Categorization training increases the perceptual separability of novel dimensions

Fabian A. Soto *, F. Gregory Ashby

Department of Psychological and Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

ABSTRACT

Perceptual separability is a foundational concept in cognitive psychology. A variety of research questions in perception – particularly those dealing with notions such as "independence," "invariance," "holism," and "configurality" – can be characterized as special cases of the problem of perceptual separability. Furthermore, many cognitive mechanisms are applied differently to perceptually separable dimensions than to non-separable dimensions. Despite the importance of dimensional separability, surprisingly little is known about its origins. Previous research suggests that categorization training can lead to learning of novel dimensions, but it is not known whether the separability of such dimensions also increases with training. Here, we report evidence that training in a categorization task increases perceptual separability of the category-relevant dimension according to a variety of tests from general recognition theory (GRT). In Experiment 1, participants who received pre-training in a categorization task showed reduced Garner interference effects and reduced violations of marginal invariance, compared to participants who did not receive such pre-training. Both of these tests are theoretically related to violations of perceptual separability. In Experiment 2, participants who received pre-training in a categorization task showed reduced violations of perceptual separability according to a model-based analysis of data using GRT. These results are at odds with the common assumption that separability and independence are fixed, hardwired characteristics of features and dimensions.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

An important task of perceptual systems is to produce a re-description of the incoming sensory input, through a representation that is useful for the tasks that are usually encountered in the natural environment. One way to characterize internal stimulus representations is to determine whether a set of "privileged" stimulus properties exists, which can be used to describe a variety of stimuli, and that are processed and represented independently from one another. In perceptual science, an important amount of effort has been dedicated to understanding what aspects of stimuli are represented in such an independent fashion (e.g., Bruce & Young, 1986; Haxby, Hoffman, & Gobbini, 2000; Kanwisher, 2000; Op de Beeck, Haushofer, & Kanwisher, 2008; Stankiewicz, 2002; Ungerleider & Haxby, 1994; Vogels, Biederman, Bar, & Lorincz, 2001).

There are many different conceptual and operational definitions of what it means for two stimulus dimensions to be independent (Ashby & Townsend, 1986), but perhaps the most widely studied and influential type of independence is dimensional separability. Separable stimulus dimensions are those that can be selectively attended and that directly determine the similarity among stimuli

* Corresponding author. Tel.: +1 805 893 7909; fax: +1 805 893 4303.
E-mail address: fabian.soto@psych.ucsb.edu (F.A. Soto).

(Garner, 1974; Shepard, 1991). This is in contrast to integral dimensions, which cannot be selectively attended and do not directly determine the similarity among stimuli. When stimuli vary along integral dimensions, their similarity is directly perceived and the notion of dimensions loses meaning.

There are two main reasons to believe that a complete understanding of complex forms of visual cognition, such as object recognition and categorization, will benefit from a good understanding of perceptual separability. The first reason is that many important questions in perceptual science can be understood as questions about perceptual separability of object dimensions.

For example, in the area of visual object recognition, the question of whether object representations are invariant across changes in identity-preserving variables (such as rotation and translation; for reviews, see Biederman, 2001; Kravitz, Vinson, & Baker, 2008; Peissig & Tarr, 2007) is essentially the same as the question of whether object representations are perceptually separable from such variables. Shape dimensions that may be important for invariant object recognition have been shown to be separable from other shape dimensions and from viewpoint information, according to traditional tests of separability (Stankiewicz, 2002).

A second example comes from the area of face perception. It has been proposed that a hallmark of human face perception is that faces are processed in a configural or holistic manner (for reviews, see Farah, Wilson, Drain, & Tanaka, 1998; Maurer, Grand, & Mondloch, 2002; Richler, Palmeri, & Gauthier, 2012). Configural or holistic face perception can be seen as non-separable processing of different face features (e.g., Mestry, Wenger, & Donnelly, 2012; Richler, Gauthier, Wenger, & Palmeri, 2008; Thomas, 2001). Similarly, influential theories of face processing have proposed that different aspects of faces, such as identity and emotional expression, are processed independently (e.g., Bruce & Young, 1986; Haxby et al., 2000) and these hypotheses are usually investigated using tests of perceptual separability (e.g., Fitousi & Wenger, 2013; Ganel & Goshen-Gottstein, 2004; Schweinberger & Soukup, 1998; Soto, Musgrave, Vucovich, & Ashby, 2015).

Casting such research questions in terms of perceptual separability is not only possible, but desirable. As we will see below, perceptual separability has a precise formal definition within multidimensional signal detection theory (Ashby & Townsend, 1986; for a review, see Ashby & Soto, 2015), which offers the advantage of providing strict definitions to rather ambiguous concepts such as independence, holistic processing, configural processing, etcetera (e.g., Fitousi & Wenger, 2013; Mestry et al., 2012; Richler et al., 2008). Furthermore, it allows us to determine whether behavioral evidence of a dimensional interaction is due to true perceptual interactions versus interactions at the level of decisional processes.

The fact that a variety of research questions in visual cognition can be characterized as special cases of the problem of perceptual separability suggests that a better understanding of this general problem, including explanations of why some dimensions are separable and how they

acquired such status, would necessarily lead to a better understanding of each of the special cases.

A second reason why an understanding of perceptual separability is important to understand visual cognition is that considerable evidence suggests that higher-level cognitive mechanisms are applied differently when stimuli differ along separable dimensions rather than along integral dimensions. Given the definition of perceptual separability, the most obvious of such mechanisms is selective attention, which is more easily deployed to separable than to non-separable dimensions (e.g., Garner, 1970, 1974; Goldstone, 1994b). Other examples of processes that might be applied differently to separable-dimension and integral-dimension stimuli are the rules by which different sources of predictive and causal knowledge are combined (Soto, Gershman, & Niv, 2014), as well as the performance cost of storing an additional object in visual working memory (Bae & Flombaum, 2013).

There is much evidence suggesting that the mechanisms used by people to categorize stimuli vary depending on whether or not categories differ along separable dimensions. Some of this evidence comes from studies using unsupervised categorization tasks, in which people are asked to group stimuli in two or more categories without feedback about their performance. When stimuli in unsupervised categorization tasks vary along separable dimensions, people rely almost exclusively on one-dimensional strategies (Handel & Imai, 1972; Handel, Imai, & Spottswood, 1980; Medin, Wattenmaker, & Hampson, 1987), even in tasks in which categories are not defined by a simple one-dimensional rule and after being explicitly told that the optimal strategy is to integrate information from two dimensions (Ashby, Queller, & Berretty, 1999). Furthermore, unsupervised learning is possible only when the categories clearly differ along a single dimension (Ashby et al., 1999). On the other hand, when stimuli vary along integral dimensions, people show limited ability to learn unsupervised categories and they do not show a strong preference for one-dimensional rules. Instead, they show a variety of strategies, including integration of information from both dimensions (Ell, Ashby, & Hutchinson, 2012).

A similar pattern of results is found in supervised categorization tasks, in which categorization choices are followed by feedback. When stimuli vary along separable dimensions, learning a category structure in which good performance requires attending to a single dimension is much easier for people than learning an equivalent category structure in which good performance requires integration of information from two dimensions (e.g., Smith, Beran, Crossley, Boomer, & Ashby, 2010). There is a large body of evidence suggesting that the one-dimensional categorization task is learned through a rule-based categorization system, whereas the information-integration task is learned through a separate procedural categorization system (for reviews, see Ashby & Maddox, 2005; Ashby & Valentin, 2005). On the other hand, when stimuli vary along integral dimensions, a one-dimensional task is not consistently easier to learn than an information-integration task (Ell et al., 2012).

Despite the importance of dimensional separability for our understanding of both perception and high-level cognition, surprisingly little is known about its origins. Specifically, an important open question is whether separable dimensions can be learned and what are the conditions that might foster such learning. Our previous review of the literature suggests that this is a foundational question in the field of object categorization and recognition. If perceptual separability of a dimension can be learned and we could understand the mechanisms by which such learning happens, then we would not only be in a better position to explain why some object dimensions are "special," in the sense of being processed independently, but also how they became special (i.e., what conditions fostered this learning) and why they should be processed in such a privileged fashion (i.e., why it is adaptive for high-level cognitive mechanisms to operate differently on these representations).

In the following section, we introduce general recognition theory (GRT), a formal framework within which perceptual separability can be defined and studied. This is followed by a review of previous literature related to the idea of separability learning.

1.1. GRT and a formal definition of perceptual separability

GRT is an extension of signal detection theory to cases in which stimuli vary along two or more dimensions (Ashby & Townsend, 1986; for a tutorial review, see Ashby & Soto, 2015). It offers a framework in which different types of dimensional interactions can be defined formally and studied, while inheriting from signal detection theory the ability to dissociate perceptual from decisional sources for such interactions. For this reason, GRT is arguably the best framework for the analysis and interpretation of studies aimed at testing different forms of dimensional independence.

GRT assumes that repeated presentations of a single stimulus produce different perceptual effects, which follow some probability distribution. The most common applications of GRT are to tasks in which stimuli vary in two dimensions, A and B, each with two stimulus components, resulting in four stimuli: A_1B_1, A_2B_2, A_2B_1 and A_2B_2. Fig. 1 is an example of a GRT model for such a 2 × 2 design. In this model, each stimulus generates perceptual effects according to a different bivariate normal distribution. Each distribution is represented in the figure by a different ellipse, which represents the set of all percepts that are elicited with equal likelihood by the stimulus. For any ellipse (and therefore any stimulus), percepts inside the ellipse are more likely than percepts outside the ellipse. After many presentations of a particular stimulus, the scatterplot of perceptual effects will take the shape of the ellipse corresponding to that stimulus. However, in some cases, presenting a stimulus will produce a percept that lies outside that stimulus' ellipse, perhaps closer to the ellipse corresponding to a different stimulus. In all cases, a decision must be made about what stimulus was actually presented. This decision process is modeled by assuming that a participant sets decision bounds that divide the perceptual space into different regions, each corresponding to the identification of a particular stimulus. The simplest
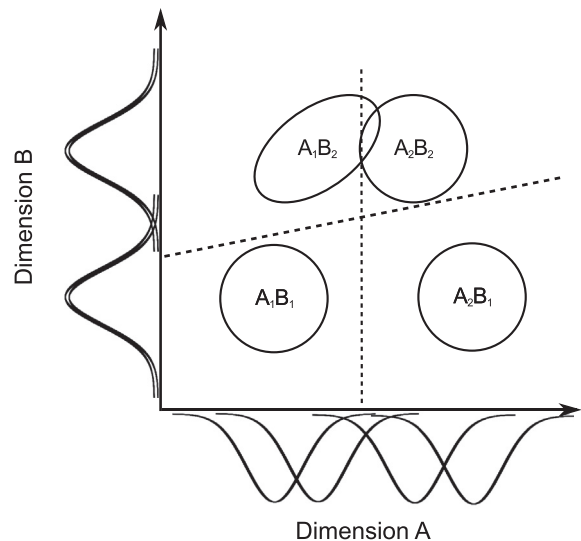

Fig. 1. Example of a multivariate normal GRT model for a 2 × 2 design (2 dimensions, each with 2 levels). Ellipses represent contours of equal likelihood for the perceptual distribution of a particular stimulus. Marginal distributions are presented in the left and bottom of the figure, next to their corresponding dimension.

decision bounds are lines, like those shown in Fig. 1, which are used to make decisions about both the level of dimension A and the level of dimension B.

In this framework, dimension A is perceptually separable from dimension B if the perceptual effects associated with dimension A do not depend on the level of dimension B. Mathematically this occurs if (and only if) the marginal distribution of perceptual effects along dimension A is the same across levels of B. Marginal distributions for dimensions A and B are depicted at the bottom and left of Fig. 1, respectively. The marginal distributions for dimension B are identical, regardless of the level of A, meaning that dimension B is perceptually separable from dimension A. Conversely, the marginal distributions for dimension A are farther apart for level 1 of dimension B than for level 2 of dimension B, meaning that dimension A is not perceptually separable from dimension B.

There are other forms of dimensional interaction that can be defined within GRT besides perceptual separability (Ashby & Townsend, 1986). One of these is decisional separability. Dimension A is decisionally separable from dimension B if the strategy used to decide the level of dimension A does not depend on the perceived value of dimension B. Mathematically, decisional separability holds if (and only if) the decision bounds are linear and orthogonal to each stimulus dimension. In Fig. 1, this means that dimension A is decisionally separable from dimension B, but dimension B is not decisionally separable from dimension A.

Finally, perceptual independence refers to dimensional interactions within a single stimulus. Perceptual independence holds for stimulus A_iB_j if the perceived value of the A component is statistically independent from the perceived value of the B component, which is true in the multivariate normal case when the correlation between

dimensions is zero. In Fig. 1, the diagonally-oriented ellipse for stimulus A_1B_2, representing a positive correlation between dimensions, is a sign of violations of perceptual independence for that stimulus.

In applications of GRT, inferences are made about these types of dimensional interactions from behavioral data. There is a number of theorems in the literature that link each type of dimensional interaction with statistics that can be computed from identification and categorization data (Ashby & Maddox, 1994; Ashby & Townsend, 1986; Kadlec & Townsend, 1992a, 1992b). Another approach is to fit one or more GRT models directly to the data; the parameter values of the best-fitting model can then be used to characterize the pattern of dimensional interactions (Ashby & Lee, 1991; Soto et al., 2015; Thomas, 2001). Here, we will use both the summary statistics and model-based approaches to study perceptual separability.

There are clear advantages to using GRT for the study of perceptual separability, instead of simply relying on traditional tests and operational definitions. The theory provides a formal definition of perceptual separability that coherently links together a number of operational definitions. This permits the consistent study of the same underlying concept through different tests and experimental designs. Furthermore, GRT allows the focus to be on perceptual separability by dissociating its influence on behavior from other forms of interactions. In particular, here we will be interested in whether training in categorization tasks produces changes in perceptual separability, independently of any changes in decision strategies. The analysis of dimensional interactions via GRT is critical to achieving this goal, as it is known that traditional tests of separability are influenced by variables such as experimental instructions (Foard & Kemler-Nelson, 1984; Melara, Marks, & Lesko, 1992), which are likely to affect decision strategies instead of perceptual interactions (Ashby, Waldron, Lee, & Berkman, 2001).

1.2. Can separable dimensions be learned?

The hypothesis that learning might have an influence on the separability of psychological dimensions is as old as the concept of separability itself (see Garner, 1970). In support of this hypothesis, developmental data have shown that the ability to selectively attend to separable stimulus dimensions develops with age. Stimulus dimensions that are perceived as integrated wholes by pre-school children are instead perceived as analytic components by older children and adults (for a recent review, see Hanania & Smith, 2010). However, it is not clear that such developmental trends are a product of learning, or even of increments in the separability of specific dimensions, as they could be the product of developmental changes in selective attention abilities. Evidence suggesting that color experts can selectively attend to at least some integral color dimensions more easily than non-experts more clearly points towards a role of learning in determining dimensional separability (Burns & Shepp, 1988).

Although little is known about what conditions might foster learning of separable dimensions, one possibility is that these conditions are met in categorization tasks. A controversial hypothesis in the field of object categorization and recognition is that these processes are often accompanied by the creation of new features (Schyns, Goldstone, & Thibaut, 1998).

There is a large body of work suggesting that categorization training does produce changes in perceptual representations of the stimuli involved (for recent reviews, see Goldstone, Gerganov, Landy, & Roberts, 2009; Goldstone & Hendrickson, 2010). Stimulus dimensions that are relevant for category discrimination become more distinctive, in what has been termed "acquired distinctiveness." Operationally, acquired distinctiveness is observed as an increase in discriminability along the category-relevant dimension after categorization training. A special case occurs when the greatest enhancement in discriminability is seen at the boundary between categories, which can be interpreted as a case of acquired categorical perception. On the other hand, stimulus components that are irrelevant for category discrimination become less distinctive, in what has been termed "acquired equivalence." Operationally, acquired equivalence is observed as a decrease in discriminability along the category-irrelevant dimension after categorization training.

Some evidence suggests that categorization training involving already-existing separable dimensions produces both acquired distinctiveness along the relevant dimension and acquired equivalence along the irrelevant dimension. On the other hand, categorization training involving integral dimensions produces acquired distinctiveness in both relevant and irrelevant dimensions (Goldstone, 1994b). These results are consistent with the possibility that categorization training alters selective attention to the category-relevant dimension. With separable dimensions, the category-relevant dimension can be selectively attended, whereas with integral dimensions, attention must be paid to both dimensions. Still, the increase in discriminability of integral dimensions was larger for the category-relevant dimension than the category-irrelevant dimension. Goldstone interpreted these results as suggesting that categorization training produces differentiation of integral dimensions.

However, using the same integral dimensions as Goldstone (saturation and brightness), Foard and Kemler-Nelson (1984) found evidence that learning effects in a sorting task transferred across different sets of stimuli only when the task-relevant dimension corresponded to the dimensions identified by the experimenter (that is, either saturation or brightness). If the task-relevant dimension was rotated 45 degrees from the original dimensions, learning did not transfer. This suggests that the integral dimensions of saturation and brightness might be primary axes in stimulus space (Smith & Kemler, 1978) despite the fact that they usually interact during perceptual tasks (see also Melara, Marks, & Potts, 1993). Thus, the results reported by Goldstone (1994b) can be interpreted as an increase in selective attention or as further differentiation of already-existing psychological dimensions.

There is also evidence suggesting that categorization training produces novel psychologically-differentiated stimulus dimensions. Goldstone and Steyvers (2001) were the first to report evidence for such an effect. In their study,

complex novel stimulus dimensions were created by taking two faces and gradually morphing one into the other in a continuous sequence. A two-dimensional face space was then created by taking two such face dimensions and factorially morphing each of their levels (see Fig. 2). In their first experiment, Goldstone and Steyvers found that training in a categorization task in which one novel dimension was relevant and another was irrelevant transferred to a new task in which either the relevant or the irrelevant dimension was replaced by a completely new dimension. Further experiments showed that these effects are not simply due to similarity-based transfer, but are better explained as the outcome of dimension differentiation. People were trained with two categorization rules using the same pair of stimulus dimensions. After learning the first categorization rule, better transfer was observed if the second rule was a 90-degree rotation from the first (i.e., the irrelevant dimension became relevant and vice versa) than if it involved a 45-degrees rotation, despite the fact that a smaller number of stimuli switched labels in the latter case. The same effects were not found if the stimuli could be described by relatively separable dimensions from the start, such as the shape of eyes and mouth. This suggests that the existence of previously available separable dimensions impairs learning of new dimensions during categorization tasks. If such dimensions do not exist, however, any direction in stimulus space can become a novel differentiated dimension, insofar as the stimulus space is created through factorial combination of stimulus sequences (Folstein, Gauthier, & Palmeri, 2012).

In contrast to these reports, there have been some failures to find any effect of categorization training on dimension differentiation when the stimuli were novel shapes created by combining sinusoidal functions (Op de Beeck, Wagemans, & Vogels, 2003). Even so, other evidence suggests that the use of special categorization training procedures that are thought to produce learning of more "robust" categories, can lead to dimension differentiation using such novel shapes (Hockema, Blair, & Goldstone, 2005).

This line of research has been driven mostly by the hypothesis that category learning is accompanied by the learning of novel features and dimensions, but the question as to whether such dimensions are truly separable has remained unanswered. Although some results have been interpreted as supporting the hypothesis of separability learning (Goldstone & Steyvers, 2001; Hockema et al., 2005), no previous experiment has directly tested this hypothesis by actually assessing dimensional separability of the relevant dimensions before and after categorization training. The rotation test performed in some of these studies (Folstein et al., 2012; Goldstone & Steyvers, 2001) is suggestive of dimension learning, but it is not usually considered a test of dimensional separability. Instead, it is better described as a test of whether two dimensions are "primary axes" (Smith & Kemler, 1978) – that is, psychologically meaningful directions in stimulus space –, even though they might combine in an integral fashion (Grau & Nelson, 1988; Melara et al., 1993). The related concepts of acquired distinctiveness and acquired equivalence are also different from separability learning, as they refer to changes in discriminability of the category-relevant and category-irrelevant dimensions, respectively, while separability learning refers to changes in the interaction between both dimensions. Thus, tests of these two concepts are also different from tests of separability learning. As we have seen earlier, perceptually separable dimensions are processed in special ways that are not true of all stimulus dimensions. Thus, an important open question is whether categorization training increases the separability of a novel category-relevant dimension.

1.3. The present study

The goal of the present study was to evaluate whether categorization training increases dimensional separability of the category-relevant dimension, using traditional tests of separability and GRT-based analyses.

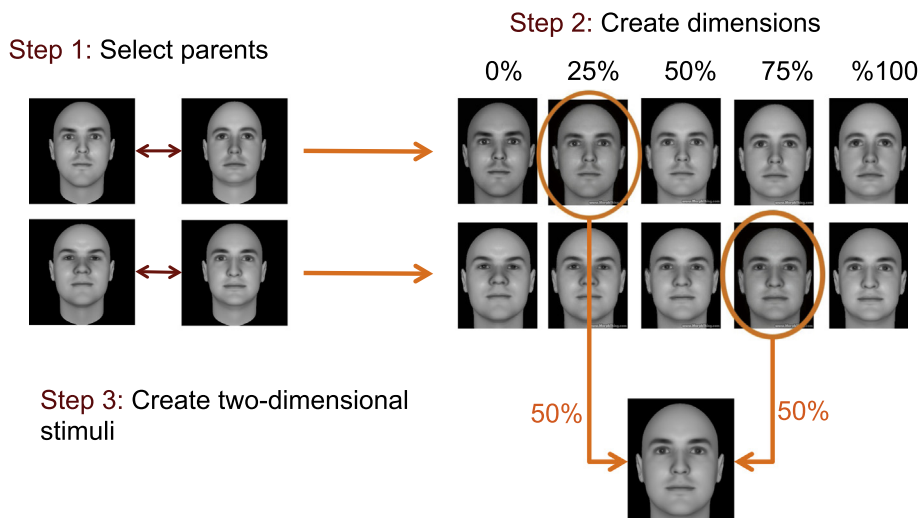We studied separability of completely novel dimensions created through face morphing, as in the seminal

Fig. 2. Schematic representation of the procedure used here to create a two-dimensional space of morphed faces.

work of Goldstone and Steyvers (2001). Using novel dimensions built through morphing is convenient because there is evidence suggesting that such dimensions are integral (Goldstone & Steyvers, 2001) and that there are no psychologically-meaningful directions in a space constructed this way (Folstein et al., 2012). Furthermore, morphing results in stimuli and dimensions that seem more ecologically valid than completely artificial stimuli.

We created a two-dimensional space of faces and trained participants in a categorization task that could be solved by selectively attending to only one of those dimensions. Next, we tested separability of the category-relevant dimension in the context of a completely novel dimension, never seen during training. We were interested in learning that is specific to the relevant training dimension, but not specific to the stimuli and irrelevant dimensions seen during training. We reasoned that when people are presented with new problems involving known dimensions, it is unlikely that such problems involve exactly the same stimuli and dimensions experienced earlier. Instead, new objects are likely to have properties that have never been seen before. Furthermore, finding that the separability of a dimension increases even when it is tested with a completely new dimension implies that there has been separability learning with the original dimensions, which generalizes across changes in the category-irrelevant dimension. This is a more general finding than observing increases in separability only for the originally-trained dimensions.

Two different tasks were used to test perceptual separability. In Experiment 1, we used the popular Garner filtering task and evaluated separability using a summary statistics approach to GRT analyses. In Experiment 2, we used an identification task and evaluated separability by directly estimating the parameters of a GRT model from the data using maximum likelihood estimation, followed by tests of separability performed on such estimates.

2. Experiment 1

Perhaps the most popular test of dimensional separability is the Garner filtering task (Garner, 1974), in which participants classify a number of stimuli according to their value on some target dimension as fast as possible. Participants complete two different conditions in different blocks. In control blocks, the stimuli vary only on the target dimension, whereas in filtering blocks, the stimuli also vary on an irrelevant dimension. If the target dimension is separable from the irrelevant dimension, then participants should be able to selectively attend to the target dimension during filtering blocks, and performance should therefore not differ significantly from that found during control blocks. On the other hand, if the target dimension is not separable from the irrelevant dimension, then performance should suffer in the filtering condition compared to the control condition, what is usually called an *interference effect*.

Although the Garner filtering task is popular and intuitive, by itself it provides only an operational test of perceptual separability, without a rigorous theoretical justification. Fortunately, the data gathered from a Garner filtering task can be used to compute a number of statistics that are known to be related to perceptual separability as defined within GRT. When the 2×2 design discussed earlier is used in a Garner filtering task, participants must report the perceived level of the relevant dimension A while ignoring the level of dimension B. Let a_i denote the event that the participant responded that the level of dimension A was i, where $i = 1$ or 2. Then a simple statistic that can be computed from the response frequencies is the proportion of correct responses for each of the four stimuli during filtering blocks, or $p(a_i|A_iB_j)$. *Marginal response invariance* holds when the probability of identifying the correct level of the relevant dimension A does not depend on the level of B, that is, when

$$p(a_i|A_iB_1) = p(a_i|A_iB_2)$$

for both $i = 1, 2$. Ashby and Townsend (1986) showed that if decisional and perceptual separability hold for dimension A, then marginal response invariance must hold as well.

Response times (RTs) are also usually gathered in a Garner filtering task. Ashby and Maddox (1994) developed an extended GRT framework to analyze such data. They assumed that classification RT decreases with the distance between the perceptual effect and the decision bound. If this RT-distance hypothesis holds then the interference effect in RT is theoretically linked to perceptual separability if two more assumptions are made: first, that decisional separability holds and second, that the perception of each stimulus is the same in the filtering and control conditions (i.e., perception is context free).

Ashby and Maddox also showed that the presence of an interference effect is generally not diagnostic of violations of perceptual separability. A much better test is given by *marginal RT invariance*: the finding that the distribution of RTs at each level of the target dimension is not affected by variations in the irrelevant dimension. Specifically, let $p(RT_i \leqslant t|A_iB_j)$ denote the probability that the RT is less than or equal to some value t on trials when the participant correctly classified the level of component A. Then *marginal RT invariance* is found when

$$p(RT_i \leqslant t|A_iB_1) = p(RT_i \leqslant t|A_iB_2),$$

for $i = 1$ and 2 and for all $t > 0$. If the RT-distance hypothesis and decisional separability both hold, then marginal RT invariance holds if and only if perceptual separability holds (Ashby & Maddox, 1994). That is, observing marginal RT invariance would indicate that perceptual separability holds and observing violations of marginal RT invariance would indicate that perceptual separability is violated. This makes marginal RT invariance the strongest test of perceptual separability that can be computed from a Garner filtering task.

Experiment 1 used the Garner filtering task to explore whether categorization training increases the separability of a category-relevant dimension. Participants in the experimental group received three sessions of training in a categorization task involving morphed faces that varied in two novel dimensions, whereas participants in the control group did not receive such training. During the test session, both groups performed a Garner filtering task with stimuli varying along the category-relevant dimension and

a new irrelevant dimension never seen before by either group. Perceptual separability was evaluated through both Garner interference effects and marginal invariance tests performed on response proportions and correct RTs.

2.1. Materials and methods

2.1.1. Participants

Forty-nine undergraduates at the University of California Santa Barbara voluntarily participated in this experiment in exchange for class credit or a monetary compensation. The experimental group consisted of nineteen participants and the control group consisted of thirty participants. More participants were included in the control group because we expected that a large proportion of them would be unable to master the difficult test task.

2.1.2. Stimuli

Stimuli were created from 6 parent images chosen from a database of 300 computer-generated Caucasian faces described by Oosterhof and Todorov (2008), created using the Facegen Modeller program (http://facegen.com), Version 3.1. From the original database, 33 male faces were chosen that had similar eyebrow color and similar levels of facial fat. The chosen images were converted to grayscale and their intensity histograms were equalized. This ensured that stimuli along the resulting dimensions varied in shape features, but not in simpler features such as skin color and brightness. Following the original study by Goldstone and Steyvers (2001), similarity measures were obtained from these stimuli using the efficient method described by Goldstone (1994a). Three pairs of faces with mean similarity values within 15% of each other were chosen as parents, to ensure that the dimensions created from them had relatively similar salience. A second criterion was that the parent pairs should not be discriminable along an easily verbalizable dimension, such as degree of femininity/masculinity or head width.

Morphs with different proportions of each parent face were generated in MATLAB using the factorial procedure of Goldstone and Steyvers (2001). The procedure is illustrated in Fig. 2. In the first step, pairs of faces are chosen to be the parents for a dimension. In the second step, each dimension is created by generating morphs with different proportions from each pair of parents. In the example shown in the figure, five levels are created for each dimension by creating morphs with 0%, 25%, 50%, 75% or 100% of the second parent. These levels were chosen to illustrate the procedure; they are not the levels used in our experiment. After two face dimensions are created, the third and final step is to generate a two-dimensional space by factorially combining each of the faces in each dimension with each of the faces in the other dimension. As shown in the figure, these two-dimensional morphs include 50% from each of the one-dimensional faces. In the example, the final two-dimensional face has a level of 25% in one dimension and 75% in the other dimension.

Dimensions were created using a continuous sequence of 19 morphs for each pair of parents, with percentages of parent 1 equal to 0%, 6%, 14%, 20%, 24%, 30%, 32%, 38%, 42%, 50%, 58%, 62%, 68%, 70%, 76%, 80%, 86%, 94%, and 100%. The three resulting dimensions were used to create two-dimensional spaces by factorially combining all levels of one dimension with all levels of a second dimension.

2.1.3. Procedure

Participants in the experimental group were exposed to 3 sessions of pre-training in the categorization task shown in Fig. 3a. This task has been used in the past to show learning of new dimensions (Folstein et al., 2012) and has the advantage that the circular arrangement of stimuli de-emphasizes the dimensional structure of the stimuli. The sessions were run within a span of three days and no more than two sessions were run on the same day. Consecutive sessions were separated by at least 1 h and at most 25 h, with the exception of a single pair of sessions that was separated by 10 min. At the beginning of each session, instructions were displayed on the screen indicating that the participant's task was to categorize faces as accurately as possible into two different categories (clubs) based purely on physical appearance. The instructions also explained the structure of each trial and how to report a categorization response. Participants were warned that they would need to guess the correct answer early in training, but they would get more accurate as the experiment progressed.

Each pretraining session consisted of 9 blocks of 72 trials each, for a total of 648 trials. Each stimulus (36 per category) was presented once in a block, with the order randomized within the block. There were voluntary breaks of 1 min between blocks, which the participant could finish by clicking on a button labeled "continue."

Each pretraining trial started with the presentation of a white cross in the middle of a black screen for 500 ms. Immediately afterwards a face stimulus was presented in the middle of the black screen until the participant pressed one of the two response buttons or a time deadline of 4 s was reached, whichever happened first. Then the participant received feedback about the correct response. For correct responses, the word CORRECT was presented for 500 ms, in green font color in the middle of the screen, accompanied by a pleasant chime presented through the headphones. For incorrect responses or if the time deadline was reached, the word INCORRECT was presented for 500 ms, in red font color in the middle of the screen, accompanied by an unpleasant buzzer presented through the headphones. This was followed by a 1 s inter-trial interval, during which the monitor was completely black.

Participants in both groups completed one session of a Garner filtering task illustrated in Fig. 3b. This task used four stimuli, which resulted from the factorial combination of two levels of the category-relevant dimension and a novel dimension. As shown in the figure, levels 5 and 15 from each dimension were used and none of the combinations of levels was used before in the categorization task. At the beginning of the session, instructions were displayed indicating that the participant's task would be to categorize four stimuli into two different groups based on physical appearance. The four stimuli were displayed in the screen, grouped in the two categories that would be used during the task. Participants were asked to study the two groups of stimuli carefully before continuing with
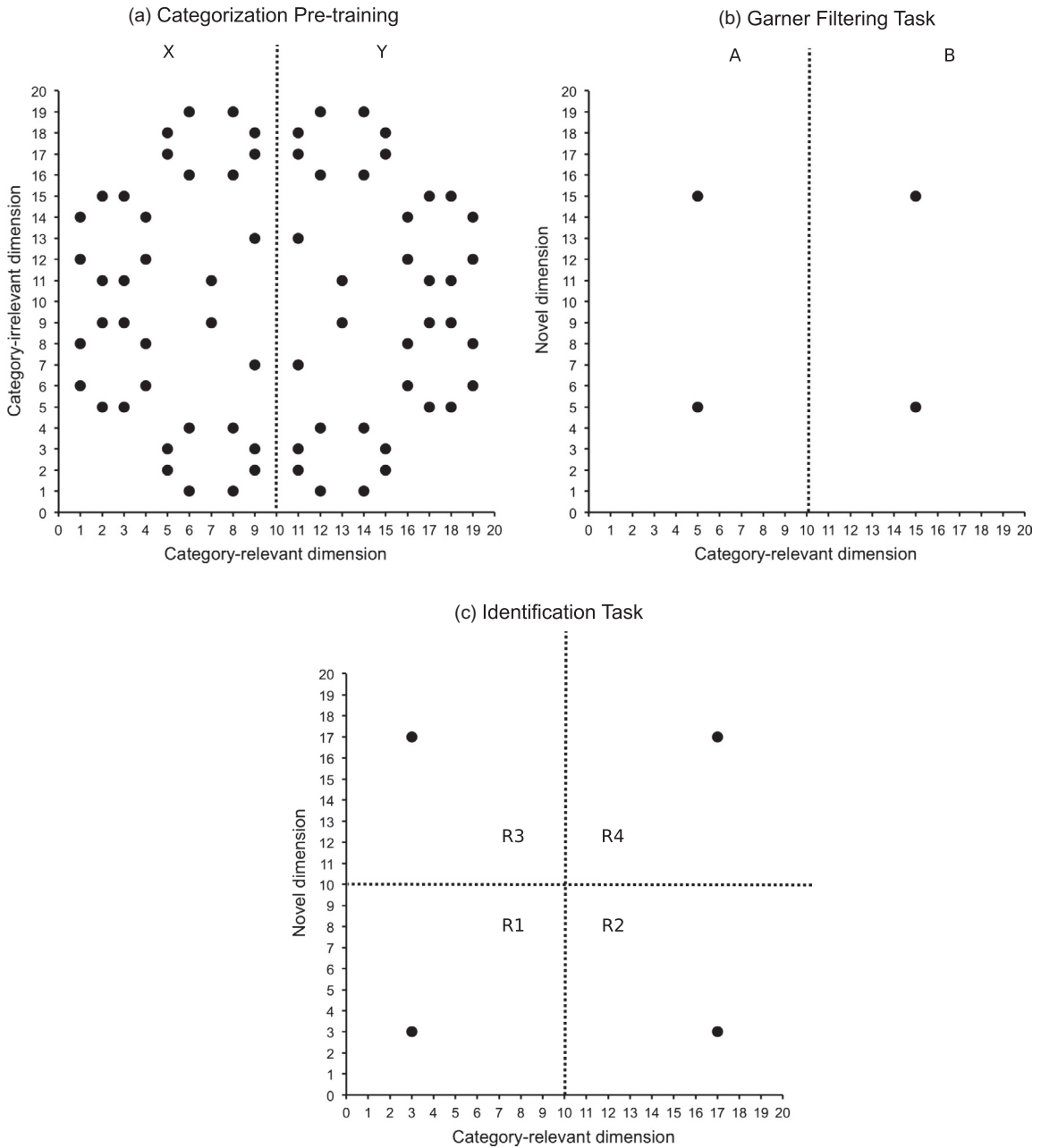
Fig. 3. Schematic representation of the tasks used in the experiments presented here. Each point represents a different stimulus presented during a task. The requirements of the task are represented by lines that divide the stimuli in response classes. Such response classes are also labeled for each task (X and Y; A and B; R1–R4).

the experiment, as a way to highlight in a non-verbal way what aspects of the stimuli were relevant and irrelevant for the task. The instructions also explained the structure of a categorization task and asked participants to respond as accurately and as fast as possible.

The session consisted of 16 blocks of 32 trials each, for a total of 512 trials. There were three block types that differed on the stimuli presented. During baseline blocks, only two stimuli were presented, which varied in the level of the category-relevant dimension, but for which the

irrelevant dimension was held constant. Each stimulus was repeated 16 times. There were two types of baseline blocks, one for each level of the irrelevant dimension. During filtering blocks, all four stimuli were presented, each repeated 8 times. Trials were randomized within blocks.

Each type of baseline block was repeated 4 times and the interference block was repeated 8 times. Blocks were grouped in pairs of one baseline block and one interference block; there were a total of 8 block pairs in the session. The

order of each block in the pair was counterbalanced and the order of presentation of the block pairs was semi-randomly determined for each participant, with the constraint of alternating the baseline block type in the sequence of block pairs.

Trials in the Garner filtering task had the same overall structure as trials in the pre-training categorization task. However, the response deadline was set to 2 s, after which the phrase TOO SLOW! was presented in the middle of the screen in red font color, accompanied by an unpleasant buzzer presented through the headphones.

2.2. Results and discussion

The experimental group showed accurate performance in the final session of the categorization pre-training task, with a mean proportion of correct responses of 0.81 ($sd = 0.1$). The Garner filtering task was quite difficult to learn, especially for participants in the control group, with some participants showing chance performance throughout the task. To minimize differences in task performance between the control and experimental groups, and because GRT is only applicable to asymptotic performance and not to learning data, for each participant we only included data starting from the first block pair in which the proportion of correct responses was at least 0.75. This excluded from the analysis all data from 14 participants in the control group (9 of these showed performance that seemed completely random) and 2 participants in the experimental group, who never reached such a high level of performance. Thus, the results presented here are based on data from 16 participants in the control group and 17 participants in the experimental group.

Analyses of RTs (measured in seconds) were performed only on data from correct trials. We used quantiles (median and deciles) to characterize the RT distribution from each participant, which were then used in group-level analyses. Quantiles are asymptotically normal and unbiased estimators, so they can be used to obtain group averages and to perform statistical analyses involving assumptions of normality (Van Zandt, 2002).

2.2.1. Garner interference effects

An interference score was computed for each participant by subtracting accuracy during control blocks from accuracy during filtering blocks. The mean of these interference scores is plotted separately for each group in Fig. 4a. It can be seen that there is a Garner interference effect in accuracy for the control group, with the mean interference score ($\bar{x} = .033$) being significantly higher than zero according to a single-sample t-test, $t(15) = 2.84$, $p < .01$, $d = .71$. On the other hand, the Garner interference effect in accuracy for the experimental group is much smaller ($\bar{x} = .002$) and not significantly greater than zero, $t(16) = .39$, $p > .1$, $d = .09$. The difference between groups ($\bar{x}_1 - \bar{x}_2 = .031$) was also found to be significant according to a two-sample t-test, $t(31) = 2.41$, $p < .05$, $d = .87$.

Traditionally, the analysis of Garner interference effects in RT is performed on the median values from the relevant RT distributions of each participant. In line with this standard procedure, we first computed the median RT for each participant for both filtering and baseline blocks. An interference score was computed by subtracting median RT during control blocks from median RT during filtering blocks. The mean of these interference scores is plotted separately for each group in Fig. 4b. The results mirror those found with accuracy: the control group shows a Garner interference effect that seems to be absent in the experimental group. The mean interference score ($\bar{x} = .023$) was significantly greater than zero in the control group, $t(15) = 2.06$, $p < .05$, $d = .52$, but not in the experimental group ($\bar{x} = .003$), $t(16) = .52$, $p > .1$, $d = .13$. However, the difference between groups ($\bar{x}_1 - \bar{x}_2 = .02$) was only marginally significant, $t(31) = 1.63$, $p = .057$, $d = .59$.

2.2.2. Marginal invariance tests

During each baseline block, participants repeatedly saw only two stimuli that shared the same value on the irrelevant dimension, so they had time to adjust their decision strategy to the demands imposed by those specific stimuli. In other words, participants might have used different decision bounds in each type of baseline block, which would violate the assumption of decisional separability required by the theorems linking marginal invariances with perceptual separability (Ashby & Maddox, 1994; Ashby & Townsend, 1986). This is less likely to happen during filtering blocks, in which participants saw all four stimuli randomly interspersed. For this reason, analyses of marginal invariances were performed only on data from filtering blocks.

In the analysis of marginal response invariance, for each participant and each level of the relevant dimension we computed the difference in accuracy across the two levels of the irrelevant dimension. The resulting value represents the deviation from marginal response invariance at one level of the relevant dimension, so there were two of these scores for each participant (one for each value of the relevant dimension). The absolute values of these two scores were added together to obtain a single score per participant, representing violations of marginal response invariance. The means of these scores are plotted separately for each group in Fig. 4c. Note that we expect these means to be above zero even if there are no violations of marginal response invariance, because in that case each individual score would be the sum of two absolute random deviations. Thus, testing whether each mean is different from zero using traditional statistical tests is not appropriate. To test whether there were significant violations of marginal response invariance, we performed a permutation test by randomly shuffling the levels of the irrelevant dimension in each participant's data 500 times. The mean scores representing violations of marginal response invariance were computed from each of these shuffled data sets, resulting in an empirical distribution of mean scores under the null hypothesis of no violations of marginal response invariance. These permutation tests revealed significant violations of marginal response invariance in both groups (both $p < .01$).

More importantly, the comparison between groups clearly showed that violations of marginal response
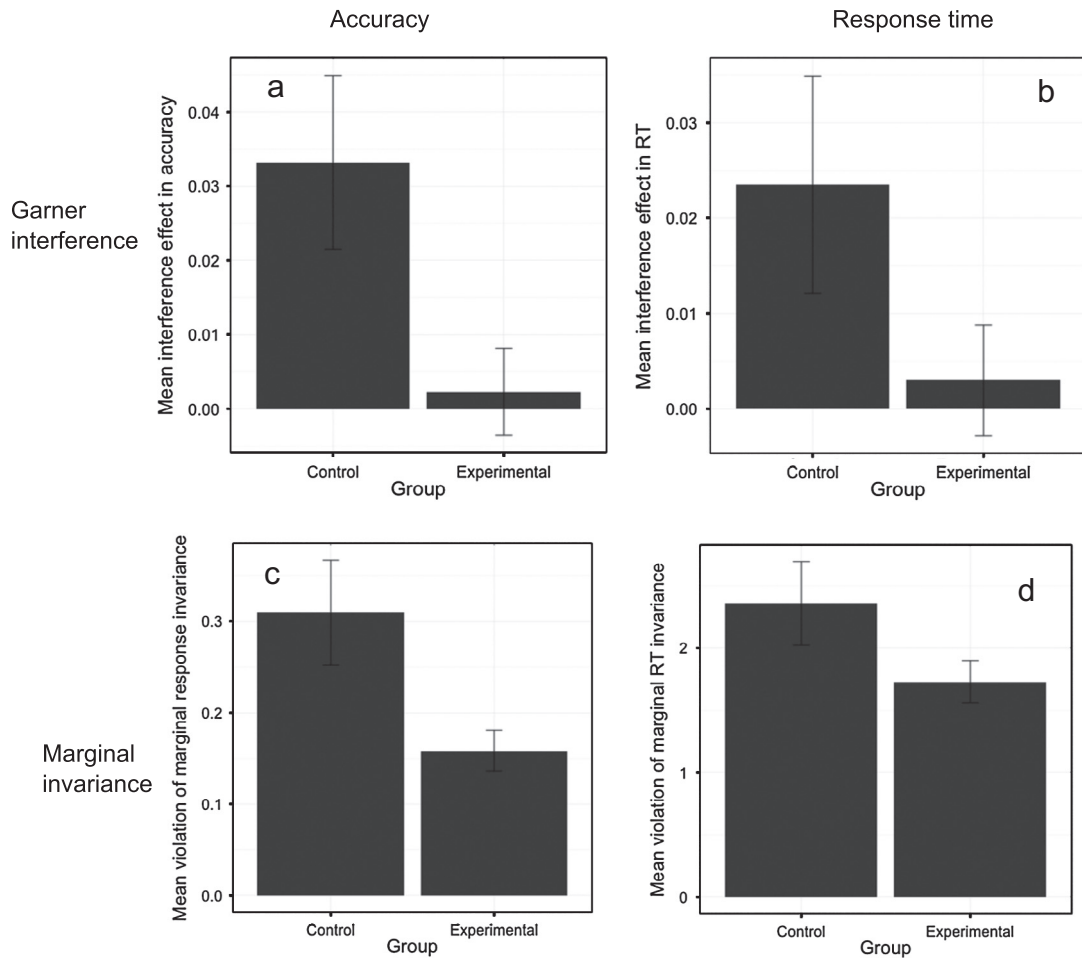
Fig. 4. Results of the Garner interference and marginal invariance tests carried out in Experiment 1. Error bars represent standard error of the mean.

invariance were higher in the control group than in the experimental group $(\bar{x}_1 - \bar{x}_2 = .15)$, $t(31) = 2.51$, $p < .01$, $d = .9$.

Remember that for marginal RT invariance to hold, the distribution of RTs for a given level of the relevant dimension should be the same across levels of the irrelevant dimension. Thus, unlike the Garner interference test, which requires a comparison of only median RTs, the marginal RT invariance test requires comparing full RT distributions. With this goal in mind, we started by computing the deciles of the RT distribution for each participant and each stimulus. As indicated earlier, quantiles (which include deciles) have been previously recommended for the analysis of RT distributions due to their good statistical properties (Van Zandt, 2002). Each set of deciles summarizes the RT distribution for one level of the relevant dimension and one level of the irrelevant dimension. If marginal RT invariance holds, we would expect that corresponding deciles for a given level of the relevant dimension would be the same across levels of the irrelevant dimension. Using d_{ijk} to represent the kth decile from the RT distribution of a stimulus with level i in the relevant dimension and level j in the irrelevant dimension, we computed a single measure of deviations

from marginal RT invariance for each participant in the following way:

$$V = \sum_{i=1}^{2}\sum_{k=1}^{9}|d_{i1k} - d_{i2k}|,$$

where $|x|$ represents the absolute value of x. The mean of these scores is plotted separately for each group in Fig. 4d. It can be seen that violations of marginal RT invariance were higher in the control group than in the experimental group $(\bar{x}_1 - \bar{x}_2 = .64)$ with the difference being statistically significant, $t(31) = 1.71$, $p < .05$, $d = 0.61$.

As in the analysis of marginal response invariance, testing whether the observed deviations from marginal RT invariance were statistically significant within each group required performing a permutation test. The test followed the same procedure as before, but computing the V statistic for each permuted data set. The results of these permutation tests revealed significant violations of marginal RT invariance in both groups (both $p < .01$).

Fig. 5 partially captures the differences between groups in RT distributions. The figure plots the mean of the decile estimates separately for the control (top panels) and experimental groups (bottom panels) and for each level
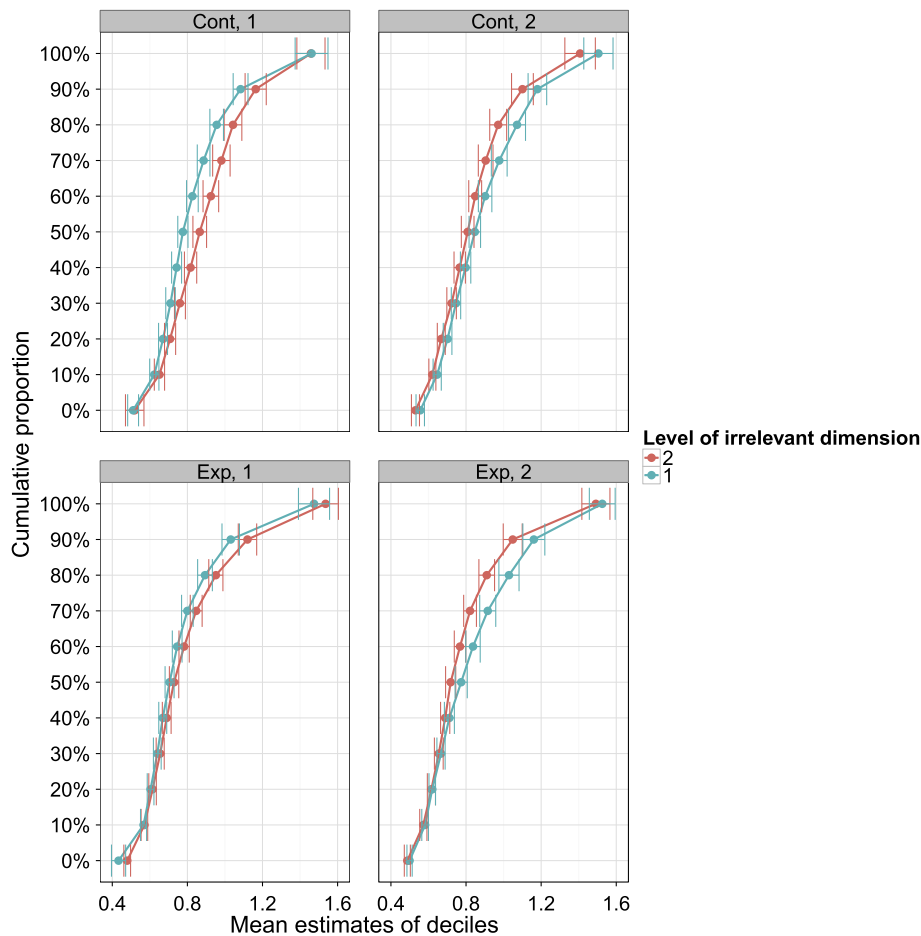
Fig. 5. Mean estimates of deciles for each group (Exp = experimental, Cont = control) and level of the relevant dimension (number in the subpanel titles). Different colors represent different levels of the irrelevant dimension. If marginal RT invariance held, lines within each panel should not differ. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of the relevant dimension (level 1 in left panels, level 2 in right panels). The distributions for the two levels of the irrelevant dimension are plotted using different colors. For each panel, the magnitude of violations of marginal RT invariance is represented by the horizontal distance between the two distributions. Note that such violations are apparent in both the experimental and control groups, even though these averaged curves can only capture deviations from marginal RT invariance that are consistent across participants. The plot also captures some of the differences between groups, especially in the lower half of the distributions (the first five deciles, corresponding to cumulative proportions of 0–50%). For level 1 of the relevant dimension (two left panels), we see that for the first five deciles (0–50%) there is almost no difference between the two distributions for the experimental group (bottom-left panel), whereas there are considerable differences for the control group (top-left panel). For the last five deciles (60–100%), differences between the distributions start being apparent in the experimental group, but these are much smaller than those found in the control group. The differences between groups are not so clear-cut for level 2 of the relevant dimension, corresponding to the two right

panels. For the first five deciles (0–50%), differences between distributions are slightly larger in the control group than the experimental group, but this pattern seems to reverse in the last five deciles (60–100%).

In summary, both the size of the Garner interference effect and the magnitude of violations of marginal invariance were smaller in the experimental group, which received categorization pre-training, than in the control group, which did not receive such pretraining. This difference between groups was found in the analysis of both accuracy and RT data and suggests that categorization training increased the perceptual separability of the category-relevant dimension.

On the other hand, permutation tests revealed violations of marginal response and RT invariance in both groups, suggesting that categorization training does not produce complete dimensional separability.

A problem with the present experiment is that the Garner interference and marginal invariance tests both rely on the assumption that decisional separability holds (Ashby & Maddox, 1994). Although it has been found that this assumption is valid in some cases (Maddox & Ashby, 1996), it is possible that categorization training had an

impact on the decision strategies used by participants rather than on perceptual separability. One way to better dissociate perceptual and decisional separability is by using an identification task, in which participants are asked to report the perceived level of both dimensions on each trial. GRT analyses of identification data do not require assuming decisional separability to test for perceptual separability, instead it is possible to simultaneously test the validity of both forms of dimensional interaction. The goal of Experiment 2 was to evaluate whether categorization pre-training increases perceptual separability using an identification task.

3. Experiment 2

The present experiment evaluated separability learning using GRT analyses of data from an identification task. Participants with experience in a categorization task performed a 2×2 identification task using stimuli created from the category-relevant dimension and a novel stimulus dimension, as illustrated in Fig. 3c. Their performance was compared to that of participants without experience with the category-relevant dimension.

The 2×2 identification task used here is similar to the classification task used in the previous experiment, in that it includes four stimuli that result from the factorial combination of two levels of dimension A and 2 levels of dimension B (compare Fig. 4b and c). The main difference is that in the identification task participants must report not only the perceived level on dimension A, but also the perceived level on dimension B – that is, they must identify the specific stimulus presented during a particular trial. The most important advantage of the identification task over the Garner filtering task used before is that it allows us to dissociate perceptual and decisional types of separability.

Currently, the best available approach within the GRT framework for the analysis of dimensional interactions, at least for 2×2 identification data, is a model-based approach using GRT-wIND (GRT with INdividual Differences; Soto et al., 2015). GRT-wIND is an extension of GRT that assumes that all participants share similar perceptual representations. More specifically, the model assumes that the same stimulus dimensions are used by all participants to represent a set of stimuli. In most cases, these dimensions are assumed to correspond to those explicitly manipulated by the experimenter in an identification task (for more on the assumption of correspondence, see Ashby & Townsend, 1986; Dunn, 1983; Soto et al., 2014), which are usually manipulated because they are suspected to be psychologically privileged. A consequence of assuming the same dimensional representation across people is that perceptual separability (and perceptual independence) in GRT-wIND either holds for all participants or is violated for all participants.

On the other hand, GRT-wIND also assumes that different participants may allocate different amounts of attention to each stimulus dimension and may use different decision strategies. Individual differences in attention are modeled by increasing or decreasing perceptual noise on each dimension. For example, attention to dimension A would result in

less perceptual noise and easier discrimination of percepts along that dimension. Although attention might change how well an individual can discriminate a dimension, it does not affect whether two dimensions perceptually interact. Individual differences in decisional strategies are modeled by having a different set of two decision bounds for each individual. For this reason, decisional separability is a phenomenon that can hold in some individuals and fail in others, or even vary for a single individual as a function of factors such as training with a task.[1]

GRT-wIND solves a number of problems that have been recently identified in traditional GRT models. Importantly, while traditional GRT models have problems dissociating perceptual and decisional forms of separability in the 2×2 identification design (e.g., Mack, Richler, Gauthier, & Palmeri, 2011; Silbert & Thomas, 2013), we have shown that GRT-wIND does not suffer from these problems (see appendix of Soto et al., 2015).

A disadvantage of using GRT-wIND to analyze data from an identification task is that the model has not been linked to RT data yet. Recent developments in GRT have provided tools to analyze RT data from identification tasks (Townsend, Houpt, & Silbert, 2012), but these tools cannot dissociate between perceptual and decisional separability, like GRT-wIND does. Because dissociating perceptual and decisional separability was one of the goals of the present experiment, here we will focus exclusively on the analysis of response frequencies using GRT-wIND.

Theorems linking dimensional interactions with summary statistics are still not available for GRT-wIND, but the full model can be fit to data from all participants using maximum likelihood estimation. Then, the analysis of dimensional interactions can be easily performed using statistical tests on maximum likelihood estimates. In Appendix, we include a more formal description of GRT-wIND, the procedures used to find maximum-likelihood estimates of its parameters, and statistical tests for the analysis of dimensional interactions. Importantly, Appendix describes a new statistical test that allows between-group tests of differences in perceptual separability, decisional separability and perceptual independence, which makes it possible to directly compare perceptual separability in the experimental and control groups of the present experiment.

3.1. Materials and methods

3.1.1. Participants

The same 19 participants from the experimental group of Experiment 1 were included in the present experiment.

---

[1] An alternative model, which makes equivalent predictions to GRT-wIND, assumes that all participants share the same decision bounds but every participant attends to different stimulus dimensions (which can be oriented in any direction of stimulus space). We did not explore this model for two reasons. First, following standard applications of signal detection theory, we assumed that participants have more control over their decision strategies than over their sensory and perceptual processing of the stimulus, whereas this alternative model assumes the opposite. Second, fitting the alternative model would be difficult, because it requires finding what "privileged" directions in space should be allowed to stretch and shrink for each participant.

An additional 25 undergraduates from the University of California Santa Barbara were recruited for the control group. They voluntarily participated in exchange for class credit.

3.1.2. Stimuli

Stimuli were the same as those described for Experiment 1.

3.1.3. Procedure

Participants in the experimental group completed 3 sessions of pre-training in a categorization task, plus one session of testing in a Garner interference task, as described previously. Participants in both groups then completed one session of the identification task illustrated in Fig. 3c. As in Experiment 1, the task used four stimuli resulting from the factorial combination of two levels of the category-relevant dimension and two levels of a novel dimension. In contrast to Experiment 1, levels 7 and 17 from each dimension were used. Note that none of the combinations of these levels was experienced before by the experimental group.

At the beginning of the session, instructions were displayed indicating that the participant's task would be to identify four faces, each one of them assigned to a unique response key. The four stimuli were displayed on the screen, together with their key assignments. The instructions also explained the structure of a trial in the task, emphasizing that stimuli would be presented very briefly, and that participants should respond as accurately and as fast as possible.

The session consisted of 45 blocks of 20 trials each, for a total of 900 trials. Each block involved 5 presentations of each of the four stimuli and trials were randomized within blocks. Each trial started with the presentation of a white cross in the middle of a black screen for 500 ms. Immediately afterwards a face stimulus was quickly flashed in the middle of the black screen for 34 ms. This short presentation time was chosen to make the task more difficult, as errors are critical to an analysis using GRT-wIND. Participants were given 2 s to report the presented face by pushing the correct button and they received feedback about the correctness of their responses as in Experiment 1. If the participant did not respond within 2 s, the trial ended with the message "TOO SLOW!" displayed in the middle of the screen and accompanied by an unpleasant buzzer through the headphones.

3.2. Results and discussion

The data from 11 participants (9 from the control group and 2 from the experimental group) were excluded from the analysis because their performance was near chance (below 27% correct) by the end of the experiment. GRT is a model of asymptotic performance, not of learning, so it is important to discard data during the learning period when estimating individual participant confusion matrices. Toward this end, learning curves were obtained by averaging performance within a moving window of 50 trials, starting with the average of trials 1–50, moving the window one trial up in each step (2–51, 3–53, and so

on), and ending with the average of the last 50 trials. An exponential function was fit to the resulting average learning curves using least-squares estimation. The point in the best-fitting exponential curve where the slope was smaller than 0.001 for the first time was used as a cutoff: only data after this point were used to estimate individual confusion matrices.

GRT-wIND was fit to the data from individual confusion matrices using the procedures outlined in Appendix. A separate fit was performed for the data of each group. To facilitate finding the global maximum of the likelihood function instead of a local maximum, the optimization was run 60 times, each time with different random starting values for the parameters and the solution with highest maximum likelihood was chosen for further analysis.

The parameter estimates from the best-fitting model were used to estimate response probabilities for each cell in each individual participant's confusion matrix. These estimated probabilities could then be compared with the corresponding observed response proportions for a quick evaluation of the model's ability to account for variability in the data (Soto et al., 2015). It was found that the model accounted for 97.43% of the variance in the data from the experimental group and 95.77% of the variance in the data from the control group.

Fig. 6 shows the group perceptual distributions obtained from the best-fitting model, separately for each group. Each ellipse represents the contour of equal likelihood for the distribution of perceptual effects elicited by one stimulus. The category-relevant dimension is represented on the abscissa, whereas the novel dimension is represented on the ordinate. Note that the scales of each dimension are not comparable across groups; the comparison of results should focus on what the figure suggests about dimensional interactions. We will focus first on the results regarding the category-relevant dimension, which are the most important given the goals of this experiment.

3.2.1. Separability of the category-relevant dimension

Violations of perceptual separability of the category-relevant dimension can be evaluated from Fig. 6 by comparing horizontal means and variances of the two left distributions and the two right distributions. The results from the control group, shown in Fig. 6a, suggest strong violations of perceptual separability for the category-relevant dimension. For example, the bottom-left distribution has a much smaller variance along the category-relevant dimension than the top-left distribution. Furthermore, the bottom-right distribution has a much smaller mean than the top-left distribution. These violations of perceptual separability were found to be statistically significant according to a Wald test (see Appendix for a description), $\chi^2(4) = 38.54$, $p < 0.001$, and they confirm the results from Experiment 1 and previous studies (Goldstone & Steyvers, 2001) suggesting that dimensions created by face morphing are not separable.

The results from the experimental group, shown in Fig. 6b, are quite different, with both means and variances along the relevant dimension being much more similar across levels of the irrelevant dimension. In other words, the figure suggests that categorization training in this group
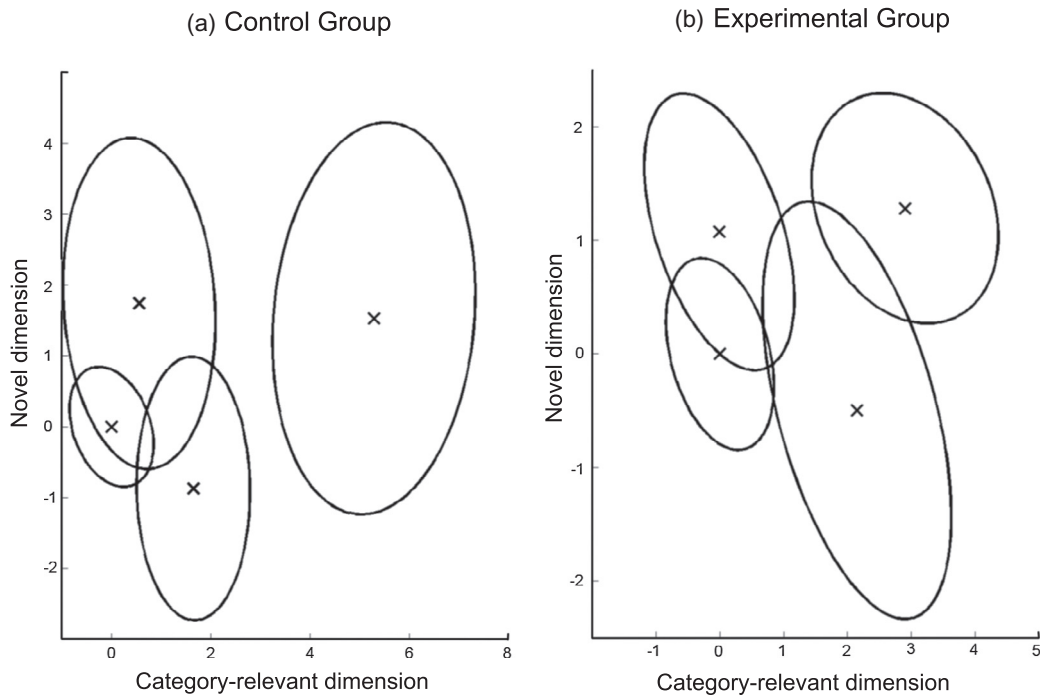
Fig. 6. Best-fitting configuration of perceptual distributions for control (a) and experimental (b) groups of Experiment 2. Ellipses are contours of equal likelihood.

did indeed increase the separability of the category-relevant dimension. This was confirmed by statistical tests, which revealed that the deviations from perceptual separability of the category-relevant dimension were not statistically significant in the experimental group, $\chi^2(4) = 1.82$, $p > 0.1$. Furthermore, a statistical test comparing the magnitude of deviations from perceptual separability between groups (which is derived in Appendix) showed that the two groups differed significantly in the separability of the category-relevant dimension, $\chi^2(4) = 12.13$, $p < 0.05$.

To visualize differences between groups in decisional separability on the category-relevant dimension, slopes of the individual decision bounds were transformed to reflect degrees of clockwise rotation from vertical. These rotation values should equal zero when decisional separability holds, with values higher or lower than zero representing deviations from decisional separability. Both individual slope values and kernel density estimates computed from those values are shown in Fig. 7, which displays the results from both groups in separate panels.

Fig. 7a shows the rotation of bounds for the control group. It can be seen that violations of decisional separability were common, with all bounds having positive rotation values and a density mode around 30 degrees of rotation. This was confirmed by a Wald test showing that, on average, the control group showed statistically significant violations of decisional separability for the category-relevant dimension, $\chi^2(1) = 5.43$, $p < 0.05$.

Fig. 7b shows the rotation of bounds for the experimental group. In this case, although some participants seem to show violations of decisional separability, the group as a whole has slope values that concentrate around zero and the density mode is only slightly below zero. Thus, it seems as if violations of decisional separability were not as strong and consistent in the experimental group as in the control group. This was confirmed by a Wald test showing that, on average, violations of decisional separability for the category-relevant dimension were not significant in the experimental group, $\chi^2(1) = 0.1$, $p > 0.1$. A between-groups statistical test comparing the magnitude of average violations of decisional separability was only marginally significant, $\chi^2(1) = 3.79$, $p = 0.051$.

In conclusion, the results from the present experiment confirm those found previously with the Garner filtering task, providing evidence that categorization training increases the perceptual separability of the category-relevant dimension. Furthermore, this experiment also provided evidence suggestive of an increase in decisional separability of the category-relevant dimension after categorization training, although the relevant statistical test comparing both groups did not reach significance at the 0.05 level.

3.2.2. Separability of the novel dimension

Violations of perceptual separability of the novel dimension can be visualized in Fig. 6 by comparing the two bottom distributions and the two top distributions with respect to their vertical means and variances. For the control group (Fig. 6a), there seems to be differences both in means and variances between the two bottom distributions, whereas the two top distributions seem to have similar means and variances. A Wald test indicated that
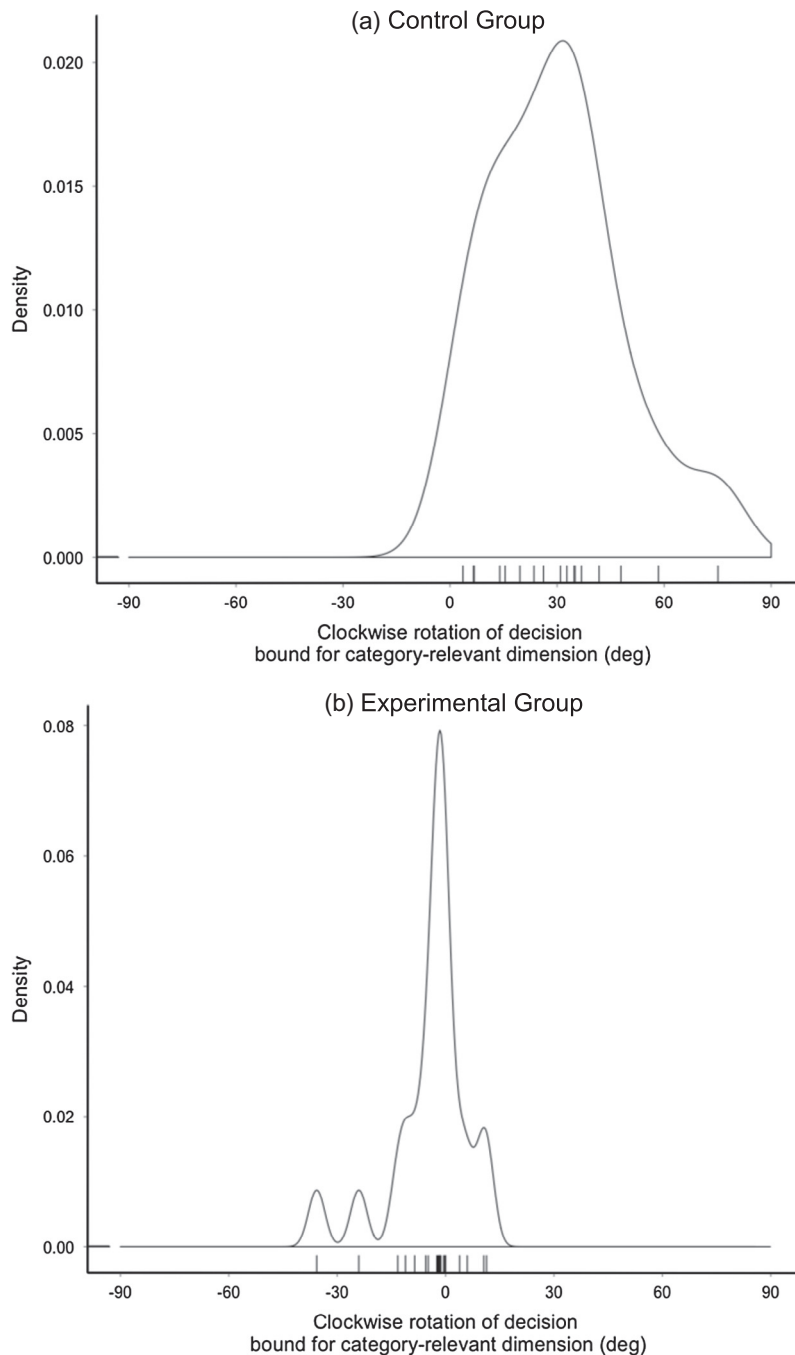
Fig. 7. Degrees of clockwise rotation from the decisional separability bound for the category-relevant dimension in Experiment 2. Both individual estimates (tick marks) and kernel density estimates are shown.

these violations of perceptual separability were statistically significant, $\chi^2(4) = 45.92$, $p < 0.001$.

Similar deviations from perceptual separability are observed in the experimental group (Fig. 6b). The means and variances of the two bottom distributions differ importantly and in the same direction as those in the control group. For both groups, the mean of the bottom-right distribution is shifted down compared to the bottom-left

distribution, and the variance of the bottom-right distribution is larger than the variance of the bottom-left distribution. There is also a difference in the variances of the two top distributions that was not observed in the control group. A Wald test indicated that these violations of perceptual separability were statistically significant, $\chi^2(4) = 12.12$, $p < 0.05$. The difference between groups in magnitude of deviations from perceptual separability in
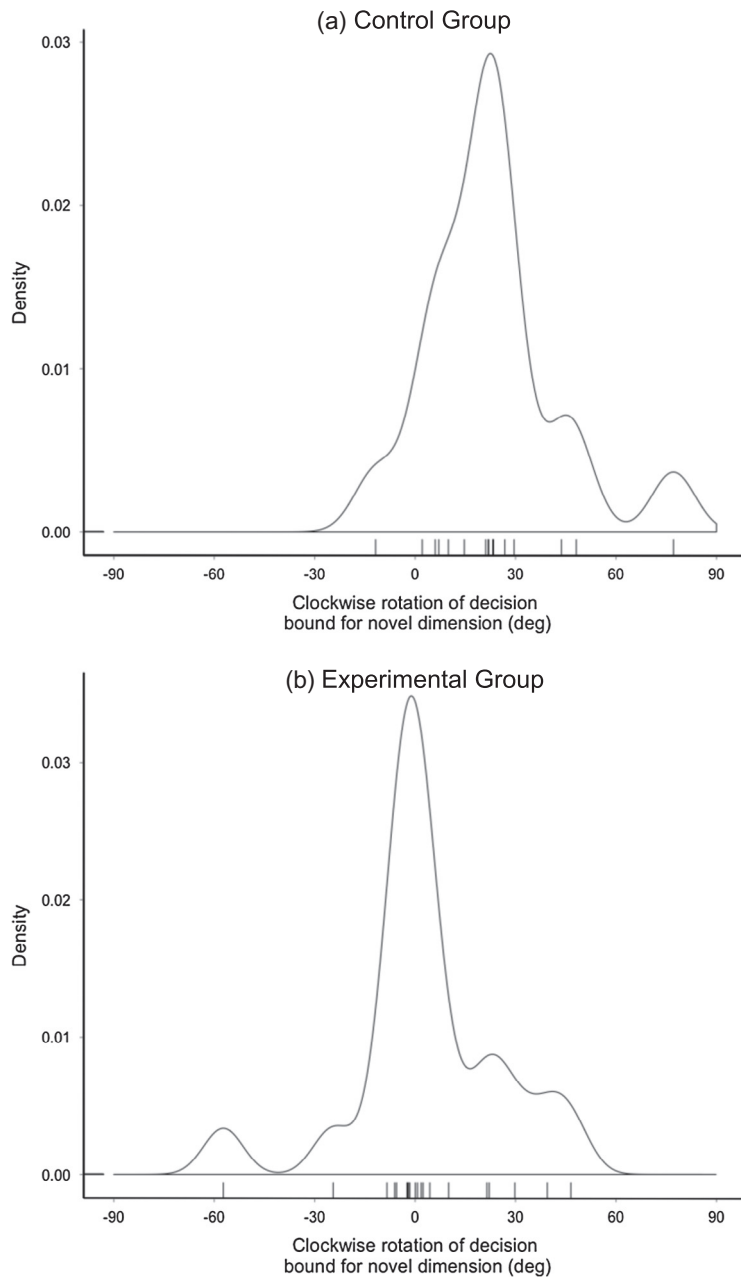
Fig. 8. Degrees of clockwise rotation from the decisional separability bound for the novel dimension in Experiment 2. Both individual estimates (tick marks) and kernel density estimates are shown.

the novel dimension was not statistically significant, $\chi^2(4) = 3.03$, $p > 0.1$.

The importance of these findings is that they indicate that separability learning is dimension-specific; that is, it involves changes in separability only for the category-relevant dimension, instead of more general changes that could affect a completely novel dimension.

The degree of clockwise rotation from horizontal in the individual decision bounds are plotted in Fig. 8. Many participants in the control group (Fig. 8a) seem to show violations of decisional separability, with all but one bound

having positive rotation values and a density mode between 20 and 30 degrees of rotation. This was confirmed by a Wald test showing statistically significant violations of decisional separability on average in the control group, $\chi^2(1) = 4.57$, $p < 0.05$.

Fig. 8b shows a different pattern of results in the experimental group. Although many participants show important violations of decisional separability, with absolute degrees of rotation larger than 20, most participants cluster around a value of 0 degrees of rotation, as reflected by the density mode. A Wald test showed that, on average, violations of

decisional separability observed in Fig. 8b were not significant, $\chi^2(1) = 0.04$, $p > 0.1$. This suggests that the effect of categorization training on decisional separability is not dimension-specific, but also affects decisional strategies for novel dimensions. However, a between-groups statistical test comparing the magnitude of average violations of decisional separability was not significant, $\chi^2(1) = 1.54$, $p > 0.1$, so the effect of category training on decisional separability is not reliable for the novel dimension.

3.2.3. Tests of perceptual independence

Although perceptual independence is not the focus of the present work, we will briefly report the results of our analyses on this form of dimensional interaction. A look at Fig. 6 reveals that none of the correlations in the control or experimental groups seem to be very high, suggesting that violations of perceptual independence were either nonexistent or very small. In line with this observation, Wald tests did not reveal significant violations of perceptual independence in either the control, $\chi^2(4) = 0.29$, $p > 0.1$, or experimental groups, $\chi^2(4) = 1.65$, $p > 0.1$. Differences between groups were also not statistically significant, $\chi^2(4) = 0.49$, $p > 0.1$.

3.2.4. Identifiability issues in GRT-wIND

Recent work (Mack et al., 2011; Silbert & Thomas, 2013) has shown that decisional separability is not identifiable in traditional GRT models for the 2×2 identification design. In some cases, perceptual separability might also show identifiability problems. Given these results, it seems appropriate to briefly discuss whether GRT-wIND might have fallen victim to such non-identifiability issues in the present study.

The most serious issue was identified by Silbert and Thomas (2013) regarding decisional separability. These authors showed that any traditional GRT model for the 2×2 identification design without decisional separability can be transformed into a model with decisional separability without affecting the predicted response probabilities. However, this non-identifiability is an exclusive feature of the specific task and model studied by Silbert and Thomas. It can be shown (see appendix of Soto et al., 2015) that for any GRT model with two or more bounds per dimension the non-identifiability of decisional separability happens if and only if all bounds for a given dimension are parallel. The family of GRT models involving two or more bounds per dimension includes the traditional GRT model for a 3×3 identification design (Ashby & Lee, 1991), the GRT model for a concurrent rating task (Ashby, 1988; Wickens, 1992), and GRT-wIND. If this condition for non-identifiability was met in the present experiment, then the values of all individual rotation parameters shown in Figs. 7 and 8 should be the same. It is easy to see that this is not the case.

A less general problem with the 2×2 design is that under some conditions a model without perceptual separability can be transformed into an equivalent model with perceptual separability (Silbert & Thomas, 2013). Although all the conditions required to observe this non-identifiability have not been worked out, one condition that is absolutely necessary is illustrated in Fig. 9. Perceptual
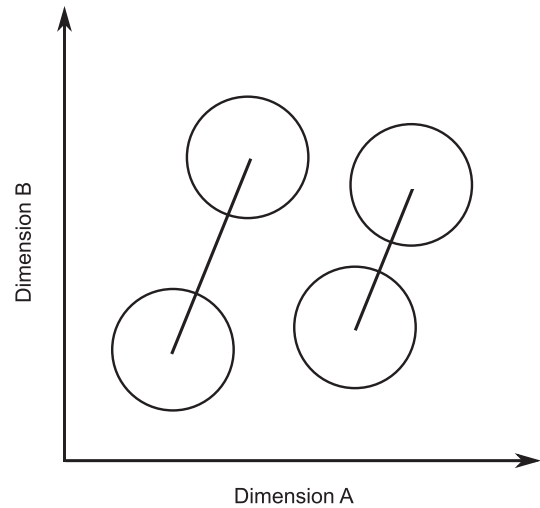

Fig. 9. Example of a GRT configuration in which perceptual separability of dimension A from dimension B is non-identifiable. A necessary condition for such non-identifiability to occur is that the line segments that join the means of distributions in the figure must be parallel.

separability of dimension A from dimension B is evaluated by comparing the distributions joined by line segments in the figure. A necessary, but not sufficient condition for non-identifiability of perceptual separability to arise is that the two line segments in Fig. 9 must be parallel. Fig. 6a shows clearly that this condition is not met for perceptual separability of the relevant dimension in our control group.

Still, the perceptually-separable configuration found in the experimental group (see right panel of Fig. 6) could be transformed into a non-perceptually-separable configuration. In the traditional individual-level GRT model, the original and transformed configurations would be equivalent, in the sense that they would predict the exact same response probabilities. On the other hand, in GRT-wIND more conditions need to be met for the transformed configuration to be equivalent to the original configuration. GRT-wIND assumes that participants might differ in the level of attention that they pay to each dimension in the task. Selective attention is modeled through an individual parameter (see Appendix) that essentially stretches and compresses the perceptual distributions in the direction of the dimensional axes. Once a transformation is applied to a GRT-wIND model, this transformation changes the position of the main axes relative to the perceptual distributions. This means that it is impossible for the new model to stretch and compress the distributions in the same directions as the original model, and therefore both models cannot be equivalent in terms of their predictions of response probabilities. Taking this into account, it is easy to see that perceptual separability can be non-identifiable in GRT-wIND only when the values of individual selective attention parameters do not stretch or compress the perceptual distributions at all, which is the case when the parameter for selective attention equals 0.5 for all participants. This was not the case in the present experiment. In the experimental group, the estimated values of the selective attention parameter ranged between 0.05 and 0.77,

with a median of 0.38. In the control group, the estimated values of the selective attention parameter ranged between 0.14 and 0.93, with a median of 0.66. In sum, GRT-wIND shows non-identifiability of perceptual separability under very specific and restricted conditions, which were not met by the maximum likelihood estimates obtained from the model in the present study.

In many signal detection theory models, means and variances of the perceptual distributions are not uniquely identifiable. Although this issue has not been systematically explored either analytically or numerically, our previous work fitting GRT-wIND to real data showed that variances seem to capture variability in the data that is not captured by the means alone, suggesting that these parameters may all be uniquely identifiable (Soto et al., 2015). Still, it might be the case that errors in our optimization algorithm led to such a result by mere chance.

Although a full study of the identifiability of means and variances in GRT-wIND is beyond the scope of the present study, we conducted two new analyses testing this possibility, which we briefly discuss here.

First, as in our previous study, we fitted a GRT-wIND model in which all variances were fixed to 1.0 to the data from both groups in the present experiment. In both cases, the fit of this restricted model was worse than that of the full model (i.e., larger AIC; for details, see Soto et al., 2015). The probability that the fixed-variances model was the best model in this comparison, computed through Akaike weights (Wagenmakers & Farrell, 2004) was 5×10^{-8} in the control group and 5×10^{-13} in the experimental group. These results are similar to those found in our previous study, making it less likely that those previous results were due to the optimization algorithm favoring one model over the other due to random chance.

Second, we performed several simulations to determine whether the response probabilities predicted by the model after a change in a single mean could be reproduced by an arbitrary change in the corresponding variances. If means and variances are non-identifiable in GRT-wIND, then when one mean changes we should always be able to find a change in variances that leads to an equivalent pattern of behavioral predictions. The results of our simulations, presented in Appendix, clearly indicate that this is not the case.

In sum, although a full study of the identifiability of means and variances in GRT-wIND is beyond the scope of this article, the available evidence from both model fits to the observed data and numerical simulations clearly suggests that these parameters are identifiable.

3.2.5. Is learned separability homogeneous across participants?

An important assumption of GRT-wIND is that perceptual representations for a set of stimuli are similar across people and differences in performance are the product of differences in individual attentional and decisional strategies. This seems like a reasonable assumption when dealing with stimulus dimensions whose separability is not malleable, because a specific level of separability has been either innately determined or learned by most adults through considerable experience with the dimensions. However, this assumption might be violated by

dimensions that are relatively novel. The category-relevant dimension might or might not be perceived similarly across participants in the Experimental group, depending on whether the limited experience with this dimension fostered similar learning across participants. Differences in learning rates could have easily led to violations of the assumptions of GRT-wIND. This issue is less critical when dimensions are completely novel, as in our Control group, where all participants are equivalently inexperienced with the dimensions.

Thus, to apply GRT-wIND in the present experiment, one must assume that the effect of categorization training was similar across participants in the Experimental group. An examination of individual performance during the last session with the categorization task reveals that this was not necessarily the case: accuracy for the participants included in the present experiment ranged between 0.60 and 0.93, meaning that they learned the task to varying degrees of proficiency.

One way to more directly determine whether the assumptions of GRT-wIND are appropriate is to analyze the fit of the model to the data from each participant. As we have argued before (Soto et al., 2015), if the assumption of common perceptual representations in GRT-wIND is violated, then the perceptual distributions estimated from data should offer a good fit to the data from some participants, but offer a poor fit to the data from other participants. This should lead to distributions of fit values that are not unimodal.

We computed the percentage of the variance in each participant's data explained by GRT-wIND as a measure of individual fit. All participants in the Experimental group showed high values of model-fit (0.92 and above), with a median of 0.97. The distribution appeared unimodal, which was confirmed by a non-significant dip test of unimodality (Hartigan & Hartigan, 1985), $D = 0.07$, $p > 0.5$.

It could be argued that differences among participants in perceptual separability of the category-relevant dimension could have been captured by GRT-wIND as differences in decisional separability (Silbert & Thomas, 2013). As indicated earlier (Section 3.2.4), this is unlikely to be an issue in GRT-wIND. Furthermore, if this was the case, then the decision bounds for the category-relevant dimension should have varied widely across participants. As seen in Fig. 7b, this was not the case: decision bounds do vary, but they are similar and close to vertical (zero degrees of rotation). In fact, a comparison of Figs. 7b and 8b reveals that bounds seemed more homogeneous for the category-relevant dimension than for a completely novel dimension. There were no significant deviations from unimodality in any of the distributions of bound parameters (rotation and location of the bounds for each dimension) in the Experimental group.

In sum, the results show no evidence that the effect of categorization training was dissimilar across participants in the Experimental group. If anything, the results seem to suggest that categorization training actually homogenized perceptual and decisional processes in the Experimental group, compared to completely novel dimensions.

The median model-fit value in the Control group was high (0.96) and comparable to that of the Experimental

group, with between-group differences in fit being non-significant according to a Kruskal–Wallis test. One participant in the Control group showed extremely poor model-fit (0.17), but a close examination of the data showed that this participant had a performance level close to chance, which was not detected prior to analysis by our exclusion criteria. Thus, the difference in fit cannot be attributed to a difference in the way this participant perceived the stimuli. The dip test revealed no violations of unimodality in the model-fit measures from the Control group, $D = 0.06$, $p > 0.5$.

The parameters from decision bounds (Figs. 7a and 8a) also seemed more heterogeneous in the Control group than in the Experimental group. However, as indicated earlier, there is little reason to expect that people would differ in their perception of completely novel dimensions. It seems more likely that the results are due to heterogeneous decision strategies, just as modeled by GRT-wIND. Importantly, even if people in the control group did perceive stimuli heterogeneously, our conclusions for the present experiment would not change. Note that there are many ways in which perceptual representations can violate perceptual separability (heterogeneity), but only one way in which they can be separable. If dimensions are perceptually separable across participants, GRTw-IND should be able to capture this pattern, as it does for the category-relevant dimension in the Experimental group. In the GRT framework, heterogeneity of perceptual distributions across participants can only mean that violations of perceptual separability and perceptual independence are present in the data of at least some participants.

4. General discussion

The experiments reported here found evidence that training in a categorization task increases the separability of the category-relevant dimension according to a variety of tests from multidimensional signal detection theory. In Experiment 1, participants who received pre-training in a categorization task showed reduced Garner interference effects and reduced violations of marginal response and RT invariance, compared to participants who did not receive such pre-training. Both of these tests are theoretically related to violations of perceptual separability when decisional separability is assumed to hold (Ashby & Maddox, 1994; Ashby & Townsend, 1986). In Experiment 2, participants who received pre-training in a categorization task showed reduced violations of perceptual separability according to a model-based analysis of data using GRT-wIND (Soto et al., 2015), the best currently available framework for the analysis of dimensional interactions using GRT. The use of an identification task and GRT-wIND allowed us to dissociate perceptual from decisional dimensional interactions in Experiment 2, leading to the finding that categorization experience increases both perceptual and decisional separabilities. We must note, however, that the effect of categorization on decisional separability was only marginally significant, and more research is required to evaluate the reliability of this result.

On the other hand, the data suggest that while the categorization training we provided seemed to increase

perceptual separability, it did not necessarily produce complete separability. Experiment 1 showed evidence of violations of perceptual separability in participants that had received categorization training, according to both the marginal response invariance and marginal RT invariance analyses. Such violations could also be observed in Experiment 2 (Fig. 6), but they were not statistically significant. One possibility is that these violations of perceptual separability might disappear with even more categorization training. Another possibility is that the violations are a consequence of testing separability of the category-relevant dimension when presented together with a completely novel dimension. It is possible that separability learning shows some level of dimensional specificity, being stronger when tested with the originally trained category-irrelevant dimension. If this is correct, then an important goal for future research is to determine under what learning conditions such specificity can be completely overcome.

As indicated in the introduction, a number of important problems in perception science can be seen as special cases of the question of whether or not two or more stimulus dimensions are perceptually separable. Notably, this includes any research testing the "independent" or "holistic" nature of perceptual processing and representations (e.g., Arguin & Saumier, 2000, 2004; Blais, Arguin, & Marleau, 2009; Fitousi & Wenger, 2013; Ganel & Goshen-Gottstein, 2004; Mack et al., 2011; Mestry et al., 2012; Richler et al., 2008; Schweinberger & Soukup, 1998; Soto & Wasserman, 2011; Soto et al., 2015; Stankiewicz, 2002; Thomas, 2001). In most research directed to answer such questions, experimenters have assumed that separability is a fixed characteristic of stimulus dimensions, in some cases determined by processing through independent brain pathways and/or representations (e.g., Andrews & Ewbank, 2004; Bruce & Young, 1986; Haxby et al., 2000; Kayaert, Biederman, & Vogels, 2005; Vogels et al., 2001; Winston, Henson, Fine-Goulden, & Dolan, 2004). If the conclusions reached by our study are correct and generalizable, then there would be three important consequences for such lines of research.

First, researchers should abandon the assumption that separability is an either/or feature of stimulus dimensions that must be hardwired in the brain. For example, some shape properties might have a special status in object recognition, being processed "independently" (Arguin & Saumier, 2000, 2004; Kayaert et al., 2005; Stankiewicz, 2002; Vogels et al., 2001), simply because they are diagnostic in object categorization tasks usually encountered in the environment, instead of being innately determined (Biederman, 2001). Whether dimensions of face stimuli are separable (Fitousi & Wenger, 2013; Schweinberger & Soukup, 1998; Soto et al., 2015) might not be the result of different brain representations that are hardwired in the brain (Andrews & Ewbank, 2004; Bruce & Young, 1986; Haxby et al., 2000; Winston et al., 2004). Instead, both behavioral and neurobiological markers of separability might be better conceptualized as the result of learning through experience in object categorization.

Second, it is important to keep in mind that conclusions about separability might depend on the methods used for

testing separability. For example, the Garner filtering task, which is one of the most widely used tests of separability or "independence" in perceptual science, is essentially a categorization task. Given the results reported here, this means that training in the task by itself might increase perceptual separability of the dimensions being tested. We currently do not know how training in any other task might affect perceptual separability. For this reason, it seems wise that any research aimed at testing notions related to separability should include controls aimed at showing that the procedures used are capable to adequately detect violations of perceptual separability.

Third, when interpreting results suggesting separability or independence of perceptual dimensions, researchers should take into account the possibility that prior category learning might have caused that separability. Given the finding that a particular dimension is processed separately or independently, the question that a researcher should ask him/herself is: Is this a dimension that people often use to divide natural objects into classes?

An important aspect of our study is that we focused on GRT analyses of dimensional interactions. One advantage of using GRT is that it dissociates perceptual from decisional factors in the study of separability and independence. The results from Experiment 2 suggest that categorization training changes the separability of perceptual representations, in addition to any changes it might also induce in decisional or attentional factors. This is in line with the results of previous studies suggesting that category learning may be accompanied by changes in perceptual representations rather than only in decisional processes (Folstein, Palmeri, & Gauthier, 2013; Goldstone, Lippa, & Shiffrin, 2001; Notman, Sowden, & Özgen, 2005).

On a more technical front, the present work also contributed to the development of new tests within the GRT framework to study differences between groups in violations of perceptual separability, perceptual independence and decisional separability (see Appendix). One reason why such tests are important is that it has been recognized for a long time that the separability/integrality of stimulus dimensions is a matter of degree rather than kind (e.g., Shepard, 1991). Our results are in line with this idea: learning can increase the separability of a dimension, even though violations of separability can still be found in the data. Given this, in many cases determining whether there are differences between two conditions in dimensional interactions might be more interesting than determining whether two dimensions do or do not perceptually interact.

How do the present results relate to the hypothesis (Schyns et al., 1998) that category learning creates new functional features? Separability learning as studied here is related to, but not the same as the creation of new functional features. An important aspect of the new features proposed by Schyns and colleagues is that they increase the representational capacity of the visual system. This is not necessarily true of separability learning, which involves a re-description of objects in terms of dimensions that might not enrich representational capacity, but might instead facilitate the use of high-level cognitive mechanisms such as selective attention and rule learning. On the other hand, one defining characteristic of new

functional features is that they are perceived independently from their components (Schyns & Rodet, 1997; Schyns et al., 1998). That is, a feature is a unit of configural or holistic perception, which is perceived directly and not analyzed into components. As indicated earlier, the difference between holistic or configural processing versus analytic processing can be cast in terms of integrality/separability of component features. Thus, at least some aspects of the problem of new feature learning, as conceived by Schyns and colleagues, are captured by the concept of separability learning.

Importantly, one problem with the controversy over whether the visual system represents objects through fixed versus flexible features is that there is no formal definition of a feature (see Schyns et al., 1998, and comments). In the object recognition literature, a feature is essentially any physical stimulus dimension that can be varied by the experimenter in a task, or any form of information that can be extracted from an image by a computational model. We propose that focusing on separable features or dimensions is both more tractable and more theoretically relevant that focusing on features in general. It is more tractable because there are clear formal definitions of perceptual separability, accompanied by operational tests to detect it. Differences in separability can be measured, which is the first step of a sound science. Focusing on separable dimensions is more theoretically relevant because, as discussed in the introduction, there are several ways in which separable dimensions are "special" compared to dimensions that do not show separability. Importantly, it seems like a number of cognitive operations (e.g., selective attention, rule learning, categorization, generalization, working memory) depend on whether or not dimensions are separable. It is such operations and their effect on behavioral choices that ultimately determine whether a set of dimensions is "special" in terms of adaptive behavior.

Author notes

Appendix

GRT-wIND

In the 2×2 identification design used in Experiment 2, stimuli vary along two dimensions A and B, each with two levels indexed by $i = 1, 2$, and $j = 1, 2$, respectively. Suppose there are N participants indexed by $k = 1, 2, \ldots N$. The GRT-wIND model for this design has four perceptual distributions, which are assumed to be bivariate normal and

common to all participants. Each distribution is described by a mean vector:

$$\underline{\mu}_{A_iB_j} = \begin{bmatrix} \mu_{A_iB_j1} \\ \mu_{A_iB_j2} \end{bmatrix}, \tag{A1}$$

and a covariance matrix:

$$\boldsymbol{\Sigma}_{A_iB_j} = \begin{bmatrix} \sigma^2_{A_iB_j1} & \rho_{A_iB_j}\sigma_{A_iB_j1}\sigma_{A_iB_j2} \\ \rho_{A_iB_j}\sigma_{A_iB_j1}\sigma_{A_iB_j2} & \sigma^2_{A_iB_j2} \end{bmatrix}, \tag{A2}$$

where σ and ρ represent standard deviations and correlations, respectively. We set $\mu_{A_1B_1} = [0,0]$ and $\sigma_{A_1B_1} = \sigma_{A_1B_2} = 1$, to fix the position and scale of the final solution.

The model also has several parameters describing processes unique to each individual. The parameters κ_k and λ_k represent global and selective attention to dimensions by participant k, respectively. The covariance matrix for the distribution of perceptual effects of A_iB_j in participant k is equal to:

$$\boldsymbol{\Sigma}_{A_iB_jk} = \begin{bmatrix} \dfrac{\sigma^2_{A_iB_j1}}{\kappa_k\lambda_k} & \rho_{A_iB_j}\dfrac{\sigma_{A_iB_j1}\sigma_{A_iB_j2}}{\sqrt{\kappa_k^2\lambda_k(1-\lambda_k)}} \\ \rho_{A_iB_j}\dfrac{\sigma_{A_iB_j1}\sigma_{A_iB_j2}}{\sqrt{\kappa_k^2\lambda_k(1-\lambda_k)}} & \dfrac{\sigma^2_{A_iB_j2}}{\kappa_k(1-\lambda_k)} \end{bmatrix}. \tag{A3}$$

Note that high values of κ_k decrease the values of all variances, leading to fewer confusion errors in general. A value of $\lambda_k = 0.5$ represents equal attention to each dimension. High values of λ_k decrease the variances on dimension A and increase the variances on dimension B, representing selective attention to A. The opposite is true for low values of λ_k.

The other four individual parameters describe the participant's linear decision bounds. Each single bound can be written as a discriminant function:

$$h_{Ak}(x_1,x_2) = b_{Ak1}x_1 + b_{Ak2}x_2 + c_{Ak}, \tag{A4}$$

where h_{Ak} represents the discriminant function used to classify component A by the kth participant. A similar equation can be used to describe h_{Bk}, the discriminant function used to classify component B by the kth participant. The parameters b_{Ak1} and b_{Bk2} were fixed to a value of 1.0.

Maximum likelihood estimation of parameters

The data from each participant in an identification experiment are summarized in a confusion matrix, with different rows for each stimulus, different columns for each response, and response frequencies reported in each cell of the matrix. Let S_i and R_j denote stimuli and responses in the task, respectively, with i and j ranging from 1 to 4. There are N participants in the experiment, indexed by $k = 1,2,\ldots,N$. Let r_{kij} denote the frequency with which participant k responded R_j on trials when stimulus S_i was presented and $P_k(R_j|S_i)$ the corresponding response probability.

Given a set of parameter values, the likelihood of the data is computed in two steps. First, the predicted confusion matrix of each participant is obtained from the model, where each $P_k(R_j|S_i)$ is computed by integrating the volume

of the S_i perceptual distribution in response region R_j (see Ashby & Soto, 2015). Second, the log of the likelihood function for each participant is computed and summed across all participants:

$$\log L = \sum_{k=1}^{N}\sum_{i=1}^{4}\sum_{j=1}^{4} r_{kij}\log P_k(R_j|S_i). \tag{A5}$$

The maximum likelihood estimates are those that maximize Eq. (A5).

Tests of dimensional interactions

Dimensional interactions can be tested by comparing maximum-likelihood parameter estimates against expected values from null hypotheses using a Wald test (Soto et al., 2015). Let $\hat{\underline{\theta}}$ be a column vector containing the maximum likelihood parameter estimates. The Wald test can be used to test any null hypothesis that can be expressed in the form of linear restrictions on $\hat{\underline{\theta}}$:

$$\begin{aligned} H_0 &: \mathbf{R}\hat{\underline{\theta}} - \underline{q} = 0 \\ H_1 &: \mathbf{R}\hat{\underline{\theta}} - \underline{q} \neq 0, \end{aligned} \tag{A6}$$

where $\mathbf{R}$ is a matrix with number of columns equal to the number of parameters and number of rows equal to the number of restrictions being tested, and $\underline{q}$ is a column vector with number of rows equal to the number of restrictions being tested. For example, if we wanted to test the hypothesis that $\hat{\theta}_1 = 0$, then $\mathbf{R}$ would have a single row (we are testing a single restriction) with a +1 in the first cell of that row and zeros in all other cells, while $\underline{q}$ would have a single cell with a zero in it. If we want to additionally test the hypothesis that $\hat{\theta}_2 - \hat{\theta}_3 = 10$, then we would add a second row to $\mathbf{R}$ with a +1 in the second column (corresponding to $+\hat{\theta}_2$) and -1 in the third column (corresponding to $-\hat{\theta}_3$), while $\underline{q}$ would now have a second cell with the value 10 in it.

The Wald statistic:

$$W = \left[\mathbf{R}\hat{\underline{\theta}} - \underline{q}\right]^{\mathrm{T}}\left[\mathbf{R}\boldsymbol{\Sigma}_{\hat{\theta}}\mathbf{R}^{\mathrm{T}}\right]^{-1}[\mathbf{R}\hat{\underline{\theta}} - \underline{q}], \tag{A7}$$

where $[]^{\mathrm{T}}$ represents matrix transpose, has a chi-squared distribution with degrees of freedom equal to the number of restrictions being tested (the length of $\underline{q}$). The covariance matrix of the maximum likelihood estimates can be estimated using the inverse of the Hessian of the log-likelihood function at the solution.

The restrictions imposed on the model by perceptual separability of dimension A from dimension B are the following:

$$\begin{aligned} \mu_{A_1B_21} &= 0 \\ \sigma_{A_1B_21} &= 1 \\ \mu_{A_2B_11} - \mu_{A_2B_21} &= 0 \\ \sigma_{A_2B_11} - \sigma_{A_2B_21} &= 0 \end{aligned} \tag{A8}$$

The restrictions imposed in the model by perceptual separability of dimension B from dimension A are the following:

$$\mu_{A_2B_12} = 0$$
$$\sigma_{A_2B_12} = 1$$
$$\mu_{A_1B_22} - \mu_{A_2B_22} = 0 \qquad \text{(A9)}$$
$$\sigma_{A_1B_22} - \sigma_{A_2B_22} = 0$$

The restrictions imposed in the model by perceptual independence in each of the perceptual distributions are the following:

$$\rho_{A_1B_1} = 0$$
$$\rho_{A_1B_2} = 0$$
$$\rho_{A_2B_1} = 0 \qquad \text{(A10)}$$
$$\rho_{A_2B_2} = 0$$

The Wald test allows tests of decisional separability for the group average or for each participant individually. Here, we test decisional separability on average to keep consistence with between-group comparisons, which can only be performed on average decisional separability (see below). The single restriction imposed in the model by average decisional separability of dimension A from dimension B can be summarized as:

$$\sum_{k=1}^{N} b_{A_{k2}} = 0 \qquad \text{(A11)}$$

The restriction imposed on the model by average decisional separability of dimension B from dimension A is the following:

$$\sum_{k=1}^{N} b_{B_{k1}} = 0 \qquad \text{(A12)}$$

Tests of differences between groups in the magnitude of dimensional interactions

In many applications of GRT, such as Experiment 2 of the present study, it is desirable to statistically compare two sets of parameter estimates, $\hat{\underline{\theta}}_1$ and $\hat{\underline{\theta}}_2$, to evaluate the reliability of observed differences in the magnitude of violations of separability or independence across experimental conditions. Here we derive a test for the case in which the two sets of parameter estimates have been obtained from independent samples.

First note that if $\hat{\underline{\theta}}_i$ is a maximum likelihood estimate, then one of its asymptotic properties is that it is approximately normally distributed:

$$\hat{\underline{\theta}}_i \overset{a}{\sim} Normal(\underline{\theta}_i, \boldsymbol{\Sigma}_i) \qquad \text{(A13)}$$

From our discussion of the Wald test, we know that many hypotheses about these parameter estimates can be expressed as linear restrictions on $\hat{\underline{\theta}}_i$. The following vector:

$$\hat{\underline{\omega}} = \mathbf{R}_i \hat{\underline{\theta}} - \underline{q} \qquad \text{(A14)}$$

measures deviations from each one of the restrictions that have been encoded in the matrix $\mathbf{R}_i$ and the vector $\underline{q}$. The previous section described what restrictions correspond to the hypotheses of perceptual separability, decisional separability and perceptual independence. Because a linear transformation of a normally distributed vector is also normally distributed, we have that

$$\hat{\underline{\omega}}_i \overset{a}{\sim} Normal\left(\mathbf{R}\underline{\theta}_i - \underline{q}, \mathbf{R}_i\boldsymbol{\Sigma}_i\mathbf{R}_i^{\mathrm{T}}\right) \qquad \text{(A15)}$$

We are interested in testing the following hypotheses:

$$H_0 : \underline{\omega}_1 - \underline{\omega}_2 = 0$$
$$H_1 : \underline{\omega}_1 - \underline{\omega}_2 \neq 0 \qquad \text{(A16)}$$

To obtain a statistic to test these hypotheses, we note that because both $\hat{\underline{\omega}}_1$ and $\hat{\underline{\omega}}_2$ are multivariate normal (Eq. (A15)), their difference is also multivariate normal and the statistic:

$$D = [\hat{\underline{\omega}}_1 - \hat{\underline{\omega}}_2]^{\mathrm{T}} \{\boldsymbol{\Sigma}_{\underline{\omega}_1 - \underline{\omega}_2}\}^{-1} [\hat{\underline{\omega}}_1 - \hat{\underline{\omega}}_2] \qquad \text{(A17)}$$

where

$$\boldsymbol{\Sigma}_{\underline{\omega}_1 - \underline{\omega}_2} = \mathbf{R}_1\boldsymbol{\Sigma}_1\mathbf{R}_1^{\mathrm{T}} + \mathbf{R}_2\boldsymbol{\Sigma}_2\mathbf{R}_2^{\mathrm{T}} \qquad \text{(A18)}$$

follows a Chi-square distribution under H_0, with degrees of freedom equal to the number of rows in the matrices $\mathbf{R}_i$, so it can be used as our test statistic. As for the Wald test, we can obtain estimates of the covariance matrices in Eq. (A18) by taking the inverse of the Hessian of the log-likelihood function at the solution. Because of the use of an estimate of the covariance matrix, the distribution of the test only approximates a Chi-square distribution. The inverse of the Hessian is a consistent estimator of the covariance matrix, so the quality of this approximation will increase as sample size increases.

Note that because slope parameters reflect individual bounds, they are not individually comparable across groups through D. The only valid between-groups comparison of decisional separability is in terms of group averages (Eqs. (A11) and (A12)). A test of individual decisional separability (used in Soto et al., 2015) does not have a corresponding between-groups test and therefore was not used here.

Identifiability of means and variances in GRT-wIND

Here we briefly report the results of simulations aimed at determining whether means and variances are both uniquely identifiable in GRT-wIND. Each of our simulations consisted of 150 repetitions of the following procedure. Starting with the best-fitting model found for the experimental group of Experiment 2, one parameter from the model was randomly selected and changed by a random value. The response probabilities from this "target model" were computed. Then, we used an optimization algorithm to find the value of a set of free parameters in a "test model" leading to predictions as similar as possible to those of the target model. The dissimilarity of predictions was measured through their sum of squared differences (SSD).

If means and variances are non-identifiable in GRT-wIND, then it should always be possible to change a mean using this procedure and find a set of corresponding variances leading to an SSD of zero. In practice, of course, the optimization algorithm might not always be able to find

the set of variances leading to the lowest SSD. To determine an upper limit of precision for our optimization algorithm, we performed several benchmark simulations in which the randomly changed parameter in the target model was a single variance (randomly changed from 0 to 1, 0 to 2, or 0 to 4 in different simulations), and the free parameters in the test model were the two variances from the same distribution. The minimal SSD found in these simulations ranged between 0 and 0.019, with a median of 2.92×10^{-6}. The maximum value of 0.019 was used as benchmark in the following analyses.

We then performed a simulation in which the randomly changed parameter in the target model was a single mean (randomly changed from 0 to 1), and the free parameters in the test model were the two variances from the same distribution. In this case, the SSD ranged between 0 and 2.302, with a median of 0.054, which was significantly higher than the maximum SSD found in our benchmark simulations, according to a Wilcoxon signed rank test ($p < 0.0001$). The only SSD equal to zero in this simulation was due to equivalent predictions in the target and test models *before* the optimization was run, which resulted from a too small random change in the mean parameter of the target model. In sum, this simulation suggests that the group means and variances are identifiable in GRT-wIND. These results are unlikely to be due to the optimization algorithm being stuck in local minima because plotting SSD as a function of the values of the variances revealed a smooth concave function.

To determine whether changes in individual variances (which are determined by the group variances and individual attention parameters) could produce the same confusion matrices as changes in a group mean, we repeated the previous simulations, but this time the free parameters in the test model were the two variances of the modified distribution plus all individual attention parameters. The minimal SSDs from this simulation ranged from 1.36×10^{-6} to 2.58, with a median of 0.049, which was significantly higher than the maximum SSD found in our benchmark simulations, according to a Wilcoxon signed rank test ($p < 0.0001$). In other words, these simulations suggest that the means and individual variances are identifiable in GRT-wIND.

Thus, all of our analyses suggest that, unlike standard signal detection theory, means and variances are uniquely identifiable in GRT-wIND. We believe this is because a change in a mean in the model affects response probabilities in *a single row of all confusion matrices*. Because each participant has different decision bounds and attention parameters, which scale variances differently, a change in a mean has a different impact in different participants, and to obtain an equivalent model by changing variances, we would have to change each individual set of variances in a specific way. This is why a change in the group variances cannot achieve the same result as a change in the means. The reason why individual variances and means are also uniquely identifiable is that individual variances depend on attention parameters that affect all perceptual distributions. Changes in individual attention parameters affect response probabilities in *all rows of a single confusion matrix*, so under no circumstances can a change in

attention parameters and global variances lead to the same predictions as a change in a single mean, which affects response probabilities in *a single row of all confusion matrices.* In other words, changes in individual attention parameters affect not only those cells in the confusion matrix affected by changes in a mean, but also several other cells.

References

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage, 23*(3), 905–913.

Arguin, M., & Saumier, D. (2000). Conjunction and linear non-separability effects in visual shape encoding. *Vision Research, 40*(22), 3099–3115.

Arguin, M., & Saumier, D. (2004). Independent processing of parts and of their spatial organization in complex visual objects. *Psychological Science, 15*(9), 629–633.

Ashby, F. G. (1988). Estimating the parameters of multidimensional signal detection theory from simultaneous ratings on separate stimulus components. *Perception & Psychophysics, 44*(3), 195–204.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General, 120*(2), 150–172.

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

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Attention, Perception, & Psychophysics, 61*(6), 1178–1199.

Ashby, F. G., & Soto, F. A. (2015). Multidimensional signal detection theory. In J. R. Busemeyer, J. T. Townsend, Z. J. Wang, & A. Eidels (Eds.), *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.

Ashby, F. G., & Valentin, V. V. (2005). Multiple systems of perceptual category learning: Theory and cognitive tests. In H. Cohen & C. Lefebvre (Eds.), *Categorization in cognitive science* (pp. 548–572). New York: Elsevier.

Ashby, F. G., Waldron, E. M., Lee, W. W., & Berkman, A. (2001). Suboptimality in human categorization and identification. *Journal of Experimental Psychology: General, 130*(1), 77.

Bae, G. Y., & Flombaum, J. I. (2013). Two items remembered as precisely as one: How integral features can improve visual working memory. *Psychological Science, 24*, 2038–2047.

Biederman, I. (2001). Recognizing depth-rotated objects: A review of recent research and theory. *Spatial Vision, 13*(2–3), 241–253.

Blais, C., Arguin, M., & Marleau, I. (2009). Orientation invariance in visual shape perception. *Journal of Vision, 9*(2), 1–23.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305–327.

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.

Dunn, J. C. (1983). Spatial metrics of integral and separable dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 9*(2), 242–257.

Ell, S. W., Ashby, F. G., & Hutchinson, S. (2012). Unsupervised category learning with integral-dimension stimuli. *The Quarterly Journal of Experimental Psychology, 65*(8), 1537–1562.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is "special" about face perception? *Psychological Review, 105*(3), 482.

Fitousi, D., & Wenger, M. J. (2013). Variants of independence in the perception of facial identity and expression. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 133–155.

Foard, C. F., & Kemler-Nelson, D. G. (1984). Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General, 113*(1), 94–111.

Folstein, J. R., Gauthier, I., & Palmeri, T. J. (2012). Not all morph spaces stretch alike: How category learning affects object discrimination. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 802–807.

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex, 23*(4), 814–823.

Ganel, T., & Goshen-Gottstein, Y. (2004). Effects of familiarity on the perceptual integrality of the identity and expression of faces: The parallel-route hypothesis revisited. *Journal of Experimental Psychology: Human Perception and Performance, 30*(3), 583–596.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist, 25*, 350–358.

Garner, W. R. (1974). *The processing of information and structure*. New York: Lawrence Erlbaum Associates.

Goldstone, R. L. (1994a). An efficient method for obtaining similarity data. *Behavior Research Methods, 26*(4), 381–386.

Goldstone, R. L. (1994b). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General, 123*(2), 178–200.

Goldstone, R. L., Gerganov, A., Landy, D., & Roberts, M. E. (2009). Learning to see and conceive. In L. Tommasi, M. A. Peterson, & L. Nadel (Eds.), *Cognitive biology: Evolutionary and developmental perspectives on mind, brain, and behavior* (pp. 163). Cambridge, MA: MIT Press.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(1), 69–78.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. *Cognition, 78*(1), 27–43.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General, 130*(1), 116.

Grau, J. W., & Nelson, D. K. (1988). The distinction between integral and separable dimensions: Evidence for the integrality of pitch and loudness. *Journal of Experimental Psychology: General, 117*(4), 347–370.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science, 13*(4), 622–635.

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

Handel, S., Imai, S., & Spottswood, P. (1980). Dimensional, similarity, and configural classification of integral and separable stimuli. *Perception & Psychophysics, 28*, 205–212.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics, 13*(1), 70–84.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223–232.

Hockema, S. A., Blair, M. R., & Goldstone, R. L. (2005). Differentiation for novel dimensions. In B. G. Bara, L. W. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the twenty-seventh annual conference of the cognitive science society* (pp. 953–958). Hillsdale, NJ: Erlbaum.

Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology, 36*(3), 325–374.

Kadlec, H., & Townsend, J. T. (1992b). Signal detection analysis of multidimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–231). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*, 759–763.

Kayaert, G., Biederman, I., & Vogels, R. (2005). Representation of regular and irregular shapes in macaque inferotemporal cortex. *Cerebral Cortex, 15*(9), 1308–1321.

Kravitz, D. J., Vinson, L. D., & Baker, C. I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences, 12*(3), 114–122.

Mack, M. L., Richler, J. J., Gauthier, I., & Palmeri, T. J. (2011). Indecision on decisional separability. *Psychonomic Bulletin & Review, 18*(1), 1–9.

Maddox, W. T., & Ashby, F. G. (1996). Perceptual separability, decisional separability, and the identification–speeded classification relationship. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 795–817.

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*(6), 255–260.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*(2), 242–279.

Melara, R. D., Marks, L. E., & Lesko, K. E. (1992). Optional processes in similarity judgments. *Perception & Psychophysics, 51*(2), 123–133.

Melara, R. D., Marks, L. E., & Potts, B. C. (1993). Primacy of dimensions in color perception. *Journal of Experimental Psychology: Human Perception and Performance, 19*(5), 1082–1104.

Mestry, N., Wenger, M. J., & Donnelly, N. (2012). Identifying sources of configurality in three face processing tasks. *Frontiers in Perception Science, 3*, 456.

Notman, L. A., Sowden, P. T., & Özgen, E. (2005). The nature of learned categorical perception effects: A psychophysical approach. *Cognition, 95*(2), B1–B14.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087–11092.

Op de Beeck, H. P., Haushofer, J., & Kanwisher, N. G. (2008). Interpreting fMRI data: Maps, modules and dimensions. *Nature Reviews Neuroscience, 9*(2), 123–135.

Op de Beeck, H. P., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: Dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General, 132*(4), 491–511.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology, 58*, 75.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(2), 328–342.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Perception Science, 3*, 553.

Schweinberger, S. R., & Soukup, G. R. (1998). Asymmetric relationships among perceptions of facial identity, emotion, and facial speech. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1748–1765.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences, 21*(01), 1–17.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(3), 681–696.

Shepard, R. N. (1991). Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In J. Pomerantz & G. Lockhead (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 53–71). Washington, DC: American Psychological Association.

Silbert, N. H., & Thomas, R. (2013). Decisional separability, model identification, and statistical inference in the general recognition theory framework. *Psychonomic Bulletin & Review, 20*(1), 1–20.

Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J., & Ashby, F. G. (2010). Implicit and explicit category learning by macaques (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Behavior Processes, 36*(1), 54–65.

Smith, L. B., & Kemler, D. G. (1978). Levels of experienced dimensionality in children and adults. *Cognitive Psychology, 10*(4), 502–532.

Soto, F. A., Gershman, S. J., & Niv, Y. (2014). Explaining compound generalization in associative and causal learning through rational principles of dimensional generalization. *Psychological Review, 121*(3), 526–558.

Soto, F. A., Musgrave, R., Vucovich, L., & Ashby, F. G. (2015). General recognition theory with individual differences: A new method for examining perceptual and decisional interactions with an application to face perception. *Psychonomic Bulletin & Review, 22*(1), 88–111.

Soto, F. A., & Wasserman, E. A. (2011). Asymmetrical interactions in the perception of face identity and emotional expression are not unique to the primate visual system. *Journal of Vision, 11*(3).

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance, 28*(4), 913–932.

Thomas, R. (2001). Perceptual interactions of facial dimensions in speeded classification and identification. *Attention, Perception & Psychophysics, 63*(4), 625–650.

Townsend, J. T., Houpt, J. W., & Silbert, N. H. (2012). General recognition theory extended to include response times: Predictions for a class of parallel systems. *Journal of Mathematical Psychology, 56*(6), 476–494.

Ungerleider, L. G., & Haxby, J. V. (1994). "What" and "where" in the human brain. *Current Opinion in Neurobiology, 4*(2), 157–165.

Van Zandt, T. (2002). Analysis of response time distributions. In J. T. Wixted (Ed.), *Steven's handbook of experimental psychology: Methodology in experimental psychology* (pp. 461–516). San Diego, CA: Academic Press.

Vogels, R., Biederman, I., Bar, M., & Lorincz, A. (2001). Inferior temporal neurons show greater sensitivity to nonaccidental than to metric shape differences. *Journal of Cognitive Neuroscience, 13*(4), 444–453.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196.

Wickens, T. D. (1992). Maximum-likelihood estimation of a multivariate Gaussian rating model with excluded data. *Journal of Mathematical Psychology, 36*(2), 213–234.

Winston, J. S., Henson, R. N. A., Fine-Goulden, M. R., & Dolan, R. J. (2004). FMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology, 92*(3), 1830–1839.