THE MANY FACES OF GARNER INTERFERENCE

Devin Michael Burns

Submitted to the faculty of the University Graduate School

in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Psychological and Brain Sciences

and the Cognitive Science Program

Indiana University

July, 2014

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Doctoral Committee

James T. Townsend

Robert Nosofsky

Jason Gold

Thomas Busey

May 28, 2014

ACKNOWLEDGMENTS

I would like to thank my wife Lara for being trapped in a tiny house with our two little girls (Catalina and Alma) during an especially cold winter while I completed this work. I would also like to thank my entire committee, the other members of the Townsend lab, Danny Algom, Rich Shiffrin, Geoff Bingham, and Ben Motz for their support.

Devin Michael Burns

THE MANY FACES OF GARNER INTERFERENCE

A series of speeded classification tasks proposed by Garner (1974) has become a well-entrenched method for identifying interactions between perceptual dimensions. The theory proposes that *integral* dimensions should produce a redundancy gain when a second dimension covaries perfectly with the attended dimension, and interference if the second dimension varies irrelevantly. This work questions the interpretation of such results as indicating interactive dimensions, reviewing independent models which naturally exhibit such effects. Furthermore, there are several methodological confounds which make the cause of Garner interference non-identifiable in the standard experimental context, the most serious of which is the conflation of changes in the number of stimuli with changes in the number of irrelevant dimensions. Here is proposed a novel three-dimensional extension of the Garner paradigm capable of disambiguating these experimental factors, which includes several conditions designed to help distinguish between various competing models of the related phenomena. This new paradigm was implemented with two stimulus sets, both composed of known integral dimensions, but from opposite sides of the complexity spectrum: color patches differing in their saturation, brightness, and hue; and faces differing in weight, age, and gender. Results show typical Garner interference effects for both stimulus sets, although the redundancy gains were rather modest. When a three-dimensional analog of the Garner filtering test is created by allowing a second irrelevant dimension to vary, however, the expected interference effects do not appear. Counter-intuitively, this additional variation often leads to an *improvement* in performance, an effect which cannot be predicted by the extant models. This effect is shown to be driven primarily by the extra dimension of variation rather than the additional stimuli. The implications for these (and other) findings are considered with regards to the utility of the Garner paradigm and the models that have attempted to describe it.

CONTENTS

CHAPTER 1

INTRODUCTION

Perception, broadly defined, is a form of measurement concerning objects or events in the world. It is tempting, therefore, to assume that when we perceive something we are measuring its qualities in roughly the same way as a physical measurement device would, though perhaps with less accuracy. This assumption is especially appealing given the popularity of the "brain is a computer" metaphor. The problem is that our perception of a given dimension is frequently affected by variations in other, presumably irrelevant dimensions. It could be said that our perceptual faculties are indeed measuring devices, but the fallacy lies in making assumptions about what exactly they are measuring.

A telling example of how our intuitions can lead us astray concerns the perception of weight. Almost anyone you ask would agree that while they are not as sensitive as a scale, they are capable of estimating the weight of an object in much the same way. This turns out to not be the case, however, as has be thoroughly documented in what is termed the "size-weight illusion." When asked to compare the weight of two objects of different sizes, but which have identical mass, people will consistently respond that the larger (less dense) object is much lighter. The effect is so robust, that even knowing exactly what is going on does not dispel the sensation that the larger object is lighter.

This illusion is only an indication of faulty perception under the premise that people are indeed measuring weight. Bingham, Schmidt, and Rosenblum (1989), however, noticed that the non-linear relation between size and weight that dictates this illusion is exactly replicated when participants are asked to select an object of variable mass and size that they expect to be able to throw the furthest. Participants are shockingly accurate at perceiving this complex dimension of "throwability," a variable that depends on a vast array of individually specific physiological variables such as muscle and tendon stiffness. Their hypothesis is that humans have evolved a "smart perceptual mechanism" capable of measuring a computationally complicated but extremely functional property (Bingham et al., 1989; Zhu & Bingham, 2011). The size-weight "illusion" is only an illusion if you make the error of assuming that weight is the basic property that we are measuring.

One way of determining how these perceptual measuring devices of ours work is to ask whether or not two physical dimensions, like size and weight, can be perceived independently of one another. There are many different ways to test for these interactions, but one of the most popular paradigms was established 40 years ago and still sees heavy use today. This method is the set of speeded categorization tasks established by Garner and Felfoldy (1970), described in more detail by Garner (1974, 1976). In the Garner paradigm, a series of comparisons between mean reaction times is used to reveal whether two dimensions are *separable* or *integral*.

1.1 THE DEFINITION OF INTEGRAL

The idea that different pairs of dimensions could combine in fundamentally different ways was first given empirical support by researchers using multidimensional scaling (MDS) theory. MDS attempts to account for an observer's reports of similarities between objects by representing the stimuli in a geometric space (see Torgerson, 1952, 1958). The similarity between any two stimuli is assumed to be a monotonic function of the distance between the two points that represent them. How this distance should be calculated, however, has been a matter of debate. The Euclidean metric, $\delta = \sqrt{\delta_x^2 + \delta_y^2}$, is most naturally imported from our geometric intuitions, and was used by the earliest developers of MDS (Young & Householder, 1938). This metric, however, requires the assumption that distances remain constant with a rotation of the axes, meaning that there are no "privileged" psychological dimensions.

Attneave (1950) argued that this was inappropriate for psychological judgments, and instead advocated for what has come to be known as the "city-block" metric: $\delta = \delta_x + \delta_y$. Under this metric, distances (and therefore similarities) are computed as a simple sum of the distances along each component dimension, meaning that the hypotenuse of a triangle is no longer shorter than the sum of its sides. In this framework distances do not remain constant under a rotation of the space, meaning that certain dimensions are the "correct" psychological dimensions. Attneave (1950) tested this hypothesis using parallelograms varying in terms of size, tilt, and color in one experiment, and squares differing with respect to area and reflectance in another. Both sets of data were fit much better by the city-block metric than the Euclidean metric, which he took to imply "...unique psychological reference-systems underlying the perception of similarity and difference between stimuli."

This characterization was soon challenged by other researchers, however. Torgerson (1958) presented data involving similarity judgments for Munsell color chips that differed in value (brightness) and chroma (saturation), and found that the data provided strong support for a Euclidean distance model. His explanation for the contrast with Attneave's results was that the two metrics might each be appropriate for different stimulus choices, and that "...Attneave's model might be appropriate in those situations where the different dimensions are obvious and compelling, whereas the Euclidean model might be appropriate otherwise."

Shepard (1964) furthered the theoretical distinction between these two classes of dimensions, using the term *analyzable* for those more similar to Attneave's example and *unitary* for those that more closely resemble Torgerson's data. Shepard suggested that the contrast could be seen in the way subjects describe the difference between two stimuli. For stimuli that differ along two analyzable dimensions, subjects will almost always describe differences in terms of those two dimensions, in accordance with the idea of "privileged" dimensions that use a city-block metric.

In contrast, the way in which two stimuli from unitary dimensions are said to differ will depend on the particular stimuli being compared. For example, if we were to vary the dimensions of hue and saturation, one color might be described as "warmer," "deeper," "pinker," or "more intense," than another. Further more, any two stimuli will be said to differ in only one way, even if they have different values of both of the experimentally manipulated dimensions. This is consistent with the use of the Euclidean metric, which favors no particular choice of dimensions. Shepard also suggests that there is likely to be a continuum between these two extremes, rather than a simple dichotomy, a conception consistent with the fact that both metrics can be represented as special cases of the more general Minkowski metric: $\delta = (\delta_x^n + \delta_y^n)^{1/n}$.

Shepard (1964) also used this distinction to explain differences in the generalizability of identification task data to a categorization task. In previous work, Shepard and Chang (1963) had found that when using eight stimuli that differed on unitary dimensions (saturation and brightness), pairwise confusions from an identification task were strongly predictive of errors in a categorization task. An earlier experiment using analytic dimensions (size, color, and shape), however, did not show the same pattern (Shepard, Hovland, & Jenkins, 1961). It was hypothesized that participants were able to use some kind of selective attention process when categorizing analytic dimensions that allowed them to ignore variation along dimensions irrelevant to the categorization task. This strategy meant that the identification data, where all dimensions were relevant, were no longer predictive of categorization performance. This type of data would only predict categorization performance if stimuli were always perceived in a unitary manner and compared in terms of all variable dimensions, regardless of their relevance to the categorization at hand.

This influence of attentional processes when perceiving analyzable stimuli was also invoked to explain some known failings of multidimensional scaling (Shepard, 1964). Since a given scaling solution is based on pair-wise similarity judgments, this solution should only be predictive of categorization performance when using unitary dimensions. When using analyzable dimensions, selective attention processes can distort the geometry of the space, with participants relying on different dimensions when making various pair-wise comparisons. This phenomenon can lead to a failure of the triangle inequality: $\delta(a, b) + \delta(b, c) \geq \delta(a, c)$.

To take a cold war era example from Tversky (1977): Jamaica is similar to Cuba via geographical location, and Cuba is similar to Russia in regards to political orientation, but Jamaica is not therefore constrained to be similar to Russia. The triangle inequality is a necessary axiom for the use of any metric distance function, not just Euclidean, and therefore casts doubt on the suitability of MDS and other geometric models of perception when using analytic dimensions. One possible way around this conclusion is to include the concept of attentional state in the multidimensional model. Shepard (1964) suggests that a separate multidimensional representation could be used for when attention is focused on each of the various dimensions. In this case, one might hope that the metric axioms are satisfied for any one representation, and Shepard presents data supportive of that belief.

Hyman and Well (1967) observed that the distinction between analyzable and non-analyzable dimensions had been almost entirely built upon circular reasoning. The difference between these classes of stimuli was asserted to be the metric distance function used to distinguish stimuli, but dimensions were only classified as analyzable or not according to the best fitting metric in an MDS scaling solution. Additionally, few if any studies had compared performance for each type of dimensions under identical methodology, and all had used distinct subjects.

To make stronger claims about these differences they used a uniform methodology to test both analyzable dimensions (tilt and size of parallelograms) and non-analyzable dimensions (value and chroma) within the same group of subjects. They analyzed the results by checking the goodness of fit for a Euclidean scaling model, and then analyzing the departures from this model in terms of the predictions of a city-block model. Results were as expected, with value and color being the only two dimensions that conformed well to the Euclidean model, strengthening the conclusion that differences between these stimulus sets were due to intrinsic properties of the stimuli themselves.

In a follow up paper, they tested whether the non-analyzability of color dimensions was dependent on the way in which the stimuli were presented (Hyman & Well, 1968). They designed a new stimulus set where each stimulus consisted of two color patches. The left patch varied in terms of chroma while always being held at a constant level of value, while the right showed the opposite pattern. These stimuli elicited behavior indicative of analyzable dimensions, being combined in the same way as the tilt and size of a parallelogram. This finding indicated that the primary factor influencing the difference in analyzability (as measured by metric distance function) is the perceptual "separateness" of the two dimensions. In other words, the reason that chroma and value are unitary dimensions, to use the term from Shepard (1964), is that they are inherently intermixed and perceived simultaneously by the observer. If steps are taken to help the observer separate them, then they can become analyzable.

Garner (1970) built upon this reasoning and hypothesized that a necessary property for unitary dimensions is their constant coexistence: if one is present the other is always also present. To use the only recognized example of integral dimensions at that time, a color cannot have a hue, brightness, or saturation without also having the other two. This is in contrast to another commonly used stimulus set at the time, defined by the size of a semi-circle and the angle of a radial line. These analyzable dimensions can obviously exist without one another. This property of coexistence is clearly not a sufficient property for integral dimensions, however, since parallelograms always must have an angle and a size, and yet these dimensions were some of the first defined as analyzable.

1.2 CONVERGENT OPERATIONS

In an attempt to get a better definition of what it means for dimensions to be unitary, Garner and Felfoldy (1970) observed that important differences between these two classes of dimensions had been observed in at least three distinct research areas. So far our discussion has been restricted primarily to multidimensional scaling theory and the appropriate choice of distance metric, but other researchers had instead been drawing distinctions using tests for redundancy gains.

Lockhead (1966) examined tasks in which absolute judgments between stimuli varying along a single dimension (e.g. line length) were difficult enough to produce errors. He then examined how the addition of redundant information aided in the discrimination of the stimuli and reduced these errors. Previous experiments showed conflicting results: sometimes this additional information led to a reduction in errors, but in other cases it did not. Eriksen and Hake (1955) varied color patches in terms of their brightness, hue, and size, and found that when these dimensions varied in a correlated manner, performance was improved in comparison with when only a single dimension varied. In contrast, Garner and Lee (1962) varied the visual positions of X's and O's, and found no improvement through the introduction of redundancy.

Lockhead (1966) argues that this difference is due to the fact that the color patch dimensions are *integral*. By this he means the same as what Shepard (1964) meant with the term *unitary*, and Lockhead uses *non-integral* in place of *analyzable*. He argues that the use of the word *analyzable* is misleading, since subjects are able to analyze (perceive) the different aspects of a Munsell color patch, which are "non-analyzable" dimensions. The defining feature of these dimensions from his point of view is that they are perceptually combined, or integrated, resulting in a redundancy gain. The positions of X's and O's were perceptually distinct, which might have led participants to ignore one of the two dimensions, thus getting no benefit from the redundancy. Lockhead proposes that integral dimensions should be defined that as those that produce a maximum redundancy gain, in an information theoretic sense.

The third type of data identified by Garner and Felfoldy (1970) as potentially useful in identifying integral dimensions (they preferred Lockhead's term) concerns selective attention. In the same way that Shepard (1964) claimed participants could selectively attend to analyzable dimensions (which Garner later came to call *separable*), Egeth (1967) noticed that the amount of interference caused by variation in an irrelevant dimension depends critically on the type of stimulus used. He used a speeded classification task, where participants were asked to rapidly sort stimuli (each displayed on their own card) into two separate categories.

In the filtering task, stimuli varied along two dimensions, but only one of them was used for classification. For an example, one stimulus set used the dimensions of shape and numerosity so that there were four distinct cards, each having one or two shapes (circles or squares). In one block of trials participants were asked to ignore the number of objects and just sort them by their shape. Another block would use the other classification rule.

The amount of interference caused by this irrelevant variation is calculated by comparing reaction times in the filtering task to reaction times in a control task that has no irrelevant variation (e.g. all cards have a single shape on them). As would be expected by the Shepard (1964) results, integral dimensions like those comprising color show much greater interference than separable dimensions like shape and numerosity.

Garner and Felfoldy (1970) saw that both the redundancy and filtering tests could be performed within the same speeded classification paradigm, allowing them to use convergent operations to better define integral and separable dimensions (the methodological details of these tests will be discussed in Chapter 2). The authors had previously argued that perception is an unobservable process intervening between stimuli and responses, and is best delimited by a set of converging operations (Garner, Hake, & Eriksen, 1956). They claim that "...a concept has no meaning beyond that obtained from the operations on which it is based," and that therefore a general perceptual process can only be described after measuring performance on a variety of tasks.

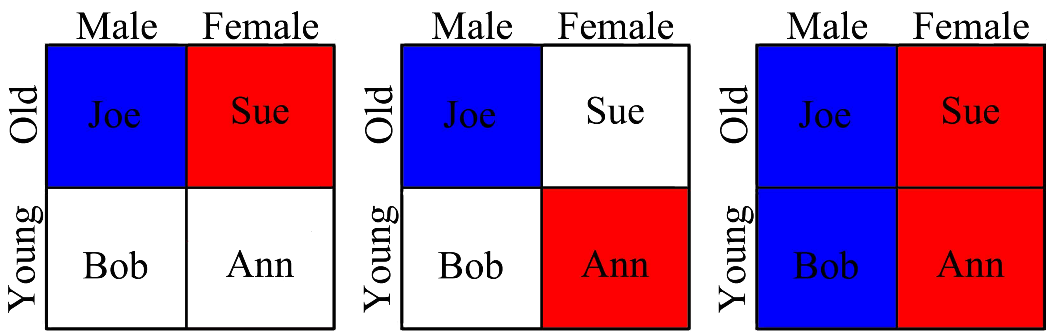
In applying this logic to the integral/separable distinction, Garner and Felfoldy (1970) argue that "...although a concept which is tied to a single experimental operation is nothing more than a restatement of an experimental result, a concept which evolves from several different experimental operations achieves a status independent of any one of the operations." With that in mind, they assert that integral dimensions are those which follow a Euclidean distance metric, produce a redundancy gain when the two dimensions are correlated, and show interference when they vary orthogonally in a selective attention task.

CHAPTER 2

THE GARNER PARADIGM

To understand the tests that Garner and Felfoldy (1970) used to identify integral dimensions, let us first consider the stimulus space. They used four different stimulus sets to compare different dimensions, but all used the same structure: a factorial combination of two dimensions which have two levels each, for a total of four stimuli.

Figure 2.1 shows an example for testing the separability of gender and age (example dimensions that were not considered in their paper). Participants are asked to selectively attend to one of each of the two dimensions in separate blocks, and for a given response dimension there are three combinations of stimuli that are presented. Let us consider the case where the response is along the gender dimension. In the *control* task, Figure 2.1a, stimuli are presented from only one level of the irrelevant decision, so the decision might be between an old man and an old woman. This is then repeated using the other level of the irrelevant dimension (young). In the *redundant* task, Figure 2.1b, which is also called the *correlated* task, both dimensions change between the two stimuli: in the case shown the participant chooses between an old man and a young woman. This task is also repeated using the opposite pairing. The *filtering* (sometimes referred to as *orthogonal*) task, Figure 2.1c, uses all four stimuli, requiring participants to "filter out" age variation to respond just to the gender. These tasks are then all repeated with subjects instructed



(a) Control (b) Redundant (c) Filtering

Figure 2.1: The three conditions for the Garner task. Blue squares represent stimuli mapped to one response (male), while red squares are mapped to the other response (female). White squares are stimuli not included in that block of trials. The control and redundant blocks are repeated using the other stimuli.

to selectively attend to age rather than gender.

Using reaction time data from these three tasks, there are two comparisons used to distinguish integral and separable dimensions. A *redundancy gain* is said to occur if the redundant task is faster than the control task, and what has come to be called *Garner interference* occurs when the filtering task is slower than the control task. Integral dimensions are predicted to exhibit both of these effects, while separable dimensions should exhibit neither, with reaction times being statistically equivalent for all three tasks. Both results follow logically from Shepard's (1964) hypothesis that with separable dimensions ("analyzable" in his verbiage), observers are able to pay selective attention to a given dimension, ignoring variation in another. This is in contrast to integral dimensions ("unitary" for Shepard), in which dimensions cannot be ignored. We can see that this ability to ignore variation with separable dimensions sometimes aids performance, as is the case in the filtering task, but can also limit performance by failing to make use of redundant information in the correlated task.

These intuitions about how integral and separable dimensions should behave in these three tasks were consistent with the data reported by Garner and Felfoldy (1970). The most commonly used examples of integral dimensions, value and chroma, showed clear redundancy gains and Garner interference, regardless of which dimension was attended to. When these dimensions were varied in two separate chips, which together composed a single stimulus (as in Hyman & Well, 1968), there were no significant differences in reaction times between the tasks, confirming the separability of these dimensions when displayed independently.

2.1 FURTHER DEVELOPMENTS

In 1976, Garner expanded his methodology by adding other convergent measures to complement the two speeded categorization tests and the choice of distance metric, also considering two new types of dimensional interaction. One of these new measures is based on free classification, as used by Handel and Imai (1972). Free classification is different from the speeded classification tasks we have already discussed in that the series of classification decisions is used as the data rather than the time taken to classify. The word "free" signifies that participants were not constrained as to which or even how many categories they could form. Despite this freedom, participants have proven to have a strong preference for using only two categories (Imai, 1966).

Handel and Imai (1972) designed a series of stimulus sets such that one pair of categories would naturally be used if the participants were basing their classifications on overall stimulus similarity, but a different pair of categories would be used if the decisions were instead based upon dimensional values. For example, in their three stimulus task, when participants use a similarity rule a given stimulus would be grouped with a second that varies only slightly on each of two dimensions. If participants use a dimensional rule, however, they would instead group that first stimulus with a third that has the same value as the first on one dimension, but is fairly distant on the other. The preferences for these two strategies are shown to be strongly dependent on the stimulus dimensions. With integral dimensions (here value and chroma), similarity is much more often chosen as a basis for classification. With separable dimensions (size and lightness), the stimuli are instead grouped by their values on one of the two dimensions, and thus this task provides another way of distinguishing the two.

Another way in which classification tasks can be used to distinguish integral from separable is to look for dimensional preferences. When participants are asked to classify the full factorial set of four stimuli, they will group them into two sets of two stimuli using one of the two dimensions, like in the filtering task (Figure 2.1c). When they are not instructed which dimension to use, however, their choice can be informative. Since we have just argued that integral dimensions are classified on the basis of their similarity, it follows that the more discriminable of the two dimensions will be used to form the groups, since it will maximize within class similarity and minimize between class similarity (Garner, 1974). As the relative discriminability approaches equivalence, we would predict that either dimension should be chosen for classification with a roughly 50% probability.

Separable dimensions, however, are supposed to be classified by their dimensional values, so relative discriminability should play a lesser effect. Imai and Garner (1965) showed that when using separable dimensions (position, distance, and orientation of a pair of dots), subjects had definite preferences for using one dimension over another, even with relative discriminability equated. This preference showed strong individual differences, with roughly half of the subjects consistently preferring distance and the other half preferring orientation (very few classified the stimuli by position). Handel and Imai (1972) present data showing that these preferences can persist even after the preferred dimension is made *less* *discriminable* than the other dimension. These results led Garner (1976) to conclude that persistent dimensional preferences were indicative of dimensions being separable.

In addition to these two new experimental methods for distinguishing dimensional interactions, Garner (1976) identifies two additional types of interaction: *configural* and *asymmetric separable*. The configural category is based on the work of Pomerantz and Garner (1973), who ran the typical speeded categorization tests on stimuli composed of two parentheses, which could each face either direction. Thus, their stimulus set was: ((, () ,)(, and)) . The unique results with these stimuli is that they exhibited a Garner interference effect, and thus participants were unable to selectively attend to the individual dimensions (in this case the left or the right parentheses), but yet there were no measurable redundancy effects. What appears to have happened for these stimuli is that an emergent dimension, closure (see Pomerantz, Sager, & Stoever, 1977), was exhibited in only one of the stimuli, () . This stimulus dominated performance, with participants being reliably faster in all three tasks distinguishing it from another stimulus, whether that stimulus differed with respect to only one parenthesis or both. It is possible that participants were more sensitive to the dimension of closure than either of the two individual dimensions the experimenters were manipulating, and therefore naturally grouped the stimuli accordingly, separating the one "well-formed" stimulus, in the Gestalt sense, from the other three. This would explain the difficulty in the filtering tasks and also the lack of redundancy. Garner (1976) hypothesized that in stimuli where the particular "configuration" of the dimensions appears more salient than the dimensions themselves, these kinds of effects can be expected.

The final type of interaction is asymmetric separable, in which case redundancy gains are present but Garner interference is only shown for one of the two dimensions. This reflects a type of dominance relation where one of the dimensions is capable of interfering with the other, but not the reverse, much like in the well known Stroop (1935) effect. Wood (1974) also presents data consistent with this type of interaction for the dimensions of pitch and stop-consonant in consonant-vowel syllables. Pitch could be selectively attended to, but stop-consonant could not. This is in accordance with the dimensional coexistence definition of integrality (Garner, 1970), since pitch can exist in a pure-tone and need not have a stop-consonant, implying separability and a lack of interference, but a stop-consonant must always have a pitch, implying integrality and therefore interference.

2.2 APPLICATIONS

Although Garner insisted on the wisdom of converging operations (Garner et al., 1956), most applications of his paradigm use only the two speeded classification tests: redundancy gain and Garner interference. These two tests alone, however, are predicted to be sufficient for discriminating between the four types of dimensional interaction discussed above (Garner, 1976), and therefore the discussion will be focused on them. These tests have often been used in situations where two dimensions or information sources are hypothesized to be processed independently. One example of such a case that has made heavy use of the Garner paradigm is the dual-route hypothesis of face recognition (Bartlett, Searcy, & Abdi, 2003). This theory posits that there are two fundamentally distinct sources of information about a face: featural and configural information. Featural information is inherently local, and describes the properties of individual pieces of a face (e.g. eye color, mouth shape, brow height). Configural information, on the other hand, is characterized by the relations between the features, like the distance between the eyes, or from nose to mouth.

Amishav and Kimchi (2010) used the Garner paradigm to test this dual-route hypothesis. They only conducted tests of Garner interference, however, choosing not to include a redundant condition. According to Garner's predictions, this would mean they would be incapable of distinguishing between configural and integral dimensions. Their stimulus set consisted of faces that had one of two sets of features (eyes, nose, and mouth) positioned according to two possible configurations, which had different inter-eyes and nose-mouth distances. As an additional between-subjects variable, they manipulated the orientation of the faces, since configural information is thought to be severely disrupted in upside-down faces. They found evidence of symmetric Garner interference for upright faces, but asymmetric interference with upside-down faces, with feature judgments being unaffected by configural variation. These results were interpreted as indicating that featural and configural information are processed in an integral fashion, but that configural information is dominated by the features with inverted faces. This is inconsistent with a strong interpretation of the dual route hypothesis, which would predict the information sources to be separable, but consistent with theories of holistic processing (Tanaka & Farah, 1993). In holistic processing, faces are thought to be perceived as non-decomposable units, with interactions between sources of information.

Holism would also predict that individual features would be processed in an integral fashion, which was also tested by Amishav and Kimchi (2010). Contrary to that prediction, they found no evidence for Garner interference between eyes and mouth or between nose and mouth, with either upright or inverted faces. Their participants were able to selectively attend to the relevant feature and were not distracted by irrelevant featural variation. This result demonstrates that not all information sources in a face are processed in an integral fashion, contradicting a strong version of holistic processing. In this way, the Garner paradigm has been used to examine how information is combined in the perceptual process, and to select between competing theories, especially with regard to the independence of processing.

Other dimensions of face processing that have been studied using the Garner paradigm include interior features vs. facial surround (Bartlett et al., 2003), emotional expression vs.

identity (Schweinberger, Burton, & Kelly, 1999; Kaufmann & Schweinberger, 2004), and expression vs. gender (Le Gal & Bruce, 2002; Ganel & Goshen-Gottstein, 2002).

CHAPTER 3

MODELS

While the question of whether dimensions are integral or separable has been asked many times throughout the literature, the question of how that difference could be accounted for in by a cognitive model has been asked relatively rarely. Here we examine some of the most successful attempts at modeling performance in the Garner tasks, highlighting the differences between their instantiations and predictions.

3.1 GENERAL RECOGNITION THEORY

Ashby and Townsend (1986), in their presentation of General Recognition Theory (GRT), describe how it could be used to further reify the concept of separability. GRT is a multidimensional extension of signal detection theory (Swets, Tanner, & Birdsall, 1961), which like its predecessor is capable of distinguishing between perceptual and decisional effects. Perceptual representations are modeled as multivariate Gaussian distributions in the space defined by the dimensions of the stimulus set, which like the Garner paradigm consists (almost always) of two dimensions with two levels each. The more the distributions for two stimuli overlap, the more confusable they are. Decision bounds are then applied to the space such that an observer responds “A” whenever a sample from a perceptual distribution is drawn from one side of the bound, and “B whenever it is drawn from the other. Although GRT is designed to be used with a full identification task, where two bounds are needed to delineate the four responses, it can also be used to model single dimensional classification decisions like those in the Garner tasks. Importantly, GRT distinguishes among several types of independence that can be defined between the two dimensions, which had been conflated in previous treatments.

As we have already seen, separability has been defined primarily in two operational contexts: the absence of interference with irrelevant variation, and the absence of a redundancy gain when the dimensions vary in perfect correlation. Ashby and Townsend (1986) claim that the first definition should primarily be understood as a statement about the workload capacity of a system. As defined by Townsend and Ashby (1983), workload capacity describes how a processing system changes with the number of items being processed. In terms of the Garner paradigm: how is the processing of brightness information changed when the system also has to process saturation, as in the filtering task? A limited capacity system predicts that the time taken to process any given source of information will increase (the system will slow down) as the number of sources is increased. When framed in this way, integral dimensions are thought to be processed in a limited capacity fashion, while separable dimensions can be processed with unlimited capacity. These authors consider the operational definition of redundancy gain to be fully expected with integral dimensions, but unduly restrictively in what dimensions can be characterized as separable. They argue that in a redundant task, an ideal observer would place at least some attentional weight on the second dimension, and that a redundancy gain will fail to occur only in the case that zero attention is allocated to this second dimension.

When considered with respect to General Recognition Theory, separability has two components: perceptual separability and decisional separability. Perceptual separability can be defined as the assumption that the perceptual representation of dimension A should not depend on the value of dimension B. In GRT the representation of dimension A is the marginal of the joint distribution function: $g(x) = \int_{-\infty}^{\infty} f(x,y)dy$, so dimension A is perceptually separable from dimension B when $g_{A_iB_1}(x) = g_{A_iB_2}(x)$. Note that this equation could be true for either value of dimension A, $i = 1, 2$, or for both. Also note that this is an asymmetric definition, and we could separately test whether dimension B is perceptually separable from dimension A. Thus there are a total of four ways in which two dimensions (which have two levels each) could be perceptually separable from each other, and the concept of Garner separability would require all four to hold.

Perceptual separability alone is not enough to guarantee Garner separability, however, because even if it holds in all four cases, a failure of decisional separability could yield data in conflict with one or both of our operational definitions. Decisional separability is defined to hold if the decision bound for dimension A does not depend on the level of B. This implies that the decision bound must be parallel to the irrelevant dimension's axis, so that the criterion for choosing between the levels of the relevant dimension remains constant for all values of the other dimension. Once again this definition is not symmetric, and A could be decisionally separable from B without the reverse being true, an analog of Garner's asymmetric separability. Another type of independence identified in GRT is perceptual independence, which is defined as the statistical independence of perceptual effects within a single stimulus, but this concept is less germane to the topic at hand.

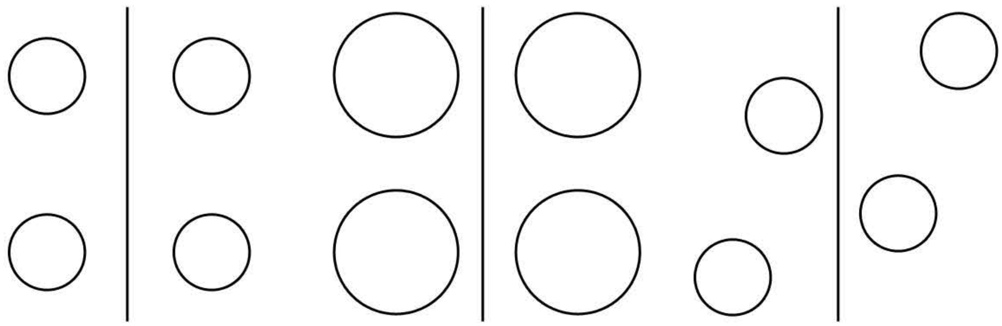
3.2 THE RT-DISTANCE HYPOTHESIS

Ashby and Maddox (1994) point out that although the Garner speeded classification tasks are based entirely upon reaction time data, GRT deals exclusively with response probabilities (accuracy), and therefore is agnostic to processing times. In order to make GRT more directly applicable to the Garner tasks, the authors introduce the RT-distance hypothesis: "On each trial, processing time monotonically decreases with the (Euclidean) distance between the percept and the decision bound." Thus stimuli near to the bound, which are more confusable, will elicit slower responses. The addition of this assumption allowed the authors to prove that if perceptual and decisional separability hold, then the reaction times will be equal for the control, filtering, and redundant tasks, implying Garner separability (also see Maddox, 1992).

The reverse inference is trickier, because Garner's tests for redundancy gain and interference will only imply perceptual separability if decisional separability is assumed. This assumption is relatively innocuous for the filtering task, since the ideal boundary between categories is parallel to the irrelevant dimensional axis, but is specious for the redundancy task. As pointed out earlier, the ideal observer will make use of both dimensions for this task, forming a diagonal decision bound and thus violating decisional separability. This strategy would lead to violation of Garner's test for separability, since the redundant task would be faster than the control task, regardless of whether or not perceptual separability was satisfied. This prediction of a failure of decisional separability in the redundancy task was subsequently shown in empirical data (Maddox & Ashby, 1996).

Ashby and Maddox (1994) found it more difficult to describe a GRT model capable of predicting Garner interference. One possibility they offered was that since the filtering task uses four stimuli and the other two tasks use only two (per block), there is an increase in stimulus uncertainty. They argue that this uncertainty could lead to larger variances for the perceptual distributions, resulting in more stimuli that are close to the boundary and therefore produce long reaction times, leading to an average slow down. A schematic showing such a model is seen in Figure 3.1b.

Another possibility involves what they call *mean-shift integrality*. Mean-shift integrality



(a) Standard Model (b) Increased Variance (c) Mean Shift Integrality

Figure 3.1: Equal likelihood contours for the four stimuli in a Garner filtering task. The vertical line represents the decision bound.

assumes an interaction between the two dimensions takes place at the level of the means of the perceptual representations, such that stimuli from the higher level of one dimension are perceived as also having a higher level of the other dimension. This results in a stimulus layout in the shape of a diamond rather than the standard square, as shown in Figure 3.1c. This effect can also lead to an average slow down in times, since (assuming decisional separability) two of the stimuli will be very close to the decision bound. In light of the reliance on decisional separability for making inferences about processing characteristics in the Garner paradigm, Ashby and Maddox (1994) recommend against allowing subjects to become well-practiced at the tasks, since increasing practice leads participants to use more optimal decision bounds that are unlikely to be separable.

Nosofsky and Palmeri (1997a) noted that Maddox and Ashby (1996) had only collected data using separable dimensions, and wanted to test their model's performance with integral dimensions: tones varying in pitch and loudness. They found that the GRT-based model was unable to predict interferenece in the filtering task when it was fit to the full distribution of reaction times, rather than just the average. Their key observation was that the two ways mentioned by Ashby and Maddox (1994) for producing such interference, mean-shift integrality or increased variance due to uncertainty, both affect the fastest reaction times in addition to the slowest ones. For simplicity, let us just consider the increased variance example, but the argument holds for mean-shift integrality as well.

When variance is increased for the filtering task, there will be more samples drawn from the perceptual distribution that lie *close* to the decision bound, and therefore are slow, but also more samples lying *far* from that bound, that therefore are fast. On its face it seems as though this would imply that the average reaction time should be unaffected, but the adoption of a non-linear function for relating distance from boundary to reaction time means that the slow times will be further from the mean than the fast times, yielding a slower average RT. When considering the full distribution of reaction times, however, this assumption demands not only that the slowest times from the filtering task will be slower than those from the control task, but also that the fastest times from the filtering task will be faster than those from the control task. This is true for mean-shift integrality as well as for a simple increase in variance. The data from Nosofsky and Palmeri (1997a) are contrary to this requirement, with the control condition producing faster responses in both tails of the distribution.

3.3 THE EXEMPLAR BASED RANDOM WALK MODEL

Nosofsky and Palmeri (1997b) propose a different model for accounting for Garner interference, the Exemplar Based Random Walk (EBRW) model. The backbone of their model is the Generalized Context Model, or GCM, which was developed by Nosofsky (1986) as an expansion of an earlier model by Medin and Schaffer (1978). The GCM proposes that categorization decisions are made by comparing a sample stimulus to previously seen exemplars for each category, and uses a ratio of summed similarities to pick the category in which to assign the new stimulus. This rule decides the probability of responding category

A for stimulus i as follows:

$$P(R_A|S_i) = \frac{\sum\limits_{j \in C_A} \eta_{ij}}{\sum\limits_{B}^{m}(\sum\limits_{k \in C_B} \eta_{ik})}, \tag{3.1}$$

with capital letters representing categories, and m being the number of those categories (the left summation in the denominator disappears for the two category case). Here, η_{ij} is the similarity from stimulus i to stimulus j, and is calculated as an exponential decay transform of the distance between them, $\eta_{ij} = e^{-d_{ij}}$. This distance is calculated in the same way as we saw earlier in MDS models, except with the addition of weighting parameters to adjust how much the different dimensions contribute to the overall distance:

$$d_{ij} = c[\sum_{k=1}^{N} w_k |x_{ik} - x_{jk}|^r]^{1/r}, \tag{3.2}$$

where $c \in [0, \infty)$ represents overall discriminability and $w_k \in [0, 1], \sum w_k = 1$ represents the weight on dimension k. This modification allows the GCM to be well suited for fitting Garner-type data, where selective attention plays such an important role.

The EBRW model uses the GCM to drive a random walk process capable of predicting accuracy and reaction time simultaneously. A random walk is a technique for modeling the noisy integration of information over time. At every time step, the evidence counter (which starts at zero) moves up a step or down a step, continuing until it hits the upper or lower boundary, at which time a response is given. The response is dictated by which boundary has been reached, and the reaction time is a simple transformation of the number of steps taken (see Luce, 1986).

Here, the two boundaries represent classifying a stimulus as belonging to either category A or category B. The evidence accumulation process is driven by exemplar similarity, as defined in the GCM. For each time step, all possible exemplars (e.g. the four stimuli in the filtering task) race against each other with rates dictated by their similarity to the item being processed (the more similar, the faster the process). The evidence counter takes a step toward the boundary representing the category to which the winning exemplar belongs. In this way, the EBRW model predicts rapid and accurate classifications for situations in which the target stimulus is highly similar to exemplars from one category and highly dissimilar from exemplars in the other category: the evidence counter should march consistently toward the first category's boundary. A target stimulus that is equally similar to exemplars from each category would produce a slow response with chance accuracy.

Although we cannot cover every detail of the implementation of the EBRW model, it is important to mention a feature inspired by the instance-based model of automaticity (Logan, 1988), which posited that processing algorithms create instances in memory that after accumulation (practice) can be used in lieu of the algorithm itself, becoming faster and automatic. This feature was incorporated into the EBRW model in that similarities between exemplars are scaled by the memory strengths of those exemplars, which are a function of repetition and practice. Thus, in the race to determine the direction of a random walk step when processing stimulus i, the probability that the process corresponding to exemplar j terminates at time t is defined as

$$f(t) = a_{ij}e^{-a_{ij}t}, \tag{3.3}$$

where the *activation value* depends on both memory strength (M_j) and similarity: $a_{ij} = M_j\eta_{ij}$. This feature allows the EBRW model to predict both practice effects and repetition effects, where response times are faster for any trial using the same target as the immediately preceding trial.

The EBRW model was proposed to be applicable to integral dimensioned stimuli, since those are theorized to be "...encoded, perceived, and represented as single, unitary wholes" (Nosofsky & Palmeri, 1997b). Exemplar level processing was hypothesized to be less efficient for separable dimensions, in which each dimension might be perceived individually, perhaps even in a serial nature (although the idea of selective attention is easily captured by the weight parameters in the distance formula, Equation 3.2). Redundancy gains are expected to occur in the correlated task for the same fundamental reason as in a GRT model (that assumes a failure of decisional separability): stimuli in this condition are further apart (less similar) than in the control condition. In this model, that leads to fewer “mis-steps” in the random-walk process, and therefore a faster response.

Slower responses in the filtering case are predicted primarily because of the number of stimuli. Because the activation value of a stored exemplar depends on its memory strength, exemplars which have been seen more frequently will be faster. Since the filtering task has four stimuli, rather than the two in each of the other conditions, each exemplar is presented only half as many times. Assuming a memory decay process requires us to rephrase this argument, but its essence remains the same: exemplars in the filtering condition are only half as likely to have been recently presented.

Nosofsky and Palmeri (1997b) also tested their model’s performance on a *stretch filtering* task, where distances along the irrelevant dimension have been exaggerated. The prediction of their model is that performance on this task should be even worse than the standard filtering task. The idea is that in a filtering task, some of the time a target stimulus will be misidentified as the other stimulus in the same category, an event which we will call a *fortuitous confusion*. This is fortuitous, because despite the wrong exemplar winning the race, the evidence counter still takes a step toward the correct category boundary. In stretch filtering, there is greater distance between stimuli of the same category, and therefore these fortuitous confusions are less numerous, slowing down the response process.

The data collected by Nosofsky and Palmeri (1997b) was fit well by the EBRW model, and show the expected reaction time ordering of correlated faster than control, which was faster than filtering, which in turn was faster than stretch filtering. In direct comparisons to the Ashby and Maddox (1994) RT-distance modification of GRT, the EBRW model produced a better fit to the data, especially in regard to the mechanism by which filtering was produced in the model. Rather than assuming an increase in variability, which we saw mandates especially fast in addition to especially slow responses for the filtering trials, a decrease in the memory strength of the exemplars predicts slower responses across the full RT distribution, which is the qualitative pattern exibited by the participants in their experiment.

3.4 LOGICAL RULE MODELS

All of the experimental tasks considered so far have used *selective attention*, in which subjects are only supposed to pay attention to one of the two dimensions. Fific and Townsend (2008) examined how multidimensional classification decisions are made under *divided attention*, where correct classifications can only be made by combining information from both of the two dimensions. They studied performance using *Systems Factorial Technology.* SFT uses reaction time data to identify system-level information processing characteristics like architecture and stopping rule, and its tests are all non-parametric and distribution-free.

Architecture refers to the structure in which two processing channels are ordered. The most common distinction is made between serial and parallel models. In a serial model, the two channels are processed one-at-a-time, with one following the other. In a parallel model, however, both are processed simultaneously. A third important model is known as *coactive*, and assumes that the two channels proceed in parallel, but that their information is pooled together and compared to a single response threshold. *Stopping rule* refers to how much information is required before the system can respond. A first-terminating "OR" rule can make a decision as soon as one of the two processing channels is completed, and is also referred to as a minimum-time rule. An exhaustive "AND" rule requires both channels to be completed before a response can be issued, and is therefore a maximum-time rule.

These stopping rules can be applied to either serial or parallel models, but not for a coactive model, since there is really only one "pooled" channel.

The tool for diagnosing these two properties is called the *Survivor Interaction Contrast*, or SIC function. Details regarding the SIC are beyond the scope of this paper, but please see Townsend and Nozawa (1995). Let us just note that Fific and Townsend (2008) found that integral dimensions (brightness and saturation) were processed coactively, while separable dimensions (color and the position of a vertical line) were processed using either a serial or parallel architecture with an exhaustive stopping rule. A coactive architecture is consistent with the idea that integral dimensions are not processed separately, but rather are inextricably bound together, pooling their information toward a common decision bound.

Fific, Little, and Nosofsky (2010) built off of the idea that architecture is an important feature of dimensional processing, and presented a *logical rule-based model* of categorization. Logical rules were one of the earliest hypotheses of how subjects categorize stimuli. An example would be that one is willing call an object a tennis ball if it is roughly fist sized AND yellow AND fuzzy. Until this paper, however, these models had never been capable of predicting response times, only categorization decisions. In order to compare this class of models with the exemplar or decision bound models we have already discussed, we need predictions at the level of reaction time distributions.

Fific et al. (2010) use logical rules to predict both choice behavior and response time by once again appealing to a geometric model. Rather than the typical Garner configuration of four stimuli, their experiment uses the factorial combination of two dimensions with three levels each, for a total of nine stimuli. The four stimuli in the upper right quadrant of this configuration are called category A, while the other five are labeled as category B. Participants are therefore instructed to respond "A" if they perceive dimension one to be at the second or third level, *and* dimension two at the second or third level. A "B" response could be initiated if either dimension one is perceived at the first level *or* dimension two is at the first level.

Just like in GRT, each stimulus gives rise to a perceptual distribution, and these distributions are sampled for each step in the random walk process (in a manner similar to the EBRW model). Samples that lie on the "A" side of the decision bound lead the process to the category A boundary, and likewise for category B. The key difference is that this model assumes that the two dimensions are processed independently of one another, with each dimension driving its own random-walk process. There will be one process operating on the marginal distributions along dimension one, and a separate process for dimension two. These random walk processes reach their own decisions, and then are combined via a logical rule. In the setup for this experiment, a response for category A will only be issued if both dimensional processes reach the "A" boundary, and will only be initiated once both have finished. We can see that this logical rule specifies the stopping rule for the system.

As mentioned earlier, the concept of processing architecture can now be incorporated into this model. The two random walk processes could be arranged in serial, so that one does not begin until the other has finished, or in parallel, where they run simultaneously. A coactive model would proceed somewhat differently, however. In this model, samples are combined from both dimensions to drive a single random walk process. On every step of the process, samples are taken from both dimensions and then combined using the appropriate logical rule. If both samples come from region "A" then the process takes a step to the "A" boundary, but if either sample comes from category B the process moves in the opposite direction.

Fific et al. (2010) conducted experiments using separable dimensions to test their model. These experiments were intended as validation tests, showing that participants are capable of using these strategies, rather than an investigation into when participants are likely to use such strategies. In the first experiment participants were explicitly instructed to use a *fixed-order serial* processing strategy, first making a decision along dimension one and then along dimension two. When the model proved to fit these reaction time and accuracy data better than other extant models, they tried relaxing their methodology. While subjects were still informed of how the categories were defined by logical rules, they were not instructed to use a particular strategy. Furthermore, some subjects were told to emphasize accuracy, while another group prioritized speed. Their model continued to provide a good fit for the data.

The applicability of this model was extended by Little, Nosofsky, and Denton (2011), who tested categorization performance when participants had to learn the categories by induction, rather than being informed of their logical structure. They also probed the difference in using separable dimensions that are spatially separate compared with those that overlap. The former class of stimuli were schematics of lamps, where the relevant information was contained at the very top and the very bottom of the lamp. Participants in this condition showed a consistent pattern of (spontaneously) adopting a serial self-terminating strategy.

The spatially overlapping stimuli consisted of a colored rectangle with a vertical line on top of it. The relevant dimensions were the saturation of the background color (always red) and the horizontal displacement of the vertical line. These dimensions have previously been shown to be separable in these stimuli (Fific & Townsend, 2008). Participants in this condition showed a mixture of serial and parallel processing of the two dimensions, and a logical rule instantiation of this mixture fit the data much better than other models.

It makes sense that separable dimensions would be processed individually and then combined via a logical rule, but what about integral dimensions? Little, Nosofsky, Donkin, and Denton (2013) used color stimuli varying in brightness and saturation to test if the logical rule model could still capture the relevant patterns in the data. As we saw earlier, Fific and Townsend (2008) presented evidence that integral dimensions are processed coactively, but they did not include accuracies in their analysis. This new experiment, however, yielded similar results, showing that a coactive rule-based model provided a strong fit to the data.

Although these rule models have proven adept at modeling performance with both integral and separable dimensions in a divided attention framework, they have not been tested in the selective attention tasks studied with the Garner paradigm. Little et al. (2013) at least offer some insights into how their model would be capable of predicting Garner interference effects with integral stimuli. A coactive architecture on its own would not be sufficient for predicting slower reaction times in the filtering condition, but can do so with the Ashby and Maddox (1994) assumption of increased variance due to increased stimulus uncertainty. Crucially, because their model only cares about which side of the decision bound a sample is drawn from, and not how far from the decision bound it is, they are not forced to make the incorrect assumption that filtering trials also produce the fastest reaction times, as detailed by Nosofsky and Palmeri (1997a).

3.5 TECTONIC THEORY

A final formal language for describing Garner effects, called *Tectonic Theory*, is laid out by Melara and Algom (2003). They ground their theory in the fundamental antagonism that exists in visual attention: the need to selectively attend to a subset of stimuli and the human adaptive propensity for integrating information and noticing patterns in what might have previously appeared irrelevant. From their point of view evidence for a "failure of selective attention" should not be taken in a pejorative sense, for this "failure" can often be a beneficial information gathering strategy. Even the classic Garner results show that selective attention is sometimes beneficial, as in the filtering condition, and sometime detrimental, as in the correlated condition.

The authors characterize the propensity for selective attention to fail as depending primarily on three properties of the source of irrelevant information. The first is surprise: the less predictable the variation in the irrelevant dimension, the more likely it will capture our attention. The second is salience: changes in the irrelevant dimension are more likely to be noticed the more discriminable they are. The final factor is correlation: the more closely associated the irrelevant dimension is with the relevant one, the more likely you are to devote some attention to it.

If we assume that the two dimensions are equated for salience, Garner filtering can be ascribed to an increase in stimulus uncertainty when using four stimuli rather than two. This increase in "surprise" will induce a failure of selective attention, and a slow down of response times. Selective attention should also fail in the redundancy task, this time due to the close association between the two dimensions. The main focus of this theory, however, was in explaining Stroop interference effects, which rely on similar attentional mechanisms as Garner interference, but include the complicating aspect of dimensional congruence.

CHAPTER 4

CRITICISM

Although the Garner paradigm has been readily taken up by many researchers, there have also been plenty of detractors. Researchers have disputed the separable-integral distinction, the methodology used in the testing, and the ways in which dimensions are defined. Some of the first criticism regarded the nature in which integral stimuli were supposed to be processed.

4.1 DIMENSIONAL CONSIDERATIONS: MELARA AND MARKS

In the standard interpretation, integral dimensioned stimuli are thought to be processed in a dimensionless manner, as a single perceptual "blob" (Garner, 1974). The component dimensions can be accessed individually, but only after some time and effort. This interpretation has been referred to as an *early-holistic* processing theory, since early in processing the perceptual object remains unitary and "whole". Melara and Marks (1990) argue instead that access to dimensional information is immediate even for integral dimensions, supporting this claim using a rotation technique established by Smith and Kemler (1978). When stimuli are processed in a dimensionless manner, then the stimulus space should be defined only up to an arbitrary rotation of the axes. As we saw in Chapter 1.1, the Euclidean metric defines distance only with regards to similarity, without caring about the orientation of the axes (in contrast to the city-block measure).

Smith and Kemler (1978) tested the psychological truth of this assumption using the dimensions of saturation and brightness. They generated two stimulus sets, one along those two primary dimensions, and another that was a 45 degree rotation of the first space. Filtering performance was equivalent for the two sets, leading the authors to conclude that no particular orientation of these two axes is primary.

Subsequent work disputed this claim, however, showing perceptual advantages for non-rotated orientations under a variety of different conditions (Foard & Kemler, 1984). Further study indicated that the auditory dimensions of pitch and loudness also show this pattern of increasing Garner interference as the stimulus space is rotated toward 45 degrees (Grau & Nelson, 1988). The authors argued that these dimensions, which should be labeled as integral due to the interference results, are in some way *less integral* due to their lack of rotational independence. Melara and Marks (1990) reason that this somewhat conflicted state of affairs can best be explained by taking a step back from the assumption of dimensionless processing.

They argue instead that stimuli are always immediately perceived along a set of primary axes, what they refer to as *attribute-level processing*. When dimensions are integral, however, *stimulus-level processing* also occurs, which accounts for the contextual effects of one dimension on the other. Testing a variety of auditory dimensions, they found consistent evidence for dimensional primacy in both selective and divided attention tasks (Melara & Marks, 1990). These results are difficult to assimilate in the classical conception of integrality.

Their argument is made more precise in a commentary against early-holistic processing models, which draws support from further experimental results (Melara, Marks, & Potts, 1993a, 1993b). In an expansive series of studies, they tested all three possible pairings of hue, saturation, and brightness at three levels of stimulus space rotation, using both selective- and divided-attention tasks. In all experiments they showed consistent evidence for the superiority of the "primary," unrotated axes. This co-ocurance of Garner interference and dimensional primacy appears incompatible with early-holistic processing, and the authors recommend an alternate explanation for interference, such as mean-shift integrality, as discussed in Chapter 3.2 (Ashby & Maddox, 1994).

An important consideration when evaluating claims of Garner interference is the question of baseline discriminability. The issue speaks to weather a finding of interference between two dimensions is indicative of the general relationship between those two dimensions, or whether conclusions must be restricted to the specific experimental context. Melara and Mounts (1993) show that the existence or direction of asymmetric Garner interference between colors and words in the standard Stroop paradigm can be systematically manipulated by altering the relative discriminabilities of the two dimensions. Their claim is that unequal discriminability causes a mandatory failure of selective attention, in that the easier dimension will interfere with the harder one. In this view, a finding of interference is not very meaningful if the dimensions are not matched for discriminability. This effect of discriminability in Stroop stimuli was replicated by Dishon-Berkovits and Algom (2000), who also present data indicating that failures of selective attention are predicted by any correlational structure in the stimulus presentation, as is the case in Stroop tasks which use an equal number of congruent and incongruent trials.

4.2 NECESSARY OR SUFFICIENT: ASHBY AND MADDOX

As pointed out in Chapter 3, Sections 1 and 2, Ashby and Maddox (1990) dispute the diagnosticity of the redundancy test. They show that a failure of decisional separability makes perceptual separability untestable. They argue that although perceptual separability can be properly construed as a property belonging to a pair of dimensions, decisional separability is better thought of as an optional strategy that is strongly influenced by the task structure. Because the correlated condition uses a stimulus set where the optimal decision bound is diagonal, practiced observers should be expected to violate decisional separability, regardless of the choice of dimensions. Later work validated these predictions, showing consistent failures of decisional separability (and therefore redundancy gains) for "separable" dimensions (Maddox & Ashby, 1996).

These authors also turned their focus to the other test of separability, Garner interference. Ashby and Maddox (1994) point out that a finding of equality between the control and filtering tasks is a necessary, but not a sufficient condition for a finding of selective attention. They claim there could be many ways in which two dimensions which cannot be selectively attended to would produce these experimental results. One example detailed by Melara and Mounts (1993) is that if the two dimensions are not matched for difficulty, the harder to perceive dimension may have no measurable effect on the easier dimension, even without true selective attention.

Another problem with the logic of the filtering test is that it assumes "The perception of a stimulus does not depend on the number or the identity of the other stimuli in the ensemble" (Ashby & Maddox, 1994). This argument was foreshadowed in Chapter 3 by the way in which the RT-distance model (Ashby & Maddox, 1994), the EBRW model (Nosofsky & Palmeri, 1997b), and the logical-rule model (Little et al., 2013) predict Garner interference effects. All models assume that performance is worse in the filtering task primarily (if not entirely) because it uses four stimuli instead of two. The EBRW instantiates this effect in terms of memory strength, whereas the other two models assume that the increase in stimulus uncertainty leads to an increase in the variance of the perceptual distributions. The problem with this is that the standard interpretation of a finding of Garner interference is that the addition of variation along an irrelevant dimension is what causes selective attention to fail, not merely the addition of more stimuli.

A final confound in the comparison of the filtering and control tasks was mentioned by Nosofsky and Palmeri (1997a). It has been shown that in a two stimulus, two response task, participants are capable of spontaneously adopting an alternative "change-detection" strategy (Fletcher & Rabbitt, 1978), where responses are based off of the previous trial. If the subsequent stimulus is the same, then the response will be as well, but if the stimulus changes, participants merely switch to the other response. This strategy is extremely efficient, since change-detection is a much faster level of processing then say, gender classification. This strategy is not available in the filtering task, since in some cases the stimulus will change but the response will remain the same, and thus would be sufficient to explain a difference between the two tasks.

CHAPTER 5

A NOVEL EXTENSION OF THE GARNER PARADIGM

So what does it mean when the processing of a second dimension affects reaction times, slowing them down if the information is irrelevant (the filtering block) and speeding them up if it is diagnostic (the redundant block)? A common conflation in the literature is that a finding of integrality implies that the two dimensions interact (Amishav & Kimchi, 2010). Adding to that confusion are the many different definitions of independence. Fitousi and Wenger (2013) have compared experimental data related to a selected variety of these definitions, finding that some are consistent with the Garner results while others are not. An important discussion largely missing from the literature is the relatively large class of models capable of predicting a finding of integrality while maintaining independence between dimensions. Here, I refer to independence in an information processing sense, as the lack of "cross-talk" between two processing channels, one for each dimension (for more precise definitions of independence, see Townsend & Ashby, 1983). In this research program I investigate what experimental factors are driving the Garner effects, and what classes of models are capable of predicting them.

5.1 CONFOUNDS

We have already talked about some potential experimental confounds within the Garner design in Chapter 4. The two primary tests used to decide if dimensions are integral or separable are for redundancy gain, where the correlated task is faster than control, and Garner interference, where filtering is slower than control. We have seen that the requirement for separable dimensions to *not* show a redundancy gain is stringent (Maddox & Ashby, 1996), and this test is often dropped in recent applications (Amishav & Kimchi, 2010). What makes a finding of Garner interference particularly difficult to interpret is that there are at least three important differences between the control task and the filtering task. The first difference, the addition of variation along an irrelevant dimension (i.e. a dimension that offers no information as to the correct response), is what is assumed to be driving the effect under the classic interpretation.

Another difference in the two tasks, however, is in the number of stimuli, changing from two to four. In addition to the effects this could have on memory strength or perceptual variance, as detailed earlier, this could impact the decision process in the amount of generalization that must be done to extract the relevant differences between stimuli. Any two faces are going to differ in an infinite number of ways (Townsend, Burns, & Pei, 2012), so there is much more flexibility in how to distinguish them than there is when trying to distinguish one pair of faces from another pair. As an example, an observer might find that the easiest way to choose between a pair of stimuli is some idiosyncratic feature like an eyebrow or earlobe. The success of such a strategy would be reduced as the number of stimuli are increased. There is also the possibility that participants are making a decision as a combination of identification judgments rather than the desired categorization: pressing button 1 for Bob or Jim and button 2 for Ann or Sue, rather than deciding male or female. This kind of strategy would also predict a slow down as the number of stimuli increases.

The final difference between these two tasks is that the response mapping also changes. Although the Garner paradigm is always presented as a series of "speeded-classification" tasks, the control and redundancy tasks have a one-to-one mapping between stimuli and responses, and could therefore be labeled as identification tasks. One problem with this is the possibility of participants using a change-detection strategy in the control task, as previously mentioned. Even if such a strategy were never adopted by participants, however, there are other reasons to believe that identification and categorization tasks might operate differently from one another.

Nosofsky (1986) describes the relationship between these two tasks when showing how they can both be fit by his Generalized Context Model, previously seen in Chapter 3.3. As proposed by Shepard et al. (1961), the simplest way to relate them is to use a mapping hypothesis to carry identification confusion data into categorical responses: the probability of a stimulus being identified as belonging to category A is merely the summed probability of it being identified as any of the members in category A (formalized for the GCM in Equation 3.1). The most basic form of this model fails to describe the data, however, primarily due to the influence of selective attention mechanisms, as shown by Shepard (1964). Nosofsky (1986) incorporates these effects into the GCM through attentional weight parameters in the distance function, as seen in Equation 3.2. What this means in the context of Garner interference is that this re-weighting of the distance function when going from the control to the filtering task could be the cause for a reaction time difference between them.

To this author's knowledge, it has not been empirically demonstrated that any of these possible confounding factors have negligible influence on the reaction time differences relied upon in this paradigm. To this end, I propose a new extension of the Garner paradigm designed to isolate these factors to test their independent effects on reaction times. Identifying the relevant factors will aid in choosing a model capable of representing the difference between separable and integral stimuli, and allow us to better define what those labels mean.

5.2 TASK STRUCTURE

In order to separate out the influences of the number of stimuli, the number of irrelevant dimensions, and the type of response mapping, I propose extending the Garner paradigm to three dimensions. By using different combinations of stimuli from this cube, tasks can be compared which differ on only one of these three factors. A schematic of all nine stimulus combinations is shown in Figure 5.1, with the top row representing the classic Garner conditions. Rather than describing each of the tasks individually, let us walk though the various comparisons between tasks that this extension was designed to enable.

Let us first note that all of the novel conditions (the lower two rows) use many-to-one response mappings, ensuring they are all true classification tasks and ruling out the utility of a change-detection strategy. A three-dimensional analog of the Garner interference test can be conducted by comparing performance in the filtering task, in which the third (depth) dimension is fixed at a constant level, to the *double filtering* task, where that dimension is now also allowed to vary. While this comparison does not entail a fundamental change in response-mapping, it still confounds a change in the number of irrelevant dimensions with an increase in the number of stimuli.

To test how the number of irrelevant dimensions itself affects reaction time, we can compare performance in the standard filtering task to what is labeled as the *correlated filtering* task. The former has one irrelevant dimension, while in the latter both the vertical and the depth dimensions (as pictured) vary irrelevantly. Importantly, both tasks are composed of four possible stimuli, thus controlling for stimulus uncertainty effects. Similarly,

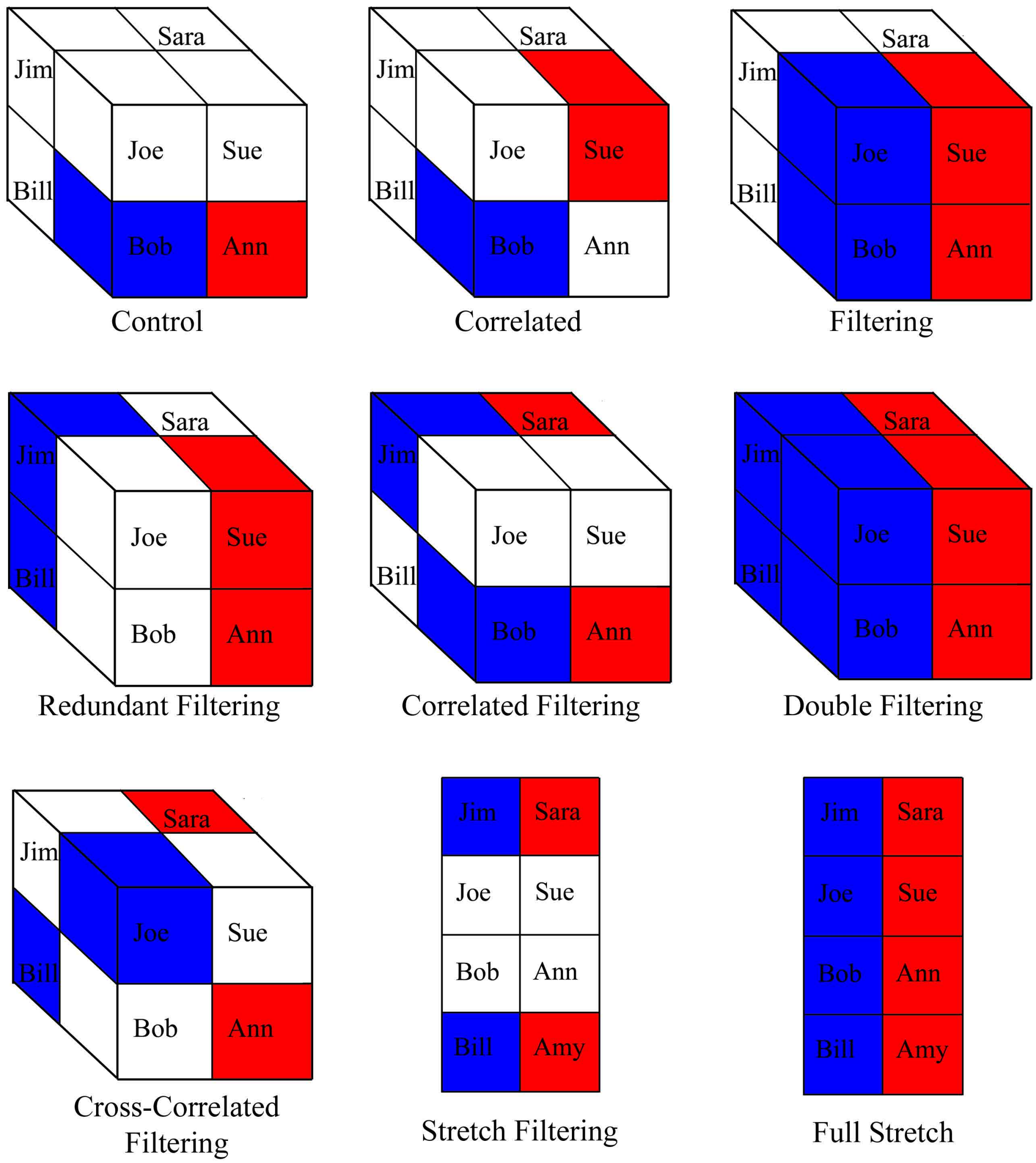


Figure 5.1: The nine stimulus combinations used in this investigation. For all tasks, attention is focused on the horizontal dimension, so blue squares represent stimuli mapped to one response, while red squares are mapped to the other.

we can measure the effect of increasing the number of possible stimuli by comparing the correlated filtering task with the double filtering task. In both tasks there are two irrelevant dimensions, but one uses four stimuli while the other uses eight.

We can now see that the correlated filtering task serves as a perfect "in-between" condition for the 3-D Garner interference comparison between filtering and double filtering. If such interference were due entirely to increasing the number of irrelevant dimensions, then correlated filtering should take as long as the double filtering task. If interference is instead caused merely by the increase in stimuli, then correlated filtering should take only as long as the standard filtering task. If both factors contribute to the interference, reaction times will lie somewhere in the middle, depending on the relative strength of the effects.

Unlike the Garner interference test, the standard redundancy comparison holds constant the number of stimuli, the response assignment, and the number of irrelevant dimensions while manipulating the number of diagnostic dimensions (here defined as dimensions that are individually sufficient for making a correct response). An issue that remains, however, is that both tasks use a one-to-one mapping, and therefore cannot be easily compared with the other tasks. Extending to three dimensions once again opens up new possibilities, as we can now test the effect of redundancy using many-to-one tasks. Recall that the classic Garner filtering task has one irrelevant dimension, four stimuli, and only one diagnostic dimension. This can be contrasted with the *redundant filtering* task, which maintains four stimuli with one irrelevant dimension, while adding a second diagnostic dimension.

A separate form of analysis that can be informative for both tests of redundancy is the measurement of workload capacity. As laid out by Townsend and Nozawa (1995), workload capacity is a measure of how processing efficiency changes when the workload is changed. In the context of this experiment, the workload is the number of relevant dimensions in a task: the control task has one, while the correlated (or redundant) task has two. In an *unlimited capacity* system, each of the two channels would continue to process information at the same rate when they are together as when they are processed alone. If the channels slow down when processed together we call it *limited capacity*, and in some cases they can even speed up, showing *super capacity*.

The Garner paradigm typically compares the correlated and control tasks solely on the basis of mean reaction times. When the correlated task is faster, a redundancy gain is said to occur, and this is taken as evidence that the dimensions are integral. This comparison has often been done only with respect to control trials from the assigned dimension, though many researchers now recognize the importance of comparing the correlated task to control conditions from each of the constituent dimensions.

As pointed out in Chapter 4.2, the absence of a redundancy gain is a strict criteria for separability, as a wide variety of models are capable of predicting such gains while maintaining complete independence between the two channels. The capacity coefficient $C(t)$, which is used to measure workload capacity, provides a more fine-grained examination of redundancy gain. A value of one indicates that performance is comparable to that of a useful baseline model: an unlimited capacity, parallel, independent model. This model assumes that each of the two dimensions are processed separately and independently in their own channels, and predicts that reactions in the correlated condition are faster due to statistical facilitation, in that the response can be issued as soon as the faster of the two channels finishes. A value of one half indicates a *fixed-capacity* system, where the correlated condition is only as fast as the average of the two controls. $C(t)$ characterizes performance across (and beyond) this spectrum of possibilities, and does so across all values of response time.

We can also use the redundant filtering task for a measure of Garner interference in the presence of two diagnostic dimensions, rather than the traditional single dimension.

Observe that the standard correlated task possesses two diagnostic dimensions, with the third dimension fixed at a single level. The redundant filtering task then allows that third dimension to vary, analogous to the comparison between filtering and control tasks. It may be that the addition of an irrelevant dimension has the same effect in both cases, but it is also possible that there is some interaction between the effects of redundant and irrelevant dimensions, and this comparison will shed light on that issue.

The bottom row of Figure 5.1 shows three conditions included to help distinguish between the various competing models. The *cross correlated* task is very similar to the correlated filtering task: they both use four stimuli, a single relevant dimension, and two irrelevant dimensions. The difference is that in the cross correlated task the direction in which the two irrelevant dimensions are correlated changes depending on the value of the relevant dimension. This change wouldn't matter for a simple distance-from-boundary model like those of Ashby and Maddox (1994) or Little et al. (2013), because the number of stimuli and their distances from the center plane are the same as in the correlated filtering task. In a similarity based model like the EBRW (Nosofsky & Palmeri, 1997b), however, the cross correlated condition should be easier. This is because while within-category distances are the same as in the correlated filtering task, between-category distances are greater. Decision models would also be capable of modeling this advantage if participants violate decisional separability and form a complex, saddle-shaped decision boundary.

The final two conditions use a different set of stimuli than the other seven. The *stretch filtering* condition was designed by Nosofsky and Palmeri (1997b) to test how Garner interference changes when the irrelevant variation is made more salient. It uses a new set of stimuli that use the same values for the relevant dimension, but double the values along the to-be-filtered dimension. As shown by Melara and Mounts (1993), there is significant evidence that interference should increase when the irrelevant dimension is more salient, a prediction also borne out in the EBRW model.

Comparing the filtering to the stretch filtering tasks is the complement of what we just saw in the cross-correlated filtering task: now the within-category distances have increased while the between-category distances remain the same, leading to a decline in performance. Again, a simple distance from boundary model would predict no effect, but could be made to do so with an added assumption that perceptual variance increases when variance between stimuli is increased, though this would run into the problems with fast responses pointed out by Nosofsky and Palmeri (1997a). This condition also allows for additional tests on how performance is affected by the incorporation of additional stimuli. The *full stretch* condition can be compared either with stretch filtering to measure the effect of adding "internal" stimuli, or with the standard filtering task to measure the effect of adding "external" stimuli.

We should note here that just like in the Garner paradigm, the addition of an extra dimension is expected to have opposite effects depending upon whether value on that dimension is indicative of the response. When the filtering task is "tilted" to yield the correlated filtering condition, we expect participants to slow down (when using integral dimensions), since the additional dimension is also irrelevant. If, however, it is rotated about its other axis to produce the redundant filtering task, participants are expected to speed up.

An important question for both of these effects, however, is whether either of these three-dimensional configurations are truly perceived as having three dimensions. It is revealing to consider the possibility that participants continue to process information as though it were coming from only two dimensions, with one of those dimensions now being a combination of the two that vary in perfect correlation with each other. Note that in the standard Garner interpretation (disputed by Melara et al., 1993a), integral stimuli are processed in a dimensionless manner, so the correlated filtering task might be better thought of as another form of the stretch filtering task: still fundamentally two dimensional, but with increased within-category distances.

5.3 MATERIALS

Since one of the goals in this work is to discern which of the aforementioned various factors contribute to the Garner interference effect, these experiments were conducted with stimuli composed of integral dimensions. The most natural choices for three mutually integral dimensions are hue, saturation, and brightness. While early work on integral dimensions focused almost exclusively on color perception, much of the recent work has focused on faces, so a second stimulus set was also used consisting of faces which differed in terms of their weight, age, and gender. These two classes of stimuli provide a useful contrast across degrees of complexity and meaningfulness, and both have been previously shown to be processed in an integral manner.

The color stimuli were chosen using the Munsell color system, which attempts to equate the discriminability of changes in saturation (referred to as *chroma*), brightness (*value*), and hue. In Munsell notation, the stimuli have a chroma of either 4 or 8, a value of either 4 or 6, and a hue of 10B or 7.5PB. These stimuli are shown in Figure 5.2(A). Each of these eight core stimuli gives rise to three "stretched" stimuli by exaggerating each of the three dimensions, creating 24 stretch stimuli in all. These exaggerated values have Munsell coordinates with chroma of 2 or 10, a value of 3 or 7, or a hue of 2.5P or 5B. Examples of these stimuli are shown in Figure 5.2(B).

The face stimuli were generated using the Basel Face Model, a 3-D morphable model based on principal components analysis (Paysan, Knothe, Amberg, Romdhani, & Vetter, 2009). The model is composed of two dissociable components: shape and albedo (color). The shape map is a triangular mesh describing the three dimensional coordinates of a

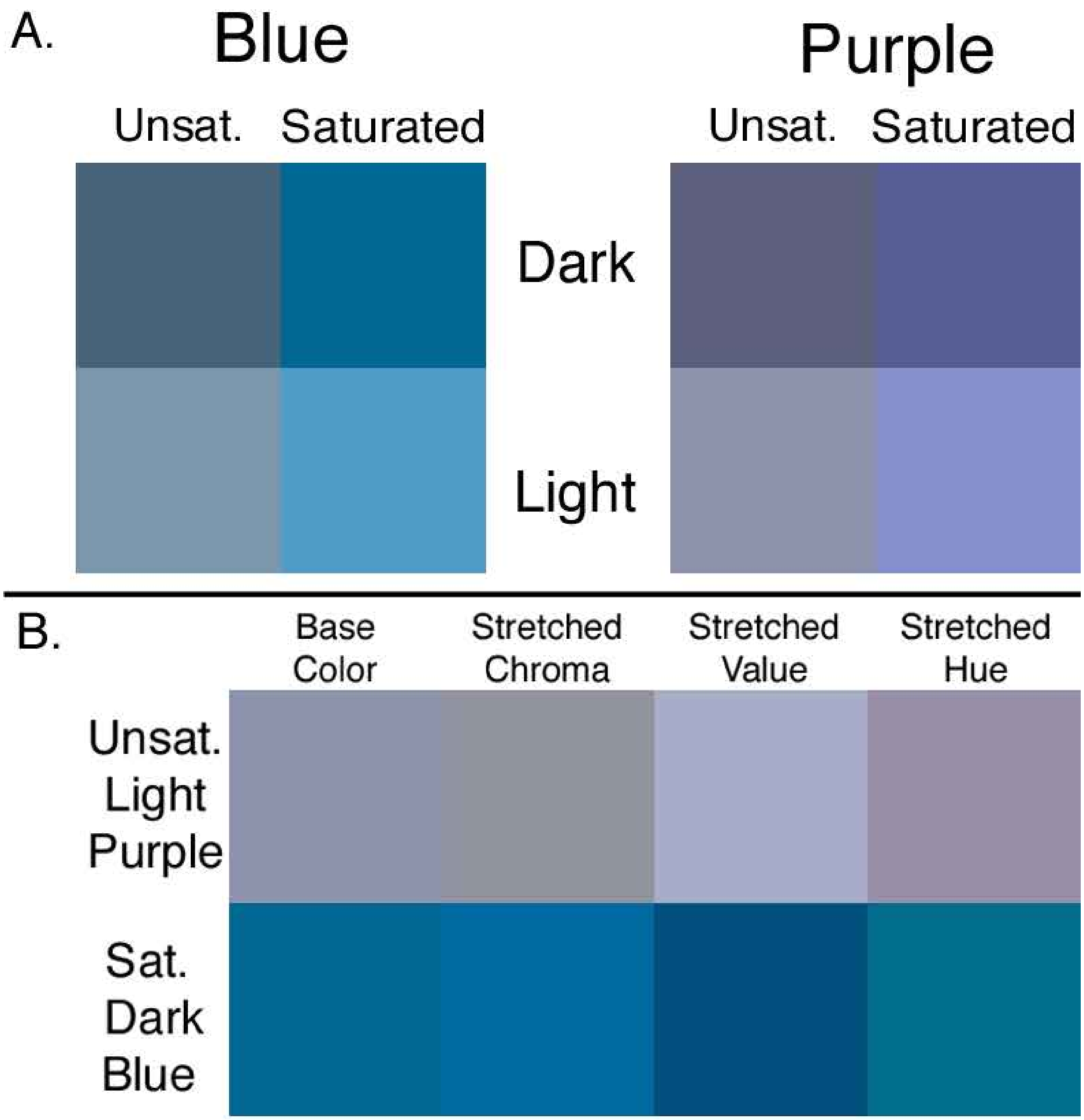


Figure 5.2: (A) The eight color stimuli, differing with respect to saturation (chroma), brightness (value), and hue. (B) Examples of color "stretch" stimuli. Two example colors are shown with versions exaggerated along each of the three dimensions.

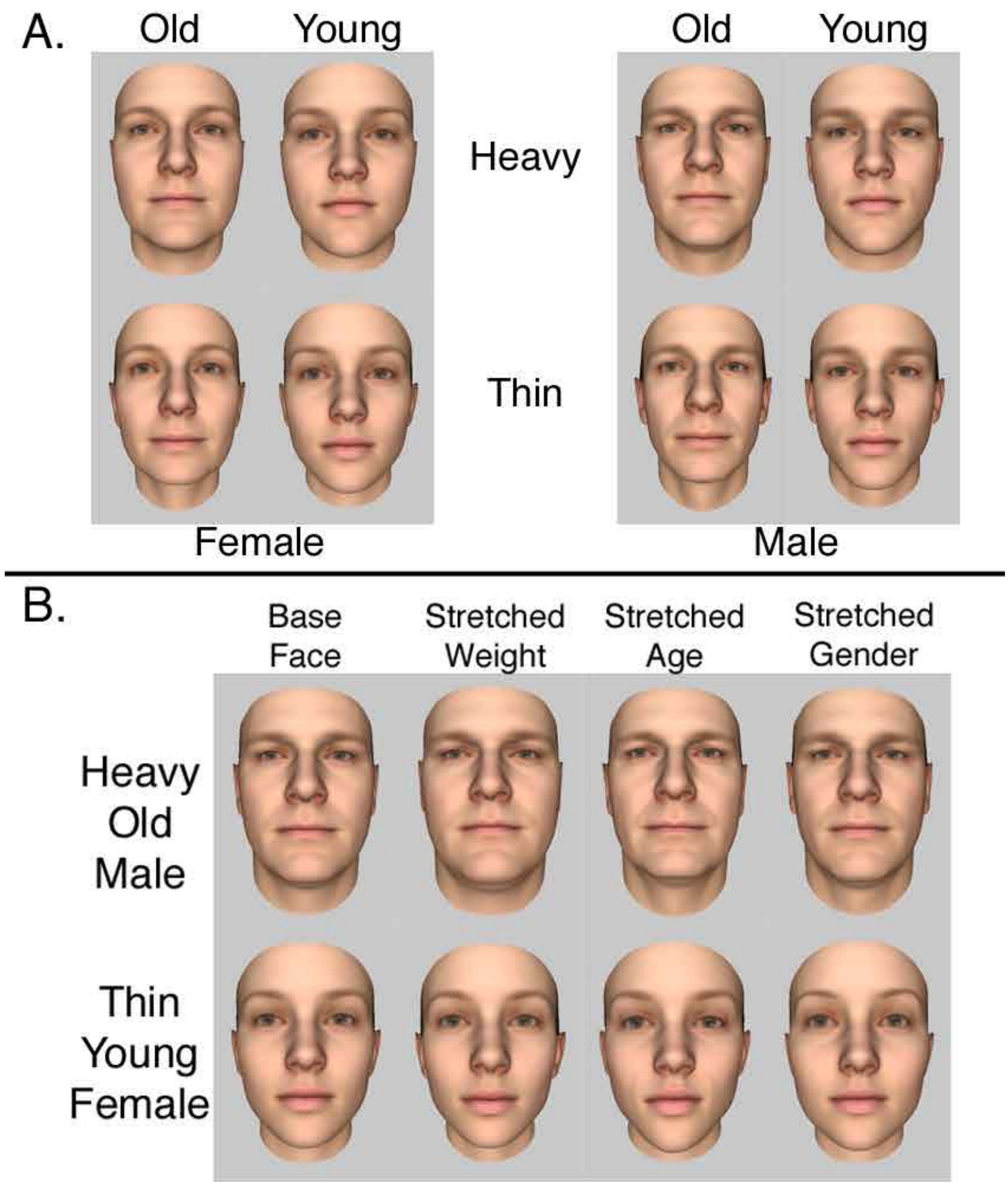


Figure 5.3: (A) The eight face stimuli, differing with respect to weight, age, and gender. (B) Examples of face "stretch" stimuli. Two example faces are shown with versions exaggerated along each of the three dimensions.

given face, while the albedo map describes the reflectance properties of every vertex in the mesh. These two components can then be combined with specifications of vantage point and lighting conditions to generate a two-dimensional image.

To create the model, the authors took three-dimensional, high-resolution laser scans of 200 individuals while simultaneously capturing color images under uniform lighting conditions. Each face was then re-parameterized through a registration procedure to ensure that fiducial points, such as the corner of the eye or mouth, share the same position in the parametrization domain. This yields the desirable property that linear combinations of faces produce other faces. Every face is specified by 53,490 vertices, with data segregated into two vectors. The shape vector specifies the three dimensional Euclidean coordinates for each vertex, while the color vector specifies the *rgb* value of that vertex under uniform lighting conditions. These data sets are treated separately by the model, with each being fit by a separate Principal Components Analysis (PCA).

Each of the 200 hundred faces used to create the model were then labelled with gender, height, weight, and age information. Directions of maximal variance for each of these attributes were then computed for the high-dimensional PCA-space, allowing for the realistic manipulation of these traits. For this experiment I used the dimensions of weight, age, and gender, and the stimuli can be seen in Figure 5.3(A). Unlike the Munsell system, the coefficients of these attribute vectors are defined computationally rather than perceptually, so it took extensive pilot testing to attempt to equate the salience of the three dimensions.

The loading values used were 30 on the weight vector for heavy faces and -18 for thin, 22 on the age vector for old faces and -17 for young, and 1.2 on the gender vector for male faces and -1.3 for female. All of the faces were constructed using the average reflectance map, with the shape determined by adding the appropriate attribute vectors to the average face shape. For the stretch stimuli, examples of which are shown in Figure 5.3(B), the coefficient of the dimension being exaggerated was doubled.

5.4 PROCEDURE

Two versions of this experiment were conducted, one with colors and one with faces, but all procedural details were identical for both stimulus sets. In all, 18 participants were recruited, with 9 classifying colors and the other 9 classifying faces. Each participant completed eight different one hour sessions. The eight days were split into two different groups of tasks, which alternated every day. The first group of tasks comprised the three classic Garner conditions: those shown in the top row of Figure 5.1. During each of the four days spent on these tasks, participants saw 9 blocks of 120 trials each. The first three blocks all used one relevant dimension (one block for each of the three Garner conditions), participants were then told to focus on a different dimension for the next three blocks, and the final dimension was used for the last three. The order in which the dimensions were chosen was randomized across days and participants, as well as the order of the tasks within those groups of three.

Participants thus experienced a block consisting of each of the 12 possible control tasks, four different stimulus pairs for each selected dimension, seeing each stimulus 60 times in each of three contexts. Similarly, every stimulus was shown 60 times in each of the 12 versions of the correlated task: the three possible dimension pairs can be correlated in two different directions and presented at each of two levels on the third dimension. There are only six different versions of the filtering task: each choice of relevant and irrelevant dimensions can be presented at two different levels of the third dimension. These tasks therefore got repeated twice, but since there are four stimuli in each task rather than two, the stimuli were again presented 60 times in each context.

The other four days of testing, which were interspersed with the former, consisted of six blocks: one for each of the novel conditions shown in the bottom two rows of Figure 5.1. For these tasks, participants always focused on a single assigned dimension: three participants always classified stimuli by weight, three by age, three by gender, and likewise for the colors. This was done in order to obtain a sufficient number of trials in each condition without placing too onerous a burden on the participants. The Garner tasks were done with each of the three dimensions in order to have a measure of each participant's dimensional preferences and the existence of Garner interference in each of the six possible dimensional pairs. This was also done because the workload capacity analysis requires measurement of each channel individually, so for instance to measure capacity in the correlated condition where age and gender are both diagnostic, we must have control conditions for both age and gender.

For the novel conditions, there were 176 trials per block. Since there is only one version of the double filtering task, this was repeated four times, and each of the eight stimuli were shown 88 times. The cross correlated task has two different versions, depending on the direction in which the two irrelevant dimensions are correlated, so each variant was repeated twice over the four days. Since it only uses four stimuli per task, once again the stimuli were shown 88 times in each condition. The correlated filtering task has the same structure as the cross correlated task, but due to a coding error the "negatively" correlated conditions were never shown, and instead the participants saw the "positively" correlated stimuli 166 times each.

The redundant filtering task, however, has four variants: two directions of correlation between the relevant dimension and each of the two other dimensions. This implies that the four stimuli were only shown 44 times in each context. The same is true for the stretch filtering task, since the stimuli can be stretched along either of the two irrelevant dimensions at one of two levels of the third dimension. The full stretch condition is the same, but since there are eight stimuli in the task, each individual stimulus was only shown 22 times for a given context.

All trials took place in a dark room, with stimuli shown against a uniform grey background on a 16" Dell Trinitron CRT monitor set to 1024 x 768 pixel resolution with a refresh rate of 75 hz. Subjects were seated 70 cm away from the monitor. Data was collected using DMDX experimental software, which is freely available through Jonathan Forster at the University of Arizona. Responses were input using a custom built response box. Regardless of the particular task, every trial unfolded in exactly the same manner.

First, a fixation cross was displayed for 400 ms, followed by a blank screen that was displayed for a random length of time uniformly distributed between 400 and 700 ms. The stimulus was then displayed in the center of the screen until a response was recorded, with a maximal allowed time of two seconds. Color stimuli were square patches 150 x 150 pixels in size, which equated to 3.8 degrees of visual angle. Face stimuli were 150 x 200 pixels, with the longer dimension spanning 5 degrees of visual angle, though the distance from forehead to chin was still only 3.8 degrees.

Auditory feedback was given on all trials, with different tones denoting correct, incorrect, or slow responses. Participants were explicitly instructed as to which dimension to pay attention to in each block, though not as to which combination of stimuli would be appearing. They were also allowed to rest between blocks of trials.

CHAPTER 6

RESULTS

Six of the original 18 participants withdrew for personal reasons or due to sub-threshold accuracy (less than 80%), and they were replaced to ensure that three participants were assigned to focus on each of the six dimensions (three per stimulus type). Some minimal cleaning was done on the data: the first 16 trials of every block were thrown out, in addition to all trials with reaction times below 300 ms and those that hit the ceiling of 2 seconds. Out of 131,400 total trials, only 37 times were too fast (and these were only 54% accurate), with 506 "timeouts". The remaining trials had an average accuracy of 96% and an average reaction time of 776 ms. The full table of accuracy data, divided by task and dimension, is shown in Table 6.1, with the reaction times for correct responses shown in Table 6.2. Reaction times for the different conditions, averaged across all participants and dimensions, can be seen for color stimuli in Figure 6.1(A) and for face stimuli in Figure 6.1(B).

6.1 BASELINE DISCRIMINABILITY

Before investigating the various planned comparisons between tasks, let us first consider the important question of baseline discriminability. Although extensive piloting was done to attempt to equate the difficulty of each triad of dimensions, individual differences in

	Colors				Faces			
	Satur.	Bright.	Hue	Ave.	Weight	Age	Gender	Ave.
Control	97.3	94.8	95.2	95.2	96.6	94.7	94.5	95.1
Correlated	95.5	98	95.9	96.3	95.3	97.3	95.9	95.9
Filtering	92.7	95.4	95.6	93.9	95.6	94.4	96.9	95.1
Redundant Filtering	95.9	98.5	98.3	97.6	97.7	96.8	98.1	97.5
Correlated Filtering	93.2	98.1	97.2	96.2	96.9	95.3	97.7	96.6
Cross Correlated	94.8	97.4	98.4	96.9	96.6	95.6	98.4	96.9
Double Filtering	93.7	97.6	97.2	96.2	95	94.5	98.1	95.9
Stretch Filtering	91.3	97.3	95.4	94.7	96.3	94.4	97.8	96.1
Full Stretch	92.4	97.8	96.3	95.5	95.6	93.7	97.7	95.6
Average	94	97.3	96.7	95.5	96.2	95.2	97.3	95.6

Table 6.1: Average accuracy percentages across participants by task and dimension.

	Colors				Faces			
	Satur.	Bright.	Hue	Ave.	Weight	Age	Gender	Ave.
Control	744	618	647	686	658	803	858	781
Correlated	774	599	624	672	685	796	810	769
Filtering	837	668	677	741	705	830	882	817
Redundant Filtering	840	608	624	688	673	764	754	730
Correlated Filtering	851	600	648	696	709	820	793	774
Cross Correlated	870	612	645	705	700	808	783	764
Double Filtering	849	626	639	701	729	805	825	787
Stretch Filtering	860	624	701	725	742	860	829	810
Full Stretch	867	629	676	720	711	841	777	776
Average	837	620	654	720	703	814	809	776

Table 6.2: Reaction times (in milliseconds) for correct trials, averaged across participants for each task and each dimension.

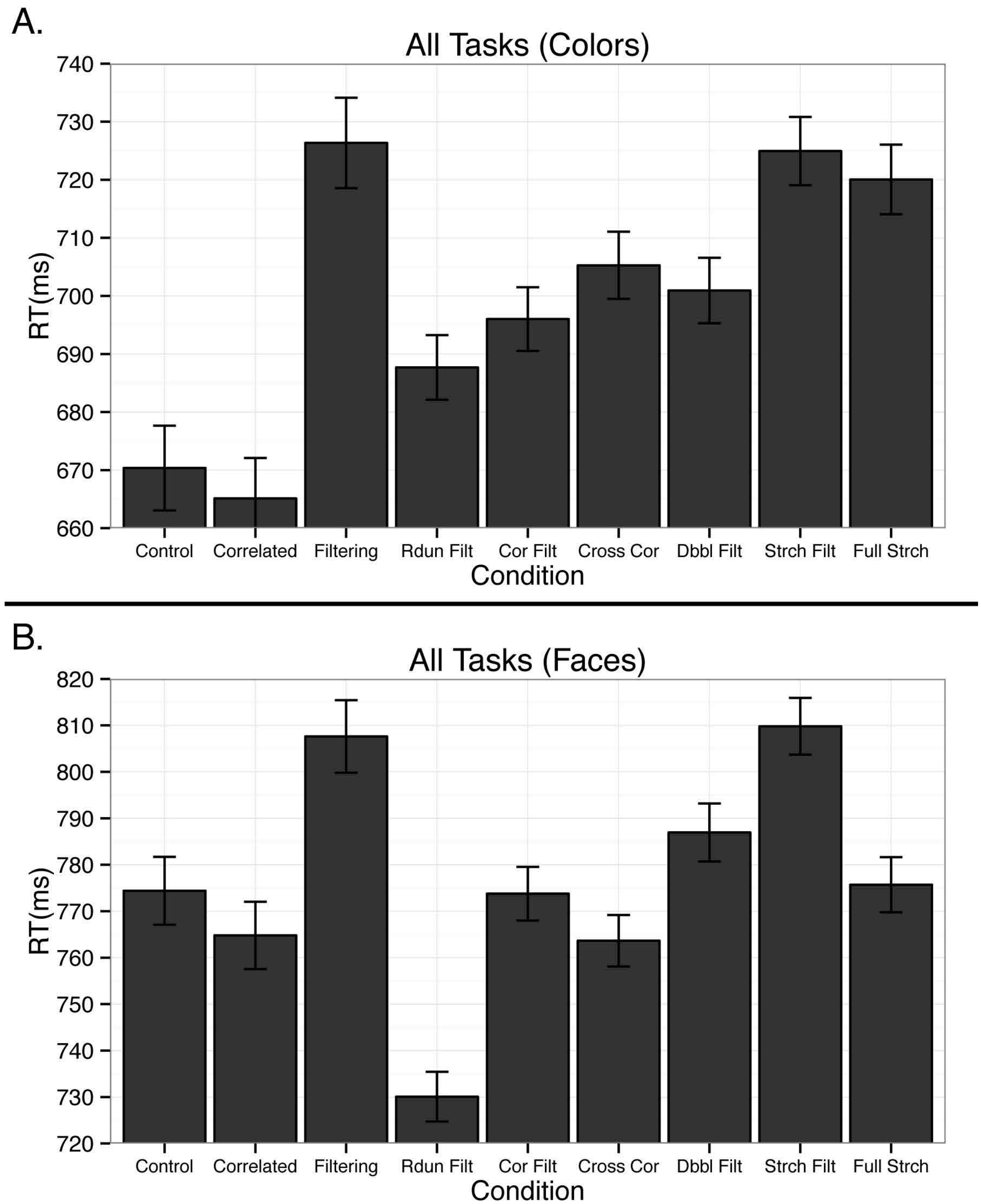


Figure 6.1: Average reaction times for each task using colors (A) and faces (B)

discriminability are pervasive, and interactions between dimensions mean the stimulus space is not a proper cube. Figure 6.2 shows reaction times for each stimulus, averaged across participants and tasks. Colors, shown in (A), are abbreviated as **S**aturated or **U**nsaturated, **D**ark or **L**ight, and **B**lue or **P**urple. Faces, shown in (B), are **H**eavy or **T**hin, **O**ld or **Y**oung, and **F**emale or **M**ale. For both stimulus sets, certain stimuli are significantly faster or slower than others, indicating congruence effects between the dimensions. These interactions were formally tested for each stimulus set using a three way ANOVA between the dimensions, with a random effect of participant.

For the color stimuli, there were main effects of both saturation and hue. Saturated colors were slower by 33.08 ms, with $p < .001$. Purple colors were slower by 32.43 ms, with $p < .001$. There was a significant interaction between brightness and hue, $p < .001$, with dark colors being 31.74 ms faster than light when they are blue, but 24.81 ms slower when purple. The final effect to reach significance was the interaction between saturation and brightness, $p < .05$, in which saturated colors were slower by 42.27 ms when dark, and only 24.05 ms slower when light.

For faces, the level of weight was significant, $p < .05$, with heavy faces taking 9.54 milliseconds longer on average. The influence of age was even stronger, $p < .001$, with old faces on average 38.13 milliseconds slower than young faces. There was also a significant two way interaction between age and gender, $p < .001$, where the difference between old and young faces was only 6.95 milliseconds with male faces, and 69.58 ms for female faces.

Although the effect of which dimension was being judged was highly significant ($p < .001$) for both stimulus sets, Figures 6.3 and 6.4 show that differences between different blocks using the same dimension were often greater than differences between dimensions, especially for the color stimuli. Group data are shown in (A), averaged across the two stimuli used in any one task. Individual data are shown in (B), averaged across both the two stimuli per task and four tasks per relevant dimension. The individual participant plots also show that participants showed different patterns of dimensional dominance, reaffirming the importance of within-subject comparisons.

6.2 TRADITIONAL GARNER TESTS

Reaction times for the three traditional Garner conditions (control, correlated, and filtering) are shown in Figure 6.5 for color stimuli, and Figure 6.6 for faces. In the top panels, Figures 6.5(A) and 6.6(A), the conditions are labeled according to the relevant dimensions: Saturation, Brightness, and Hue for colors, and Weight, Age, and Gender for faces. The label "AG-" denotes a negative correlation between Age and Gender, with "SxH" meaning Hue varies irrelevantly while participants judge Saturation. For the individual participant plots, Figures 6.5(B) and 6.6(B), the nine participants are grouped by row according to the dimension they focused on. "+ Crltd" represents the condition where their assigned dimension was positively correlated with the "next" dimension (e.g. weight/age or gender/weight). Similarly, "Filt +1" is when the "next" dimension is irrelevant, where as "Filt +2" requires filtering out the "previous" dimension.

Group level tests were conducted by dimension using the data from all nine participants using each stimulus set, since all participants completed a version of the traditional tasks using each of the three dimensions in turn. Repeated measures anova tests were run using task as a fixed effect and participant as a random effect. Due to strong individual differences present in the data, t-tests were also conducted for each participant's data.

Table 6.3 summarizes these results by displaying the number of participants showing statistically significant ($p < .05$) differences between the control and filtering tasks. If the data was significant *and* contrary to expectations, in that the control task was *slower* than the filtering task, the result is notated in the "Negative" column, and in the rare event that

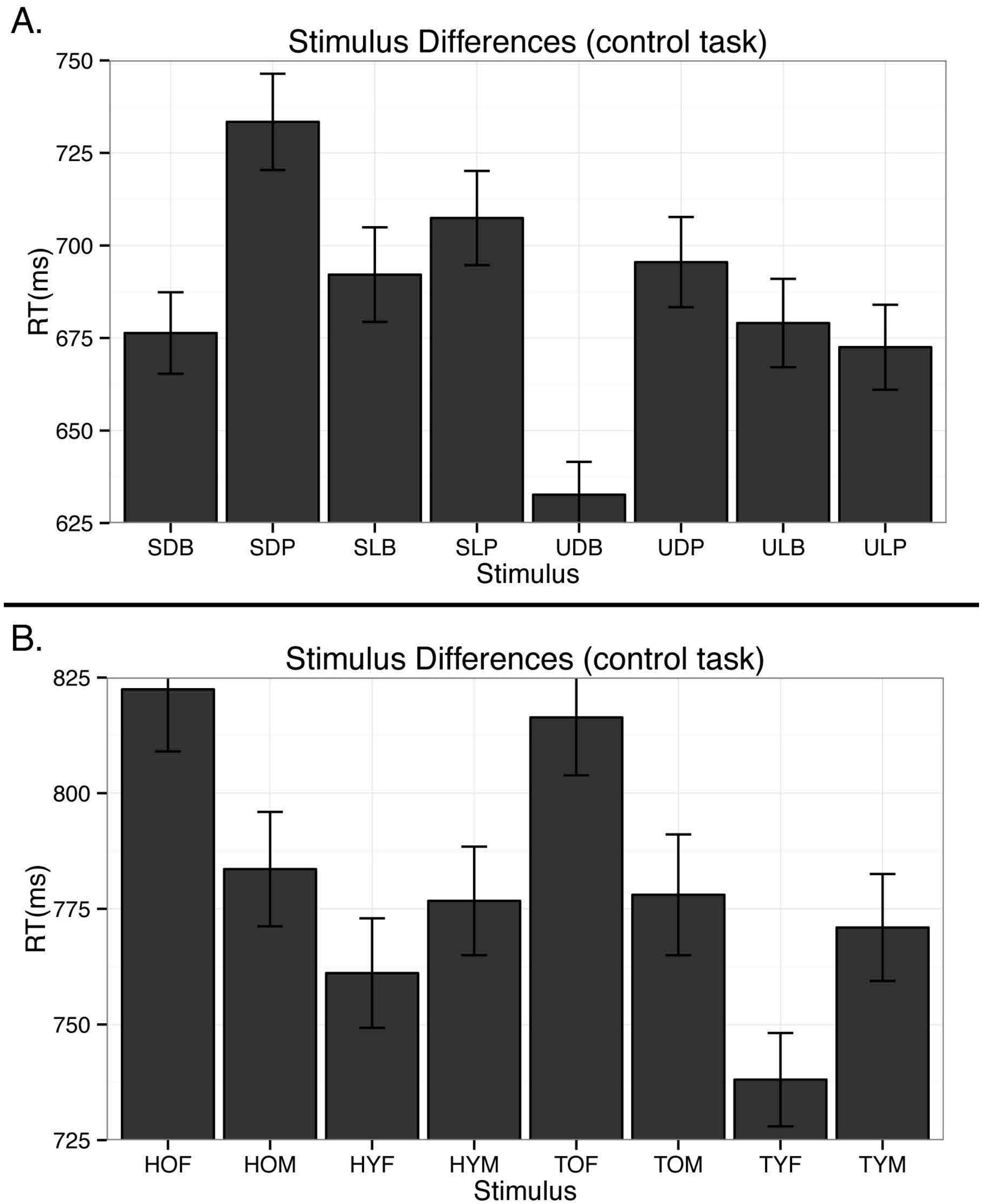


Figure 6.2: Reaction times in control tasks for each stimulus, averaged across participants and conditions. Color stimuli are shown in (A), with faces in (B).

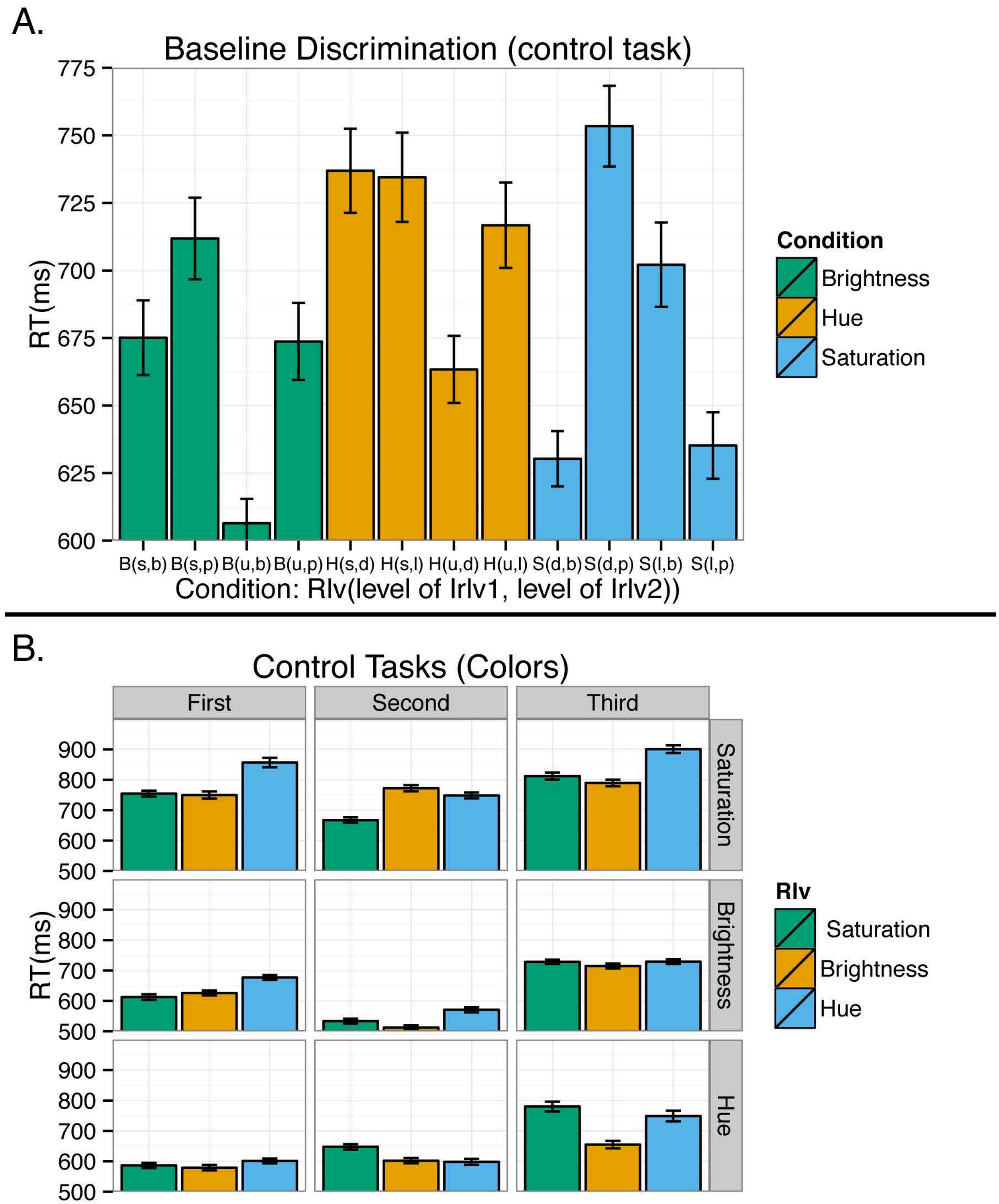
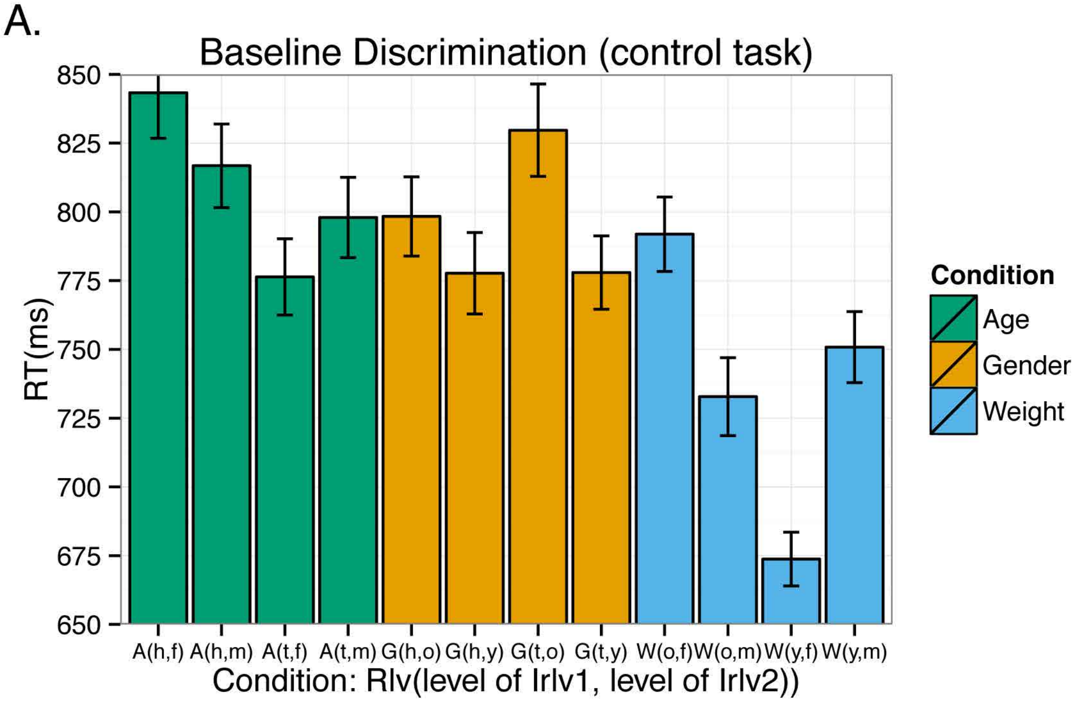


Figure 6.3: Reaction times for the various controls tasks for colors.

A.



B.

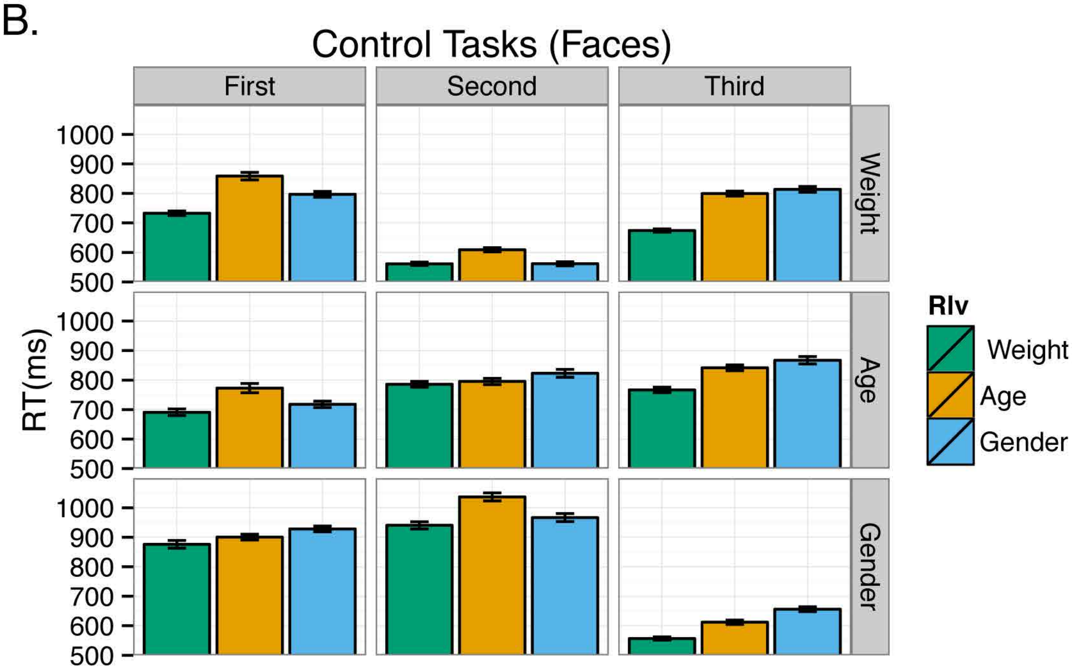


Figure 6.4: Reaction times for the various controls tasks for faces.

Relevant	Irrelevant	Positive	Negative	Speed/Acc.	Not Sig.	Group
Saturation	Brightness	8	0	0	1	Positive
Saturation	Hue	9	0	0	0	Positive
Brightness	Hue	3	4	1	1	Not Sig.
Brightness	Saturation	9	0	0	0	Positive
Hue	Saturation	6	1	0	2	Positive
Hue	Brightness	4	0	0	5	Positive
Weight	Age	5	0	0	4	Positive
Weight	Gender	7	1	0	1	Positive
Age	Gender	3	3	0	3	Not Sig.
Age	Weight	7	1	0	1	Positive
Gender	Weight	3	1	1	4	Positive
Gender	Age	2	4	0	3	Not Sig.

Table 6.3: Number of individual participants reporting statistically significant results (or otherwise) when comparing filtering and control tasks (Garner interference).

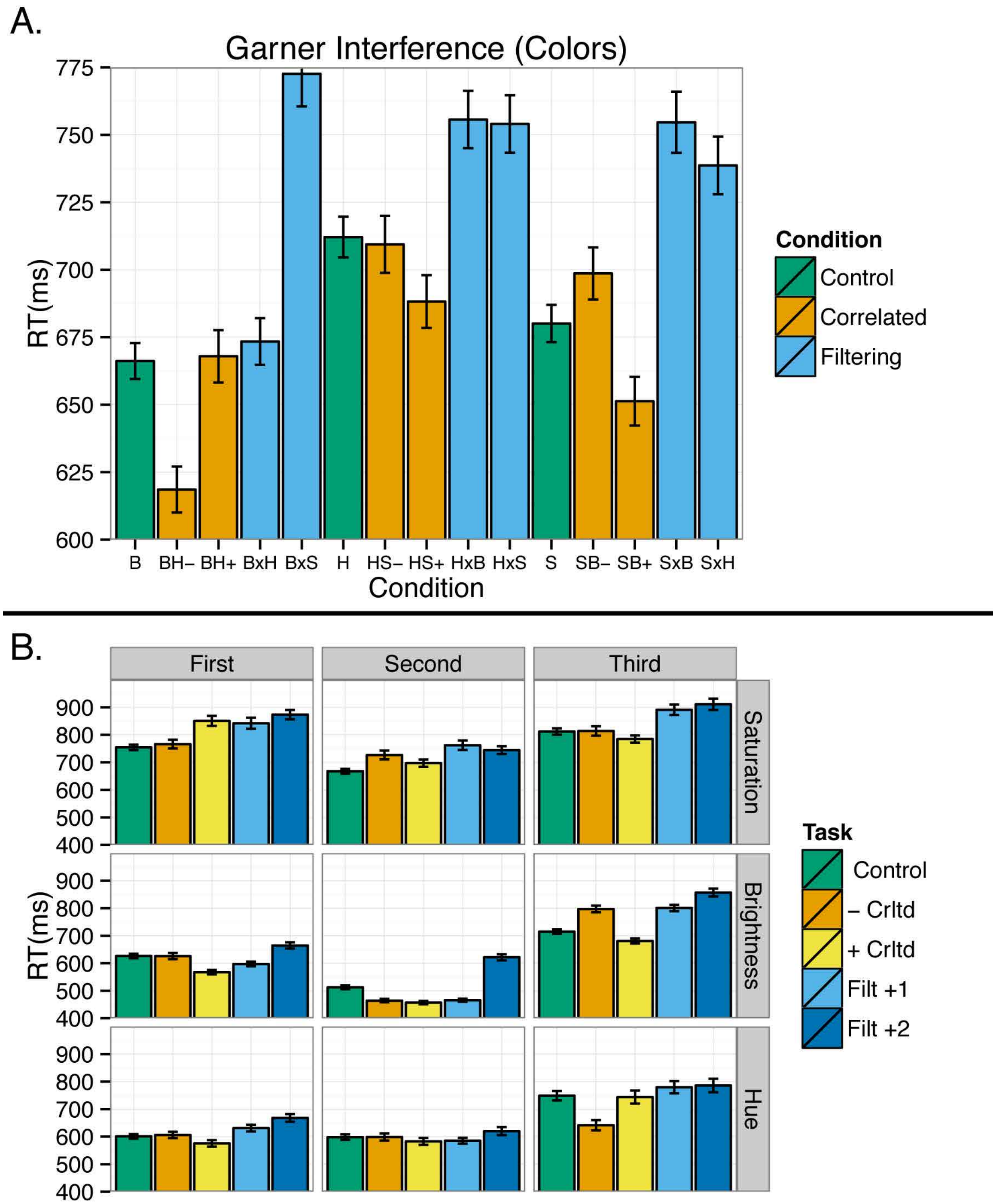


Figure 6.5: Reaction times from traditional Garner conditions with color stimuli averaged across participants (A) and for each individual (B).

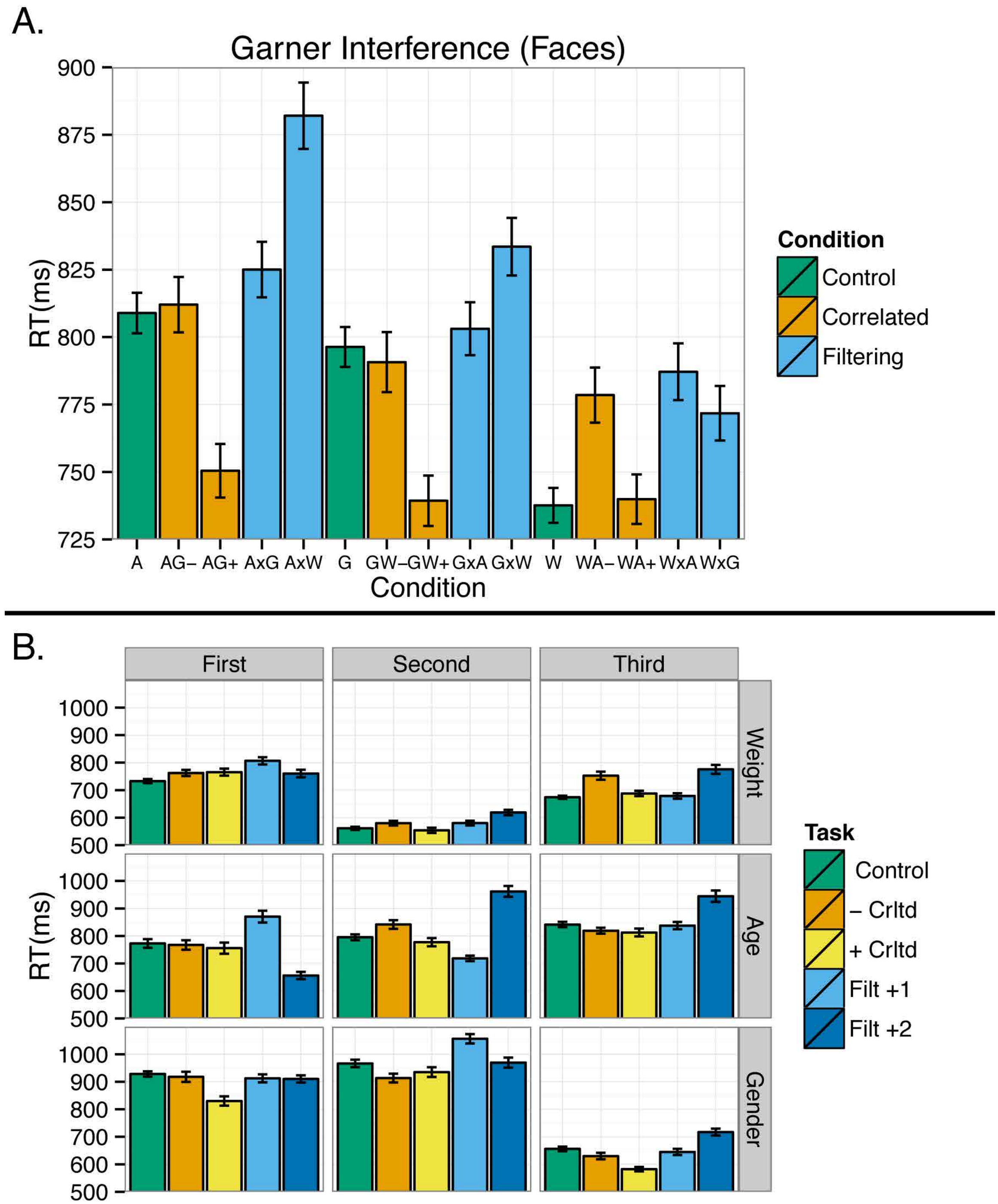


Figure 6.6: Reaction times from traditional Garner conditions with face stimuli averaged across participants (A) and for each individual (B).

Relevant	Correlated	Direction	Positive	Negative	Speed/Acc.	Not Sig.	Group
Saturation	Brightness	+	4	1	0	4	Positive
Saturation	Brightness	-	0	4	0	5	Negative
Brightness	Hue	+	5	0	3	1	Not Sig.
Brightness	Hue	-	7	1	0	1	Positive
Hue	Saturation	+	6	1	1	1	Positive
Hue	Saturation	-	2	2	0	5	Not Sig.
Weight	Age	+	3	0	1	5	Not Sig.
Weight	Age	-	1	6	0	2	Negative
Age	Gender	+	8	0	0	1	Positive
Age	Gender	-	4	2	0	3	Not Sig.
Gender	Weight	+	8	0	0	1	Positive
Gender	Weight	-	1	3	0	5	Not Sig.

Table 6.4: Number of individual participants reporting statistically significant results (or otherwise) when comparing correlated and control tasks (redundancy gain).

both speed and accuracy differences were significant *and in opposite directions*, results are considered a speed/accuracy tradeoff. As can be seen, findings of Garner interference were rather robust, with the main exceptions being filtering hue while reporting brightness, and either ignoring age in favor of gender or vice versa. More detailed information about levels of both statistical and practical significance can be seen in Appendix A, Figures A.1-A.6.

Results for redundancy gains, the comparisons between the control and correlated tasks, are shown in Table 6.4. As a point of clarification, "positive" and "negative" directions of correlation are only definitional, since these are all categorical dimensions. The "positive" label was used when pairing the values of unsaturated, dark, or blue for the color stimuli, and heavy, old, or male for the faces. Only around half of the dimensional pairings showed consistent gains for participants. Notably, there were consistent redundancy *losses* (control was faster than correlated) for two conditions: when unsaturated colors were light, or when heavy faces were old. Once again more detailed information about levels of both statistical and practical significance can be found in Appendix A, Figures A.7-A.12.

Capacity analyses were also done to further test redundancy gains. The $C(t)$ functions are shown in Figure 6.7. As can be seen, these functions are all fairly limited in capacity, laying close to the dashed line indicating fixed capacity. For each stimulus set, there are two conditions which lie below this threshold for almost all values of reaction time, indicating that the correlated trials were slower than the average of the corresponding control trials. These pairs were the same ones with the most reported instances of significant redundancy losses: when unsaturated colors were light, blue colors were saturated, female faces were heavy, or when heavy faces were old. The only pairing to show super-capacity for an extended range of reaction times was for when male faces were old, and that was only for the fastest responses.

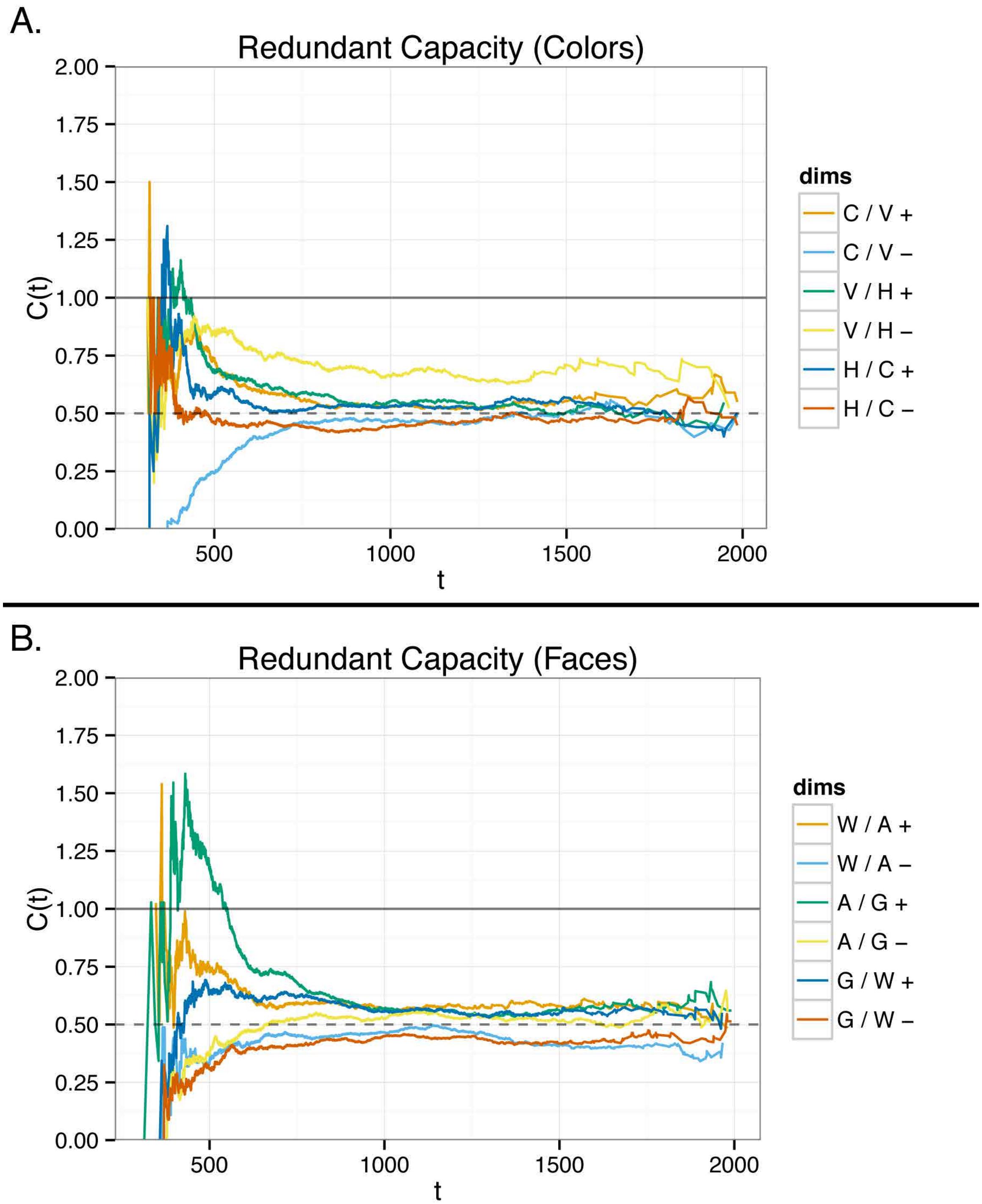


Figure 6.7: Workload capacity functions comparing redundant and control trials for colors (A) and faces (B).

6.3 THREE-DIMENSIONAL GARNER FILTERING

The comparison between the filtering and the double filtering tasks can be seen as a three-dimensional analog of the Garner interference test, where the third dimension was fixed at one level in the former and allowed to vary in the latter. As mentioned previously, the correlated filtering task serves as a perfect middle ground capable of dissociating between the previously confounded effects of a change in the number of stimuli and a change in the number of irrelevant dimensions. Reaction times for all three tasks are shown in Figures 6.8 and 6.9. As also mentioned previously, a coding error resulted in the correlated filtering task only being run with one direction of correlation between irrelevant dimensions. For unbiased comparisons, therefore, these analyses were performed with using only that subset of four stimuli for all three tasks.

Similar to the analysis of the traditional Garner tasks, group level tests were conducted using repeated measures anova tests with task as a fixed effect and participant as a random effect. Since these comparisons use the novel tasks that were only done using the assigned dimension for each participant, there are only three participants to be grouped for each comparison. T-tests were also conducted using each participant's data individually.

As can be seen, 3-D Garner interference was rarely positively significant and frequently negatively significant, meaning that the double filtering task was actually *faster* than the standard filtering task. The comparison between correlated filtering and standard filtering, to establish the effect of the number of irrelevant dimensions, showed similar patterns to the double filtering trials, in that the test was almost never significantly positive and frequently significantly negative. The test for the influence of the number of stimuli, comparing correlated filtering to double filtering, was rarely significant and showed no consistent trends.

Relevant	Test	Positive	Negative	Speed/Acc.	Not Sig.	Group
Saturation	3D GI	1	0	0	2	Not Sig.
Brightness	3D GI	0	3	0	0	Negative
Hue	3D GI	1	2	0	0	Negative
Weight	3D GI	1	1	0	1	Positive
Age	3D GI	1	1	0	1	Negative
Gender	3D GI	0	2	0	1	Negative
Saturation	Stimuli	0	0	0	3	Not Sig.
Brightness	Stimuli	1	0	0	2	Positive
Hue	Stimuli	0	0	0	3	Not Sig.
Weight	Stimuli	1	0	0	2	Positive
Age	Stimuli	0	1	0	2	Not Sig.
Gender	Stimuli	1	1	0	1	Positive
Saturation	Irlv. Dims.	1	1	0	1	Not Sig.
Brightness	Irlv. Dims.	0	3	0	0	Negative
Hue	Irlv. Dims.	0	1	0	2	Negative
Weight	Irlv. Dims.	0	0	0	3	Not Sig.
Age	Irlv. Dims.	1	1	0	1	Not Sig.
Gender	Irlv. Dims.	0	2	0	1	Negative

Table 6.5: Number of individual participants reporting statistically significant results (or otherwise) for 3D Garner interference tests.

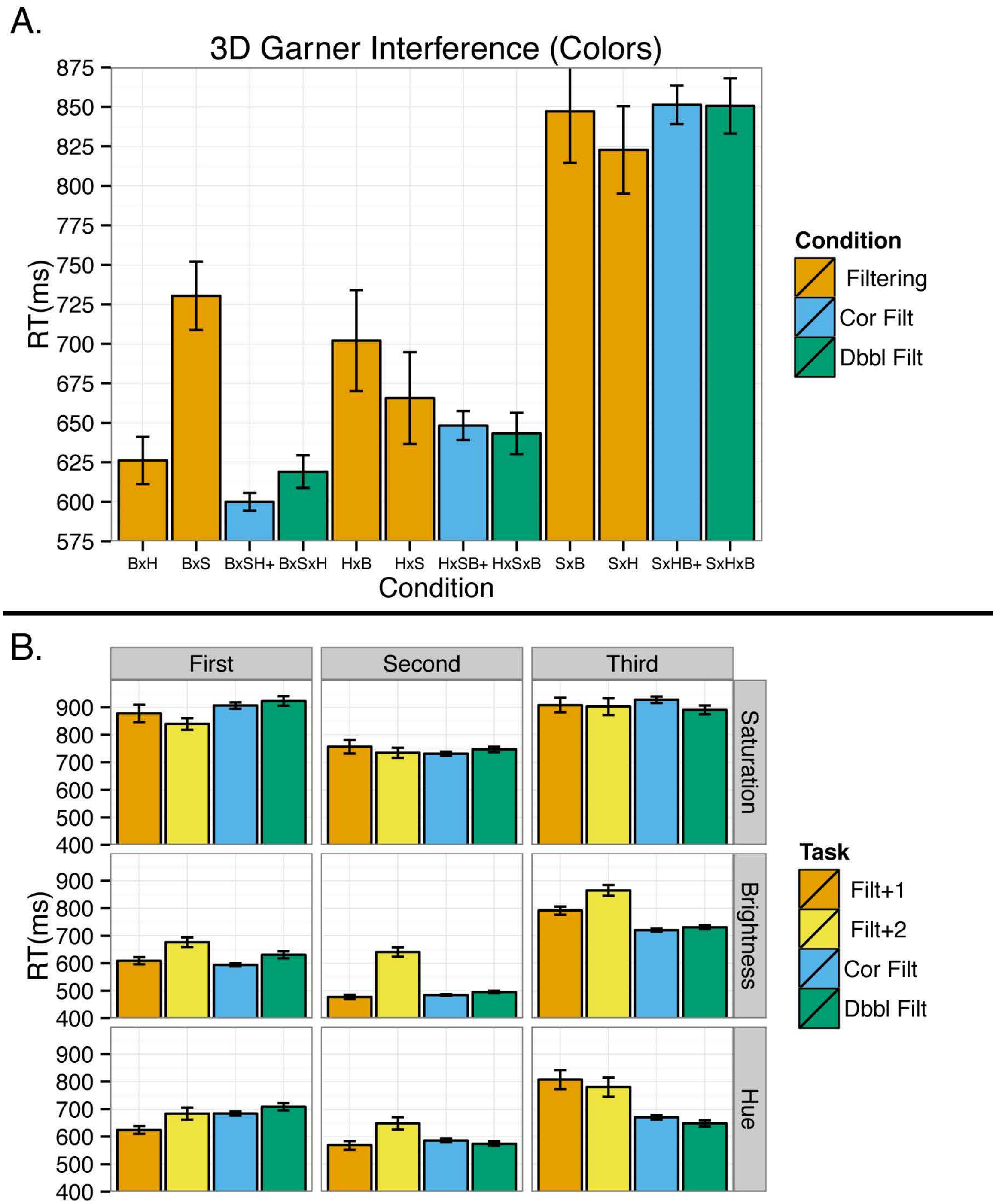


Figure 6.8: Reaction times for 3-D Garner interference with color stimuli averaged by task (A) and for individual participants (B).

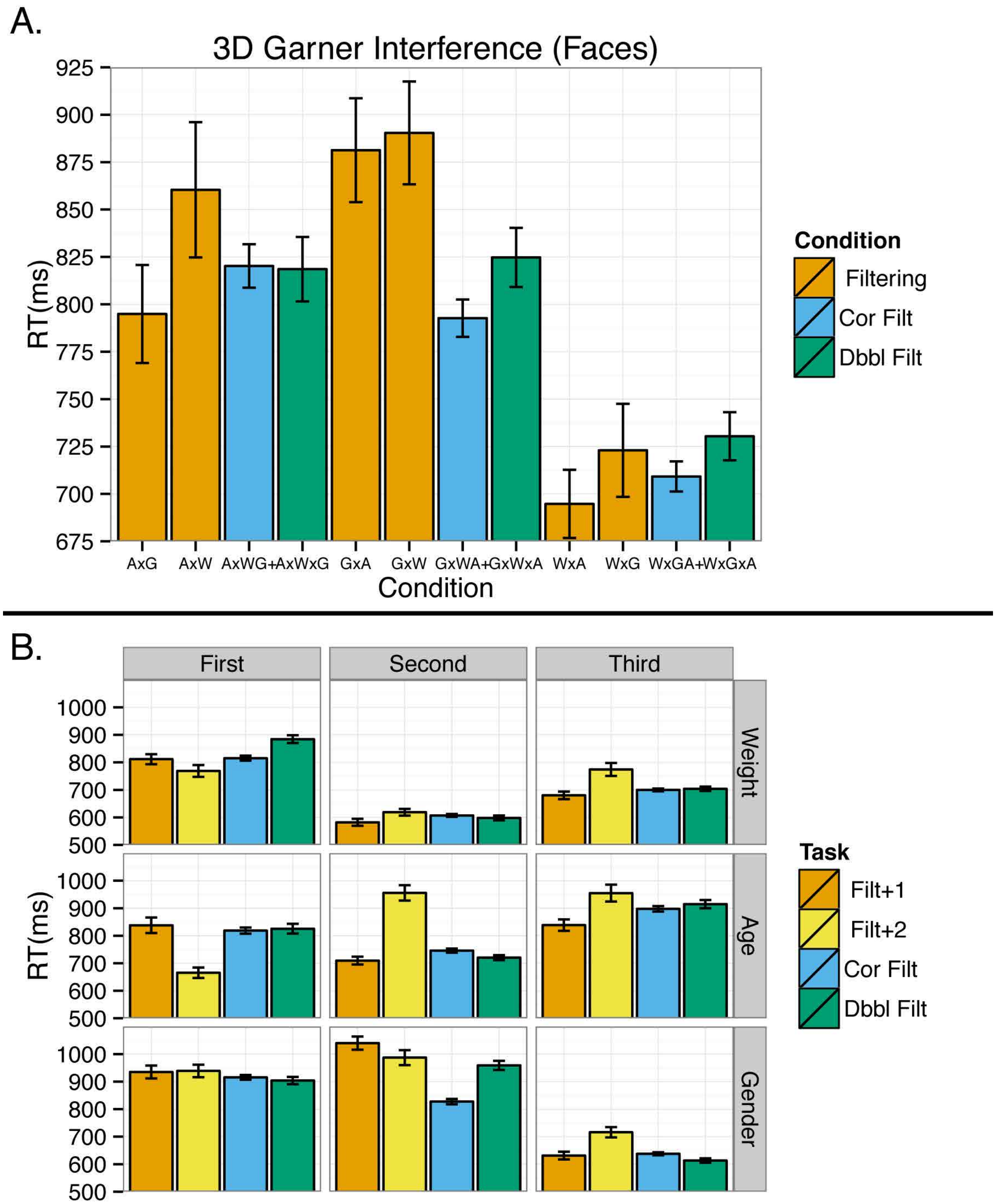


Figure 6.9: Reaction times for 3-D Garner interference with face stimuli averaged by task (A) and for individual participants (B).

Relevant	Irrelevant	Test	Positive	Negative	Speed/Acc.	Not Sig.	Group
Saturation	Brightness	Irlv. Dist.	1	0	0	2	Positive
Saturation	Hue	Irlv. Dist.	1	0	0	2	Not Sig.
Brightness	Hue	Irlv. Dist.	1	1	0	1	Negative
Brightness	Saturation	Irlv. Dist.	0	3	0	0	Negative
Hue	Saturation	Irlv. Dist.	2	0	0	1	Positive
Hue	Brightness	Irlv. Dist.	1	1	0	1	Not Sig.
Saturation	Brightness	Ext. Stim.	2	0	0	1	Positive
Saturation	Hue	Ext. Stim.	1	0	0	2	Not Sig.
Brightness	Hue	Ext. Stim.	1	2	0	0	Negative
Brightness	Saturation	Ext. Stim.	0	3	0	0	Negative
Hue	Saturation	Ext. Stim.	1	1	1	0	Positive
Hue	Brightness	Ext. Stim.	1	2	0	0	Negative
Saturation	Brightness	Int. Stim.	1	1	0	1	Not Sig.
Saturation	Hue	Int. Stim.	0	0	1	2	Not Sig.
Brightness	Hue	Int. Stim.	1	2	0	0	Not Sig.
Brightness	Saturation	Int. Stim.	1	1	0	1	Not Sig.
Hue	Saturation	Int. Stim.	0	2	0	1	Negative
Hue	Brightness	Int. Stim.	0	2	0	1	Negative

Table 6.6: Number of individual participants reporting statistically significant results (or otherwise) for the stretched color stimuli.

6.4 STRETCH FILTERING

The stretched stimuli were used for several comparisons, with results detailed in Tables 6.6 and 6.7. The first comparison examines the effect of increased distance along the irrelevant dimension by contrasting reaction times in the standard and stretched filtering conditions, with the expectation that increased distance will lead to poorer performance. The data are inconclusive, with roughly equal numbers of positively and negatively significant results, for both faces and colors.

The other two comparisons both concern the number of stimuli used. The full stretch condition, which uses all eight stimuli, was compared with the standard filtering condition to observe the effect of adding "exterior" stimuli to the stretched condition, and also compared with the stretch filtering task to measure the effect of adding "interior" stimuli. The expectation was that more stimuli would produce slower reactions, but both effects were more often negative, meaning that the full stretch condition was often faster than the other two. This effect was especially pronounced when adding exterior stimuli with face stimuli. Reaction time plots for all three tasks can be seen in Figures 6.10 and 6.11.

6.5 REDUNDANT FILTERING

In the same way that the double filtering condition created a three-dimensional analog to the Garner filtering test, the redundant filtering condition supplies a 3-D analog to the test for redundancy gains. As can be seen by the significance results in Table 6.8, the redundant filtering task was reliably faster than the standard filtering task. Reaction time differences were significantly positive for the majority of participants in all conditions except when participants focused on saturation, where the group level results were non-significant.

Capacity analyses were also done to further test redundancy gains in the presence of

Relevant	Irrelevant	Test	Positive	Negative	Speed/Acc.	Not Sig.	Group
Weight	Age	Irlv. Dist.	3	0	0	0	Positive
Weight	Gender	Irlv. Dist.	0	1	0	2	Negative
Age	Gender	Irlv. Dist.	1	0	0	2	Positive
Age	Weight	Irlv. Dist.	1	1	0	1	Not Sig.
Gender	Weight	Irlv. Dist.	0	1	1	1	Negative
Gender	Age	Irlv. Dist.	0	1	0	2	Negative
Weight	Age	Ext. Stim.	1	0	0	2	Positive
Weight	Gender	Ext. Stim.	0	1	1	1	Negative
Age	Gender	Ext. Stim.	2	0	0	1	Not Sig.
Age	Weight	Ext. Stim.	2	1	0	0	Not Sig.
Gender	Weight	Ext. Stim.	0	3	0	0	Negative
Gender	Age	Ext. Stim.	0	2	0	1	Negative
Weight	Age	Int. Stim.	0	2	1	0	Negative
Weight	Gender	Int. Stim.	0	1	0	2	Not Sig.
Age	Gender	Int. Stim.	0	1	0	2	Negative
Age	Weight	Int. Stim.	1	1	0	1	Not Sig.
Gender	Weight	Int. Stim.	0	3	0	0	Negative
Gender	Age	Int. Stim.	0	0	0	3	Not Sig.

Table 6.7: Number of individual participants reporting statistically significant results (or otherwise) for the stretched face stimuli.

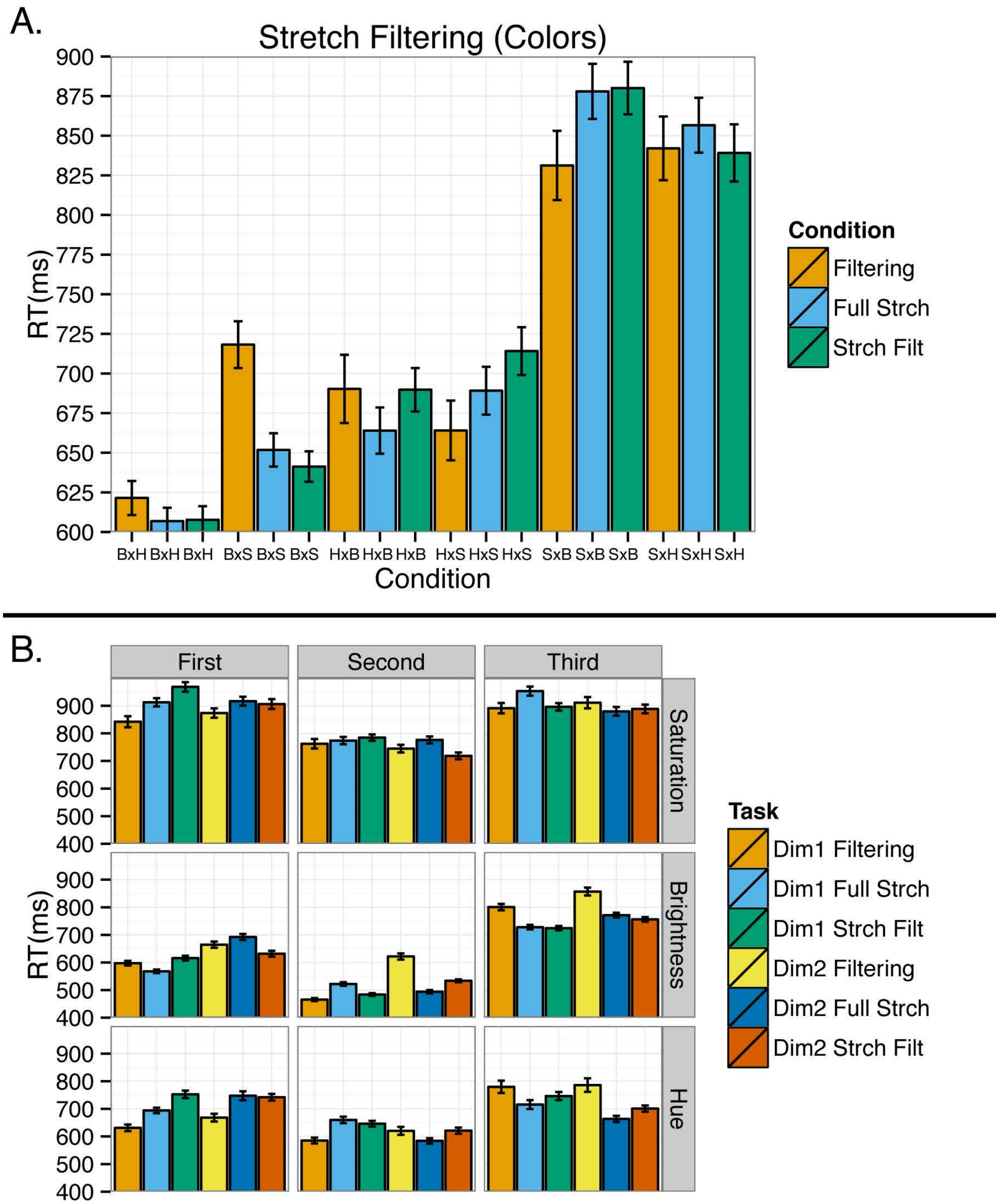


Figure 6.10: Reaction times for stretched conditions with color stimuli averaged by task (A) and for individual participants (B).

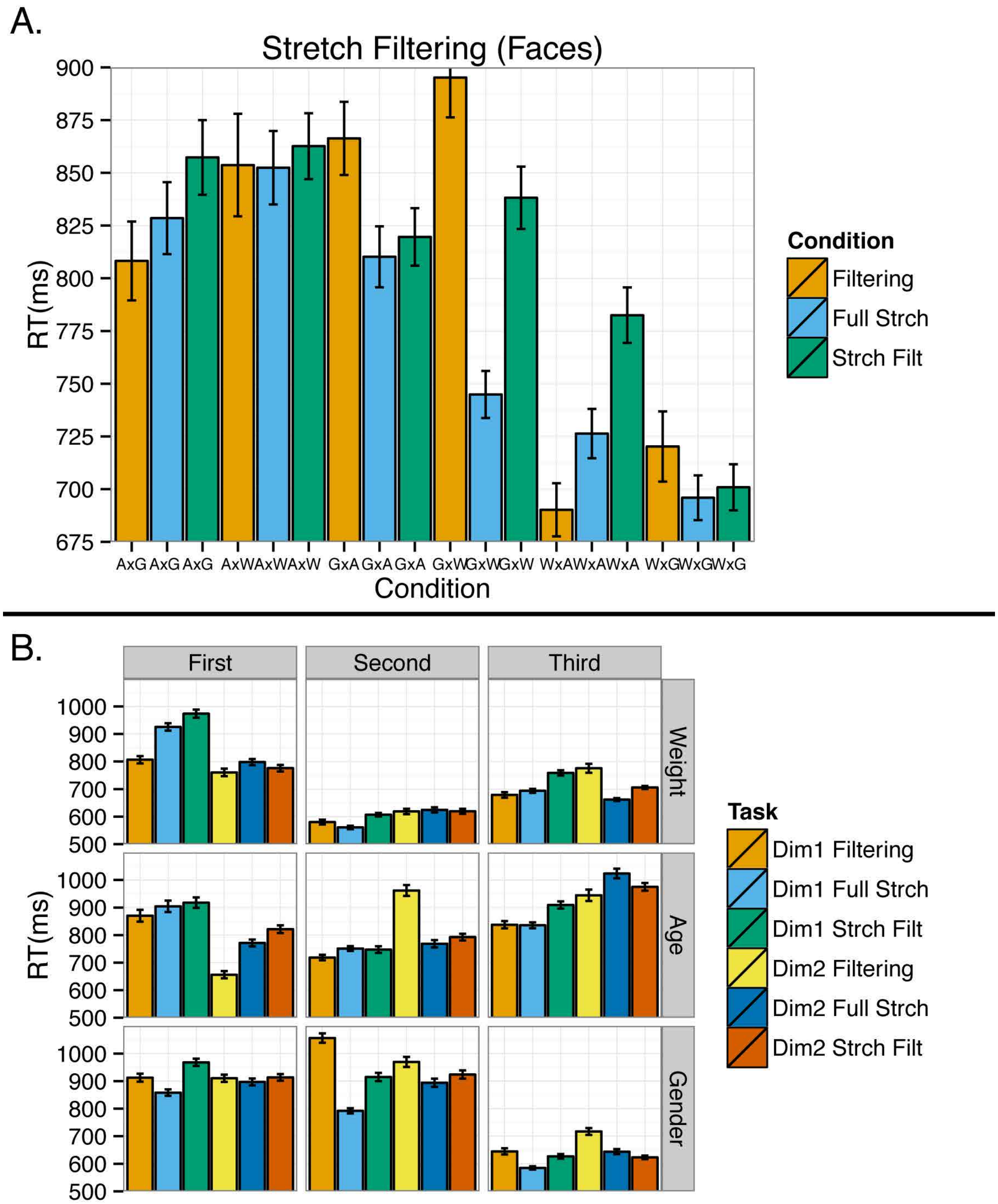


Figure 6.11: Reaction times for stretched conditions with face stimuli averaged by task (A) and for individual participants (B).

Relevant	Correlated	Direction	Positive	Negative	Speed/Acc.	Not Sig.	Group
Saturation	Brightness	+	1	1	0	1	Not Sig.
Saturation	Brightness	-	1	1	1	0	Not Sig.
Brightness	Hue	+	3	0	0	0	Positive
Brightness	Hue	-	3	0	0	0	Positive
Hue	Saturation	+	3	0	0	0	Positive
Hue	Saturation	-	3	0	0	0	Positive
Weight	Age	+	2	1	0	0	Positive
Weight	Age	-	2	0	0	1	Positive
Age	Gender	+	2	0	0	1	Positive
Age	Gender	-	2	1	0	0	Not Sig.
Gender	Weight	+	3	0	0	0	Positive
Gender	Weight	-	2	0	0	1	Positive

Table 6.8: Number of individual participants reporting statistically significant results (or otherwise) when comparing redundant filtering and standard filtering tasks (3D redundancy gain).

Relevant	Correlated	Direction	Positive	Negative	Speed/Acc.	Not Sig.	Group
Saturation	Brightness	+	1	0	0	2	Positive
Saturation	Brightness	-	1	0	1	1	Positive
Brightness	Hue	+	2	1	0	0	Positive
Brightness	Hue	-	1	1	0	1	Not Sig.
Hue	Saturation	+	1	0	0	2	Not Sig.
Hue	Saturation	-	0	3	0	0	Not Sig.
Weight	Age	+	1	2	0	0	Not Sig.
Weight	Age	-	0	2	1	0	Not Sig.
Age	Gender	+	0	2	0	1	Negative
Age	Gender	-	2	1	0	0	Positive
Gender	Weight	+	0	3	0	0	Negative
Gender	Weight	-	0	3	0	0	Negative

Table 6.9: Number of individual participants reporting statistically significant results (or otherwise) when comparing redundant filtering and correlated tasks (redundant Garner interference).

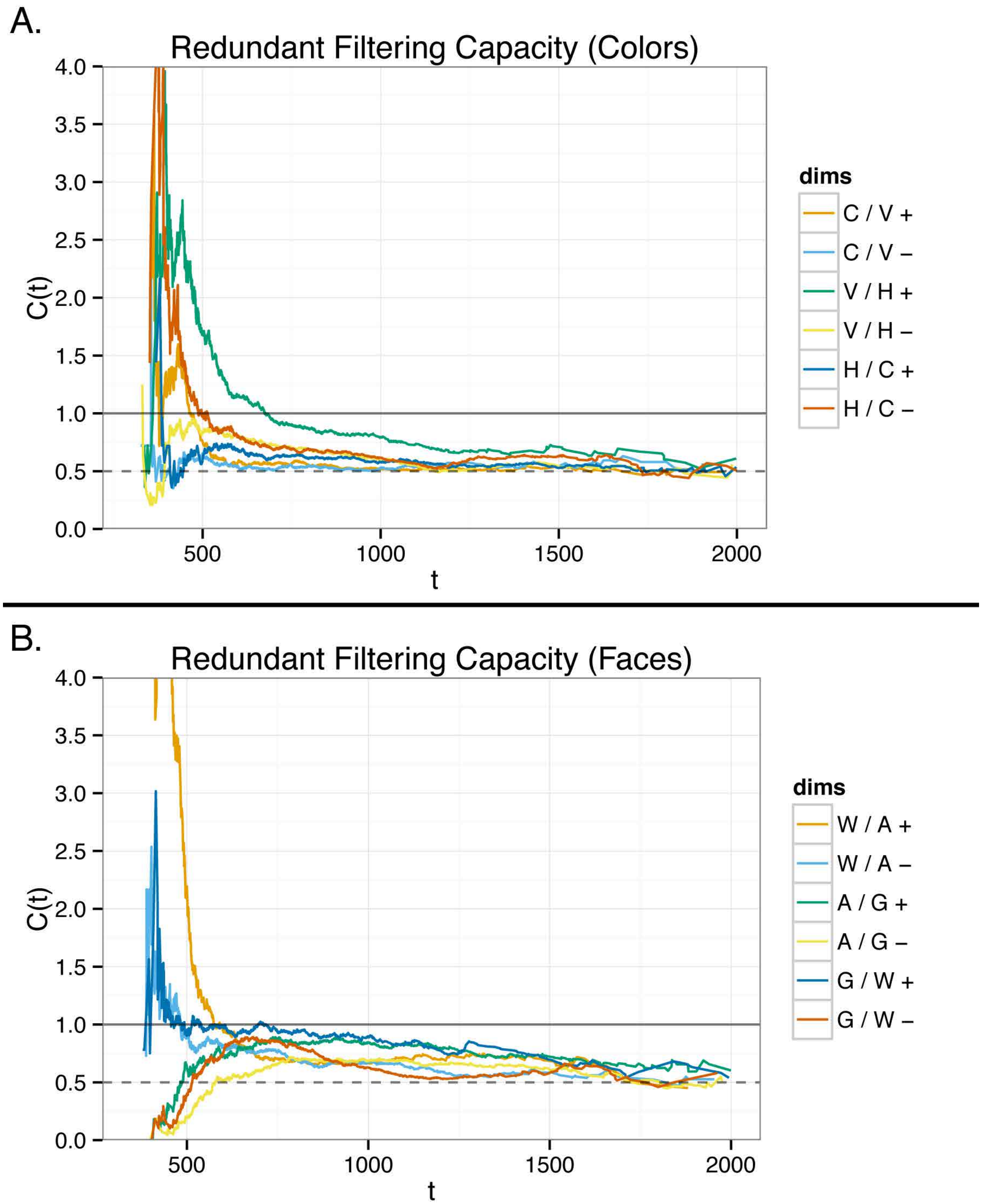


Figure 6.12: Workload capacity functions comparing redundant filtering and standard filtering trials for colors (A) and faces (B).

an irrelevant dimension. The $C(t)$ functions are shown in Figure 6.12. Consonant with the positively significant reaction time comparisons, all of the $C(t)$ functions stay above the fixed capacity boundary of one half (except for a few of the earliest reaction times). Capacity values are in general higher than for the 2-D redundant condition (Figure 6.7), with several conditions displaying super-capacity for early reactions. On the whole, however, capacity is still predominantly limited, hovering between one half and one.

The two and three-dimensional redundancy tasks can also be directly compared as another alternate form of the Garner interference test: the redundant filtering task is simply allowing a third dimension to vary (irrelevantly) that was previously fixed in the standard correlated task. While the previous tests (Table 6.8) probed how the addition of a redundant dimension changes reaction times in the presence of an irrelevant dimension, this comparison can reveal how the addition of an irrelevant dimension affects responses in conditions that each have the same correlation between two relevant dimensions. The significance values comparing these two redundant tasks are shown in Table 6.9.

The results are mixed for the color stimuli, but definitely favor negative significance for the faces. This means that the redundant filtering task was significantly *faster* than the correlated task. Therefore, adding irrelevant variation led to improved performance, the same pattern seen in the three-dimensional Garner filtering results.

6.6 CROSS CORRELATED

The final task to be analyzed is the cross correlated condition. Comparing this task with the correlated filtering condition gives us an idea of how increasing between-category distances influences performance. The expectation was that participants should do better in the cross correlated task.

T-tests for 12 out of the 18 individual participants were insignificant. Out of the nine participants who used color stimuli, there were significantly negative results (cross correlated was *slower* than correlated filtering) for two of the participants focusing on brightness and for one focusing on saturation. For face stimuli, one of the weight participants produced a negative significant difference, while one from each of the other two dimensions, age and gender, had a significantly positive difference. These data indicate a slight trend toward an advantage for correlated filtering over the cross correlated task, especially with regards to the color stimuli.

6.7 SUMMARY

In order to look at a broader picture of this plethora of comparisons, Figure 6.13 plots the significance values for all of the tests discussed in this section. To avoid overplotting, data was pooled across both participants and dimensions before being subjected to a repeated measures ANOVA which used participant as a random effect. This yielded a single data point for each stimulus set for each test. Values of statistical and practical significance (p-values and Cohen's D, respectively) are plotted against each other for concise presentation.

As can be seen, almost all comparisons achieve statistical significance, which should not be too surprising given that approximately 16,000 data points are used for each test. The least significant tests are those of comparing distances (both stretch filtering vs. filtering and cross correlated vs. correlated filtering) or the number of stimuli (double filtering vs. correlated filtering). The addition of exterior stimuli (full stretch vs. filtering) was non-significant for colors, but sizably negative for faces.

The strongest effects were those of three-dimensional redundancy, which compared the redundant filtering and standard filtering tasks, and the traditional measure of Garner interference. After those, no other test had a Cohen's D value greater than .1 in the expected direction: redundancy gains were consistent but weak. On the other side of the

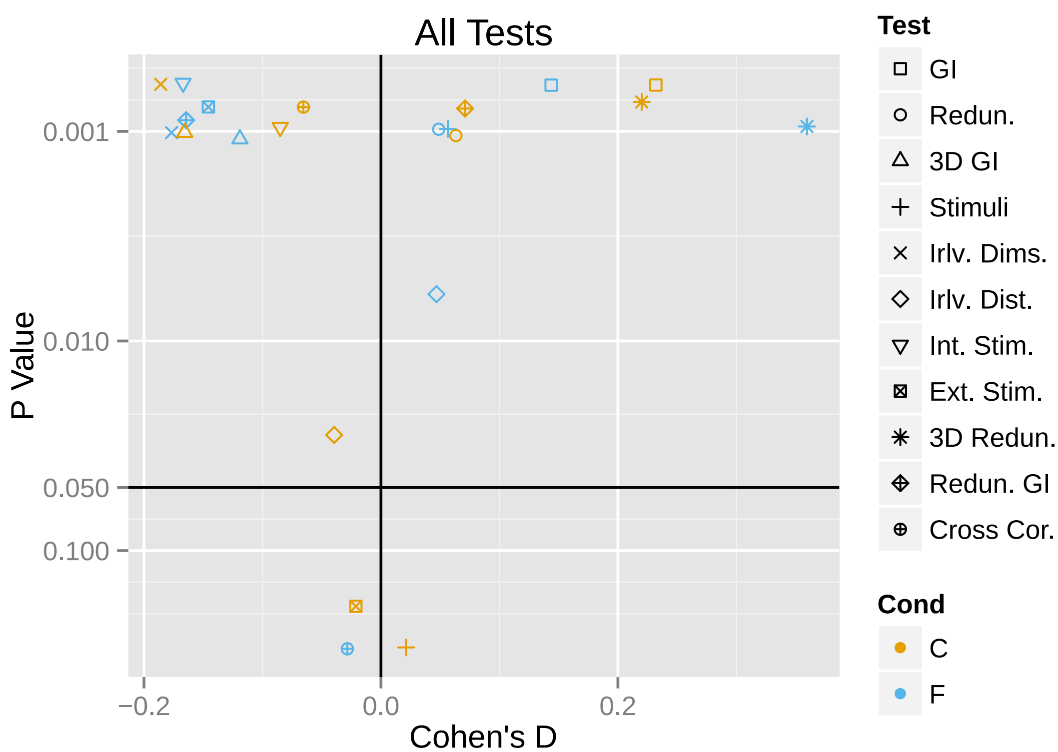


Figure 6.13: Significance plot for all of the major comparisons. RT differences that are opposite from expected are coded as negative values of Cohen's D. Data is averaged across dimensions and participants, as in Figure 6.1.

chart, there were some fairly strong negative effects. The addition of an irrelevant dimension (correlated filtering vs. filtering) was the strongest of these, for both classes of stimuli. 3D Garner interference also followed this trend as expected, since it incorporates a change in the number of irrelevant dimensions. The addition of interior stimuli also tended to lead to faster responses. The final test to mention is redundant Garner interference, which was negative for faces and weakly positive for colors.

CHAPTER 7

DISCUSSION

The Garner paradigm has been hugely influential in the study of how dimensions interact in human perceptual processing, but the conclusions many have drawn from it are weakly supported by the evidence it gathers. The first problem in this paradigm is the availability of alternate strategies that allow participants to perform with perfect accuracy without employing the targeted processing systems. This can be corrected by having a many-to-one stimulus to response mapping for all tasks. Secondly, while a finding of Garner interference is claimed to show an effect of increasing the number of irrelevant dimensions, this effect is confounded by a simultaneous increase in the number of stimuli. By transitioning to a three-dimensional stimulus set, these effects can be separated from one another. Finally, the manipulation of redundancy should be studied in a many-to-one mapping condition without changes in the number of irrelevant dimensions or in the number of stimuli presented, an opportunity also provided by the three-dimensional configuration.

7.1 TRADITIONAL GARNER TESTS

The Garner interference results were largely as expected, with most participants reporting significantly slower reactions in the filtering task than in the control task, validating these dimensions as integral. Several of the dimensional pairs, however, showed significant differences across participants, with roughly equal numbers reporting significantly negative and positive results. This finding of strong individual differences calls into question the routine practice of averaging across observers in studies of Garner interference. The overall results, however, confirmed that the general experimental design was largely consistent with previous work.

The redundancy gains for the correlated condition were more variable than expected, perhaps indicating distortions of the stimulus space. As seen in the analysis of baseline discriminability in Figures 6.3 and 6.4, the values of even non-varying dimensions had a strong influence on reaction times, and the correlated trials followed the same patterns. The counter intuitive findings of negatively significant results, or redundancy losses, can be explained through such stimulus space effects as mean-shift integrality (Ashby & Maddox, 1994). If four stimuli form a diamond shaped configuration rather than a square, as seen in Figure 3.1c, the diagonal distance between one pair of stimuli can be less than the distance along one of the sides of the diamond, meaning that the redundant pair can be less discriminable than the control pair. In this situation, the other diagonal distance will be forced to be even larger, however, leading to the prediction that the oppositely correlated condition should have a relatively large redundancy gain. The data from Table 6.4 support this theory, in that for any pair of dimensions that led to a redundancy loss for one direction of correlation, there were significant redundancy gains for the other direction. Such distortions should be expected with integral dimensions, and can indicate congruency effects across dimensions, where young faces are more naturally associated with being heavy than old faces are, for example.

The workload capacity results from Figure 6.7 are even more surprising. They show that although redundancy gains were found in many conditions, they were rather meager in size, with almost all of the $C(t)$ functions hovering around the level of fixed capacity. The task structure of the correlated condition is such that information from either of the two dimensions is sufficient for making a response. Using the information-processing language introduced in the discussion of logical rule models on page 27, this is a first-terminating stopping rule. An unlimited capacity, independent, parallel model is a useful benchmark in capacity analyses. This model predicts that in the correlated condition, both channels (one for each dimension) continue processing exactly as they had when they were operating on their own in the control conditions. This time, however, a response can be issued as soon as either one finishes, making this condition faster due to statistical facilitation.

Our capacity data, on the other hand, shows consistently limited capacity, just barely above the boundary for fixed capacity, where the correlated task is only as fast as the average control task. This level of performance could be predicted by a serial model, where dimensions are processed one at a time. Since this is a first-terminating task, processing would stop after the first dimension was processed, no matter their order. This hypothesis is in conflict with the typical interpretation of integral dimensions, but actually would also be capable of predicting Garner interference, given that the order in which the dimensions are processed varies from trial to trial. In filtering trials, this model would predict performance to be identical to control trials if the relevant dimension is processed first. If, however, the irrelevant dimension is processed first, a response could not be issued until both dimensions had finished processing, leading to slower average RTs. A problem with this model is that it would predict a bimodal distribution of reaction times depending on processing order, which there was no evidence for in this data.

A different processing assumption capable of explaining these data is a parallel model with limited capacity. In this model, each channel would slow down in multidimensional conditions (everything except the control task), but quite not enough to over power the effect of statistical facilitation, leaving the correlated trials slightly faster than the controls. This could also be used to straight-forwardly explain the effects of Garner interference, since the addition of an irrelevant channel does not contribute any facilitatory effects, and still would reduce the speed of the relevant dimension due to capacity limitations.

7.2 THREE-DIMENSIONAL GARNER FILTERING

The results posing the stiffest challenge to the extant models come come from the three-dimensional Garner filtering results. The correlated filtering task was designed to dissociate the effects of number of stimuli and number of irrelevant dimensions, and appears to have done so. Reaction times from that condition were far more similar to the double filtering condition than to the standard filtering condition, indicating that the difference in the number of irrelevant dimensions was more important than the difference in the number of stimuli, which was largely non-significant.

These results appear to indicate that the standard interpretation, that Garner interference effects are due primarily to in increase in the number of irrelevant dimensions, is the correct one. The difficulty presented by these data is that the effect was in the *opposite direction* of what was expected: adding a second irrelevant dimension led to *faster* responses. Out of all of the models presented in Chapter 3 (GRT, RT-Distance, EBRW, or Logical-Rule), none is capable of simultaneously predicting that the first irrelevant dimension slows responses and the second speeds them back up. I should be clear that this pattern was not seen globally for all participants and dimensions, but was common enough to demand explanation (five out of nine color participants and four out of nine face participants). A successful model for these data would require a mechanism by which the effect of the number of irrelevant dimensions was non-monotonic: the first hurts performance, but the second actually helps. An important question to address in future work is whether this trend continues for a third or even more dimensions.

A closer examination of the individual participant results shown in Figures 6.8(B) and 6.9(B) reveals that in many cases the correlated filtering and double filtering tasks were only as fast as the faster of the two filtering conditions (e.g. the second participants to focus on brightness, hue, and age, as well as the third participant focusing on weight). While this means that the two tasks are faster than the *average* of the filtering tasks, it is a weaker finding than if they were faster than *both* of the filtering conditions (which was the case for the third participants focusing on brightness and hue). It is also worth mentioning that in these cases, the faster filtering task, and therefore also the double filtering task, was as fast or faster than the control task, not showing the standard Garner interference effect. The other filtering task, however, did show interference for these participants, and it remains to be explained how this interference could disappear in the presence of additional irrelevant variation.

7.3 STRETCH FILTERING

The stretched filtering results were inconsistent: for example all three participants showed the expected decline in performance with stretched stimuli when weight was relevant and age was irrelevant, and yet all three participants focusing on brightness while attempting to ignore saturation actually performed better with the stretched stimuli. The general lack of significant results may indicate that the stimuli were not stretched far enough to achieve the robust results found by Nosofsky and Palmeri (1997b). Alternately, because stretch filtering has not been subjected to the same amount of research as, say, Garner interference, it is possible that the stimulus dimensions used in this experiment do not exhibit those effects. The dimensions found to exhibit stretch filtering effects by both Melara and Mounts (1994) and Nosofsky and Palmeri (1997b) were auditory pitch and loudness. It is possible that the strong congruence effects between those dimensions might be important for their stretch filtering results.

Although the EBRW model naturally predicts performance to decrease with reduced within-category similarity, several of the other models do not. In an independent information processing model, the salience of the irrelevant dimension should have no effect on the speed of the relevant channel, and therefore on the response times in a filtering task. This remains true even if Garner interference effects are caused by capacity limitations. The RT-distance and logical-rule models also do not predict that discriminability along the irrelevant dimension should influence reaction times. Further research on the contexts under which these stretch filtering effects occur could prove useful for model selection.

The full stretch condition, which used the eight stimuli from combining the standard filtering and stretch filtering tasks, was compared to both of those tasks in order to probe the effect of adding stimuli with the same values of the relevant dimension. These comparisons combine the effects of number of stimuli with the effects of differing distances along the irrelevant dimension. The addition of exterior stimuli to the filtering condition was hypothesized to decrease performance in two ways: more stimulus uncertainty could slow participants down, and the stretched stimuli themselves should provide slower reactions. Perhaps unsurprisingly, given our previous null results for both the number of stimuli and for stretched stimuli, this was not the case. More surprising was the trend towards negatively significant results, with consistently faster responses in the full stretch condition for the dimensions of brightness and gender.

The addition of interior stimuli had an even stronger negative effect, with participants reliably performing faster in the full stretch condition than the stretch filtering task. Eight out of the eleven significant differences for color stimuli were negative, and eight of the nine for faces. This effect has been previously shown in the literature, however, so is less surprising.

Melara and Mounts (1994) ran experiments testing how the range of the irrelevant dimension (amount of stretch) and the number of stimuli along that dimension (which were always equally spaced in the interior of the range) influenced Garner interference effects. Similar to Nosofsky and Palmeri (1997b), they found that increased range corresponded to increased interference. Adding interior stimuli, however, led to a decrease in interference. They hypothesized this to be the case because increasing the number of stimuli between the extremes on the irrelevant dimension increases the likelihood that a given stimulus will be similar to the previously presented stimulus in terms of the irrelevant dimension, thus eliciting less interference on a trial-to-trial basis. One way to test this hypothesis would be to measure whether reaction times are reliably influenced by the amount of change along the irrelevant dimension from one trial to another.

Because the results from the present experiment were coded under the simple assumption that more stimuli would produce slower reactions, the prediction of Melara and Mounts (1994) that adding internal stimuli should improve performance counted as a "negative" effect in Tables 6.6 and 6.7. Their logic does not apply to the addition of exterior stimuli, however, so the negative results there are still unexplained.

7.4 REDUNDANT FILTERING

Adding a redundant dimension to the filtering task led to faster responses, almost without fail. The comparison between redundant filtering and standard filtering produced two of the largest effect sizes out of all of the tests considered here, with Cohen's d between .2 and .4, as seen in Figure 6.13. When these redundancy gains were examined in further detail using the capacity coefficient, Figure 6.12, they were seen to be reliably stronger than those in the two-dimensional redundant condition, Figure 6.7.

Although these functions were still mostly limited capacity, there were several conditions showing super capacity for significant portions of the reaction time distribution. This means that performance in the redundant filtering was even better than would be predicted by an unlimited capacity, parallel, independent model. Such a model would predict that the redundant trials would be a simple "minimum-time" combination of the two corresponding filtering trials. If age and weight were relevant and gender was irrelevant, responses would be issued as soon as either one of two systems finished: one processing age while filtering gender, the other processing weight while filtering gender. Performance greater than the predictions of this model is typically only seen with strongly configural stimuli (Wenger & Townsend, 2006), and it could be an indication of interactions between the dimensions.

The result that redundancy gains were stronger in the presence of an irrelevant dimension could be explained in the framework of a ceiling-effect: because the filtering task is slower than the control task, there is more room for improvement when a redundant dimension is added. One way to analyze this hypothesis is to compare the two redundant tasks directly, as was done in Table 6.9. In this test of "redundant Garner interference," the results for color stimuli were fairly evenly spread between positive and negative differences, indicating that the two tasks were similar in speed and the addition of an irrelevant dimension did not have a reliable effect. With face stimuli, however, the redundant filtering task was significantly faster, almost without exception. This would rule out the ceiling-effect explanation, and contribute to the mystery of the 3-D Garner interference tests by providing another situation in which adding irrelevant variation actually *improves* performance.

7.5 CROSS CORRELATED

The cross correlated condition is interesting in that it increases the distance between categories (compared with correlated filtering), and yet it uses the same stimulus set and maintains two irrelevant dimensions. The EBRW model is able to predict a benefit from this condition, since response times are based explicitly on the similarities between stimuli: since the members of the two categories are less confusable with each other, responses are faster. From the point of view of an information processing model, however, nothing has changed: there is still the same amount of information in each of the three channels, one relevant and two irrelevant. Distance-from-boundary models lie somewhere in between, since they could be capable of predicting faster processing if participants were to adopt a complex saddle shaped boundary between the categories, rather than the more natural plane down the middle.

Results from this comparison were largely non-significant, however, so these data cannot play a major role in model selection. The few participants that did show significance tended to perform faster in the correlated filtering task, slowing down for the cross correlated condition. While none of the models predict this pattern, it most violates the predictions of the EBRW model, which predicts facilitation.

One possible alternate hypothesis is that the correlated filtering condition is not perceived as varying along two irrelevant dimensions, since those two dimensions vary in perfect synchrony. The task can instead be seen as involving only a single irrelevant dimension, be it a non-standard dimension composed of a combination of the two. The cross correlated condition, on the other hand, cannot be reduced to a two dimensional stimulus space. This could explain a decrease in performance for this condition, even with the EBRW model, since attention would need to be spread across more dimensions. This returns us to the assumption that adding a second irrelevant dimension is detrimental to performance, which is specious given our 3-D Garner interference data. Another issue with this explanation comes when reconciling it with a different comparison from the 3-D filtering results: if the correlated condition is perceived as being two dimensional, than why should it be easier than the standard filtering task? The EBRW model predicts that performance should decrease, like in a stretch filtering task, since there is now a greater distance between members of the same category.

The arguments of Melara and Marks (1990) about the primacy of perceptual dimensions could possibly help resolve this issue (see Chapter 4.1). If rotations of the dimensional axes are more difficult to perceive than the primary dimensions, then the correlated filtering task would in effect be making the irrelevant information *less* salient, since it now lies along a non-standard axis. If this effect were to dominate the effect of the greater distance along that dimension, which on its own makes the irrelevant information *more* salient, then we could correctly predict that correlated filtering is better than standard filtering. However, a theory of this nature would also have to explain why the redundant filtering task, which now places the *relevant* information on a non-standard, rotated dimension, is able to achieve such impressive redundancy gains.

7.6 CONCLUSIONS

This work was intended as a relatively straight forward extension of the Garner paradigm designed to deconfound the effects of number of stimuli and number of irrelevant dimensions, while simultaneously removing the possible utility of a change-detection strategy and ensuring all tasks are true categorization judgments rather than identification tasks. The results from these three-dimensional tasks, however, call into question the very goal of characterizing the interactions between two dimensions. If a given dimension, such as saturation, reliably produces interference when allowed to vary irrelevantly for either brightness or hue judgments, and yet produces facilitation when added to both brightness-by-hue and hue-by-brightness filtering tasks, what are we to say about its influence upon them? It might be more appropriate to frame our investigations in the manner of Melara and Mounts (1994), who claim that "...interactive effects are mainly a characteristic of stimulus relations and stimulus changes, rather than a quality intrinsic to a pair of dimensions."

Even under the assumption that interactive effects are more contextual and less "intrinsic to a pair of dimensions," the reason why variation along a second irrelevant dimension is capable of improving performance still begs explanation. The fact that this pattern occurred with both stimulus sets in two completely disjoint tests, both when comparing filtering to correlated filtering and the comparison between the correlated and redundant filtering tasks, signifies that this is a real effect which demands an explanation. It would be difficult to explain all instances of this finding by mere congruence effects or distortions of the stimulus space.

There is one case of similar findings reported in the literature. Ganel (2011) used a double filtering task in order to test the effect of head orientation on the relationship between gaze and expression judgments. Previous research (Ganel, Goshen-Gottstein, & Goodale, 2005) had shown the existence of symmetric Garner interference when participants are asked to perceive the emotional expression of a face or the direction of its gaze. However, the information used to perceive gaze direction crucially depends on the orientation of the head: if someone's head is pointed straight toward you, a full circular iris designates that they are looking at you. If their head is pointed elsewhere, on the other hand, that very same cue of a full iris now signifies they are looking away from you. Thus Ganel (2011) tested whether the interference between these two dimensions still held when head orientation was allowed to vary.

The tasks used in this experiment included four control blocks and two filtering blocks for both eye-gaze and emotion, with each using either front or side facing heads. Head orientation was then allowed to vary within blocks for a second experiment (using separate participants), yielding two gaze-by-orientation filtering blocks (at different levels of emotion), two emotion-by-orientation filtering blocks, and two double filtering blocks, one where gaze was relevant and one for emotion. Head orientation was never used as a relevant dimension. Comparisons between filtering and control blocks from the first experiment showed standard interference effects, but in the second experiment the filtering and double filtering tasks were the same. This means that the inhibitive effects of irrelevant eye gaze variation on emotion judgments (and vice versa) disappeared when in the presence of irrelevant head orientation variation.

It is important to note that this conclusion only compares the double filtering trials to one of the relevant filtering tasks, however. The figures presented by Ganel (2011) indicate that the double filtering trials were in fact slower than the either the gaze-by-expression or expression-by-gaze filtering tasks, but unfortunately these tasks were done by a separate pool of participants, making valid conclusions difficult. The author goes on to conduct experiments testing the effect of head orientation on the relationship between identity and expression, but in this case the double filtering task was slower than all of the filtering tasks (though once again some of these used a different group of participants).

Although the findings of Ganel (2011) are certainly related to those presented here, there are important differences we should note. There are theoretically justifiable reasons to expect that the perception of gaze direction should change in the presence of head orientation variability: while the size of the iris alone is sufficient to identify gaze direction when orientation is constant, this is no longer the case in the presence of variability. In fact, the value of orientation, direct or off to one side, *reverses the mapping* between iris shape and gaze direction. Because of this inherent relation between these two dimensions, it is no longer fair to say that the introduction of head orientation variation is truly irrelevant to the task. It is therefore no surprise that the filtering tasks where head orientation varied were slower than those where it was fixed, although once again the use of different subjects does not allow for proper comparisons. It is believable, however, that the effect of varying head orientation would dominate variations in emotional expression in the double filtering task, since it has such a profound effect on the interpretation of the information relevant for gaze judgements.

Thus we are still left wondering how the double filtering task could ever be reliably *faster* than the single filtering tasks it is composed of. This question surely merits further research, and will hopefully promote the development of new models capable of explaining the non-monotonic effects of irrelevant dimensions.

Another important avenue of exploration is the effect that attentional demands have on perceptual interactions. Traditional Garner tasks always instruct participants to pay selective attention to a single dimension. Sometimes this is the only variable dimension, or sometimes there is another that is either helpful or not, but they are always instructed to attend to a single dimension. What if participants were told to divide their attention between several dimensions? Eidels, Townsend, and Algom (2010) conducted a standard Stroop test, which implements a Garner filtering task using stimuli composed of the words "red" or "green" printed in either red or green ink. Crucially, they included conditions where participants were instructed to respond using a self-terminating rule: is either the word "red" or the ink color red? This rule could either be congruent, as in the example, or incongruent, where the word "green" in red ink would be a redundant target.

Their results found standard Stroop interference, in that when told to respond only to the ink color the congruent stimuli were reliably processed faster than the incongruent stimuli. This is usually taken to indicate both a failure of selective attention and the existence of interactions between the two dimensions. In stark contrast to these conclusions, the divided attention tasks showed evidence of the two channels being processed in parallel, independently of one another, with capacity values between one half and one. Even more surprisingly, whether the redundant target in these conditions was congruent or incongruent *had no effect* on the reaction times!

The bewildering results of the Eidels et al. (2010) experiments seem to indicate that when subjects are instructed to separate the two dimensions, completely ignoring one in favor of the other, the two dimensions interact. When they are instructed to combine information from the two dimensions to reach a common decision, however, the two dimensions are processed independently of one another! This indicates that there is still much to learn with regards to how attentional focus affects dimensional interactions. Perhaps a more complete theory of this relationship would also be capable of explaining why adding multiple irrelevant dimensions can sometimes aid performance.

The promise of a method for easily and quickly diagnosing the interactions between perceptual dimensions and separating them cleanly into the two neat groups of *separable* and *integral*, has proven quite tempting to researchers over the past 40 years. Unfortunately, it is a promise that the Garner paradigm cannot fully deliver on. Garner himself, with his emphasis on the need for convergent operations (Garner et al., 1956), would likely agree that the two tests most commonly employed, interference and redundancy, can only provide an incomplete picture at best. By building and contrasting quantitative models of these phenomena, expanding test conditions to greater numbers of stimuli and dimensions, and studying the influence of attentional state, we can hope to more fully capture the nuances and contextual effects that give depth to the concept of dimensional interactions.

REFERENCES

Amishav, R., & Kimchi, R. (2010). Perceptual integrality of componential and configural information in faces. *Psychonomic Bulletin and Review*, *17*(5), 743–748.

Ashby, F. G., & Maddox, W. T. (1990). Integrating information from separable psychological dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 598–612.

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

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

Attneave, F. (1950). Dimensions of similarity. *The American journal of psychology*, *63*(4), 516–556.

Bartlett, J. C., Searcy, J. H., & Abdi, H. (2003). What are the routes to face recognition. In M. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes: analytic and holistic processes* (pp. 21–52). Oxford: Oxford University Press.

Bingham, G. P., Schmidt, R. C., & Rosenblum, L. D. (1989). Hefting for a maximum distance throw: A smart perceptual mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 507–528.

Dishon-Berkovits, M., & Algom, D. (2000). The Stroop effect: It is not the robust phenomenon that you have thought it to be. *Memory and Cognition*, *28*(8), 1437–1449.

Egeth, H. E. (1967). Selective attention. *Psychological Bulletin*, *67*(1), 41–57.

Eidels, A., Townsend, J. T., & Algom, D. (2010). Comparing perception of Stroop stimuli in focused versus divided attention paradigms: Evidence for dramatic processing differences. *Cognition*, *114*(2), 129.

Eriksen, C. W., & Hake, H. W. (1955). Multidimensional stimulus differences and accuracy of discrimination. *Journal of Experimental Psychology*, *50*(3), 153–160.

Fific, M., Little, D. R., & Nosofsky, R. M. (2010). Logical-rule models of classification response times: A synthesis of mental-architecture, random-walk, and decision-bound approaches. *Psychological Review; Psychological Review*, *117*(2), 309–348.

Fific, M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology*, *34*(2), 356–375.

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.

Fletcher, B. C., & Rabbitt, P. M. A. (1978). The changing pattern of perceptual analytic strategies and response selection with practice in a two-choice reaction time task. *Quarterly Journal of Experimental Psychology*, *30*, 417–427.

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

Ganel, T. (2011). Revisiting the relationship between the processing of gaze direction and the processing of facial expression. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(1), 48–57.

Ganel, T., & Goshen-Gottstein, Y. (2002). Perceptual integrality of sex and identity of faces: Further evidence for the single-route hypothesis. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 854–867.

Ganel, T., Goshen-Gottstein, Y., & Goodale, M. A. (2005, April). Interactions between the processing of gaze direction and facial expression. *Vision Research*, *45*(9), 1191–1200.

Garner, W. R. (1970). The stimulus in information processing. *American Psychologist*, *25*(4), 350–358.

Garner, W. R. (1974). *The Processing of Information and Structure.* Potomac: Erlbaum Associates.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *8*(1), 98–123.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*(3), 225–241.

Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review; Psychological Review*, *63*(3), 149–159.

Garner, W. R., & Lee, W. (1962). An analysis of redundancy in perceptual discrimination. *Perceptual and motor skills*, *15*(2), 367–388.

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; Journal of Experimental Psychology: General*, *117*(4), 347–370.

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception and Psychophysics*, *12*(1), 108–116.

Hyman, R., & Well, A. (1967). Judgments of similarity and spatial models. *Perception and Psychophysics*, *2*(6), 233–248.

Hyman, R., & Well, A. (1968). Perceptual separability and spatial models. *Perception and Psychophysics*, *3*(3), 161–165.

Imai, S. (1966). Classification of sets of stimuli with different stimulus characteristics and numerical properties. *Perception and Psychophysics*, *1*(1), 48–54.

Imai, S., & Garner, W. R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology*, *69*(6), 596.

Kaufmann, J. M., & Schweinberger, S. R. (2004). Expression influences the recognition of familiar faces. *Perception*, *33*(4), 399–408.

Le Gal, P. M., & Bruce, V. (2002). Evaluating the independence of sex and expression in judgments of faces. *Attention, Perception, & Psychophysics*, *64*(2), 230–243.

Little, D. R., Nosofsky, R. M., & Denton, S. E. (2011). Response-time tests of logical-rule models of categorization. *Journal of experimental psychology. Learning, memory, and cognition*, *37*(1), 1–27.

Little, D. R., Nosofsky, R. M., Donkin, C., & Denton, S. E. (2013). Logical rules and the classification of integral-dimension stimuli. *Journal of experimental psychology. Learning, memory, and cognition*, *39*(3), 801–820.

Lockhead, G. R. (1966). Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, *72*(1), 95–104.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review; Psychological Review*, *95*(4), 492.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Maddox, W. T. (1992). Perceptual and decisional separability. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 147–180). Hillsdale: Laurence Erlbaum Associates.

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.

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

Melara, R. D., & Algom, D. (2003). Driven by information: A tectonic theory of Stroop effects. *Psychological Review; Psychological Review*, *110*(3), 422–471.

Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 398.

Melara, R. D., Marks, L. E., & Potts, B. C. (1993a). Early-holistic processing or dimensional similarity? *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1114–1120.

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

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory and Cognition*, *21*(5), 627–645.

Melara, R. D., & Mounts, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Attention, Perception, & Psychophysics*, *56*(1), 73–90.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*(1), 39–57.

Nosofsky, R. M., & Palmeri, T. J. (1997a). Comparing exemplar-retrieval and decision-bound models of speeded perceptual classification. *Attention, Perception, & Psychophysics*, *59*(7), 1027–1048.

Nosofsky, R. M., & Palmeri, T. J. (1997b). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*(2), 266–300.

Paysan, P., Knothe, R., Amberg, B., Romdhani, S., & Vetter, T. (2009). A 3D face model for pose and illumination invariant face recognition. In *Proceedings of the 6th ieee international conference on advanced video and signal based surveillance (avss) for security, safety and monitoring in smart environments* (pp. 296–301). Genova, Italy: IEEE.

Pomerantz, J. R., & Garner, W. R. (1973). Stimulus configuration in selective attention tasks. *Perception and Psychophysics*, *14*(3), 565–569.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(3), 422–435.

Schweinberger, S. R., Burton, A. M., & Kelly, S. W. (1999). Asymmetric dependencies in perceiving identity and emotion: Experiments with morphed faces. *Attention, Perception, & Psychophysics*, *61*(6), 1102–1115.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.

Shepard, R. N., & Chang, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology*, *65*(1), 94–102.

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

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

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Swets, J. A., Tanner, W. P., Jr, & Birdsall, T. G. (1961). Decision processes in perception. *Psychological Review; Psychological Review*, *68*(5), 301.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Torgerson, W. S. (1952). Multidimensional scaling I—theory and method. *Psychometrika*, *17*, 401–419.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge, UK: Cambridge University Press.

Townsend, J. T., Burns, D., & Pei, L. (2012). The prospects for measurement in infinite dimensional psychological spaces. In B. Berglund, J. T. Townsend, G. B. Rossi, & L. Pendrill (Eds.), *Measurement in persons: Theory and methods.* New York: Taylor & Francis Group.

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

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

Wenger, M. J., & Townsend, J. T. (2006). On the costs and benefits of faces and words: Process characteristics of feature search in highly meaningful stimuli. *Journal of Experimental Psychology*.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination. *Perception and Psychophysics*, *15*(3), 501–508.

Young, G., & Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances. *Psychometrika*, *3*(1), 19–22.

Zhu, Q., & Bingham, G. P. (2011). Human readiness to throw: the size–weight illusion is not an illusion when picking the best objects to throw. *Evolution and Human Behavior*, *32*(4), 288–293.

APPENDIX A

SIGNIFICANCE PLOTS

Figures A.1-A.6 plot significance values, both statistical and practical, for the Garner interference tests. Similar plots for redundancy gains are shown in Figures A.7-A.12.

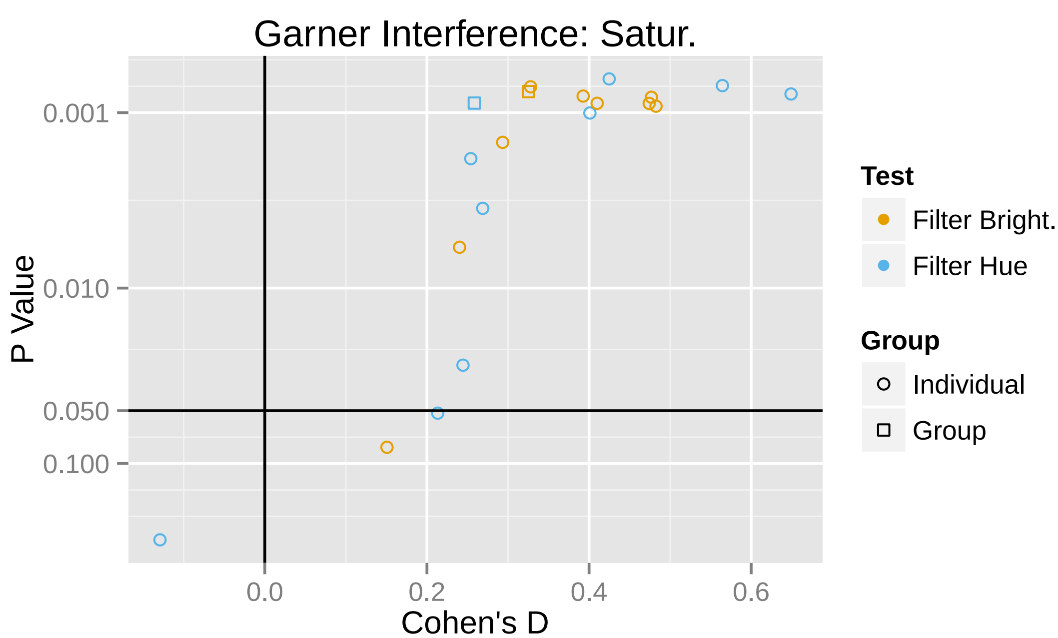


Figure A.1: Significance values for Garner interference as measured by Cohen's D and p-values. Results in the opposite from expected direction (Control slower than filtering) are coded as having negative values of Cohen's D. All results with $p < .001$ are capped at this value and then randomly jittered for clearer visualization.

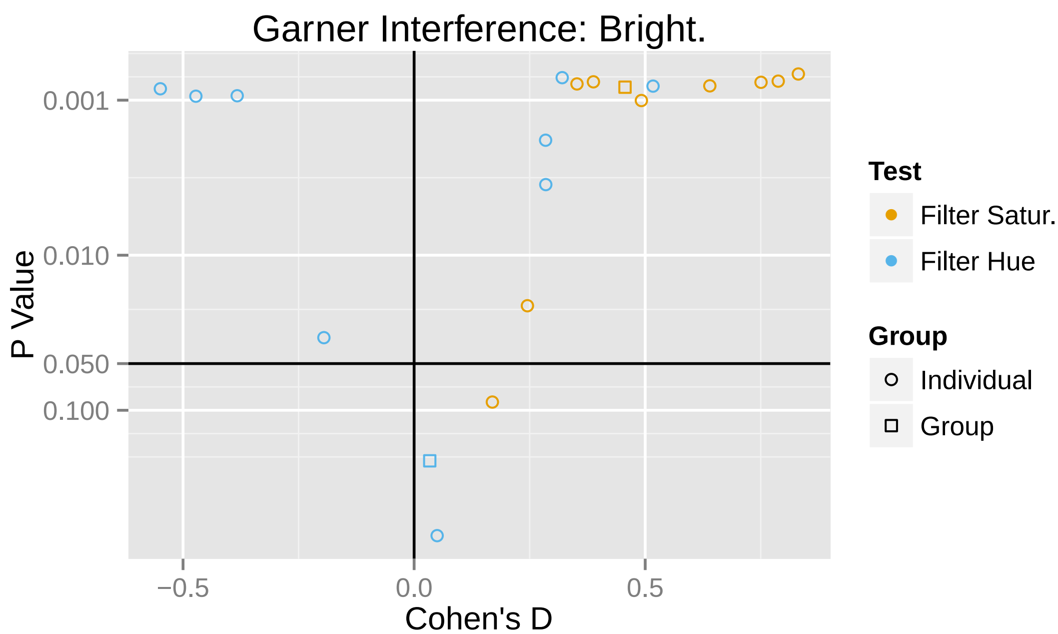


Figure A.2: Garner interference when responding to brightness

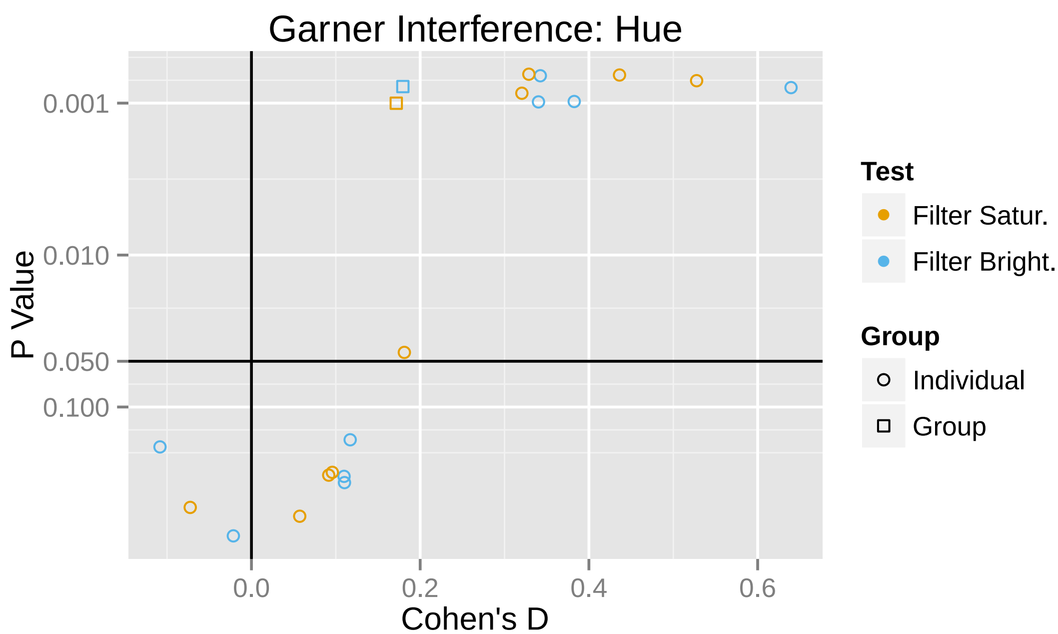


Figure A.3: Garner interference when responding to hue

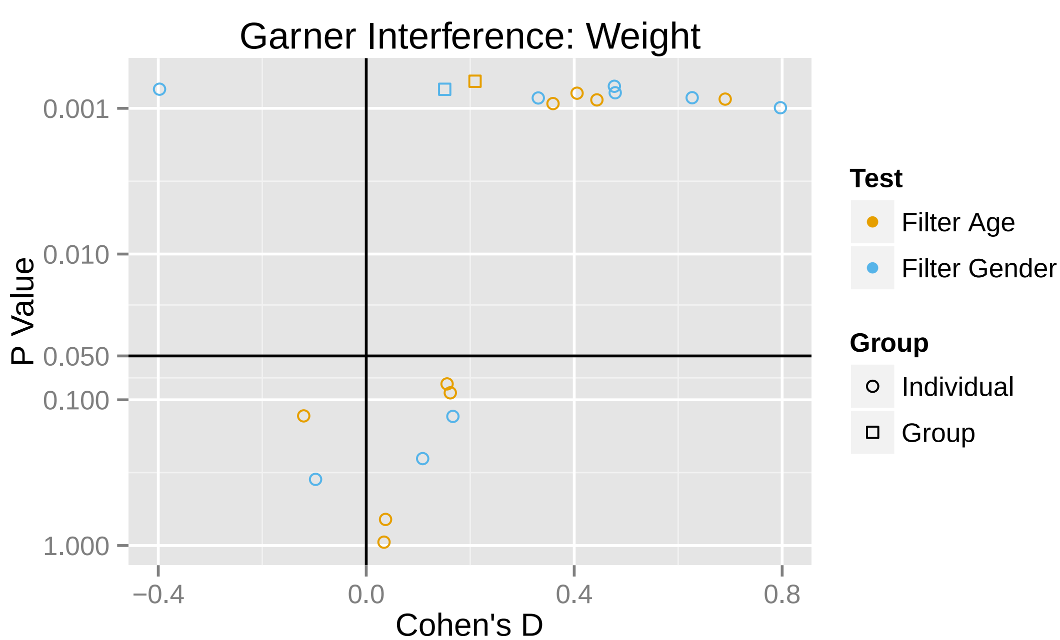


Figure A.4: Garner interference when responding to gender

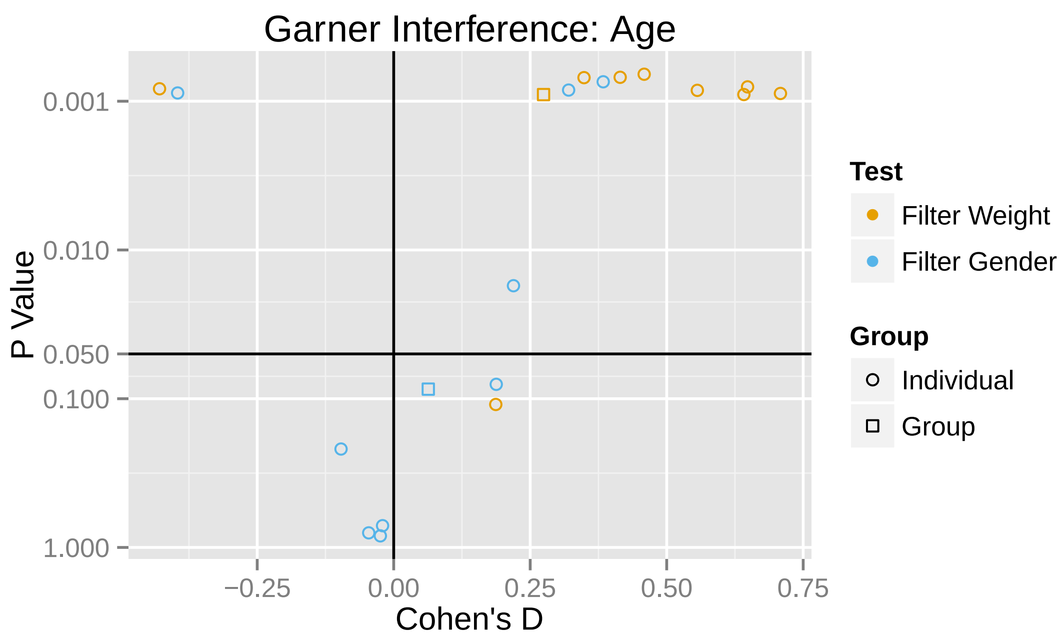


Figure A.5: Garner interference when responding to age

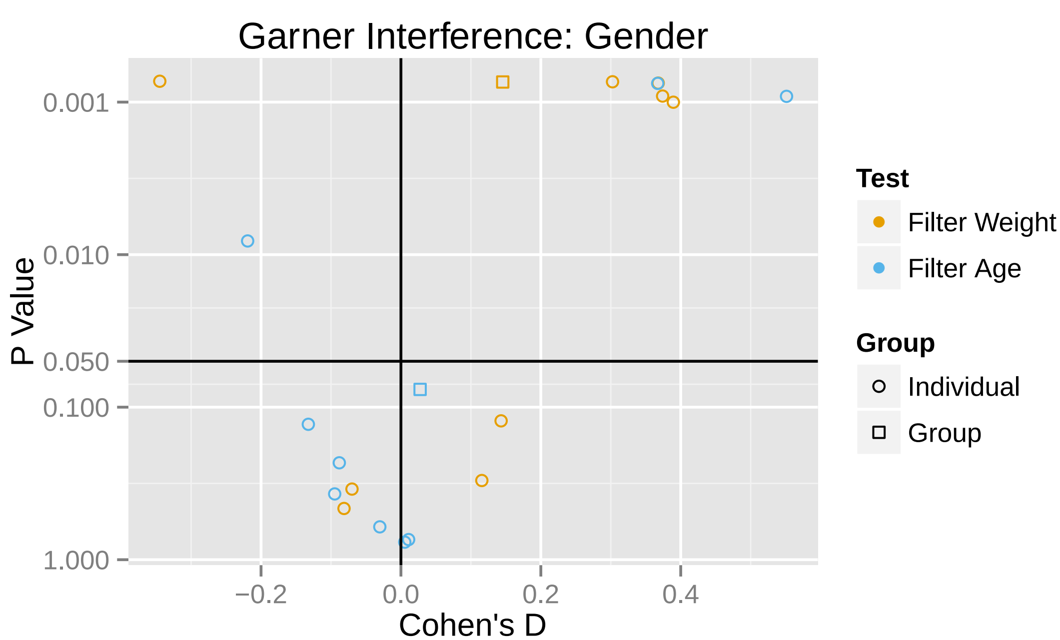


Figure A.6: Garner interference when responding to gender

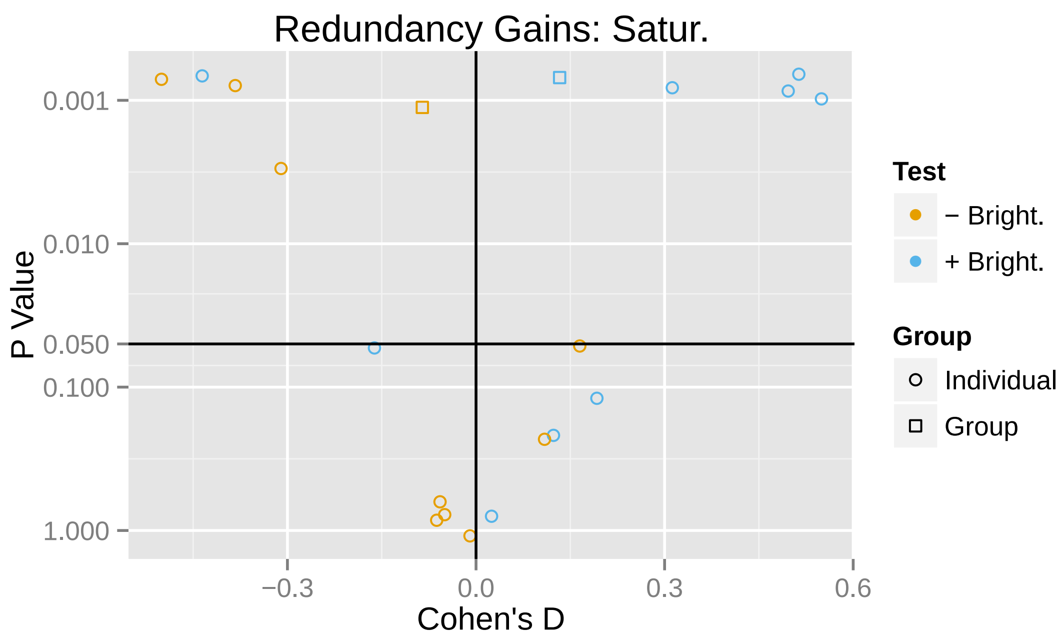


Figure A.7: Significance values for redundancy gains.

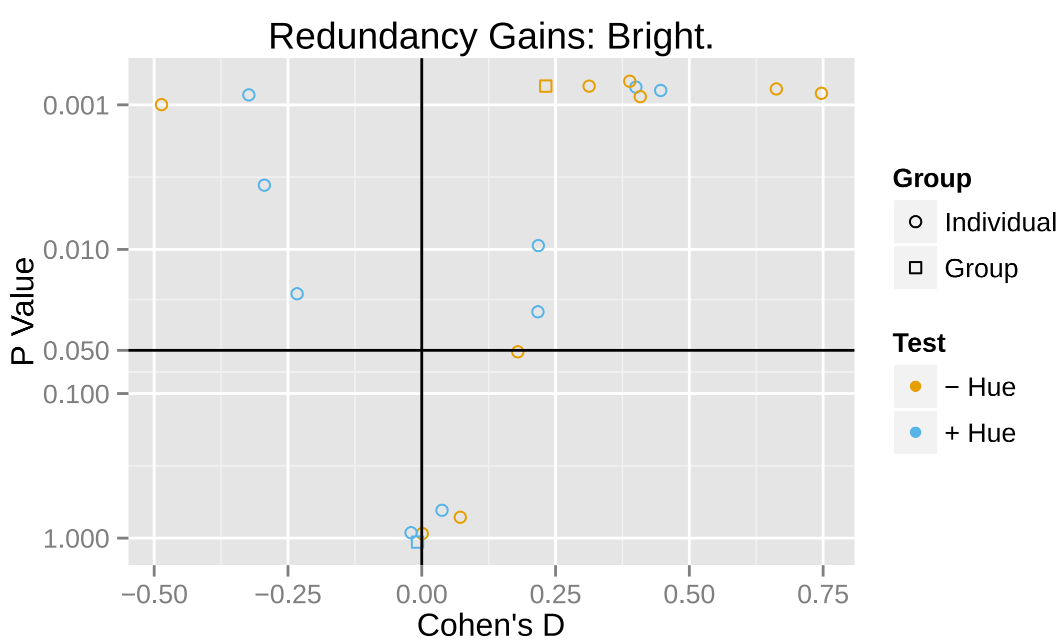


Figure A.8: Redundancy Gains when responding to brightness

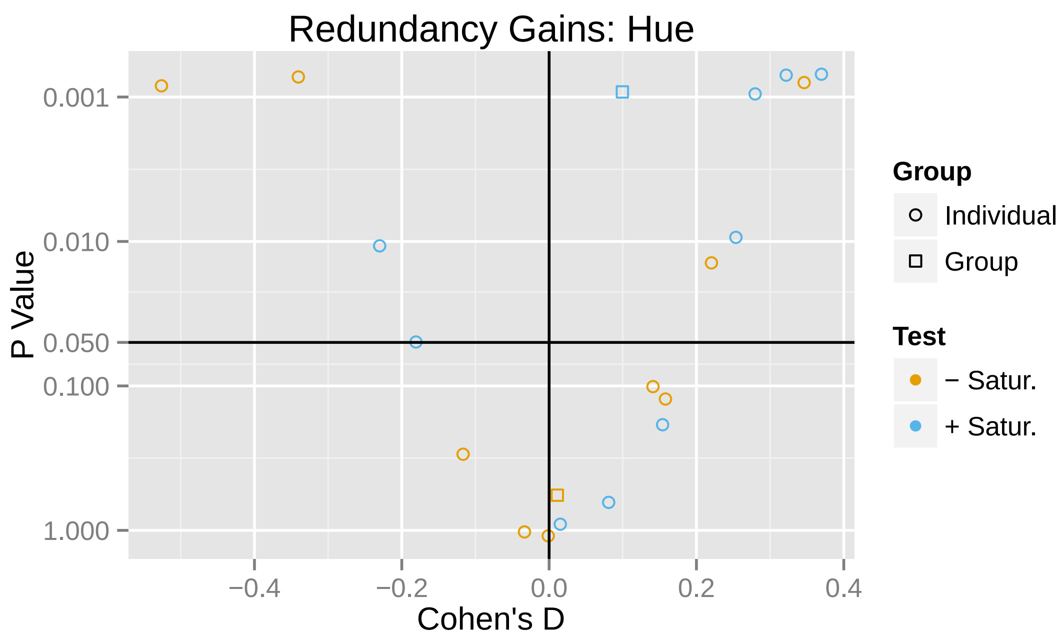


Figure A.9: Redundancy Gains when responding to hue

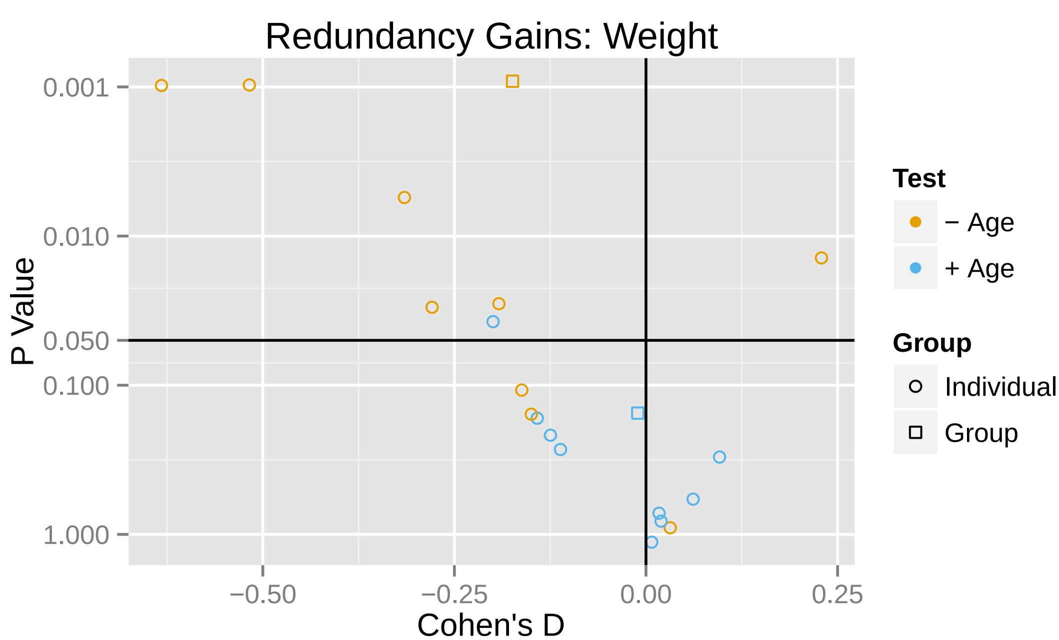


Figure A.10: Redundancy Gains when responding to gender

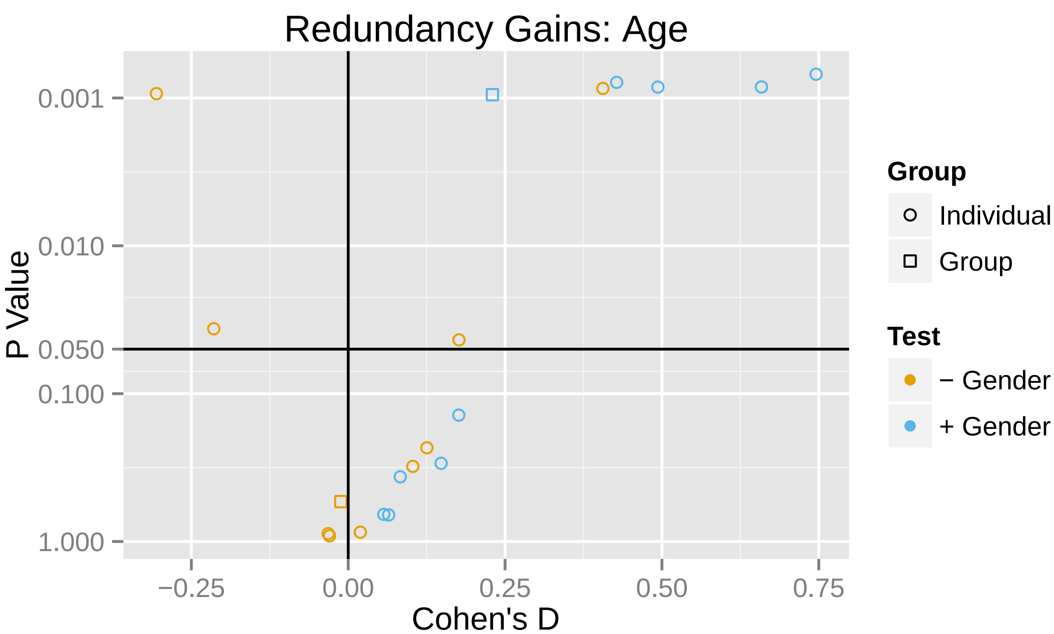


Figure A.11: Redundancy Gains when responding to age

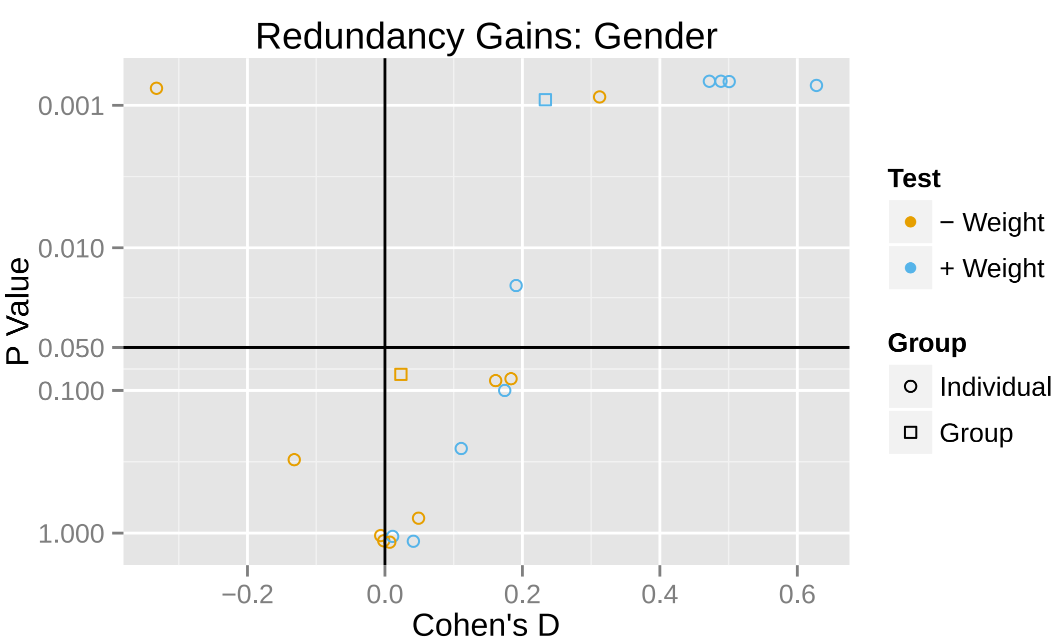


Figure A.12: Redundancy Gains when responding to gender

Devin M. Burns

Contact Information

Psychological and Brain Sciences
Indiana University, Bloomington
1101 E 10th Street
Bloomington, IN 47405 USA

Cell: (505) 504-8474
Lab: (812) 855-1554
E-mail: devburns@indiana.edu

Education

Indiana University, Bloomington, IN USA

Ph.D., Cognitive psychology and cognitive science **May 2014**

- Thesis Topic: *The Many Faces of Garner Interference*
- Advisor: James T. Townsend
- 3.97 GPA

Bradley University, Peoria, IL

B.S., Mathematics **May 2008**

- Named the outstanding student in Mathematics **2007 & 2008**
- 4.0 GPA as a Mathematics Major

Teaching Experience

Instructor

P335: Cognitive Psychology **Fall 2013**

- Conducted as a flipped class, using Just in Time pedagogy

P211: Methods of Experimental Psychology **Springs 2011 & 2012**

- Taught three sections total

LSAT Prep Course **2010-2013**

- Taught eight different month-long classes
- Offered at the request of the IU Law School

Teaching Assistant

Graduate Statistics **Fall 2011**

Statistical Techniques in Psychology **Summer 2011**

Embodied Cognition and Ecological Psychology **Fall 2010**

History and Systems of Psychology **Fall 2010**

Guest Lecturer

Intro to Mathematical Psychology	**Falls 2011-2013**
Differential Geometry and Psychology	**Springs 2012 & 2013**
Human Face Perception	**Spring 2014**

AWARDS

- Cognitive Science Program Summer Research Fellowship **2012**
- Received commendation for qualifying exam performance **2011**
- Society for Mathematical Psychology Student Award **2009-2012**
- Honorable Mention for NSF Graduate Research Fellowship **2009**
- NIMH Cognitive Modeling Training Grant **2008**

PUBLICATIONS

Burns, D. M., Houpt, J. W., & Townsend, J. T. (2013). Functional Principal Components Analysis of Workload Capacity Functions. *Behavioral Research Methods*, 45, 1048-1057.

Houpt, J. W., Blaha, L. M., & Burns, D. M. (2013). *Latest Developments in Systems Factorial Technology with R*. Manuscript submitted for publication.

Houpt, J. W., Hawkins, R.D., Burns, D. M., & Townsend, J.T. (2013). Measuring Configural Superiority with the Capacity Coefficient. *Journal of Vision*, 13, 73.

Townsend, J. T., Burns, D. M., & Lei, P. (2012). The prospects for measurement in infinite dimensional psychological spaces. In B. Berglund, J. T. Townsend, G. B. Rossi, and L. Pendrill (Eds.), Measurement with Persons: Theory and Methods. New York: Taylor & Francis Group.

Townsend, J. T.,Yang, H., & Burns, D. M. (2011). Experimental discrimination of the world's simplest and most antipodal models: The parallel-serial issue. In Hans Colonius and Ehtibar Dzhafarov (Eds), Descriptive and Normative Approaches to Human Behavior in the Advanced Series on Mathematical Psychology. Singapore: World Scientific.

CONFERENCE PRESENTATIONS

Burns, D. M., & Townsend, J. T. (2013). The Many Faces of Garner Interference. Talk presented at: Cognitive Lunch talk series; Indiana University, Bloomington, IN.

Houpt, J. W., Blaha, L. M. & Burns, D. M. (2013). The latest systems factorial technology in R developments. Talk at the Society for Computers in Psychology Annual Meeting; Toronto, ON.

Burns, D. M., Houpt, J. W., & Townsend, J. T. (2012). Functional Principal Components Analysis of Workload Capacity Functions. Talk presented at: Society for Mathematical Psychology 45th Annual Conference; Columbus, OH.

Burns, D. M., & Savion, L. (2010). Hijacking social networks in the service of pedagogy. Talk presented at the Faculty Colloquium on Excellence in Teaching (FACET) annual Associate Faculty and Lecturers Conference.

Burns, D. M., Houpt, J. W., Endres, M. J., & Townsend, J. T. (2010). Functional principal components analysis and the capacity coefficient. Poster presented at: Society for Mathematical Psychology 43rd Annual Conference; Portland, OR.

Burns, D. M., Houpt, J. W., & Townsend, J. T. (2010). Capacity Analysis of Facial Perception. Poster presented at: Vision Sciences Society Conference; Naples, FL.

Burns, D. M., Pei, L., Houpt, J. W., & Townsend, J. T. (2009). Facial Perception as a Configural Process. Poster presented at: Annual Summer Interdisciplinary Conference; Aosta, Italy.

Burns, D. M., Pei, L., Houpt, J. W., & Townsend, J. T. (2009). Facial Perception as a Configural Process. Talk presented at: Society for Mathematical Psychology 42nd Annual Conference; Amsterdam, Netherlands.

Burns, D. M., Pei, L., Houpt, J. W., & Townsend, J. T. (2009). Facial Perception as a Configural Process. Poster presented at: Annual Meeting of the Cognitive Science Society; Amsterdam, Netherlands.

Invited Talks

Burns, D. M.(2010). Dealing with Public Speaking Anxiety in the Classroom. Talk presented at campus-wide orientation for new AIs at IU.

Professional Service

- Mentored Undergraduate Research Assistants
  - Robert Hawkins **Fall 2009 - Spring 2014**
  - Daniel Baggerman **Fall 2012 - Spring 2014**
  - Kamila Salibayeva **Spring 2012 - Spring 2013**
  - Hannah Leeper **Spring 2010 - Spring 2011**
- Ad-hoc reviewer
  - *Behavioral Brain Research*
  - *Journal of Experimental Psychology: Human Perception and Performance*
  - *Journal of Mathematical Psychology*

- Helped organize and run the *Mathematical Models of Perception and Cognition* conference at Indiana University **April 2013**

PROFESSIONAL AFFILIATIONS

- Society for Mathematical Psychology
- Cognitive Science Society
- Vision Sciences Society
- Association for Psychological Science

SKILLS

Experienced in LaTeX, PsychoPy, R, Matlab, DMDX, C++, and Perl

REFERENCES

James T. Townsend	Email: jtownsen@indiana.edu
Robert Nosofsky	Email: nosofsky@indiana.edu
Benjamin Motz	Email: bmotz@indiana.edu

Address:
Psychological and Brain Sciences
Indiana University, Bloomington
1101 E 10th Street
Bloomington, IN 47405