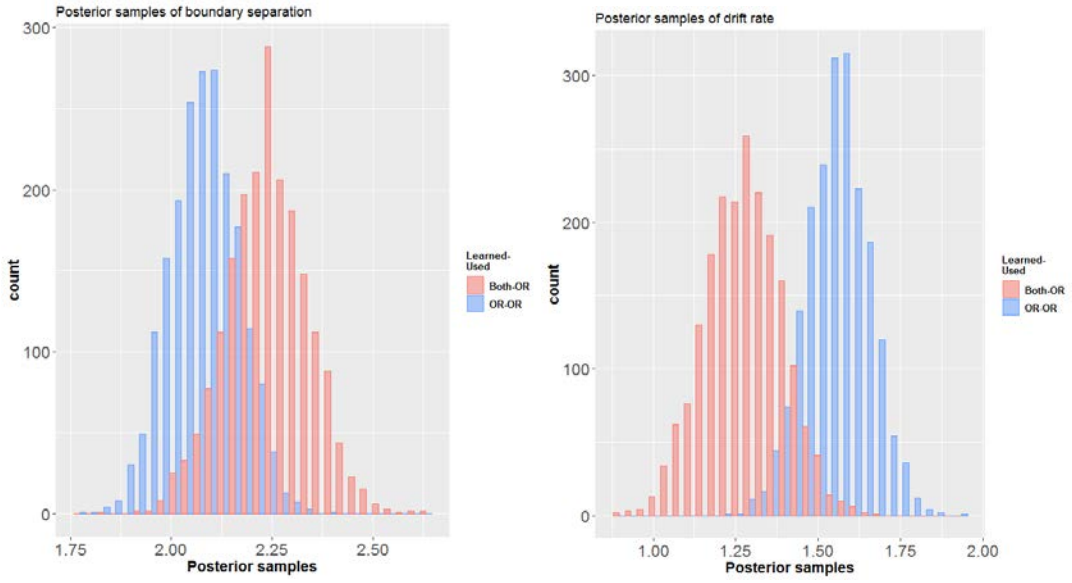


Figure 38. Comparing OR-OR and Both-OR participants. a) Mean accuracy of participants in each block (each circle represents a participant). b) Mean RT of participants in each block (each circle represents a participant). c) Posterior samples of each group, when dependent variable is mean accuracy. d) Posterior samples of each group, when dependent variable is mean RT.

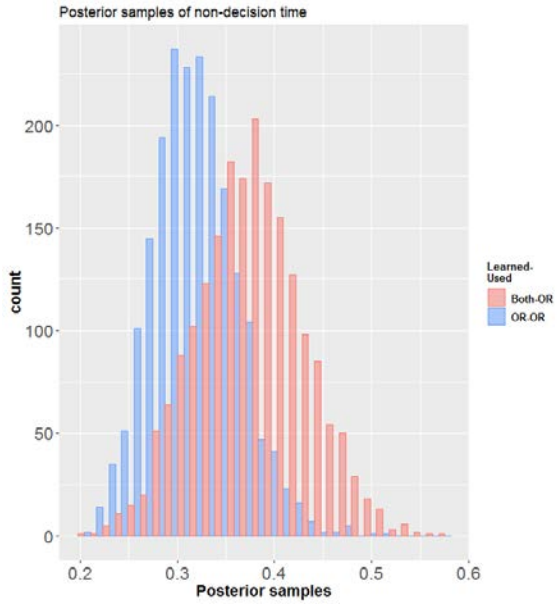
posterior samples of the Both-OR coefficient are bigger than OR-OR coefficients. Figure 36d shows the posterior samples of group factor's coefficient in the model with RT as dependent variable. Overall, Figure 38 suggests that OR-OR participants are faster than Both-OR participants are, but not more accurate.

The differences between the two groups are explored using a hierarchical DDM. Figure 39 shows the posterior samples of three parameters of the DDM. As shown by the figure, the main difference seems to be between the drift rates (Figure 39b; over 96% of OR-OR posterior samples were bigger than the Both-OR posterior samples). Additionally, around 90% of boundary separation posterior samples from Both-OR are bigger than OR-OR (Figure 39a) and around 85% of non-decision time posterior samples from Both-OR are bigger than OR-OR (Figure 39c).

It seems that since both OR-OR and Both-OR are using orientation to categorize stimuli, there should not be any difference in difficulty, so it is difficult to explain the differences in drift rates. The expected result would be differences in non-decision time component (due to additional load caused by latently learning the bar width dimension). The differences in boundary separation and non-decision time are also not small. Overall, it is important to emphasize again that since the groups are observed (and not assigned randomly to participants) these results are going to tell very little about the effect of latently learning a dimension.



(a) (b)



(c)

Figure 39. Posterior samples of DDMs for OR-OR and Both-OR participants. a) Posterior samples of boundary separation. b) Posterior samples of drift rate. c) Posterior samples of non-decision time.

APPENDIX B

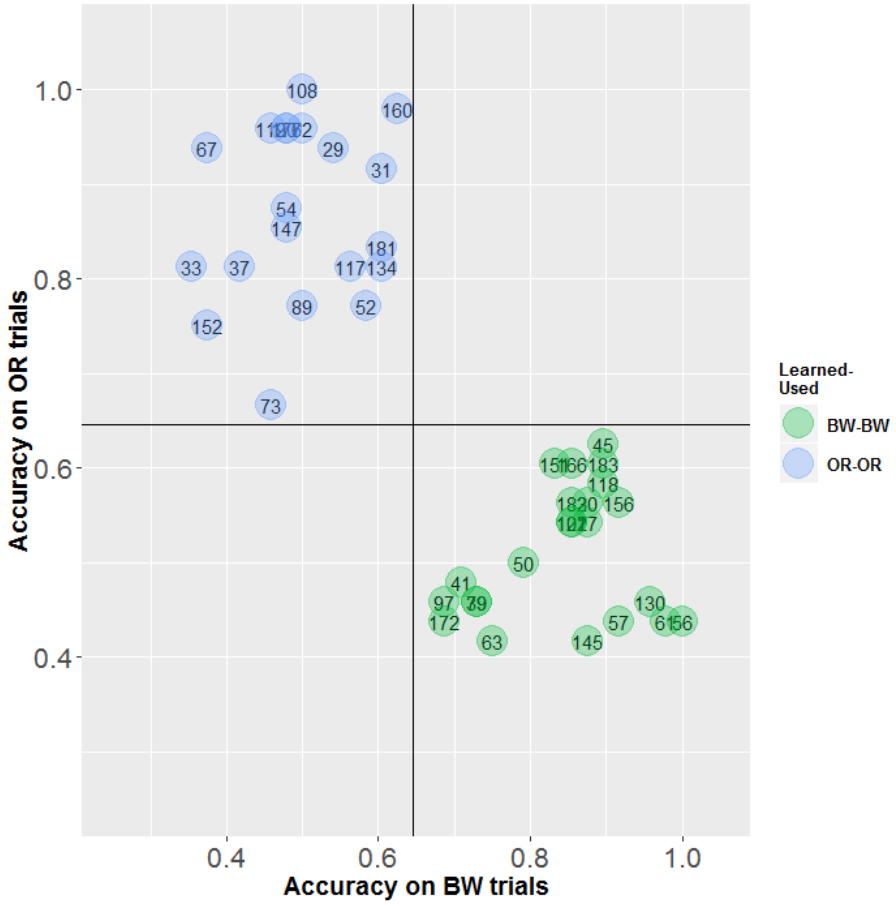


Figure 40. Comparing participants that learned and used bar width (green circles) and participants that learned and used orientation (blue circles).

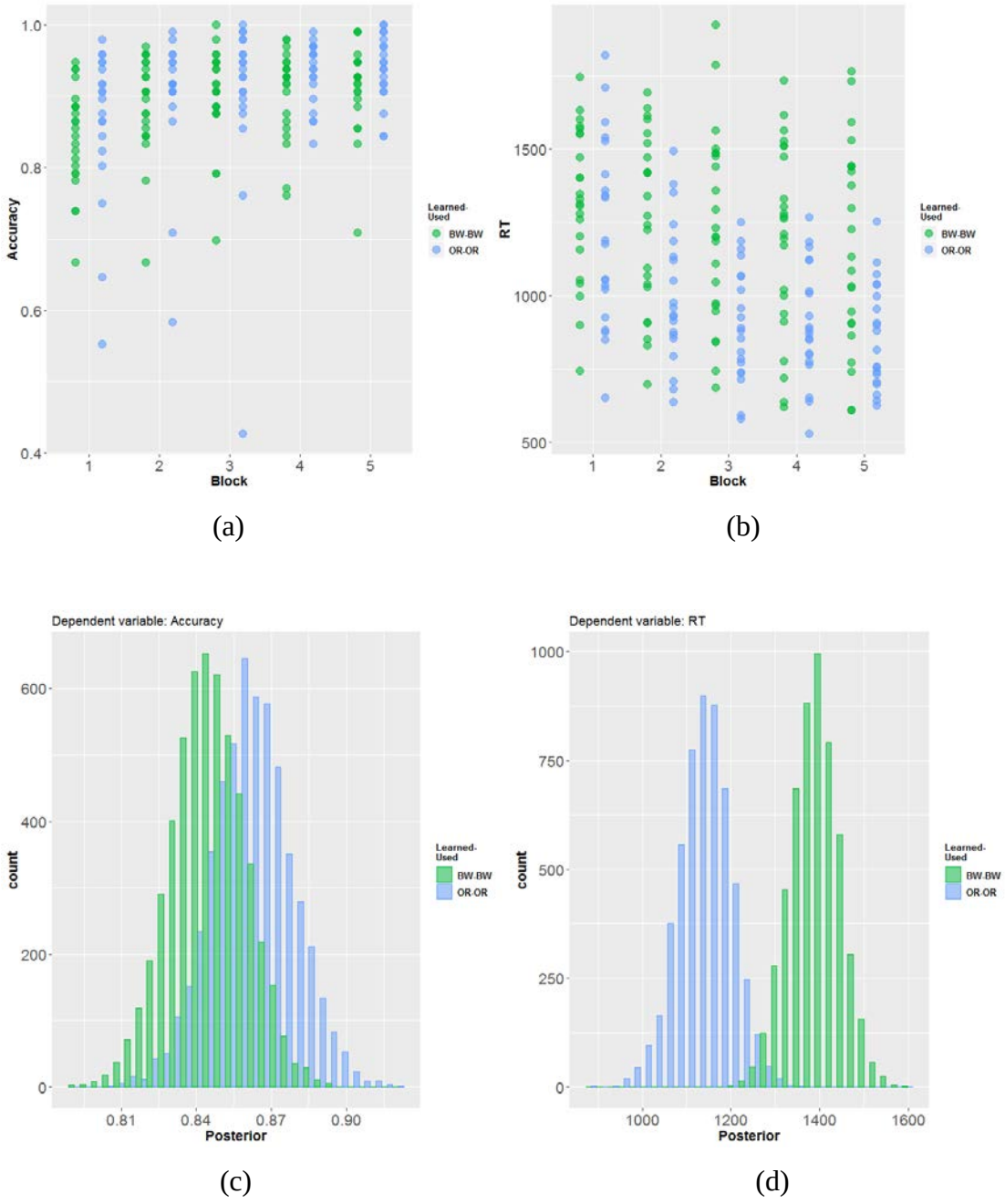


Figure 41. Comparing BW-BW and OR-OR participants. a) Mean accuracy of participants in each block (each circle represents a participant). b) Mean RT of participants in each block (each circle represents a participant). c) Posterior samples of each group, when dependent variable is mean accuracy. d) Posterior samples of each group, when dependent variable is mean RT.

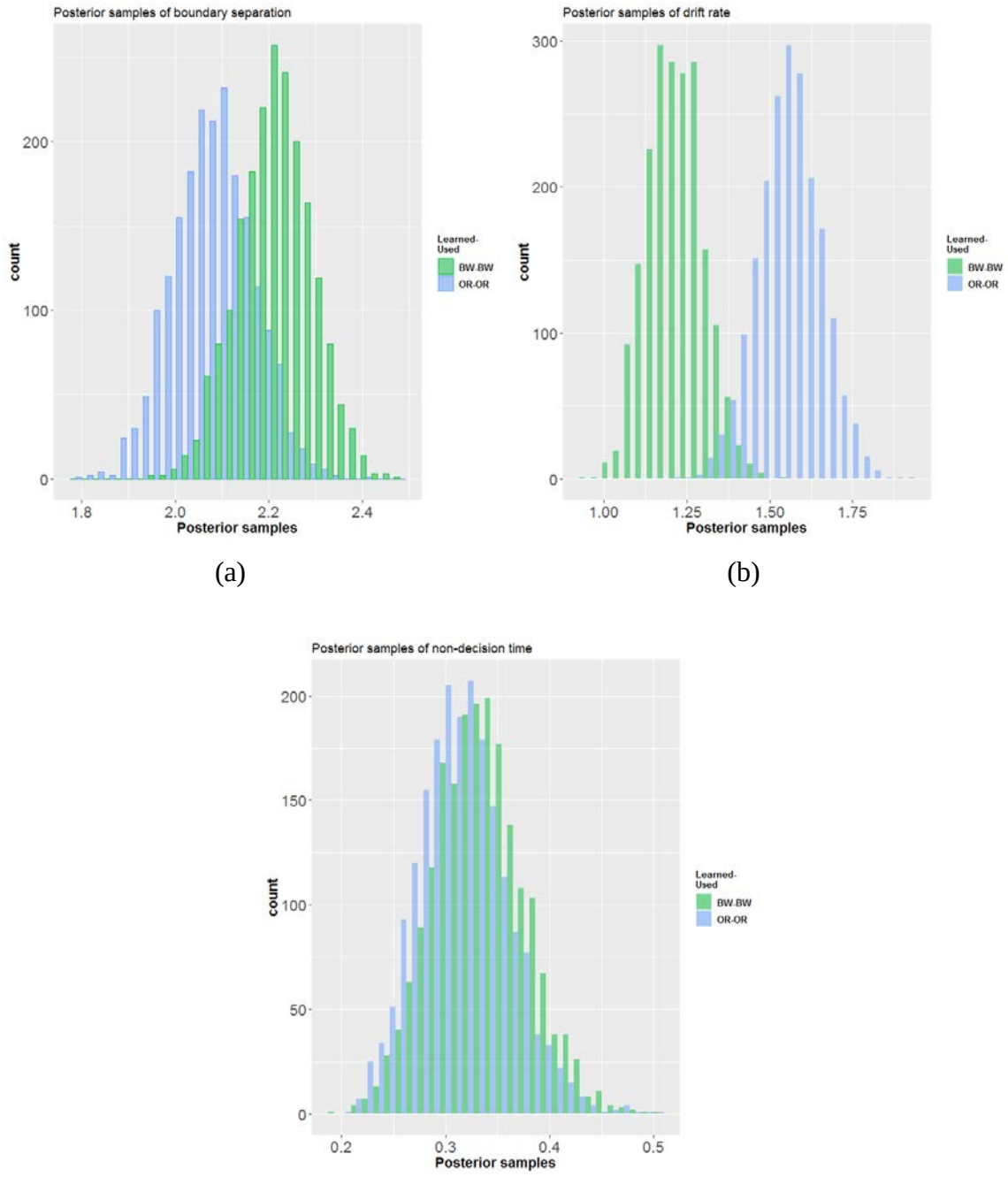


Figure 42. Posterior samples of DDMs for BW-BW and OR-OR participants. a) Posterior samples of boundary separation. b) Posterior samples of drift rate. c) Posterior samples of non-decision time.

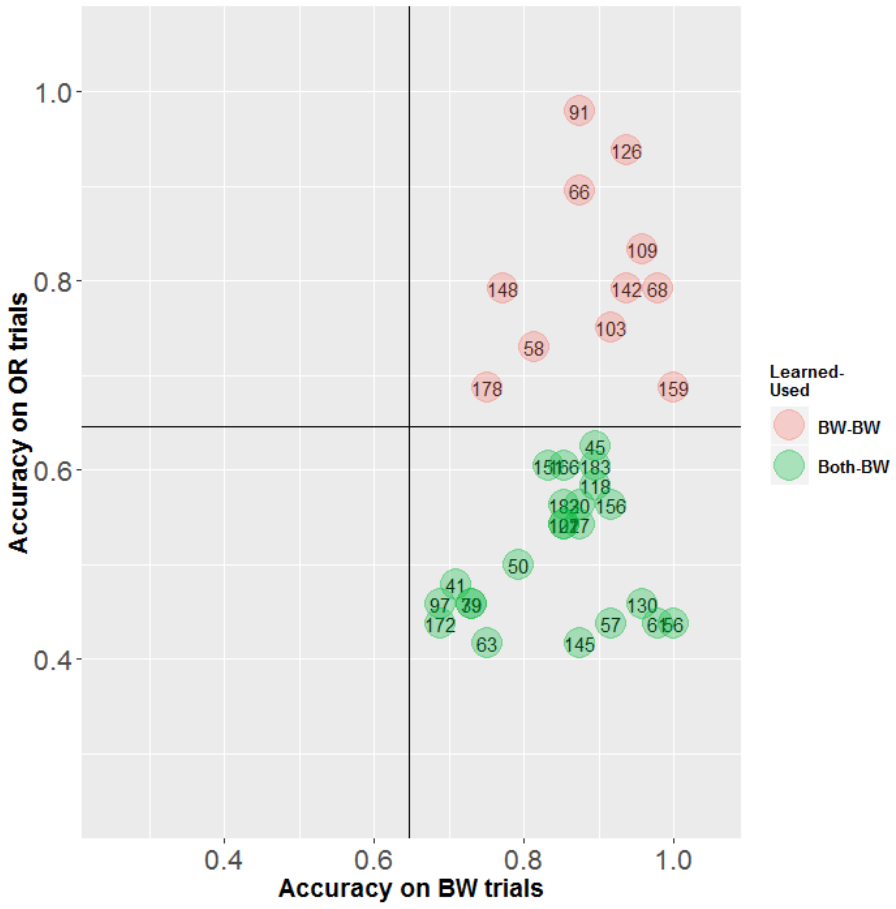


Figure 43. Comparing participants that learned and used bar width (green circles) and participants that learned both dimensions but used only bar width (red circles).

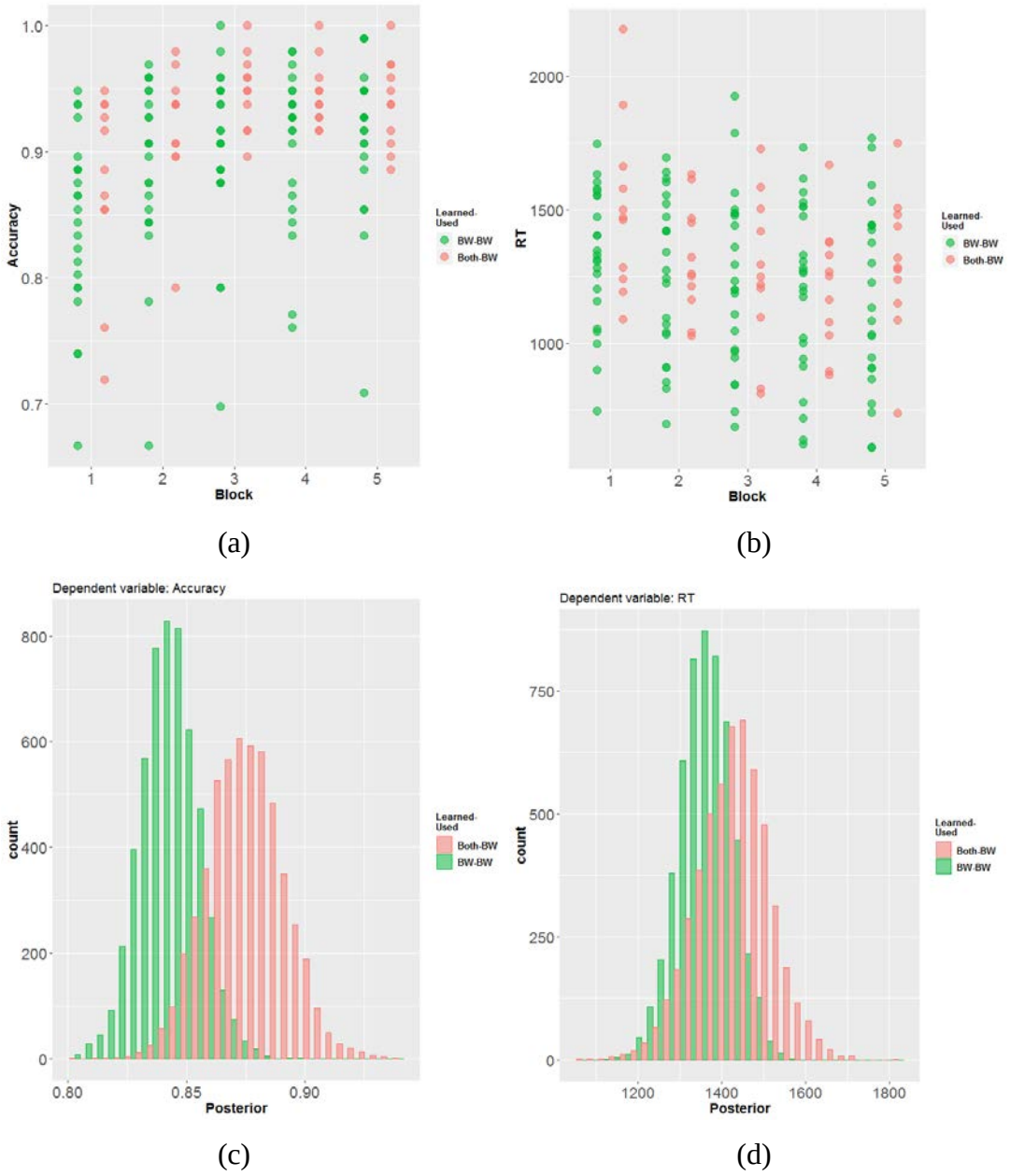


Figure 44. Comparing BW-BW and Both-BW participants. a) Mean accuracy of participants in each block (each circle represents a participant). b) Mean RT of participants in each block (each circle represents a participant). c) Posterior samples of each group, when dependent variable is mean accuracy. d) Posterior samples of each group, when dependent variable is mean RT.

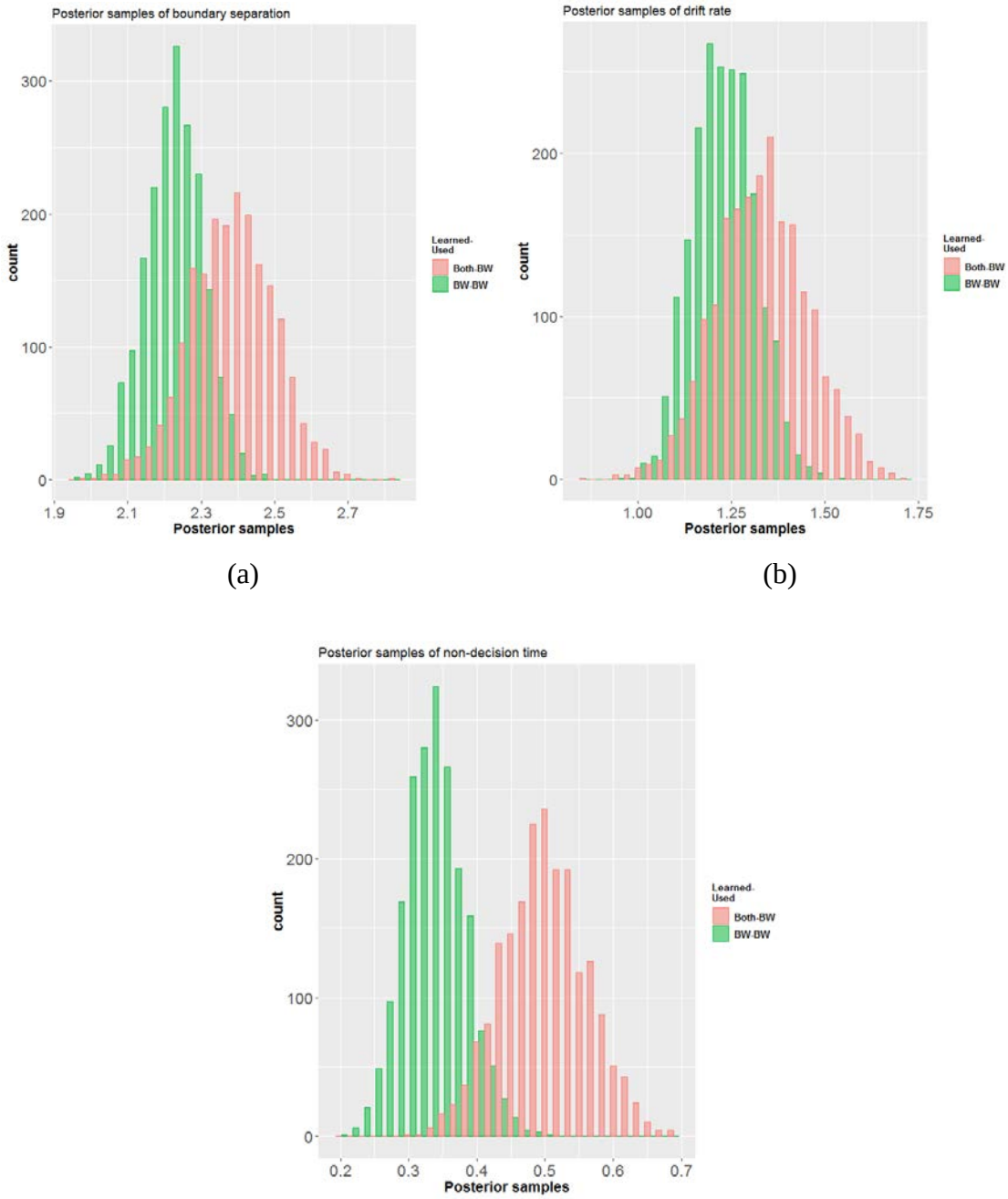


Figure 45. Posterior samples of DDMs for BW-BW and Both-BW participants. a) Posterior samples of boundary separation. b) Posterior samples of drift rate. c) Posterior samples of non-decision time.

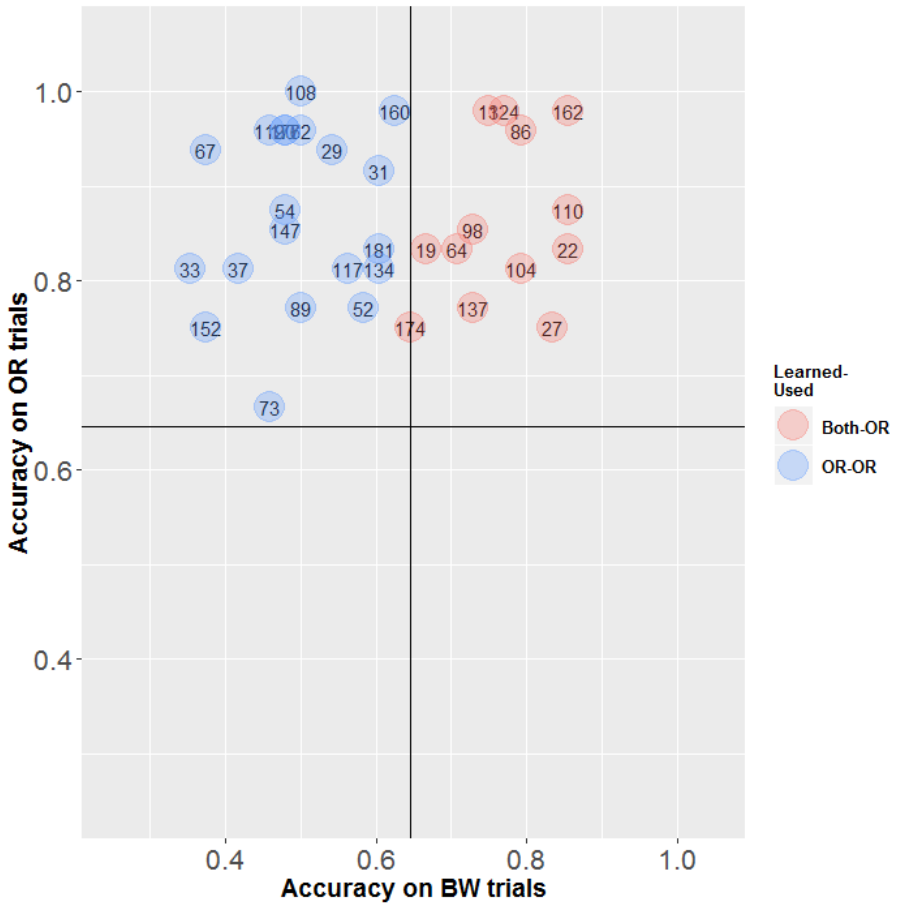


Figure 46. Comparing participants that learned and used orientation (blue circles) and participants that learned both dimensions but used only orientation (red circles).

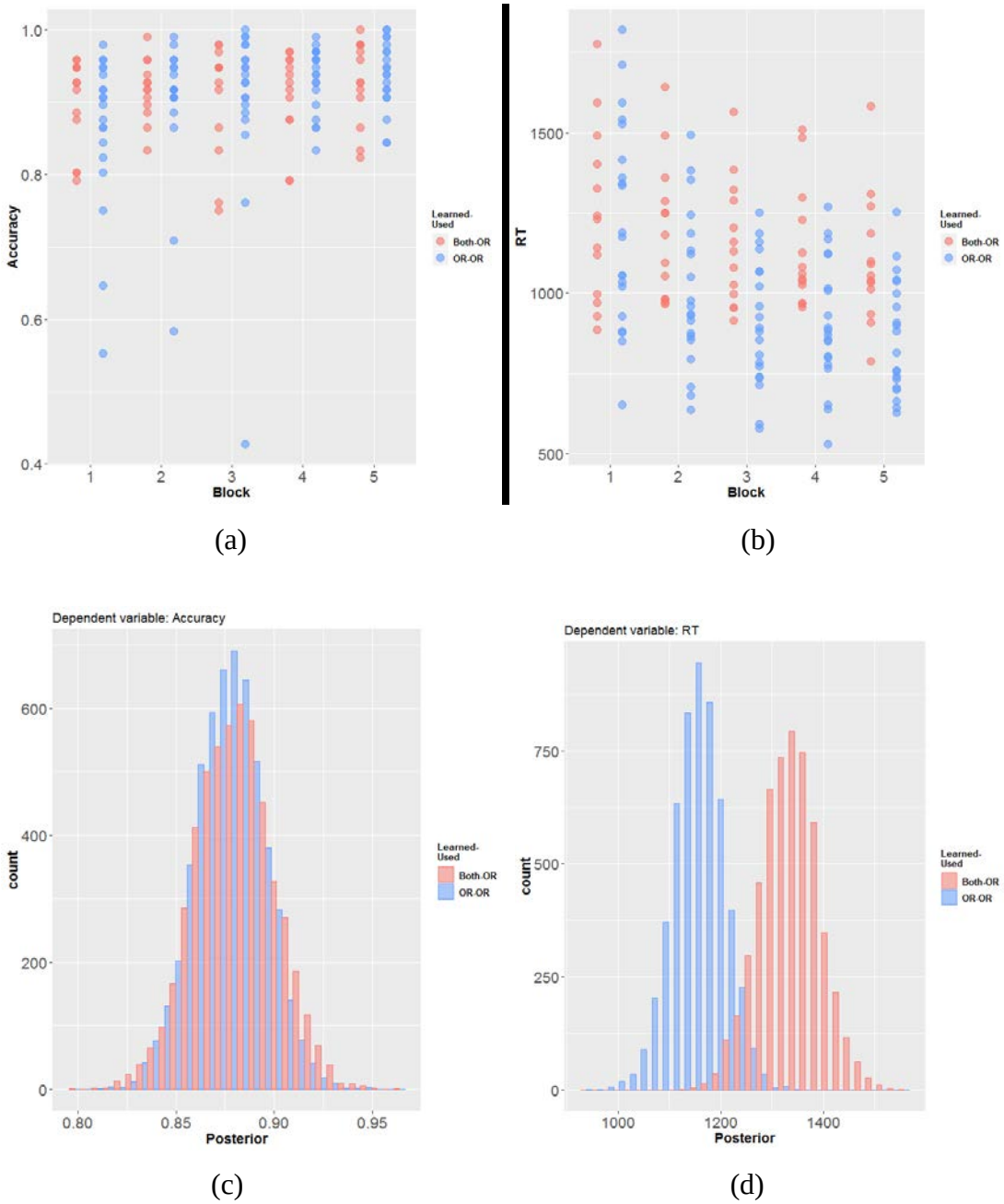


Figure 47. Comparing OR-OR and Both-OR participants. a) Mean accuracy of participants in each block (each circle represents a participant). b) Mean RT of participants in each block (each circle represents a participant). c) Posterior samples of each group, when dependent variable is mean accuracy. d) Posterior samples of each group, when dependent variable is mean RT.

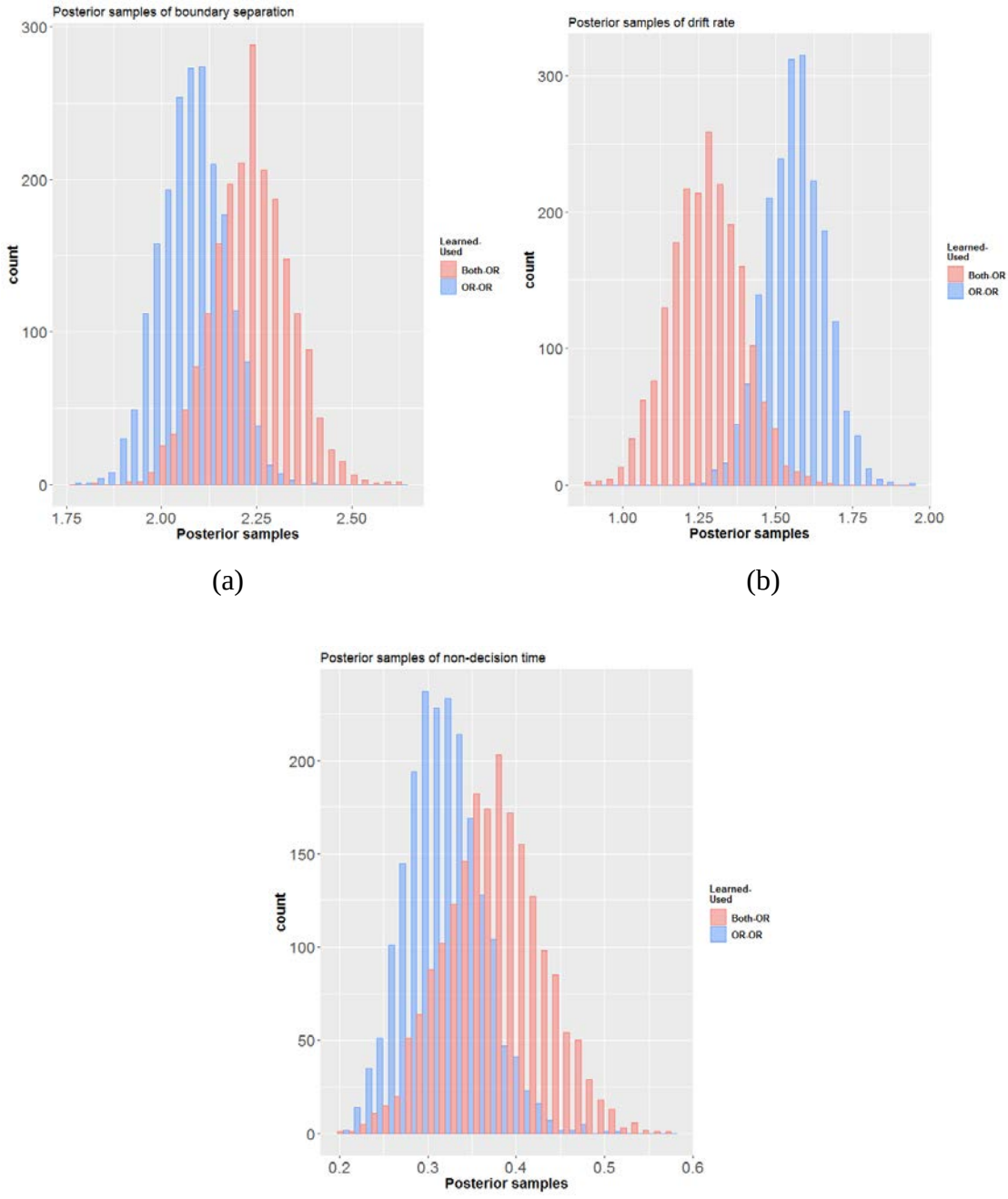


Figure 48. Posterior samples of DDMs for OR-OR and Both-OR participants. a) Posterior samples of boundary separation. b) Posterior samples of drift rate. c) Posterior samples of non-decision time.