INTEGRATION OF MULTIDIMENSIONAL SIGNAL DETECTION THEORY WITH FUZZY SIGNAL DETECTION THEORY

by

MAUREEN ELIZABETH O'CONNELL

B.S. University of Illinois Urbana-Champaign, 2005

M.S. Northeastern University, 2007

M.A. University of Central Florida, 2014

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Science
at the University of Central Florida
Orlando, Florida

Fall Term
2015

Major Professor: James L. Szalma

ABSTRACT

Signal detection theory (SDT) has proven to be a robust and useful statistical model for analyzing human performance in detection and decision making tasks. As with many models extensions have been proposed in order capture and represent the real world to a greater degree. Multidimensional Signal Detection Theory (MSDT) has had success in describing and modeling complex signals, signals that are comprised by more than one identifiable component dimension. Fuzzy Signal Detection Theory (FSDT) has had success in modeling and measuring human performance in cases where there exist ambiguity in the signal or response dimension characteristics, through the application of fuzzy set theory to the definition of the performance outcome categories. Multidimensional Fuzzy Signal Detection Theory (MFSDT) was developed to accommodate simultaneously both the multidimensionality of a signal and the fuzzification of outcome categories in order to integrate the two extensions. A series of three studies were performed to develop and test the theory. One study's purpose was to develop and derive multidimensional mapping functions, the aspect of MFSDT where MSDT and FSDT were integrated. Two receiver operating characteristic (ROC) studies were performed, one simulated and one empirical. The results from both ROC analysis indicated that for perceptually separable and perceptually integral complex stimuli that MFDST is a viable methodological approach to analyzing performance of signal detection tasks where there are complex signals with ambiguous signal characteristics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Signal Detection Theory (SDT) is a statistical model for measuring and describing performance for detection and decision making tasks. It is used in a variety of research domains including but not limited to psychology, engineering and medicine. For example, SDT has been used to analyze performance of individual baggage screener in detecting threat items passing through x-ray baggage screening with the use of threat image projection (TIP), in which realistic looking x-ray images of threat objects (i.e. guns, knives) are inserted into images during normal baggage screening operation. (Hofer & Schwaninger, 2005). In the medical sciences SDT has been used to evaluate different magnetic resonance techniques, blood oxygenation level dependent (BOLD) vs. echo-planar imagine and signal targeting (EPISTAR), with use for functional brain imaging (Siewert, Bly, Schlaug, Thangaraj, Warach, & Edelman, 1996) as well as different medical imaging techniques in general (Swets, 1979). SDT is also applicable to the evaluation of diagnostic systems (Swets & Pickett, 1982), alarms (Parasuraman, Hancock, & Olofinboba, 1997), and in the design and operation of automation (Parasuraman, Sheridan, & Wickens 2000; Sorkin & Woods, 1985).

The Traditional SDT Model

Signal Detection Theory emerged from engineering and Decision Theory. It was originally applied to the detection of psychophysical stimuli in noise, such as auditory signals in communication systems (Peterson & Birdsall, 1953) or visual signals on radar displays (Tanner & Swets, 1954). In SDT the psychophysical stimulus to be detected is referred to as the signal and all other environmental events are defined as noise. Green and Swets (1966) delineate the decision making event as comprised of three parts; 1) the state-of-the-world, 2) the information presented to the observer, i.e., the magnitude of the evidence variable and, 3) the decision

regarding the classification of the stimulus as a signal or non-signal event. The state-of-the-world is essentially the ground truth regarding whether the stimulus presented is a member of the category 'signal' or 'noise'. The information refers to stimulus properties presented to the observer, which comprises the evidence variable. The decision is the response by the observer regarding whether the signal is present or absent or, more generally, the category (as defined in the specification of the state of the world) to which the stimulus event belongs. While typically dichotomous, the response can be multicategorical (MacMillan & Creelman, 2005). For each stimulus presentation the observer responds affirmatively (i.e. signal present) if the stimulus magnitude is larger than an internal value, referred to as the criterion, and the observer responds negatively regarding signal presence if the stimulus magnitude is less than the criterion. Four possible outcomes result from the combination of the two states-of-the-the world and the two response alternatives.

		Response Alternatives	
		S	N
State of the World Alternatives	s	P(S|s) Hit (H)	P(N|s) Miss(M)
	n	P(S|n) False Alarm (FA)	P(N|n) Correct Rejection (CR)

Figure 1. Truth table for SDT that maps the state-of the-world to observer response. Adapted from Green and Swets (1966). S = Affirmative response regarding signal presence, N = Negative response regarding signal presence, s = the signal is present, n = the signal is absent ('noise' or non-signal).

These outcomes can be specified as the conditional probabilities of each possible response by state of the world combination. For example, a correct detection or hit (H) is the probability of responding that a signal is present given that the state of the world was in the

category signal. The four outcomes provide performance information regarding the accuracy of the observer, but in SDT these scores can be used to compute indices of perceptual sensitivity and response bias (criterion setting). Sensitivity refers to the capacity of the observer (or non-human detector) to distinguish signal from noise events. An observer who achieves higher signal detection performance will have higher sensitivity and therefore will have a higher rate of correct rejections and a lower rate of false alarms. For example, Bonnel, Mottron, Peretz, Trudel, Gallun, and Bonnel (2003) investigated pitch sensitivity in high functioning autistics using two auditory perceptual tasks and found compared to neuro-typical individuals, high functioning autistics exhibited higher pitch sensitivity (i.e. they were able to detect tones others were unable to, they were more sensitive to the detection of tones).

Response bias refers to the tendency of the observer to respond 'signal present' versus a 'signal absent', and it is determined by the magnitude of the evidence variable at which the observer sets the decision criterion. The response bias of an observer can be described as lenient, unbiased, or conservative. A lenient criterion would indicate a lower magnitude value of the stimulus for the observer to respond in the affirmative to the signal presence. A lenient criterion results in more hits but at the cost of more false alarms. A conservative criterion would indicate a higher magnitude of the stimulus (the evidence variable) that is required for the observer to respond in the affirmative regarding signal presence. A conservative criterion results in fewer false alarms but at but at the cost of fewer correction detections. An example of response bias, Thomas and Houge (1976) used SDT analysis in their construction of model for jury decision making, and they reported that among other decision criteria, that the juror instructions regarding the definition of reasonable doubt affected response bias, the more lenient definitions were associated with more lenient the response biases. Response bias in this case was defined as

the jurors' willingness to render a guilty vote. The study also found that the severity of the case and of its punishment influenced response bias. Jurors adopted a more conservative response bias (i.e., they adopted a stricter criterion for a "guilty" decision") if it was a criminal case with severe punishment attached.

SDT being a statistical model of human performance in decision making and detection is based on five statistical assumptions (for more detailed descriptions of the statistical model, see Green & Swets, 1966, MacMillan & Creelman, 2003):

1) Noise is always present in a detection system and it is normally distributed with a variance, $\sigma_n^2 = 1$
2) When a signal is added to the noise, the distribution shifts along the sensory dimension as shown in Figure 2. The equal variance model assumes that the variance of the distribution remains unchanged when shifted, $\sigma_s^2 = \sigma_n^2 = 1$
3) The observer is both the sensor and decision maker
4) The observer adopts a decision criterion in order to determine whether a signal is present
5) Sensitivity and bias are independent of one another.

These assumptions are reflected in the decision space defined by the model and illustrated in Figure 2.

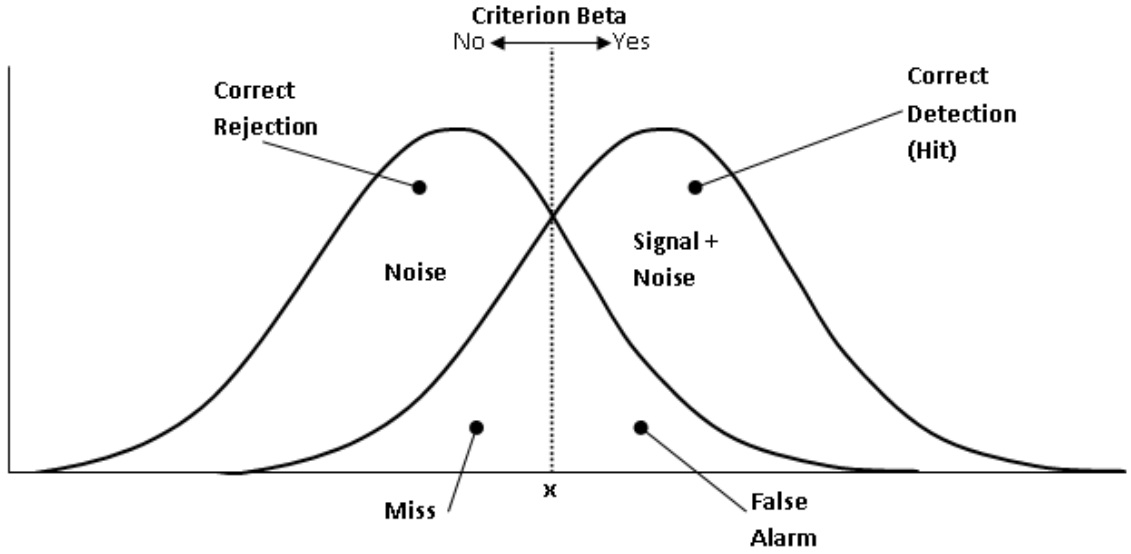


Figure 2. Representation of the Signal Detection Theory decision space. The x-axis represents the magnitude of the evidence variable.

The decision space consists of two normal distributions, the right-most normal distribution is the distribution of noise (assumption 1) and the left-most normal distribution is the distribution of the signal embedded in the noise (assumption 2). The line labeled criterion beta, is the minimum magnitude of the evidence variable at which the observer responds affirmatively. If the magnitude of x is greater than this criteria then the observer decides the signal is present, if the magnitude is less than the observer decides the signal is not present. Note that the criterion beta line in Figure 2 is the SDT measure response bias. The distance between a two distributions is the metric for sensitivity.

Extensions of Signal Detection Theory

One strength of SDT is that it provides measures of performance that represent different components of detection and decision making tasks. As with many mathematical or statistical models modifications of the initial theory have been proposed in order to refine the model into a better representation of real world decision making. Two variants of SDT have shown potential

for achieving this goal, Multidimensional Signal Detection Theory (MSDT) for stimuli that vary on more than one stimulus dimension and Fuzzy Signal Detection Theory (FSDT) for cases in which the state of the world on stimulus or response alternatives are not mutually exclusive categories.

Multidimensional Signal Detection Theory

Multidimensional Signal Detection Theory (MSDT) was derived from General Recognition Theory (GRT; Ashby & Townsend, 1986) and it may be considered a multivariate generalization of signal detection theory. GRT was developed to in order to explain how the different perceptual dimensions, e.g. orientation, shape, hue, brightness, etc, are combined in the perceptual processing of a stimulus. Ashby and Townsend (1986) argued that because the perceptual process itself is non-observable, the perception of the stimulus must first undergo some form of decision processes, which based on the direct perception of the stimulus selects an appropriate response. This argument is similar to the assumption of SDT in which the observer adopts a decision criterion in order to determine whether a signal is present or not. By assuming that the perceptual dimensions of the stimulus are multivariate normal distributions, the normal distributions being an assumption of SDT, GRT can be used as a multivariate extension of SDT (Ashby & Townsend, 1986), thus creating MSDT.

In GRT, for a single detection or discrimination event the observed stimulus is represented by a point in the multidimensional perceptual space and variability is introduced in an aggregation of perceptual response over multiple repetitions of the stimulus (Ashby, 2000). This aggregation results in a multivariate probability distribution, i.e. $f_{A_iB_i}(x, y)$ for a presentation of a complex stimulus A_iB_i which is a specific combination of two perceptual dimensions A and B, as well as marginal distributions of perceptual effects on each perceptual

dimensions, i.e. $g_{A_iB_i}(x)$ and $g_{A_iB_i}(x)$ for presentation of a complex stimulus A_iB_i, as seen in Figure 3.

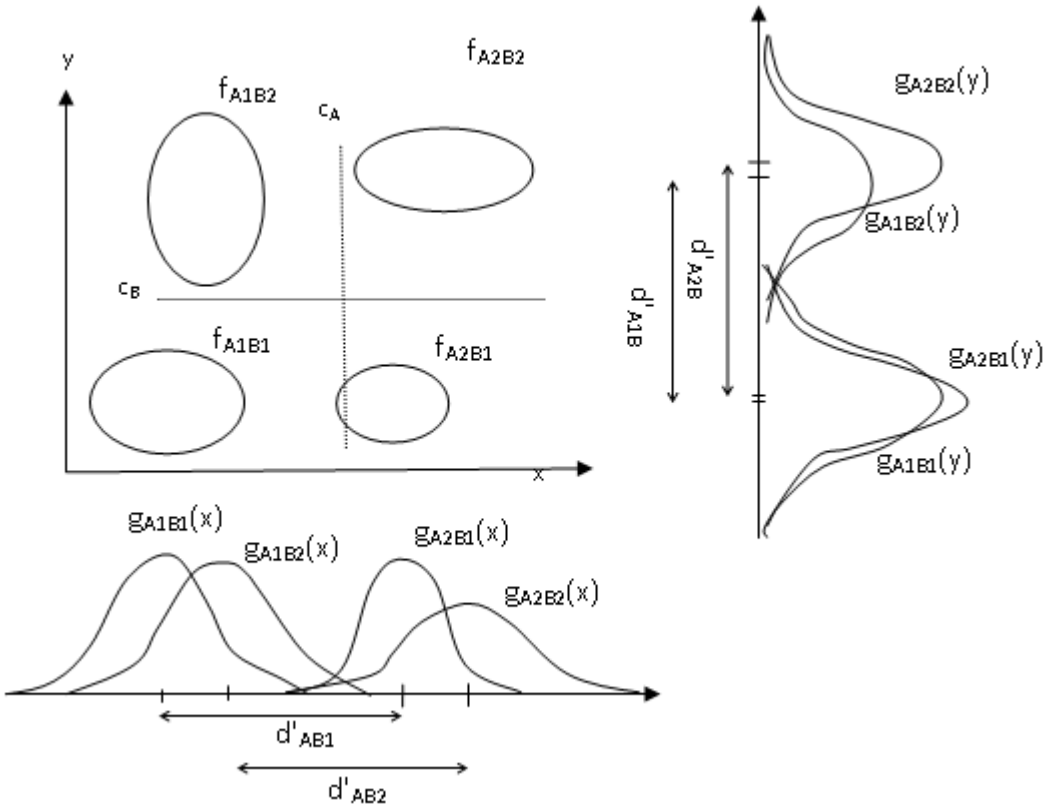


Figure 3. Example of GRT with unequal variance probability distributions. Illustration based on Ashby and Townsend (1986).

The circles shown in Figure 3 are a plane cutting of the four bivariate probability distributions f(x, y) at a given height creating an equal probability contour for each distribution. Figure 4 illustrates the transformation from bivariate probability distributions to equal probability contours. A plane cutting takes a 3-dimensional probability distribution such as f_1 in Figure 4a and transforms it into a 2-dimensional equal probability contour, such as $f_1(x,y)$ in Figure 4b or $f_{A2B1}(x,y)$ in Figure 5. Ashby and Townsend (1986) note that for many common probability distributions, such as the bivariate normal distributions seen in Figures 4 and 5, all contours have the same shape so it is sufficient to represent each distribution as a single contour.

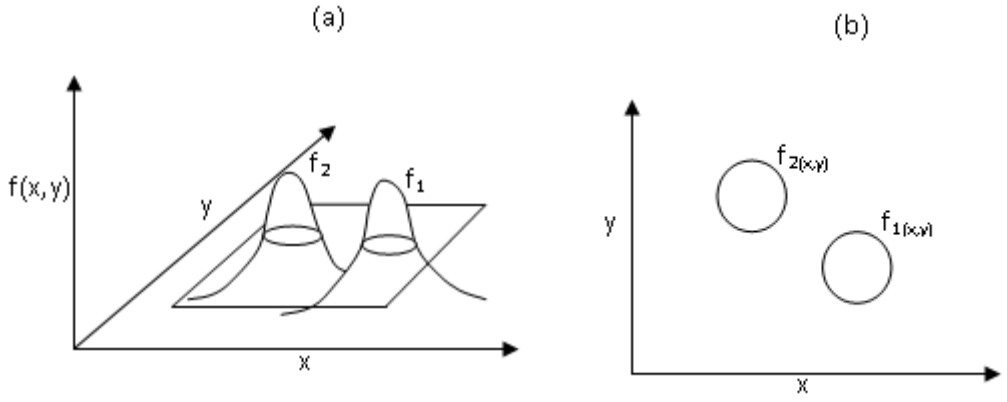


Figure 4. (a) Graph of bivariate probability distributions $f_1(x, y)$ and $f_2(x, y)$ for two stimulus S_1 and S_2 (b) Cross sections of Figure 4(a) resulting in contours of equal probability. Illustrations based on Ashby and Townsend (1986).

GRT assumes the observer partitions the perceptual region along each dimension and associates a response label with each partition, so that if prompted with a specific stimulus the response of the observer would be determined by which partition the stimulus is perceived to represent in the multidimensional space. An example of one type of these partitions can be seen in Figure 3 as delineated by two response criterion c_A for determining signal on the dimension A along the x-axis and c_B for determining signal on the dimension B along the y-axis.

An important aspect of MSDT studies is the use of complete identification experiments, as it provides a methodology to determine the relationships between the multiple component dimensions. These experiments create stimuli by combining factorially all levels of the different perceptual dimensions. For the simple case, assume two levels for both a brightness dimension (A_1 and A_2) and a loudness dimension (B_1 and B_2). Combining factorially the two perceptual dimensions would result in four stimuli, A_1B_1, A_1B_2, A2B1, and A_2B_2, with four possible response sets for an observer to choose from when identifying the stimulus, a_1b_1, a_1b_2, a_2b_1, and a_2b_2.

As an example of how GRT and the multidimensional space are structured, consider a hypothetical study examining the perceptual dimensions comprising a siren, specifically the

brightness of the siren and the loudness of the siren In terms of the multidimensional perceptual space, assume the component A_i of the brightness dimension is associated with x, and the component B_i of the loudness dimension is associated with y. The function $f_{A_iB_i}(x, y)$ is the perceptual distribution associated with each stimulus A_iB_i. By taking a cross section of each of the four f(x,y) functions mapped in the same multidimensional space, an equal probability contour is graphed in the two dimensional, x and y, for each f(x,y) function by taking a cross section of the f(x,y) functions. The result would look something like Figure 5.

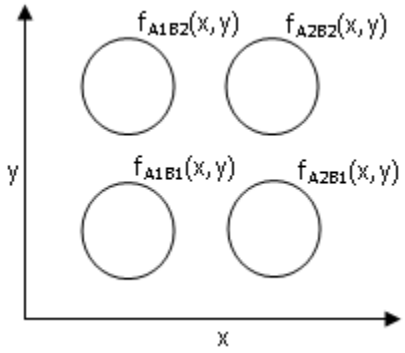


Figure 5. Equal probability contours for each of the four A_iB_i created by taking a cross section from each of the probability distribution functions$f_{A_iB_i}(x, y)$. Illustration based on Ashby and Townsend (1986).

For any of the stimuli presented to the observer, the perceptual event lies on the point (x, y) in the multidimensional space, for example the point in the f_{A2B2}(x, y) contour in Figure 5. Based on the internal criteria the observer decides which response to assign to the perceptual event. For example if the observer is shown the stimulus A_2B_2, the perceptual point may lie as shown as a dot in Figure 6 and based on their internal criteria (the dotted lines in Figure 6) the point is perceived as lower in magnitude than c_A and greater in magnitude than c_B, and therefore would respond a_1b_2, a hit in terms of signal detection. The location in the two-dimensional space of a stimulus in relation to the criteria set along the dimensions, will determine the categorical response of the observer. If the observer were to set their criteria, c_A and c_B, as shown in Figure

6, their response would be A_2B_2, as the magnitude of the stimulus event is greater than both the c_A and c_B criteria. For the present example, c_A would represent the criterion label for the magnitude of brightness dimension and c_B would be the criterion for the level of the loudness dimension.

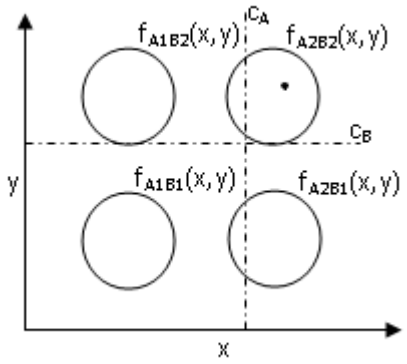


Figure 6. Equal probability contours with a point representing the perceptual event which occurs. Dotted lines are the criteria set by the observer. Illustration based on Ashby and Townsend (1986).

Perceptual independence, perceptual separability, and decisional separability. One of the strengths of GRT is that it specifies the relationships among the dimensions that define a complex stimulus. Specifically, the dimensions may be perceptually independent of one another, they may be perceptually separable (but not independent), or the decision regarding category membership along the dimensions may be separable or integral. These concepts from GRT are referred to *as perceptual independence, perceptual separability* and *decisional separability*, respectively. Perceptual independence is defined as the statistical independence of the perceptual effects of the components of a single stimulus. For instance, in the case of the two components of a siren, that the perceptual effect of the brightness of the siren does not affect the perceptual effect of the loudness of the siren indicates that the dimensions are perceptually independent.

This can be tested by showing that the associated covariance parameter between the two components is equal to zero.

Perceptual separability refers to whether the perceptual effect of one dimension is affected by the perceptual level of other dimension. That is, two dimensions A and B would be perceptually separable if all perceptual effects of A were unaffected by variations in the perceptual levels of B. An example of two perceptually separable dimensions are shape and hue (Maddox, 1992). Variations in the hue dimension do not affect the perception along the shape dimension. If the perceptual effects of dimension A are affected by variations in the perceptual levels of dimension B and the perceptual effects of dimension B are affected by the perceptual levels of dimension A, then the two dimensions are perceptually *integral*, also known as *perceptually integral*. An example of two perceptual integral dimensions are hue and brightness (Maddox, 1992). Variations along the hue dimension influence perception along the brightness dimensions.

If dimension A is perceptually separable from B and B is perceptually separable from A, the dimensions are said to be mutually perceptually separable. However if dimension A is not perceptually separable from dimension B but B is perceptually separable from A, (or vice versa, if dimension B is not perceptually separable from dimension A but dimension B is perceptually separable from dimension A), then the dimensions are said to have asymmetric perceptual separability. An example of asymmetric perceptually separable dimensions was reported by Ashby and Lee (1991), who observed that the orientation and size of a line had asymmetrical perceptual separability. When judging the size of the line the orientation did not affect the observer's perception however when judging orientation the size of the line did influence the

observer's perception of the orientation. Note however, that previous work by Garner and Felfoldy (1970) demonstrated that the size and orientation of lines were perceptually separable.

Decisional separability refers to whether the decision regarding the level of a stimulus on a dimension is dependent on the perceptual effects of the any other dimension. If for the bivariate case both levels of dimension A (A_1 and A_2, see Figure 7) the only decision boundary used by the observer was y_0, the decision criterion for B, then there would be decisional inseparability. However if for decisions about A, decision criteria x_0 was used and for decisions about B decision bound y_0 was used then the dimensions are considered decisionally separable. Figure 6 illustrates a hypothetical two level-two perceptual dimension multivariate mapping where decisional separability exists, the fact that the two decision criteria, x_0 and y_0, are parallel to the coordinate axis is evidence of decisional separability.

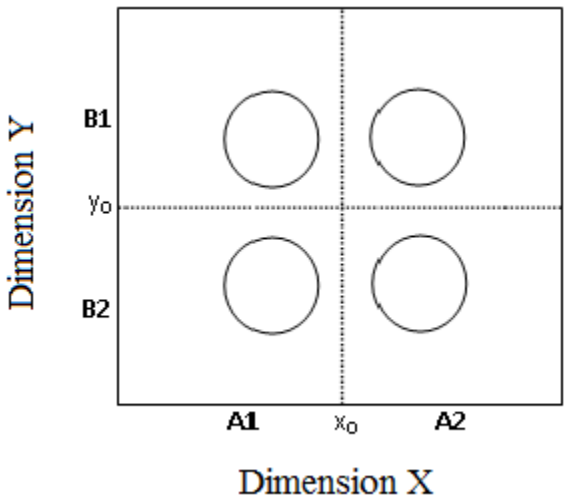


Figure 7. Example of two dimensional MSDT. Each circle represents a cross section of a three dimensional distribution of two normally distributed stimuli. The dotted lines in the figure are the decision criteria for dimension X and dimension Y. Illustration based on figure from Macmillan and Creelman (2005).

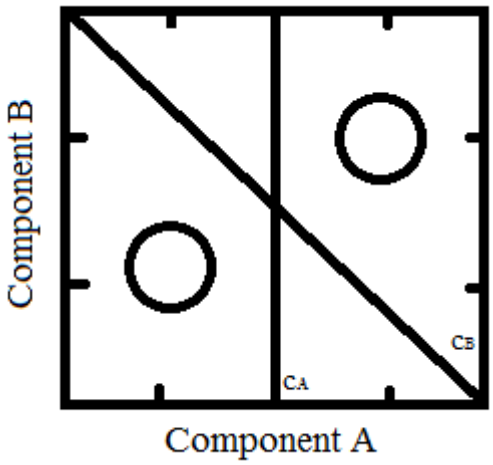


Figure 8. Based on figure from Maddox and Ashby (1996), showing a hypothetical set of equal probability contours and decision bounds

However, in the hypothetical example illustrated in Figure 8, decisional separability has been violated, evident in the fact the decisional criterion (c_B) for component B is diagonal rather than parallel with the x-axis.

Example of MSDT vs. SDT. MacMillan and Creelman (2005) describe an example of how MSDT can provide a better model that one-dimensional SDT for complex or compound stimuli. Using the combination of two perceptually independent dimensions, brightness and loudness (Figure 9a), MacMillan and Creelman describe various representations of the decision boundary of the observer in the two dimensional space. The null (non-signal) stimulus S1's equal probability contour is centered at the point (0, 0) on the graph and the signal stimulus S2's equal probability contour is center at the point (d'_x, d'_y). Figure 9b illustrates the use of a single component to make the decision, essentially transforming the compound problem back into a one-dimensional SDT problem. Using this decision bound (or criterion) the observer would respond affirmatively regarding signal presence if the stimulus magnitude exceeds the criterion level on the loudness dimension across all levels of brightness, in essence a single criterion result in one-dimensional SDT problem by removing the brightness dimension from the decision

processes. Figure 9c illustrates the maximum decision rule in which any stimulus greater in magnitude than both single decision bounds are included. The observer would respond 'yes' if both components (x, y) of the stimulus magnitude was greater than its respective decision bound (x decision bound, y decision bound). Figure 9d illustrates the minimum decision rule in which the area less than either or both decision lines is categorized as a non signal. If the observer perceives the stimulus magnitude across either dimension to be greater than their internal criterion for that dimension they will respond affirmatively. Figure 9e illustrates the Optimal or Diagonal rule which shows only the signal distribution, S2, not the null distribution, S1, centered at 0,0 as in the other illustrations. With the optimal rule the decision axis runs between the means of the signal distributions and has the criterion perpendicular to the decision axis. In the optimal rule both perceptual dimensions contribute to the decision process in a way that strong evidence in one dimension can compensate for weaker evidence in the other dimension. Of these decision rules the optimal rule performs the best. This is evident when examining the ROC curves of the four decision rules (Figure 10).

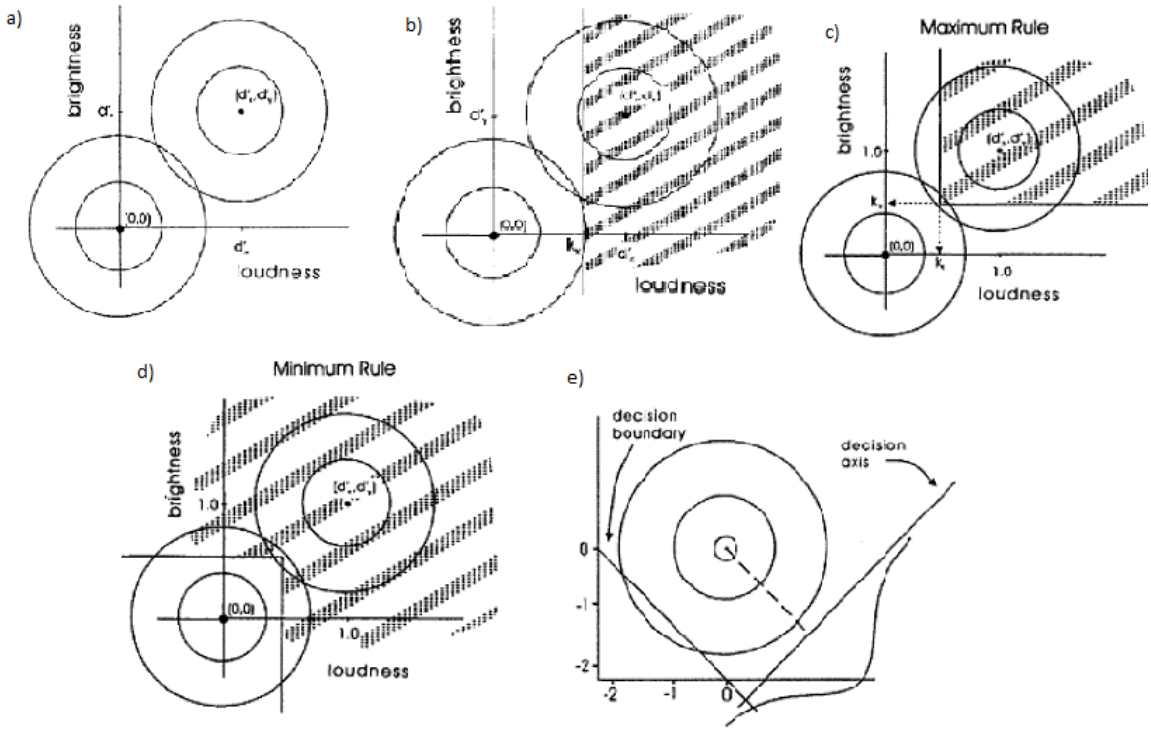


Figure 9. Multidimensional Decision Rules. The shaded area in diagrams above represent the area of multidimensional space in which the observer would respond a signal is present. a. Illustration of complex problem. b. Single Component c. Maximum Rule d. Minimum Rule e. Optimal (Diagonal) Rule. Illustrations from Macmillan and Creelman (2005).

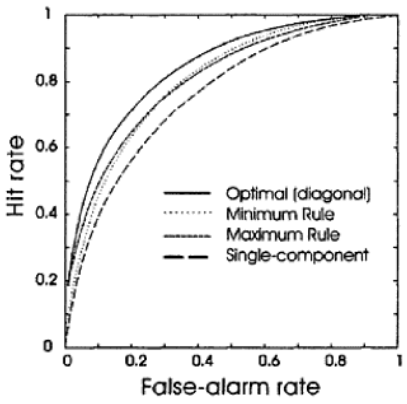


Figure 10. ROC Analysis of the four decision rules. Illustration from MacMillan and Creelman (2005).

For the ROC curves shown in Figure 10 the closer the bend in the curve is to the top left corner, indicating a high hit rate and low false alarm rate, the more ideal the signal detection performance of the observer. As evident in the figure, the three multidimensional rules

(optimal/diagonal, maximum rule, minimum rule) outperformed the single component or traditional SDT case, indicating that for complex stimuli MSDT may provide a better representation than SDT of how the observer categorized the multidimensional stimuli. The present investigation extended this logic to the case of the fuzzy signal detection model.

Fuzzy Signal Detection Theory

In most research and application of SDT the stimulus and response categories are mutually exclusive such that the state of the world is categorized as either present or absent, and the response from the observer is a decision that the signal is present or absent. Hancock, Masalonis and Parasuraman (2000), suggested that in many real world applications it is difficult define the true state of the world, and that this ambiguity adds to the difficulty of evaluating detection performance. Defining the categories of signal and non-signal can depend on contextual factors including the person(s) who define the possible state of the world and the decision alternatives. In some cases characteristic features of the signal itself is ambiguous and thus difficult to define because its perceptual properties contain elements of both signal and noise.

Szalma and Hancock (2013) describe one example of this phenomenon in the case of threat detection. Knowledge regarding the characteristics of threat items, such as a potential weapon hidden in a bag passing through x-ray baggage screening devices or an improvised explosive device (IED) placed on the side of a road, are often difficult to know in operational settings because the items are often ambiguous. Take for example a long thin tapered object viewed in a handbag passing through x-ray, this may be the threat object, a knife blade, or it may be a nail file, a non-threat object, as both share similar perceptual properties which indicate a signal. It is possible to know the identity of the item upon removing the object from the handbag

(i.e. a posteriori), however in operational settings the observer must make a decision for action (i.e. whether to manually search the bag) almost immediately upon inspection of the ambiguous object on the x-ray screen In situations such as these the forced categorization of stimulus definition into signal or noise required by SDT may result in loss of information regarding the true nature of the stimulus events. Parasuraman, Masalonis and Hancock (2000) proposed Fuzzy Signal Detection Theory (FSDT) to address this problem.

Fuzzy Signal Detection theory combines principles of fuzzy set theory with those of SDT. Fuzzy set theory, is a mathematical theory which permits elements of sets to simultaneously belong to multiple sets within the set universe. In contrast to traditional set theory, in which a set element has either full (one) or no (zero) membership in a set, fuzzy set theory allows for an element to belong simultaneously to a certain degree to one or more sets in the defined universe. In the case of FSDT, a stimulus in the environment can simultaneously belong to a degree to both the set 'signal' (s) and the set 'noise' (not s), indicating that the variable has elements or properties of both. Thus, FSDT allows for the non-mutually exclusive categorization of the signal and response dimensions, thereby eliminating the limitation regarding crisp categorization. That is, rather than a mutually exclusive classification of a stimulus as either a signal (s = 1) or noise (s = 0) event, in FSDT the stimulus has a degree of "signalness" represented by a value between 0 -1. For example, if the signal membership value of the stimulus is of 0.8, it would indicate that the stimulus has a moderately high degree of signal-ness but also retains characteristics of "noise-ness". Additionally, observer response can also defined as a fuzzy set, such that a response can belong to both the response sets "yes, signal represent" or "no, signal absent" to varying degrees (Parasuraman, Masalonis, and Hancock, 2000).

Mapping functions. The membership value in a fuzzy stimulus set is defined through the use of mapping functions. Mapping functions transform the physical value of the stimulus (or response options provided to the observer) into a membership value in the set signal **s** (or the set response **r**). Mapping functions can be derived based upon theoretical considerations, established empirical evidence, or regulatory standards (Parasuraman et al., 2000). As an example application, Parasuraman et al. (2000) described a mapping function for detection of potential midair aircraft collisions, a task that confronts that air traffic controllers (see Figure 11). In traditional SDT (crisp) terms, a signal is defined as an instance in which two aircraft are within 5 nautical miles horizontally from one other, as defined by the FAA. However, air traffic controllers often use different separation distances based on current context to make decisions regarding corrective action. The example mapping function which could be used for this scenario was $s = 1/ [1+ (a/10)^5]$, in which 'a' was the separation of the aircraft in nautical miles. For example if the two aircraft are 10 nm apart their signal membership0 would be s = 0.5 or if the aircraft were 9 nm apart their signal membership would be s = 0.65. Figure 11 illustrates this particular mapping function as well as its crisp counterpart.

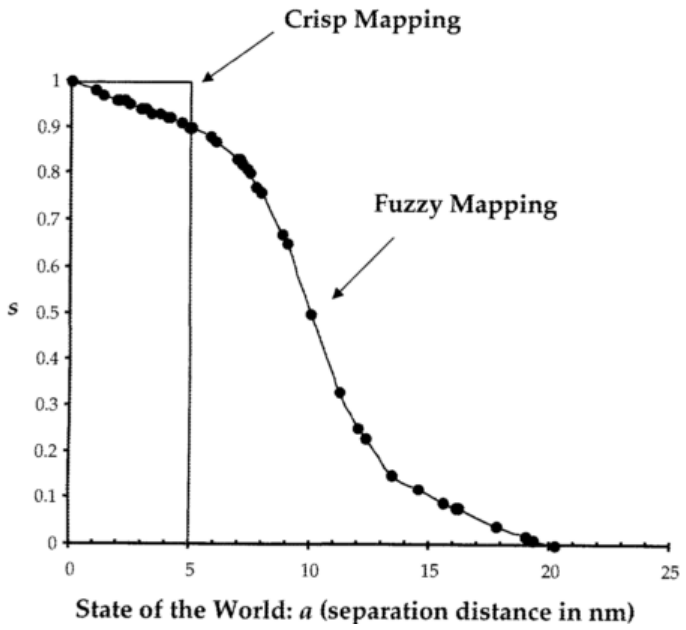


Figure 11. Fuzzy signal mapping function from Parasuraman et al. (2000).

FSDT procedure. Upon establishing mapping functions for the stimulus and response sets, the resulting s-r pairs are subsequently transformed using mixed implications functions that assign each s-r pair membership values in the four fuzzy criteria sets. In SDT the four potential performance outcomes are by transforming the observer response and stimulus value data via the truth table in Figure 1. In FSDT, the transformation occurs using the aforementioned mapping functions in order to define membership values in both the signal and response sets. These memberships values are then used to calculate the four possible performance outcomes for each trial (hits, misses, false alarms, and correct rejections) using a set of mixed implication functions described by Parasuraman et al. (2000). The definitions for each outcome are summarized in Equations 1-4, in which s is the membership value of the stimulus in the set signal, and r is the membership value of the response in the set 'response' ('yes' response). While membership values are bound by the range [0-1], they are not to be confused with proportions. Consider a single s-r pair with outcome membership H=0.11, M=0.24, FA = 0, and CR = 0.65, the hit

membership value of 0.11 is not the same as saying hits were 0.11 (or 11%) of the signal trials but rather that signal and response pair corresponding to a given stimulus presentation belongs 0.11 to the set hit, as it also belongs 0.24 to the set miss, does not belong at all to the set false alarm and belongs 0.65 to the set correct rejection.

Hit	$H = min\ (s, r)$	(1)
Miss	$M = max\ (s- r, 0)$	(2)
False Alarm	$FA = max\ (r- s, 0)$	(3)
Correct Rejection	$CR = min\ (1- s, 1- r)$	(4)

Note that as fuzzy sets a signal-response pairing can result in simultaneous membership in more than one outcome category. It should be noted that using the mixed implication functions with mutually exclusive membership (s = 1 or r = 1) or non-membership (s = 0 or r = 0) values, results in the four possible outcomes corresponding to the crisp SDT's truth table. Additionally, the outputs of Equations 1-4 can be used to calculate rates (proportions) for each fuzzy outcome category by summating over multiple trials using the formulas summarized in Equations 5-8 (Parasuraman, et al., 2000).

Hit Rate	$HR = \sum(min(s_i, r_i))/ \sum(s_i)$ for $i = 1$ to N	(5)
Miss Rate	$MR = \sum(max(s_i- r_i, 0)) / \sum(s_i)$ for $i= 1$ to N	(6)
False Alarm Rate	$FAR = \sum(\max(r_i - s_i,. 0))/ \sum(1-s_i)$ for $i = 1$ to N	(7)
Correct Rejection Rate	$CRR = \sum(min(1 - s_i, 1 - r_i))/ \sum(1-s_i)$ for $i = 1$ to N	(8)

Parasuraman, et al. (2000), suggest that the fuzzy hit and false alarm rates obtained using these procedures can be used with the well-established formulas for computation of sensitivity and bias. Several empirical studies have supported the application of these procedures (Masalonis & Parasuraman, 2003; Murphy, Szalma, & Hancock, 2004; Szalma et al., 2006; Szalma & Hancock, 2013).

ROCs and FSDT. In analyzing signal detection and diagnostic system data, estimation of a Receiver Operating Characteristic (ROC) is a useful technique for evaluation of decision

making systems (Swets & Picket, 1982). ROCs provide information regarding how the conservatism of the operator's decision criterion varies with hit and false alarm rates for a given sensitivity (Swets, Dawes and Monahan, 2000). Additionally the transformation of ROCs in to z-score form, $z_H = bz_F + a$, provides traditional performance measures of a detection or decision making task (e.g. β, d') and it can be used to determine whether data meets the statistical assumptions of SDT. The z-score form of an ROC can inform us whether the data meets the assumptions of normality and equal variance. A linear ROC function (z-score form) that indicates that the normality assumption of SDT has been met and if the slope of the function b = 1, then the equal variance assumption has been met.

ROC analysis has been of particular value in validating the use of FSDT. The FDST decision space was unspecified when the model was proposed, however practitioners of FSDT are to use traditional SDT performance measurements that are derived from SDT's decision space and thus require that SDT assumptions be met. It is assumed the structure of the FSDT decision space is of the same form as that of the traditional SDT decision space (Figure 2), but it is unclear how the fuzzy set membership in both signal and noise can be integrated into the traditional decision space representation, as the latter defines the categories as which use mutually exclusive distributions. Given ROC analysis can be used to test SDT assumptions from the outcomes of a SDT task, ROC has been used to validate application of SDT performance measures to FSDT as well as whether the outcome rates calculated via FSDT methodology result in data that conforms to traditional SDT assumptions.

Since the proposal of the FSDT in 2000, several studies have tested the tenability of the model by examining the underlying assumptions of FSDT. Empirically, FSDT has been successfully implemented in experiments ranging from discrimination of morphed images

(Szalma, Oron-Gilad, Saxton, & Hancock, 2006), temporal discriminations (Szalma & Hancock, 2013) to vigilance tasks (Stafford, Szalma, Hancock, & Mouloua, 2003). It has also been applied to archive air traffic control data (Masalonis & Parasuraman, 2003). Many of these studies used the ROC analysis approach to determine whether FSDT, as an extension of SDT, conformed to the fundamental statistical assumptions of the latter model thereby validating the use of traditional performance measures.

Szalma and O'Connell (2011), conducted a series of Monte Carlo simulations in order to test the assumptions of FSDT in a statistical simulation rather than an applied experimental setting. The results of the Monte Carlo simulations, indicated that FSDT analysis is capable of meeting both the normality and equal variance assumption required to use traditional SDT performance measures. Further, the studies indicated that sensitivity increases when applied to situations using categorical or continuous signal and/or responses, indicating FSDT may be a preferential choice compared to SDT for these situations.

Development of Multidimensional Fuzzy Signal Detection Theory

Through MSDT's modeling of a complex stimuli by its multiple component dimensions there is an improvement on the traditional SDT model because additional information about the stimuli is captured that would otherwise be lost by treating the unidimensional. Adding a multidimensional aspect to FSDT should provide a more accurate representation of the real world than SDT when a complex stimuli is used in conjunction with categorical or continuous signal or response sets.

The general premise of adding a multidimensional component to FSDT is to change the mapping function from a univariate to multivariate and to use this function to define the fuzzy

signal and response sets as specified in FSDT. For present work, this multivariate mapping function will be of the form:

$$S = \sum_{i=1}^{n} w_i \, s_i(a_i) + \mathrm{C} \qquad (9)$$
$$w_1 + \ldots + w_n = 1$$

in which $s_i(a)$ is the mapping function for each individual dimension, w_i is the weight associated with each dimension which determine the magnitudes of the dimensions that contribute to category membership, and *C* is a constant (or an interaction term in the case of inseparability). The term (a) is the stimulus value used to compute the signal membership value from the mapping function $s_i(a)$ or the response value used to compute the response membership value from the mapping function$r_i(a)$. Note that while $s_i(a)$ is used here to show the signal mapping function *S*, the same equations can be applied using the set $r_i(a)$ to develop a multivariate response mapping function *R*. Additionally the weights are bound by the value of one in order to preserve the scaling of the membership value of S between the values of 0 and 1.

MFSDT example. The procedure used with MFSDT is the same method used by FSDT with an additional step for developing the weights to use for each function. Using the same combination of two perceptually independent dimensions, brightness and loudness that were used in the section discussing MSDT, the first step is to derive unique signal membership mapping functions for each of the two dimensions. Figure 12 shows two plausible cases of mapping functions where one where the individual dimension mapping functions are similar and one where the mapping functions differ.

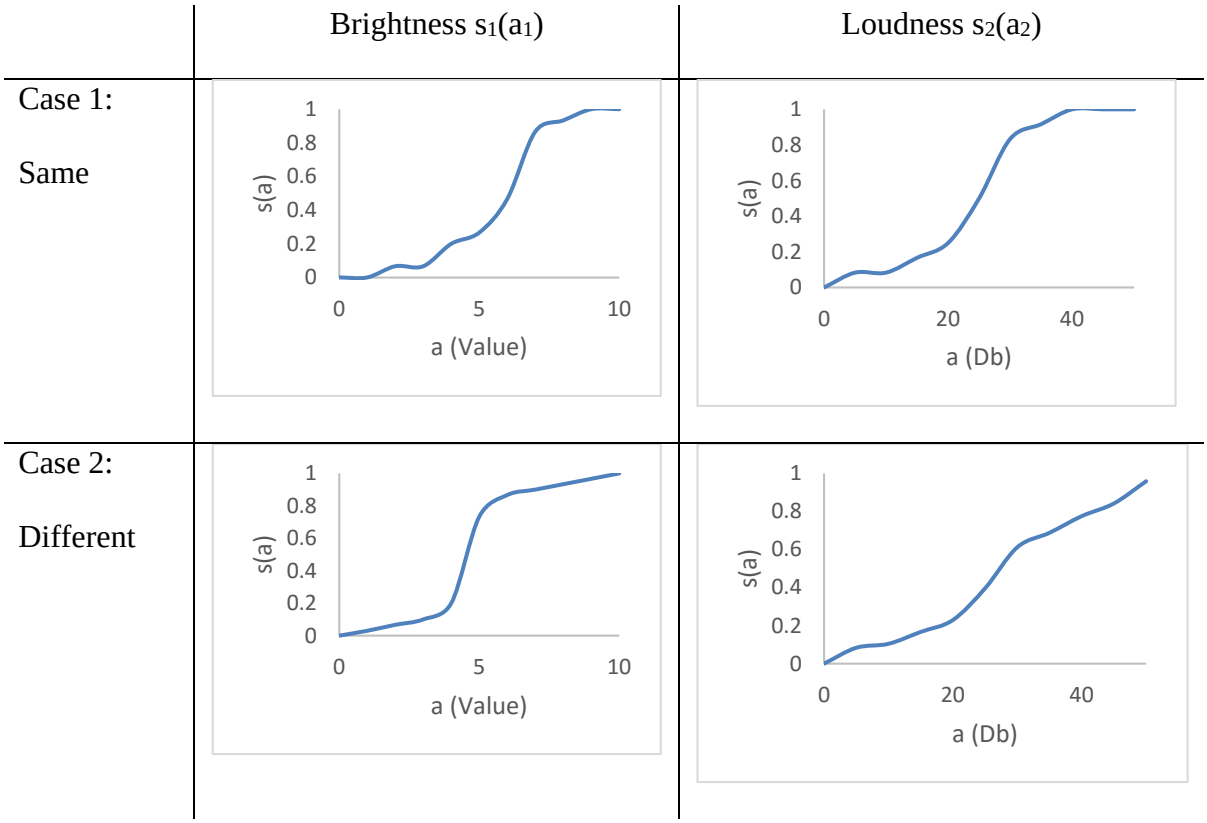


Figure 12. Two plausible scenarios of mapping functions for a complex signal comprised of brightness and loudness to be used by MFSDT.

In both scenarios the brightness function $s_1(a_1)$, a_1 is measured by value from the Munsell Color System and the loudness function $s_2(a_2)$, a_2 is measured by decibels. All of these mapping functions are similar to functions derived by other basic stimuli using psychophysical measurement methodologies (i.e. Method of Limits) in previous FSDT research (Szalma & Hancock, 2013; Szalma & O'Connell, 2011; Murphy, Szalma, & Hancock, 2004).

After defining the individual dimension mapping function, MFSDT requires the additional step of determining the weights of each dimension to be used in generating a combined mapping function of the two dimensions. These weights may be derived theoretically or empirically. Three different weight combinations were used for the purposes of this example. The first weight combination assumes that both dimensions have an equivalent effect on the signal perception and assign both individual functions a weight of 0.5, resulting in a combined

mapping function $S_{(a1,a2)} = 0.5s_1(a_1) + 0.5s_2(a_2)$, where $s_1(a_1)$ is the mapping function for the brightness dimension and $s_2(a_2)$ is the mapping function for the loudness dimension. The second weight combination assume the brightness dimension has a weight of 0.7 and the loudness dimension has a weight of 0.3, resulting in a combined mapping function of $S_{(a1,a2)} = 0.7s_1(a_1) + 0.3s_2(a_2)$. The third weight combination assume the brightness dimension has a weight of 0.3 and the loudness dimension has a weight of 0.7, resulting in the combined mapping function of $S_{(a1,a2)} = 0.3s_1(a_1) + 0.7s_2(a_2)$.

Since both individual dimension mapping functions are simply discretely defined signal membership values for eleven different values of a, the combined mapping function is the 121 possible combinations of pairings of the 11 brightness values (a_1) with the 11 loudness values (a_2) with their signal membership values defined by the combined mapping function equation. Figure 13 illustrates the three different weight combined mapping functions for the two individual mapping function cases.

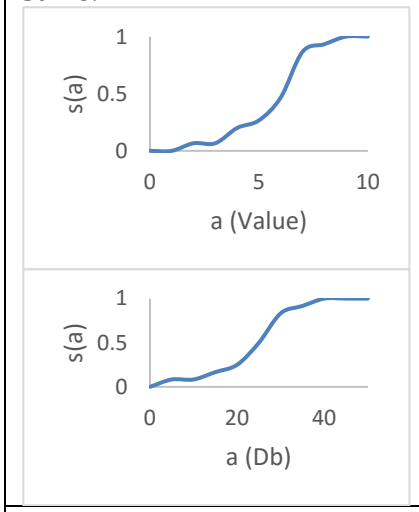
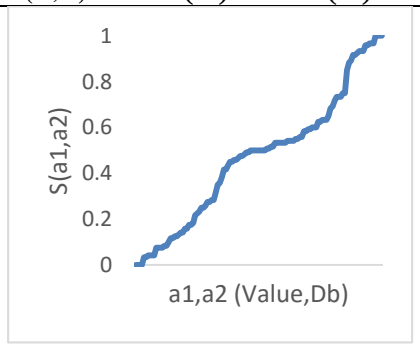
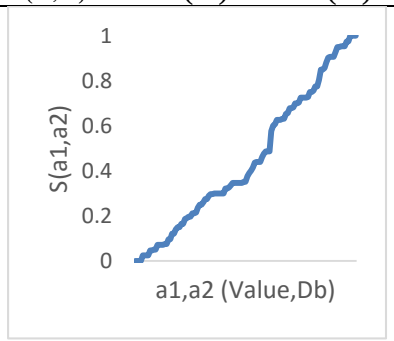
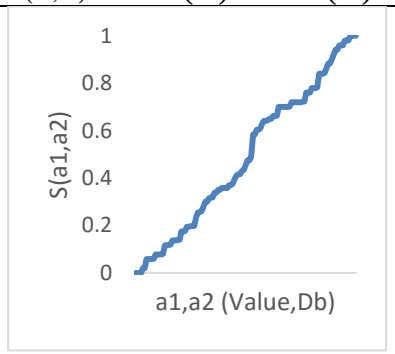
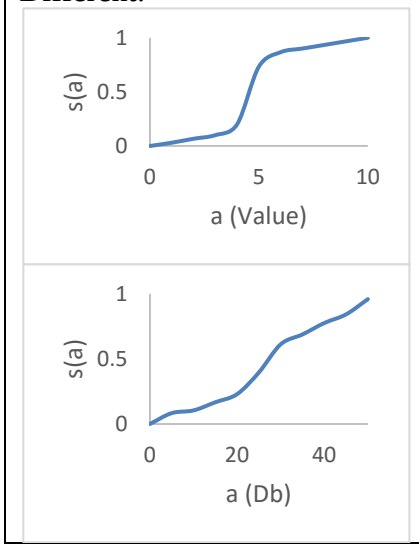
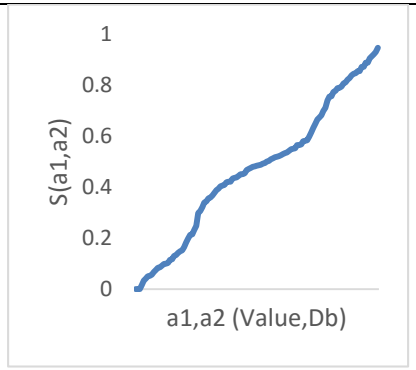
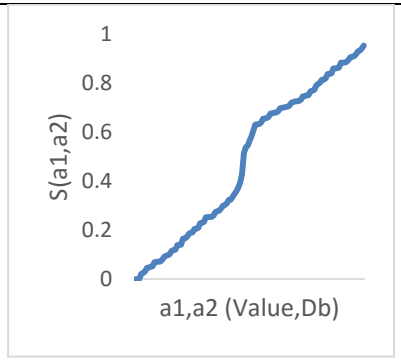
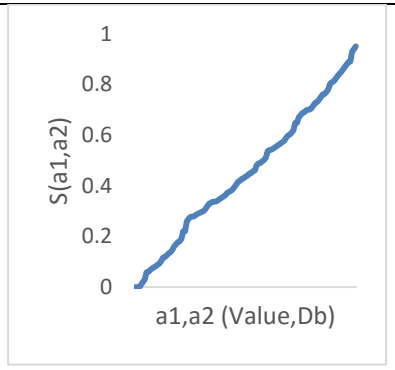
Individual Functions	$S_{(a1,a2)} = 0.5s_1(a_1) + 0.5s_2(a_2)$	$S_{(a1,a2)} = 0.7s_1(a_1) + 0.3s_2(a_2)$	$S_{(a1,a2)} = 0.3s_1(a_1) + 0.7s_2(a_2)$
Same: 			
Different: 			

Figure 13. Combined mapping functions for three potential weight options, 50/50, 70/30, and 30/70 for both individual functions scenarios.

With a combined mapping function defined, MFSDT is now used in the same as FSDT. In order to proceed with a FSDT/MFSDT analysis of data a response mapping function must also be defined. For this example, 10 categories of response were used, resulting in the mapping function shown in Figure 14.

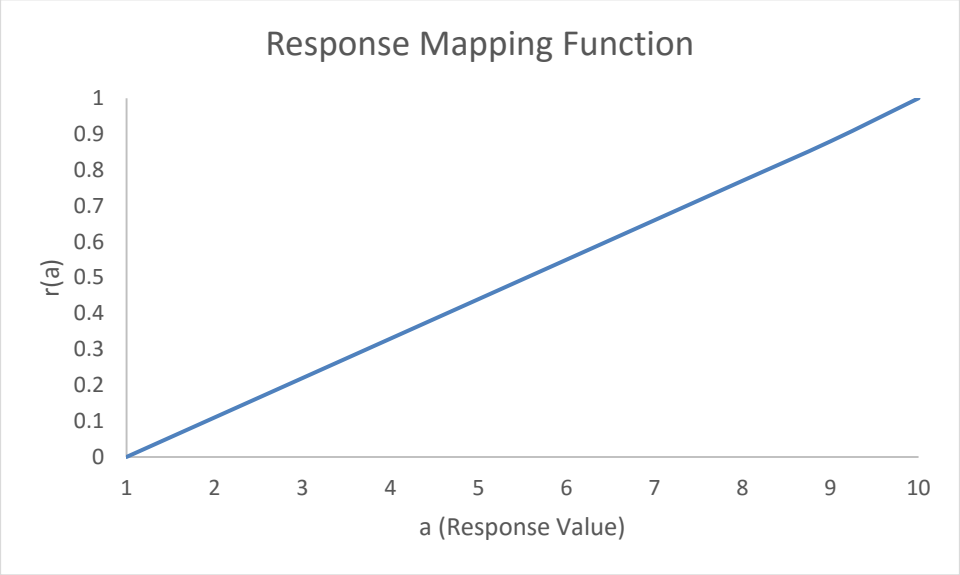


Figure 14. Response mapping function for 10 categories of response.

With the decision space unspecified in FSDT and therefore MFSDT as well, ROC analysis is a useful tool for understanding what effects manipulations may have when using either model. In order to compare the five possible models for this combined stimulus pair (Brightness only FSDT, Loudness only FSDT, and three Combined Brightness and Loudness MFSDT), an ROC was calculated using three different possible response bias types. The first response type (lenient) entailed an observer who responded always one or two categories higher than the what the "true" state of the world was for each possible stimuli The second response bias type (unbiased) entailed an observer responding always one category higher than the true state of the world or one category less than the true state of the world for each possible stimuli. The final response bias type (conservative) entailed an observer who responded always one or

two categories less than what the true state of the world was for each possible stimuli. Table 1 shows the resulting outcome rates using the rate functions defined in Parasuraman, Masalonis, and Hancock (2000) for the similar individual mapping function case and Table 2 shows the resulting outcome rates for the differing individual mapping function case. Note that both cases were computed using hypothetical data in order to illustrate the MFSDT procedures.

Table 1. *The four possible outcome rates for each of the five models using similar individual dimension mapping functions, calculated using FSDT for the individual dimensions and MFSDT for the three combined dimensions.*

	HR	MR	FAR	CRR
BRIGHTNESS				
Lenient	0.984589	0.015411	0.289946	0.710054
Unbiased	0.878425	0.121575	0.120652	0.879348
Conservative	0.710274	0.289726	0.010326	0.989674
LOUDNESS				
Lenient	0.897143	0.102857	0.194839	0.805161
Unbiased	0.745143	0.254857	0.065484	0.934516
Conservative	0.546857	0.453143	0	1
50/50 COMBINED				
Lenient	0.997026	0.002974	0.184071	0.815929
Unbiased	0.833475	0.166525	0.044409	0.955591
Conservative	0.585981	0.414019	0	1
70/30 COMBINED				
Lenient	0.989863	0.010137	0.21729	0.78271
Unbiased	0.838823	0.161177	0.065796	0.934204
Conservative	0.607848	0.392152	0.000091	0.999909
30/70 COMBINED				
Lenient	0.979241	0.020759	0.173335	0.826665
Unbiased	0.80401	0.19599	0.046385	0.953615
Conservative	0.565544	0.434456	0	1

Table 2. *The four possible outcome rates for each of the five models using differing individual dimension mapping functions, calculated using FSDT for the individual dimensions and MFSDT for the three combined dimensions.*

	HR	MR	FAR	CRR
BRIGHTNESS				
Lenient	0.934158	0.065842	0.221973	0.778027
Unbiased	0.789534	0.210466	0.084241	0.915759
Conservative	0.604658	0.395342	0.002883	0.997117
LOUDNESS				
Lenient	0.960337	0.039663	0.256914	0.743086
Unbiased	0.828429	0.171571	0.004032	0.995968
Conservative	0.65728	0.34272	0	1
50/50 COMBINED				
Lenient	0.99925	0.00075	0.197901	0.802099
Unbiased	0.839156	0.160844	0.050484	0.949516
Conservative	0.591549	0.408451	0	1
70/30 COMBINED				
Lenient	0.975804	0.024196	0.193906	0.806094
Unbiased	0.805496	0.194504	0.05719	0.94281
Conservative	0.571308	0.428692	0	1
30/70 COMBINED				
Lenient	0.999412	0.000588	0.223543	0.776457
Unbiased	0.851803	0.148197	0.064781	0.935219
Conservative	0.613278	0.386722	0	1

These rates are then used to calculate ROC statistics as well as traditional SDT measures of bias and sensitivity. Table 3 summarizes the results of an ROC on the models for the similar individual mapping function case and Table 4 summarizes the results of the ROC on the models for the differing individual mapping function case.

Table 3. *Displays the sensitivity and criterion bias calculate for the Brightness FSDT, Loudness FSDT and the three combined Brightness and Loudness MFSDTs for the similar individual mapping function case.*

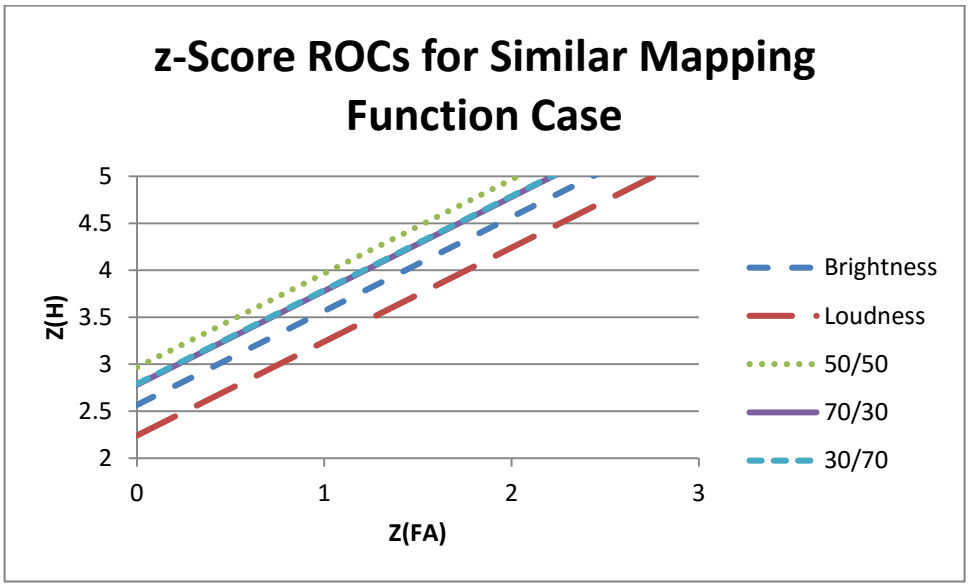
Model	A_z	a	b	Conservative β_{ln}	Unbiased β_{ln}	Lenient β_{ln}
Brightness FSDT	0.965	2.567	1.0	2.061	1.284	0.536
Loudness FSDT	0.943	2.241	1.0	2.185	1.560	0.911
50/50 Combined MFSDT	0.982	2.966	1.0	2.761	1.902	0.829
70/30 Combined MFSDT	0.975	2.780	1.0	2.533	1.687	0.720
30/70 Combined MFSDT	0.976	2.789	1.0	2.642	1.859	0.897

Table 4. *Displays the sensitivity and criterion bias calculate for the Brightness FSDT, Loudness FSDT and the three combined Brightness and Loudness MFSDTs for the differing individual mapping function case.*

Model	A_z	A	b	Conservative β_{ln}	Unbiased β_{ln}	Lenient β_{ln}
Brightness FSDT	0.95	2.320	1.0	2.112	1.468	0.883
Loudness FSDT	0.968	2.620	1.0	2.484	2.029	0.768
50/50 Combined MFSDT	0.982	2.972	1.0	2.756	1.877	0.775
70/30 Combined MFSDT	0.970	2.666	1.0	2.516	1.729	0.825
30/70 Combined MFSDT	0.979	2.876	1.0	2.614	1.712	0.690

The ROC of the similar mapping function case showed an increase in sensitivity for the 50/50 combined MFSDT analysis (A_z = 0.982), 70/30 combined MFSDT (A_z = 0.975) and the 30/70 combined MFSDT (A_z = 0.976) compared to the individual FSDT analysis of the brightness dimension (A_z = 0.965) and the individual FSDT analysis of the loudness dimension (A_z =

0.943). The ROC of the differing mapping function case also showed an increase in sensitivity for the three combined mapping MFSDT analysis (50/50 combined A_z = 0.982, 70/30 combined A_z = 0.97, 30/70 combined A_z = 0.979) compared to the individual FSDT analysis of the brightness dimension (A_z = 0.95) and the loudness dimension (A_z = 0.968). This result mirrors the Macmillan and Creelman example for MSDT, by treating the two dimensions together as a complex stimuli rather than analyzing each dimension individually results in an increase in an increase in sensitivity. Figures 15 and 16 illustrate how the combined MFSDT z-score form ROCs move toward the top left hand corner for similar and different mapping function cases respectively, in the same vein that the Macmillan and Creelman ROC (Figure 10) indicates an increase in sensitivity. Again it should be noted that these data are hypothetical and intended to illustrate the MFDST procedures.



\

Figure 15. Z-score form ROC for the 5 model types (Brightness FSDT, Loudness FSDT, 50/50 Combined MFSDT, 70/30 Combined MFSDT, and 30/70 Combined MFSDT)) for the similar individual mapping function case.

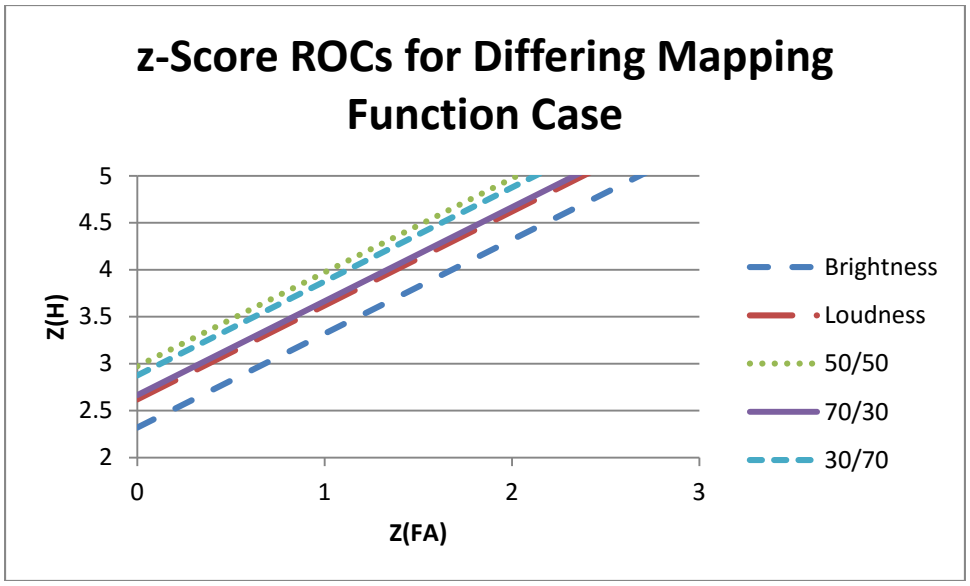


Figure 16. Z-score form ROC for the 5 model types (Brightness FSDT, Loudness FSDT, 50/50 Combined MFSDT, 70/30 Combined MFSDT, and 30/70 Combined MFSDT) for the differing individual mapping function case.

Assuming the observer uses decision strategy that is based on both dimensions rather than relying solely on a signal dimension of a complex stimuli, MFSDT may result in a better model for analyzing decision making tasks appropriate for FSDT analysis when complex stimuli are involved.

Creation of stimuli. To evaluate the influence of the separability of the dimensional stimuli on the derivation of mapping functions, two complex stimuli were created using basic psychophysical dimensions which are known from previous research to have decisional separability and in one case known to be perceptual separable and in the other case perceptually integral. Complex stimulus, C_1, will combine the two simple dimensions, hue and shape, which are known to be perceptually separable (Garner, 1977) and complex stimulus C_2 will combine the two simple dimensions, hue and saturation, which are known to be perceptually integral (Burns & Shepp, 1988). In order to facilitate the mapping functions from FSDT the use of the Munsell color system to specify hue and saturation allows for a single hue value to be used rather

than a multi-value combination of hue used in other color specifications, for example RGB. Additionally with the use of the Munsell color system, a single increase in any of the Munsell values is the same perceptual increase anywhere along the scale.

Munsell Color System. The Munsell Color System was developed by Albert H. Munsell as a method way to describe and categorize color. During multiple revisions of the color solid, a 3-d representation of all possible perceivable hue, saturation and brightness combinations, Munsell used professional experts in color, such as psychologists and optometrists, in order to measure perceptual differences of the three aspects of color in his system (Nickerson, 1976). The resulting irregular color solid has three axis representative of the different qualities of color, hue, saturation and brightness with a perceptually uniform step for any increase along an axis.

Hue was divided into five main colors, red (R), yellow(Y), green(G) ,blue(B) and purple(P) and five intermediary colors, red-yellow(RY), yellow-green(YG), green-blue(GB), blue-purple(BP), purple-red(PR). There are a total of 10 steps between each principal color, and the numerical value of a hue increased by 2.5 for each perceptual step. For example going from 5B (named blue) to 5P (named purple), has the following hues specified in between: 5B, 7.5B, 10B, 2.5BP, 5BP, 7.5BP, 10BP, 2.5P and 5P.

Brightness, referred to as value in the Munsell color system, ranges from 0 (black) to 10 (white). Numbers between 0 and 10 are various shades of grey which get lighter the higher the value.

Saturation, referred to as chroma in the Munsell color system, uses a scale step of 2 to indicate a perceptual step in saturation. The higher the chroma the more vivid the color. A color with 0 value/ 0 chroma would be black, a color with a 10 value/ 0 chroma, would be white, and any value between 0 and 10 with a chroma of 0 would be grey. While no upper limit exists for

chroma, there are however limitations to what can be reproduced on computer displays. Reproducible chroma values very between hue and value combinations, thus resulting in irregular shape of the color solid. For example, for 5P (purple) with value equal to 5 there are 14 possible chroma values (0-26), compared to 5B where at the same value level there are only 4 (0-6).

Newhalll, Judd, Nickerson (1943) developed what is known as the Munsell renotation system, which consists of colors which fall within the Macadam limits (MacAdam, 1935), or the limits of what colors humans can actually perceive. Additionally they defined each Munsell color in terms of the 1931 Commission International de l'Eclairage (CIE) color specification. With this data the Munsell renotation system can be reproduced on an electronic display using transformations of the CIE values into RGB vales needed for electronic color specification.

Overview of Sequence of Studies

A series of three studies were conducted in order to test the viability of Multidimensional Fuzzy Signal Detection Theory. Each study was conducted with two stimulus sets, one in which the stimuli were perceptually separable, i.e., shape and hue, and one in which the stimuli were perceptually integral, i.e., saturation and hue. The first study examined the perceptual properties of the dimensions in order to derive each individual component dimension's mapping function, as well as how the component dimensions interacted with one another, in order to derive potential weights for the complex mapping function. The second study focused on model fit and simulation. Model fit was used to determine what the individual component dimension weights for the complex mapping function and an ROC simulation similar to that reported by Szalma and O'Connell (2011) was conducted to test the viability of the derived complex mapping functions.

The final study was an empirical ROC conducted to test the tenability of the overall theory in an experimental setting.

CHAPTER 2: THE DERIVATION OF MAPPING FUNCTIONS (STUDY 1A)

Method

Participants. Thirty-two participants (11 males, 21 females) ranging in age from 17 to 28 (M = 19.15, SD = 2.13), participated in this study. All participants were students at the University of Central Florida and received course credit as compensation for participation.

Materials. The three individual dimensions of the two complex stimuli were evaluated separately, shape, hue, and saturation. The shape stimuli consisted of polygons ranging in number of sides from 6 to 36 (30 total) with the size of 3 inches (7.26 cm) at the maximum height and width. These were selected based on a previous FSDT study in which these stimuli were employed. In order to display the hue and the saturation, a 3 inch (7.62 cm) by 3 inch (7.62 cm) square was used, as it is perceptually separable from both perceptual properties and not in the range of shapes used for the polygon stimulus. The hue stimuli ranged from 2.5BP and 5P in the Munsell color system and was displayed at a brightness value of 5 and saturation value of 10 chroma. The saturation stimuli ranged from chroma 2 to 10 in the Munsell color system and were displayed at a brightness of 5 and hue of 7.5BP. Figure 17 displays the range in stimulus magnitude for the hue and saturation dimensions used in this study. This particular range was chosen as it resulted in the highest number of hue-saturation combinations which are displayable on a standard computer monitor.

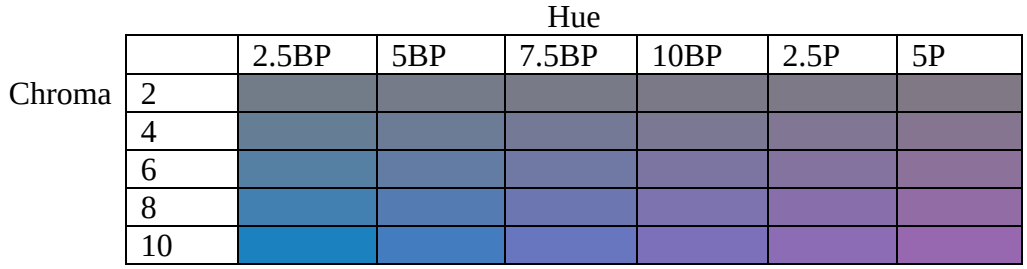


Figure 17. Grid display of the Munsell color system to be used in the experiments. The columns are the changes in hue from values 2.5BP to 5P. The rows are the changes in chroma (saturation) from 2 to 10. The value (brightness) is held constant at 5.

The LCD monitor displaying the stimulus was calibrated to Illuminant D50 (CIE, 2004) using a X-Rite CMUNDIS ColorMunki Display. A visual basic program was created to orchestrate the presentation of stimuli and the recording of participant responses. The same computer and monitor was used for each participant.

Procedure. Prior to beginning the experimental tasks each participant was tested for normal color vision using the Ishihara Color Vision Test (Ishihara, 1993). After evaluating the participants color vision the participant was instructed to view pairs of stimuli and respond as to which of the pair of stimuli best answered a question regarding similarity to the criterion stimulus (e.g., a circle). The stimulus pairs remained on the screen until the participant selected a response.

For the hue and saturation dimensions the stimulus pairs were comprised of a fully factorial combination of levels along each dimension. Thus, each level of hue was paired with each other level of hue for a total of fifteen pairs. Each level of saturation was paired with each other level of saturation for a total of ten pairs. However for the shape stimulus because the fully factorial combination of shape pairs was 465 pairs, the shape stimuli were grouped into sets of similarly sided polygons. The sets were as follows: {6,7,8}, {10,11,12}, {13,14,15}, {16,17,18}, {19,20,21}, {22,23,24}, {25,26,27}, {28,29,30}, {31,32,33}, {33,34,35,36}. In each case the number indicates the number of sides of the polygon stimulus. For each participant one element

of each set was randomly selected to be paired with a randomly selected element from each of the other sets. This resulted in a total of forty-five stimulus pairs.

Each stimulus was paired with questions which determined which of each presented pair was closer in similarity to the extreme values of the stimulus. Each question was asked both in the form "more like" and "less like". For the shape pairs, participants were asked either "Which shape is more like a circle?" or "Which shape is less like a circle?" For hue pairs, participants were asked "Which color is more blue?" or "Which color is less blue?". For saturation pairs, participants were asked "Which color is more vivid?" or "Which color is less vivid?" Each of the six stimulus/question combinations were presented three times each and presented separately. Participants were able to take a small break after each stimulus/question combination if needed.

Results

Two of the thirty two participants were color blind, as a result those responses were excluded from analysis of the hue and the saturation conditions, and are thus not included in the following calculations. Additionally, one participant did not understand the question for the saturation condition even after multiple attempts at clarification by the experimenter; therefore these responses are also not included in calculations for saturation.

Two methods were used to calculate potential mapping functions, one being the proportion of paired comparison trials on which a given stimulus was selected as more similar to the target value, and the second being Thurstone scaling which applies a normalization and scaling process to the paired comparison judgments.

Similarity judgment frequencies. For each participant the frequency of responding that a given stimulus level was determined to be greater in similarity to the extreme high values (circle for the shape stimulus, purple for the hue stimulus and vivid for the saturation stimulus) was

tallied. These frequencies were then summed over all participants. The summed frequencies were then divided by the total number of times that stimulus was presented in a pair. Table 5 shows for the stimulus saturation the frequency count for each level of saturation for all participants, the number of times the saturation level was presented, and the proportion of presentations where that saturation level was determined to be more vivid than the saturation level it was presented with for each level of saturation.

Table 5. *Frequency count, number of presentations, proportion presentations perceived as more vivid for each level of saturation.*

Chroma	**Frequency**	**# of Presentations**	**Proportion more vivid**
2	34	744	0.045
4	191	744	0.257
6	355	744	0.478
8	549	744	0.737
10	731	744	0.983

To transform these data this into a mapping function the $s_{saturation}(a)$ value for each a (chroma level) value was set equal to the proportion of presentations the chroma level was judged to be more vivid. The resulting mapping function is illustrated in Figure 18.

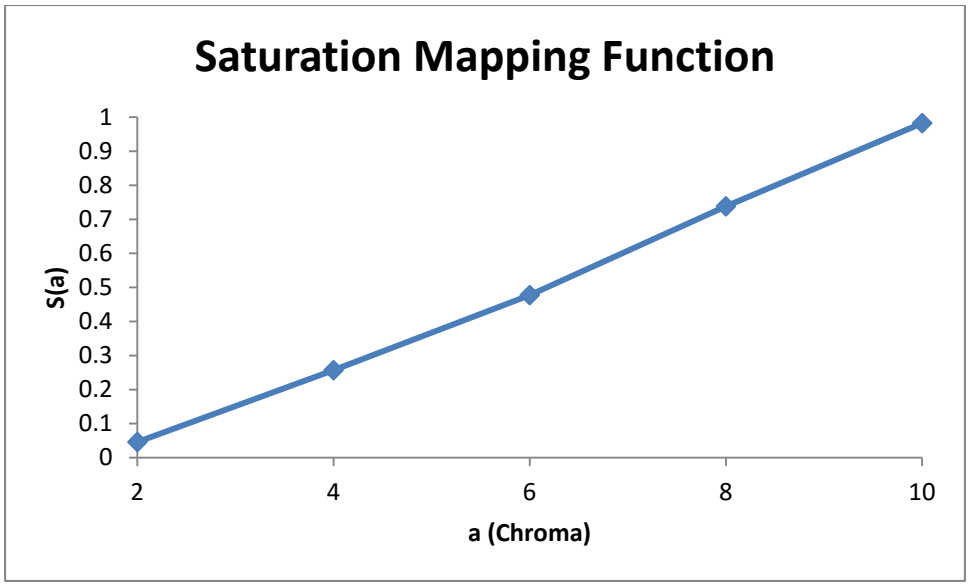


Figure 18. Saturation mapping function derived from the proportion each level was perceived as greater than another level.

Table 6 shows the frequency count for each level of hue for all participants, the number of times the hue level was presented, and the proportion of presentations of stimulus pairs on which the respective hue level was judged to be more purple.

Table 6. *Munsell value, frequency count, number of presentations, proportion presentations perceived as more purple for each level of hue (mapping function a value).*

Hue	**Munsell Value**	**Frequency**	**# of Presentations**	**Proportion more purple**
1	2.5BP	91	960	0.095
2	5BP	173	960	0.180
3	7.5BP	353	960	0.367
4	10BP	578	960	0.602
5	2.5P	758	960	0.789
6	5P	901	960	0.939

The mapping function was determined by setting the $s_{hue}(a)$ for each a (hue level) equal to the proportion of stimulus pairs on which a given hue value was judged to be more purple. Figure 19 illustrates the resulting mapping function.

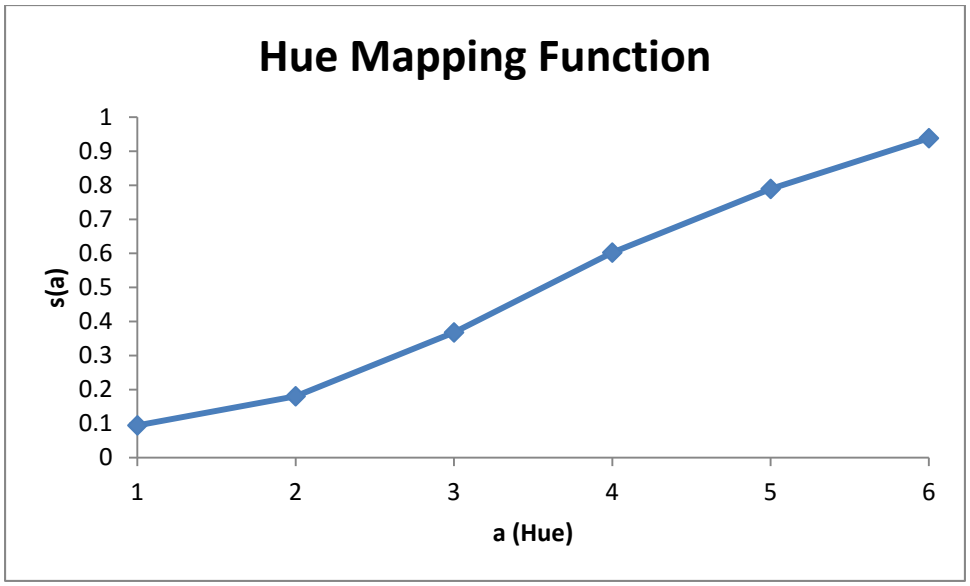


Figure 19. Hue mapping function derived from the proportion each level was perceived as greater than another level.

Table 7 shows for the stimulus shape the frequency count for each level of shape for all participants, the number of times the shape level was presented, and the proportion of presentations where each shape level was determined to be more like a circle than the stimulus to which it was compared.

Table 7. *Frequency count, number of presentations, proportion presentations perceived as more like a circle for each level of shape*

Number of Sides of Polygon	**Frequency**	**# of Presentations**	**Proportion more like a circle**
6	8	660	0.012
7	4	392	0.01
8	4	610	0.007
9	54	558	0.096
10	51	356	0.143
11	96	778	0.123
12	94	456	0.206
13	*144*	*591*	0.247
14	150	676	0.222
15	139	425	0.327
16	218	590	0.369
17	228	736	0.310
18	278	630	0.441
19	260	573	0.454
20	227	580	0.391
21	268	503	0.533
22	423	800	0.529
23	285	511	0.558
24	322	480	0.671
25	333	572	0.582
26	530	788	0.673
27	499	691	0.722
28	466	650	0.717
29	384	533	0.720
30	650	809	0.803
31	383	494	0.775
32	494	599	0.825
33	308	336	0.917
34	561	662	0.847
35	210	286	0.734
36	569	717	0.794

The mapping function was determined by setting the $s_{shape}(a)$ for each a (number of sides of shape) equal to the proportion of presentations where the level of shape was determined to be more like a circle. Figure 20 illustrates the resulting mapping function.

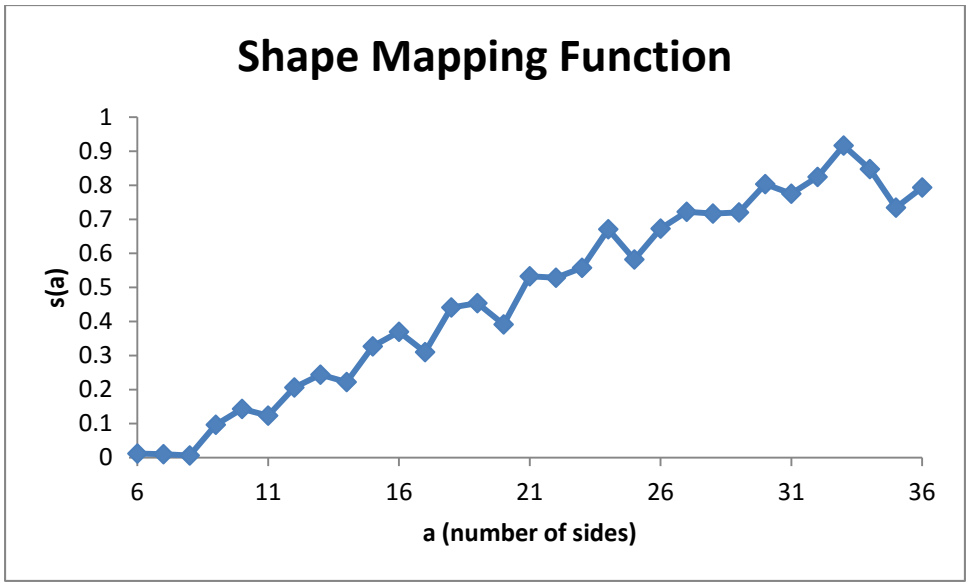


Figure 20. Shape mapping function derived from the proportion each level was perceived as greater than another level.

A problem with mapping function in Figure 20 is that as you increase in number of sides you may decline in signal membership, where logically the more sides a polygon has the more it looks like a circle. The decline in signal membership is likely due to differences in sampling of each stimulus, as the proportion estimate for each individual side is unlikely to have reached stability. In order to address this problem the analytic procedure was replicated but with the stimuli organized into the aforementioned sets. Table 8 shows the resulting table which includes the frequency each set was said to be greater than another set, the number of times each set was presented, and the proportion of times the set was determined to be more like a circle than the other sets.

Table 8. *Frequency count, number of presentations, proportion presentations perceived as more like a circle for each set*

Set	Frequency	# of Presentations	Proportion more like a circle
{6,7,8}	16	1728	0.009
{9,10,11}	201	1728	0.116
{12,13,14}	388	1728	0.225
{15,16,17}	585	1728	0.339
{18,19,20}	765	1728	0.442
{21,22,23}	975	1728	0.564
{24,25,26}	1183	1728	0.684
{27,28,29}	1340	1728	0.775
{30,31,32}	1514	1728	0.876
{33,34,35,36}	1648	1728	0.954

The mapping function was determined by setting $s_{shape}(a)$ equal to the proportion of trials the level of shape was perceived to be more like a circle for each a (number of sides) as determined by the set that a belongs. Figure 21 shows the resulting mapping function.

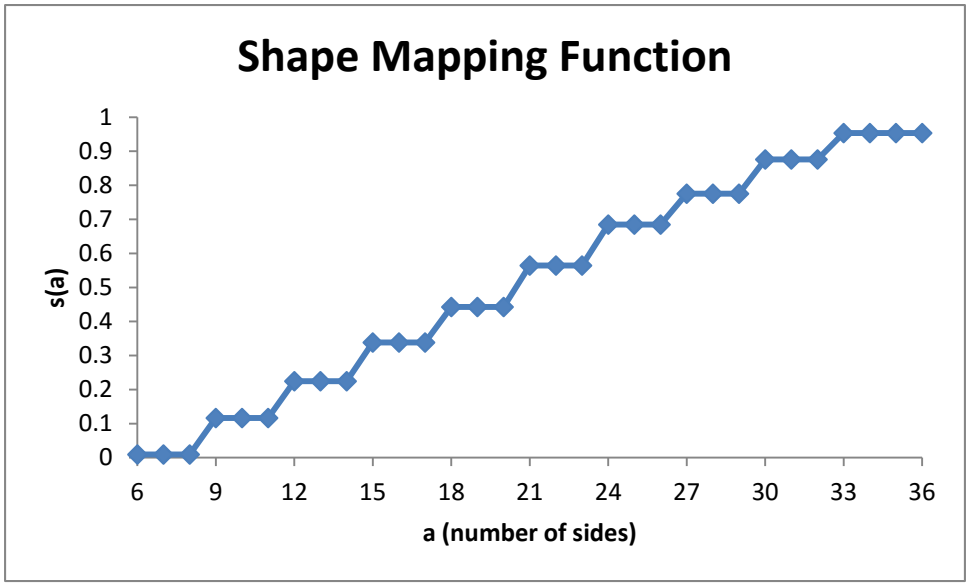


Figure 21. Shape mapping function derived from proportion each set was perceived as greater than another set.

Thurstone-Scale Method. For each participant the frequency for each stimulus level pairing was tallied and this frequency was then summed over all participants. The total frequency

for each stimulus pairing is then divided by the total number of trails on which that pairing was presented. This proportion is then converted into z-score using the cumulative normal probability function. The z-scores are then summed over all pairings for each stimulus level. The average z-score for each level is then adjusted by subtracting the lowest obtained z-score average from each of the other averaged z-scores. In cases where the proportion computed for a stimulus pairing was zero or one, an arbitrary adjustment of 0.0001 was added or subtracted respectively, before converting into z-score form, because the z-score of zero and the z-score of one evaluates into infinity. Table 9 shows for each saturation level the average z-score, z-score after scale adjustment, and proportion of highest z-score.

Table 9. *Average Z-score and Scaled Z-score produced by Thurstone scaling for each level of saturation*

Chroma	**Average Z-score**	**Z-score adjusted**	**Proportion of highest Z-score**
2	-1.715	0	0
4	-1.024	0.692	0.162
6	-1.288	0.427	0.100
8	1.467	3.183	0.744
10	2.56	4.276	1

The mapping function is then generated by setting $s_{saturation}(a)$ equal to the proportion of each saturation levels adjusted z-score out of the highest adjusted z-score. Figure 22 shows the resulting mapping function.

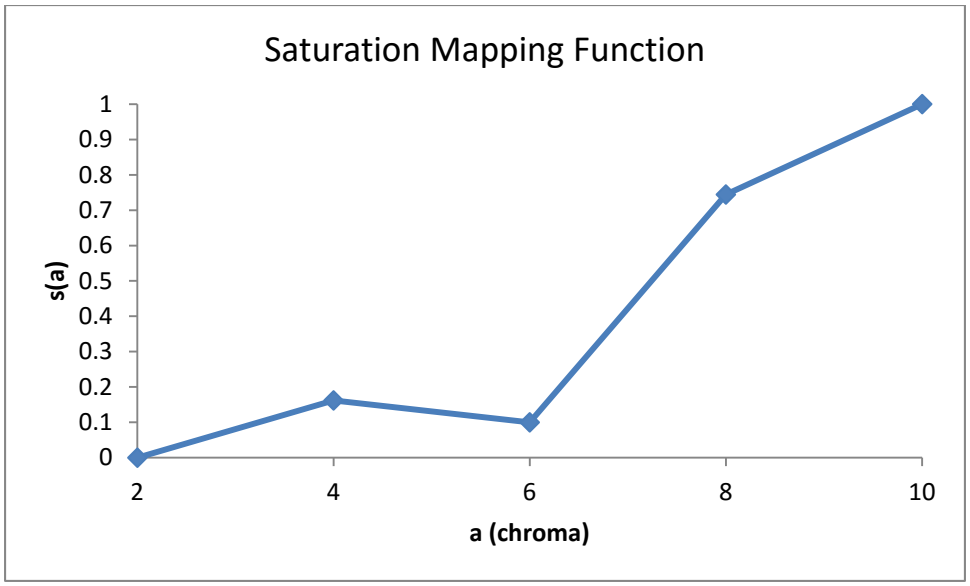


Figure 22. Saturation mapping function produced by Thurstone scaling.

Table 10 shows for each hue level the Munsell value for hue, average z-score, z-score after scaling adjustment and the proportion of highest z-score.

Table 10. *Munsell value, average z-score, adjusted z-score and the proportion of highest z-score for each level of hue (mapping function a value).*

Hue	**Munsell Value**	**Average Z-Score**	**Z-score adjusted**	**Proportion of highest Z-score**
1	2.5BP	-1.50	0.00	0.00
2	5BP	-1.25	0.247	0.079
3	7.5BP	-0.700	0.800	0.256
4	10BP	0.454	1.954	0.624
5	2.5P	1.201	2.702	0.863
6	5P	1.631	3.131	1.00

The mapping function is generated by setting $s_{hue}(a)$ equal to the proportion of each hue level's adjusted z-score to the highest adjusted z-score. Figure 23 shows the resulting mapping function.

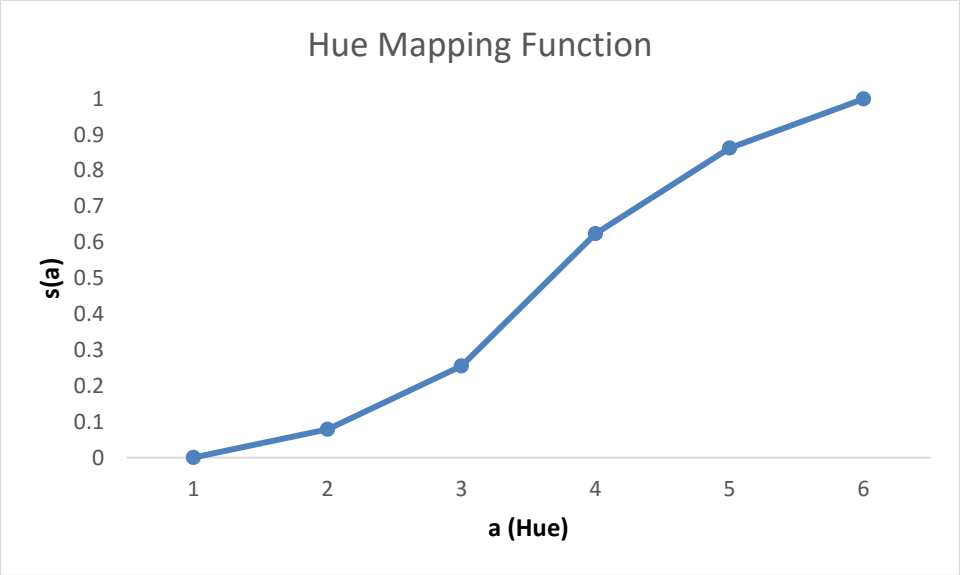


Figure 23. Hue mapping function produced by Thurstone scaling.

For shape stimuli treating each level pairing separately is problematic because of the large stimulus set for this dimension. While structural zeros and ones appear in the case of both of the hue and saturation dimensions, the majority of the data is neither. In the case of shape, only 254 of the 868 specific-pair responses are neither a one nor a zero. This is due to the fact some pairings were not sampled and presented to the thirty-two participants, and participants achieved a high success rate for determining correctly which shape in a pair was more like a circle. Collapsing across the shape stimulus sets helps mitigate the number of zero and one cells since each combination of *sets* were observed by all participants. The problem with high success rate still exists, as there are no errors for 40% of the set combinations, but this is preferable to the 71% zeros and ones when treating the shapes levels separately (see Appendix A for non-collapsed data). Table 11 shows for each shape set the average z-score, z-score after scale adjustment, and proportion of highest z-score.

Table 11. *Average z-score, adjusted z-score and the proportion of highest z-score for each shape set.*

Set	Average Z-score	Adjusted Z-score	Proportion of highest z-score
{6,7,8}	-2.776	0.00	0.00
{9,10,11}	-2.401	0.375	0.068
{12,13,14}	-2.063	0.712	0.129
{15,16,17}	-1.184	1.591	0.289
{18,19,20}	-0.678	2.098	0.381
{21,22,23}	0.635	3.411	0.619
{24,25,26}	1.403	4.179	0.759
{27,28,29}	1.930	4.707	0.855
{30,31,32}	2.299	5.074	0.921
{33,34,35,36}	2.732	5.508	1.00

The mapping function for each $s_{shape}(a)$ equal to the corresponding proportion of highest z-score for each a (number of sides) as determined by the set the a belongs to. Figure 24 plots the resulting mapping function.

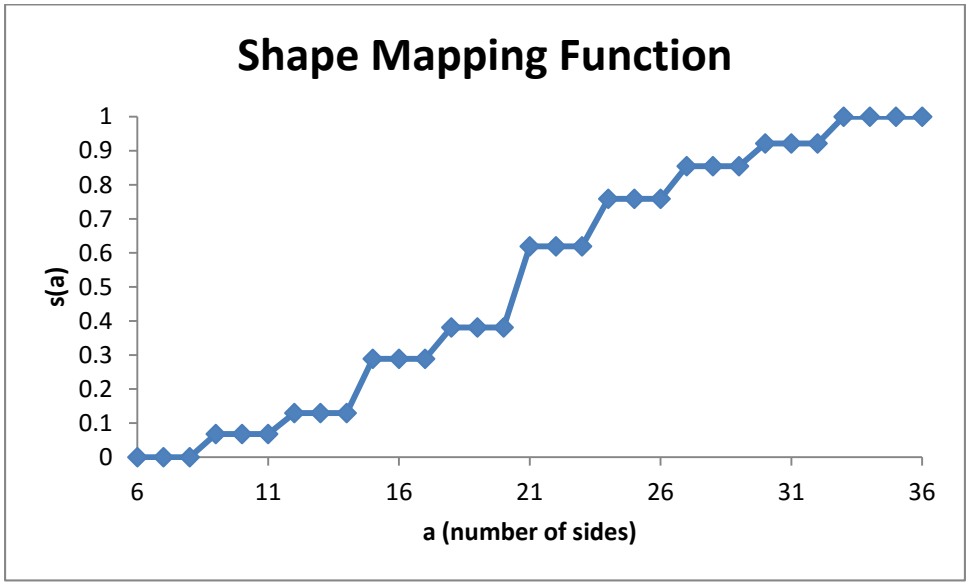


Figure 24. Shape mapping function determined by Thurstone scaling.

Discussion

Parasuraman et. al (2000) notes that mapping functions should be determined either empirically, theoretically or through regulatory standards, however there are currently no guidelines in existence as to how to choose between valid empirical methods that result in vastly differing mapping functions. While establishing guidelines for empirical methods is outside the scope of this work, decisions had to be made regarding which mapping functions to use for each of the individual component dimensions. Guilford (1928) suggested a shortcut in the steps of calculation required for Thurstone scaling by tabulating the total number of times a given response was selected regardless of what it was paired with and then proceeding with the analysis, rather than treating each response/comparison pair separately. This tabulating across responses is the same as the similarity judgment frequencies, thus the difference between the two methods of analyzing paired comparison data lies mainly in the fact that the Thurstone scaling does a z-score transformation and then scales the z-scores based on the lowest observed z-score value. In previous FSDT research the empirically derived mapping functions have been in the form of an ogive function, which is the form of the cumulative normal distribution function. Hence, a useful initial guideline for selecting a method for empirical mapping function derivation to use Thurstone scaling unless there is evidence indicating the process is not appropriate. This guideline is not suggested as the best way to determine mapping functions in general, but rather as an approach to mapping function selection between these two specific methods in this specific case.

The two generated hue functions shown in Figure 19 and Figure 23 are both of the ogive form. The difference between the two shapes is in the exaggeration of the inflections. Figure 19 has shallower changes at inflection points than Figure 23. Following the aforementioned

guideline for determining which function is more appropriate, the hue function developed by Thurstone scaling was therefore used.

The form of the shape function is difficult to determine from Figure 20 and Figure 24 since each individual shape's signal membership is determined by the its set value therefore it shares the same signal membership as at least two other shapes. By plotting the set values only once the shape the overall shapes of the functions are easier to observe. Figures 25 and 26 plot this for Figures 20 and 24 respectively.

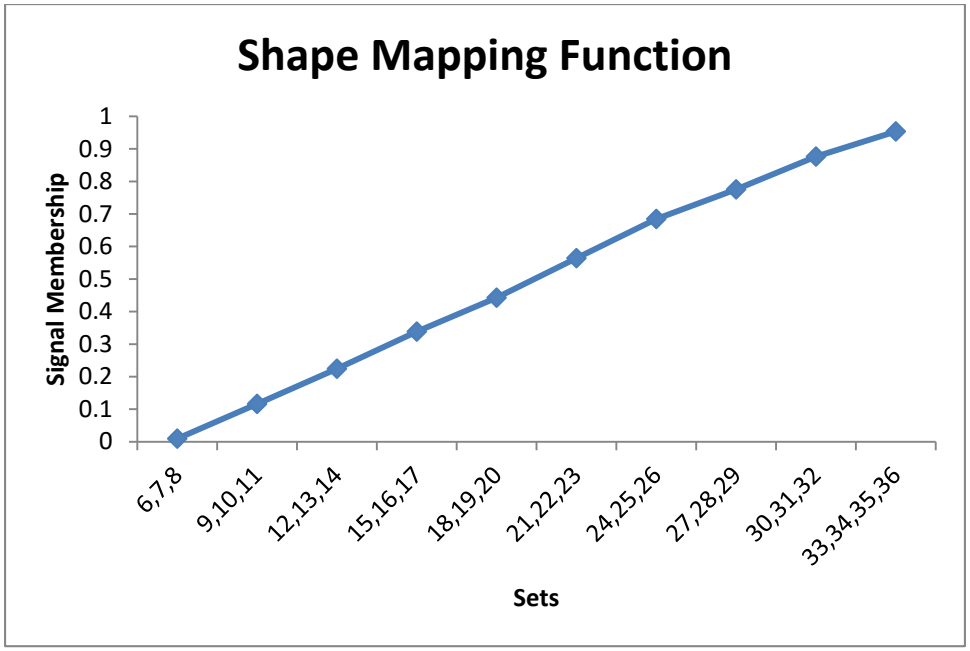


Figure 25. Single point form of Figure 20, the proportion greater method's mapping function for shape.

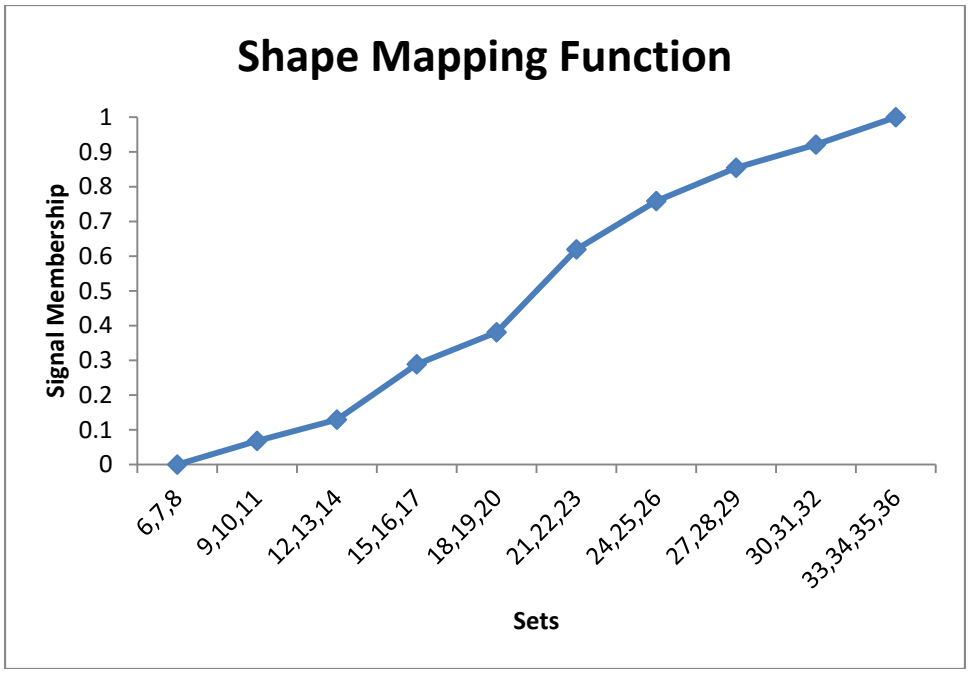


Figure 26. Single point form of Figure 24, the Thurstone scaling method's mapping function for shape.

As shown by Figure 25 the function appears to only have a slight noticeable curve at the higher shape values. Any other changes in slope of the line are difficult to discern visually, giving the appearance of a straight line until set {24,25,26}. Figure 26 shows a very shallow o-give function with the point for the set {15,16,17} slightly higher than one would expect of a typical o-give form, though still less than the next higher set up {18,19,20}. The less than ideal o-give function shape could be a product of the 40% perfect discrimination in the observed data. A set with less than perfect discrimination between comparisons with other sets would vary to a greater degree with a point following along an o-give function, than sets with more perfect discriminations between comparisons with other sets. With no evidence indicating that it is inappropriate to use Thurstone scaling, our guideline indicates the shape function should also be derived via Thurstone scaling should be used.

The saturation functions shown in Figure 18 and Figure 21 have the largest discrepancy in describing how saturation values translate into signal membership values. The data in Figure

18 can easily be described accurately with a straight line. The linear equation y = 0.1177x - 0.2065 fits the data with a $R^2 = 0.9983$. Furthermore, 81% of participants had perfect performance, and this subset data can be described by the linear equation y = 0.125x - 0.25 with a fit of $R^2 = 1$. Unlike the other two dimensions, hue and shape, Figure 21 shows that with saturation that the resulting mapping function from the Thurstone Scaling method fails to conform to the o-give function shape that is expected with data that is normally distributed. It predicts that a color with saturation of chroma = 6 has less signal membership than a color with more saturation (chroma = 4), which when dealing with saturation while controlling other color dimensions (brightness and hue) does not make logical sense. All things being equal an increase in saturation should result in an increase in signal-ness. This combined with the fact that the data can be described with a very high degree of accuracy by a linear equation, suggests that applying a normalization process inappropriate. Therefore using our decision guideline, the saturation function should be the function determined by similarity judgment frequencies.

CHAPTER 3: THE SPECFICATION OF STIMULUS DIMENSONALITY (STUDY 1B)

Method

Participants. Three women ranging in age between 22 and 45 volunteered to participate in this study. All had normal or corrected 20/20 vision and normal color vision. One woman was an undergraduate lab assistant in the same lab as the author and the other two women were acquaintances of the author.

Materials. Two complex stimuli were created for this experiment. Complex stimulus C_1 combined hue and shape, and complex stimulus C_2 combined hue and saturation. Since this experiment uses traditional MSDT tasks a factor complete factorial design of the extreme high and low levels of the individual dimensions of the complex stimuli comprised the four stimulus combinations A_1B_1, A_1B_2, A_2B_1, and A_2B_2 used. Figure 27 shows the combined stimuli used for this portion of the experiment. The combined stimuli used for C1 are defined as follows: A_1B_1 is low hue-hexagon, A_1B_2 is low hue-circle, A_2B_1 is high hue-hexagon, A_2B_2 is high hue-circle. The combined stimuli used for C2 are defined as follows: A_1B_1 is low hue-low saturation, A_1B_2 is low hue-high saturation, A_2B_1 is high hue low saturation A_2B_2 is high hue high saturation.

C1

	A1 Low Hue	A2 High Hue
B1 Hexagon		
B2 Circle		

C2

	A1 Low Hue	A2 High Hue
B1 Low Sat.		
B2 High Sat.		

Figure 27. Combined stimuli C1 and C2 used for weight generating portion of experiment. Note: not actual presentation size.

Each shape was 3inches x 3inces at its greatest width and height for both complex stimuli. The LCD monitor displaying the stimulus was calibrated to Illuminant D50 (CIE, 2004) using a X-Rite CMUNDIS ColorMunki Display. A visual basic program was developed to display stimuli and record participant responses. Each participant completed the experimental protocol on the same computer and monitor.

Procedure. The conditions employed in this experiment were based on those used in Maddox and Ashby (1996). Each complex stimulus was tested in ten separate conditions. The conditions were as follows:

1. Complete-Identification Condition: All four stimuli (A_1B_1, A_1B_2, A_2B_1, and A_2B_2) are used. The participants are asked to identify which of the four stimuli was presented.
2. Filtering A: All four stimuli (A_1B_1, A_1B_2, A_2B_1, and A_2B_2) are used. The participants are asked to identify the level of stimulus on dimension A.
3. Control Condition A-1: The stimuli used are the complex stimuli where dimension B is at non-signal level (A_1B_1, A_2B_1). The participants are asked to identify the level of the stimulus on dimension A.
4. Control Condition A-2: The stimuli used are the complex stimuli where dimension B is at signal level (A_1B_2, A_2B_2). The participants are asked to identify the level of the stimulus on dimension A.
5. Redundancy Condition A-: The stimuli used are the complex stimuli A_1B_1 (both dimensions are at non-signal level) A_2B_2 (both dimensions are at signal level). The participants are asked to identify the level of the stimulus on dimension A.
6. 4: Redundancy condition A+: The stimuli used are the complex stimuli A_1B_2 (dimension A is at non-signal level and dimension B is at signal level) and A_2B_1 (dimension A is at

signal level, dimension B is non-signal level). The participants are asked to identify the level of the stimulus on dimension A.

7. Filtering B: All four stimuli (A_1B_1, A_1B_2, A_2B_1,and A_2B_2) are used. The participants are asked to identify the level of stimulus on dimension B.
8. Control Condition B-1: The stimuli used are the complex stimuli where dimension A is at non-signal level (A_1B_1, A_1B_2). The participants are asked to identify the level of the stimulus on dimension B.
9. Control Condition B-2: The stimuli used are the complex stimuli where dimension A is at signal level (A_2B_1, A_2B_2). The participants are asked to identify the level of the stimulus on dimension B.
10. 7-8: Redundancy Condition B+ and B-: The stimuli are the same as used in conditions A- and A+ but participants are asked to identify the level of the stimulus on dimension B.

Table 12 summarizes the generic structure of the conditions by showing which stimuli are used for each condition as well as what the participant is asked to identify.

Table 12. *Chart indicating which stimuli combinations are used for each condition as well as the identification to be made by the participant.*

Condition	**A1B1**	**A1B2**	**A2B1**	**A2B2**	**Identify**
Identification	X	X	X	X	Complete Identification
Filtering A	X	X	X	X	Level of Dimension A
Control Condition A-1	X		X		Level of Dimension A
Control Condition A-2		X		X	Level of Dimension A
Redundancy Condition A-	X			X	Level of Dimension A
Redundancy Condition A+		X	X		Level of Dimension A
Filtering B	X	X	X	X	Level if Dimension B
Control Condition B-1	X	X			Level of Dimension B
Control Condition B-2			X	X	Level of Dimension B
Redundancy Condition B-	X			X	Level of Dimension B
Redundancy Condition B+		X	X		Level of Dimension B

Table 13 replicates Table 12 but specified for the complex stimulus C_1, hue and shape.

Table 13. *Experimental stimuli used in each condition for Complex stimuli C_1. The x indicates that pairing of stimuli is used for that condition. Blue is defined as the low-hue and purple as the high-hue.*

Condition	**Blue Hexagon**	**Blue Circle**	**Purple Hexagon**	**Purple Circle**	**Identify**
Identification	X	X	X	X	Blue Hexagon, Blue Circle, Purple Hexagon, or Purple Circle
Filtering A	X	X	X	X	Color(Blue or Purple)
Control Condition A-1	X		X		Color (Blue or Purple)
Control Condition A-2		X		X	Color (Blue or Purple)
Redundancy Condition A-	X			X	Color (Blue or Purple)
Redundancy Condition A+		X	X		Color (Blue or Purple)
Filtering B	X	X	X	X	Shape (Hexagon or Circle)
Control Condition B-1	X	X			Shape (Hexagon or Circle)
Control Condition B-2			X	X	Shape (Hexagon or Circle)
Redundancy Condition B-	X			X	Shape (Hexagon or Circle)
Redundancy Condition B+		X	X		Shape (Hexagon or Circle)

On any given trial a randomly selected stimulus appropriate for its given condition was displayed on the screen for 300ms after which response options appeared on screen. The response options remained on screen until the participants selected a response. The identification made by the participant and the participant's response time was recorded for all conditions. The Identification condition and the Filtering conditions consisted of 90 trials each. The control and redundancy conditions consisted of 45 trials each. In each session the participants completed all ten conditions for complex stimulus C1 and all ten conditions for complex stimulus C2. Short breaks

occurred between each condition. For each participant the first session consisted of 10 extra practice trials for the identification and filtering tasks and 5 extra practice trials for the control and redundancy conditions. Participants completed a total of eight sessions. The condition order and session order (where session refers to a set of conditions) was randomized across participants.

Results

Complex Stimulus 1 - Shape/Hue. Analysis of the identification condition for complex stimulus C_1 resulted in the confusion matrices for Participants 1-3 in tables 14-16, respectively.

Table 14. *Confusion Matrix on Complex Stimulus C_1 (Shape/Hue) for Participant 1*

Stimulus	Response			
	a_1b_1 (Blue Hexagon)	a_1b_2 (Blue Circle)	a_2b_1 (Purple Hexagon)	a_2b_2 (Purple Circle)
a_1b_1 (Blue Hexagon)	153	1	5	1
a_1b_2 (Blue Circle)	0	172	0	2
a_2b_1 (Purple Hexagon)	4	0	198	0
a_2b_2 (Purple Circle)	0	7	2	175

Table 15. *Confusion Matrix on Complex Stimulus C_1 (Shape/Hue) for Participant 2*

Stimulus	Response			
	a_1b_1 (Blue Hexagon)	a_1b_2 (Blue Circle)	a_2b_1 (Purple Hexagon)	a_2b_2 (Purple Circle)
a_1b_1 (Blue Hexagon)	180	0	4	0
a_1b_2 (Blue Circle)	0	168	1	0
a_2b_1 (Purple Hexagon)	1	1	162	0
a_2b_2 (Purple Circle)	0	1	1	201

Table 16. *Confusion Matrix on Complex Stimulus C_1 (Shape/Hue) for Participant 3*

Stimulus	Response			
	a_1b_1 (Blue Hexagon)	a_1b_2 (Blue Circle)	a_2b_1 (Purple Hexagon)	a_2b_2 (Purple Circle)
a_1b_1 (Blue Hexagon)	166	4	5	0
a_1b_2 (Blue Circle)	2	162	1	0
a_2b_1 (Purple Hexagon)	1	0	175	1
a_2b_2 (Purple Circle)	0	4	5	194

All three confusion matrices are fairly sparse with respect to errors, indicating participants achieved a high degree in accuracy for identifying which of the four shape/hue stimuli was presented. Grtools package (Soto & Zheng, 2015) in the software R was used to analyze the fit of the identification data to a hierarchy of traditional GRT models which indicate the separability and independence of the data. Table 17 lists the various models fitted as well as their AIC fit statistic for all 3 participants. The model with the lowest AIC indicates the best fit. Figures 28-30 plot the best fitting model generated by grtools for each of the participants.

Table 17. *Traditional GRT models and fit statistics for All Participants.*

Model	Participant 1 AIC	Participant 2 AIC	Participant 3 AIC
{PI, PS, DS}	238.615	132.447	255.242
{1_Rho, PS, DS}	238.447†	122.261†	251.309†
{PS,DS}	256.118	142.081	270.393
{PI, PS(A), DS}	240.754	144.720	265.194
{1_Rho, PS(A), DS}	244.918	132.741	262.180
{PS(A), DS}	279.139	175.219	296.233
{PI,PS(B),DS}	244.813	142.678	261.504
{1_Rho, PS(B), DS}	247.002	133.518	259.551
{PS(B), DS}	280.545	181.922	294.716
{PI,DS}	252.831	154.370	279.203
{1_Rho, DS}	261.198	166.285	291.168
{DS}	339.314	235.209	356.211

Note: PI - Perceptual Independence, PS - Perceptual Separability, DS - Decisional Separability, †- Best fitting model

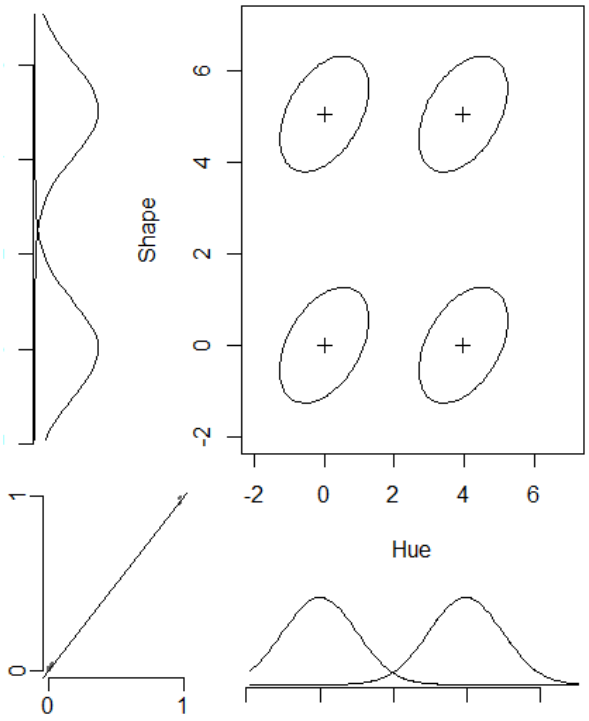


Figure 28. Best GRT Model Fit - {1_Rho, PS, DS} of Hue and Shape for Participant 1.

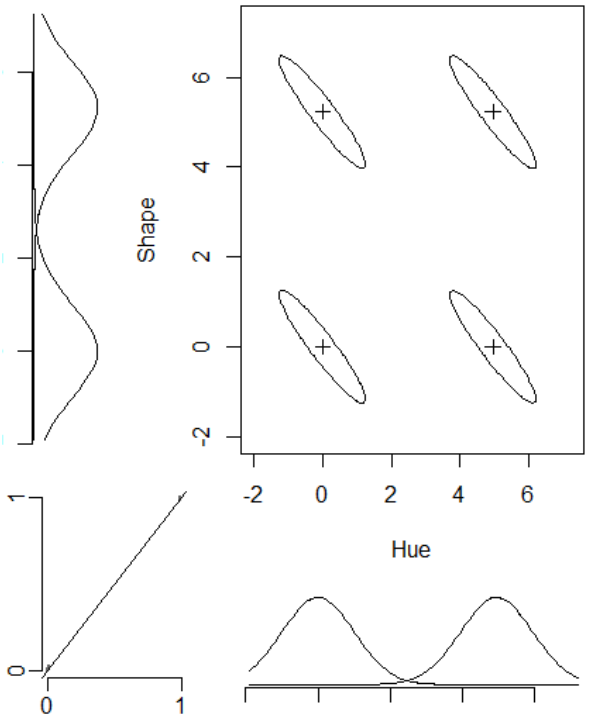


Figure 29. Best GRT Model Fit - {1_Rho, PS, DS} of Hue and Shape for Participant 2.

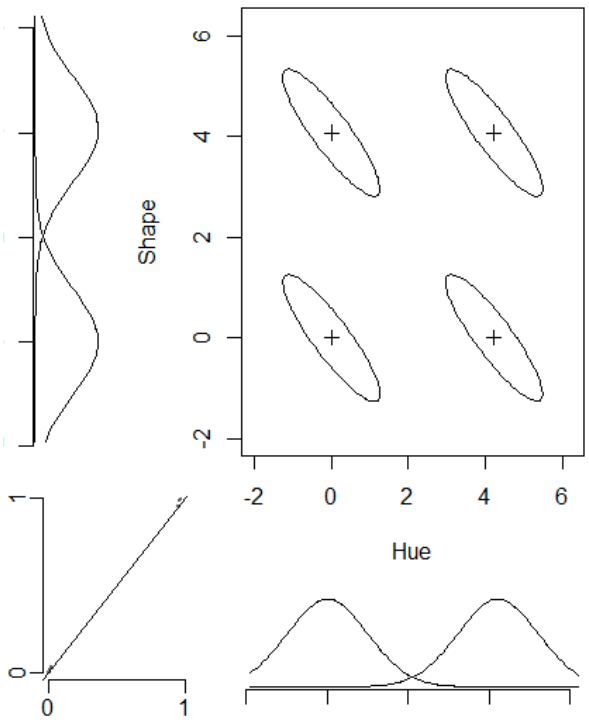


Figure 30. Best GRT Model Fit - {1_Rho, PS, DS} of Hue and Shape for Participant 3.

For all three participants the best fitting GRT model was {1_Rho, PS, DS}, indicating that for all three there is perceptual separability (PS) and decisional separability (DS). The perceptual separability is also evident when examining the plots and the fact the paired z-score means, the center (x,y) coordinates for each $f(x, y)$, create a square pattern. In other words, both low level shape means (y-coordinates on bottom pair) are equivalent, both high level shape means (y-coordinates on top pair) are equivalent, both low hue means (x-coordinates on left pair) are equivalent and both high level hue means (x-coordinates on right pair) are equivalent. The 1_Rho (which reflects the relationship between the two dimensions) in the model indicates that perceptual independence was violated. The violation to perceptual independence is evident in the slanted $f(x, y)$s in the plots rather than circular $f(x, y)$ functions.

Means and standard deviations for both correct and error frequencies as well as response times for the filtering, control and redundancy tasks for all three participants are provided in Appendix B.

Complex Stimulus 2 - Saturation/Hue. Analysis of the identification condition for complex stimulus C_2 resulted in the confusion matrixes for participants 1-3 in Tables 18-20, respectively.

Table 18. *Confusion Matrix on Complex Stimulus C_2 (Saturation/Hue) for Participant 1*

Stimulus	Response			
	a_1b_1 (Dull Blue)	a_1b_2 (Vivid Blue)	a_2b_1 (Dull Purple)	a_2b_2 (Vivid Purple)
a_1b_1 (Dull Blue)	178	1	1	0
a_1b_2 (Vivid Blue)	0	183	1	0
a_2b_1 (Dull Purple)	8	0	173	0
a_2b_2 (Vivid Purple)	0	3	1	171

Table 19. *Confusion Matrix on Complex Stimulus C_2 (Saturation/Hue) for Participant 2*

Stimulus	Response			
	a_1b_1 (Dull Blue)	a_1b_2 (Vivid Blue)	a_2b_1 (Dull Purple)	a_2b_2 (Vivid Purple)
a_1b_1 (Dull Blue)	165	1	22	0
a_1b_2 (Vivid Blue)	0	169	1	0
a_2b_1 (Dull Purple)	11	0	167	1
a_2b_2 (Vivid Purple)	0	0	0	183

Table 20. *Confusion Matrix on Complex Stimulus C_2 (Saturation/Hue) for Participant 3*

Stimulus	Response			
	a_1b_1 (Dull Blue)	a_1b_2 (Vivid Blue)	a_2b_1 (Dull Purple)	a_2b_2 (Vivid Purple)
a_1b_1 (Dull Blue)	168	0	13	0
a_1b_2 (Vivid Blue)	2	174	0	1
a_2b_1 (Dull Purple)	5	0	166	2
a_2b_2 (Vivid Purple)	0	2	3	152

Similar to complex stimulus C_1 the error frequencies all three confusion matrixes are fairly low, indicating participants had a high degree in accuracy for identifying which of the four shape/hue

stimulus combinations was presented. The largest source of confusion appears to be between the two hues at low saturation. Grtools package in R was used to run a model fit of a hierarchy of traditional GRT models on the confusion matrixes in order to assess separability and independence. Table 21 shows the results of model fitting. Figures 31-33 plot the best fitting model for each participant.

Table 21. *Traditional GRT models and fit statistics for All Participants.*

Model	Participant 1 AIC	Participant 2 AIC	Participant 3 AIC
{PI, PS, DS}	177.521	323.079	277.768
{1_Rho, PS, DS}	168.942†	319.304	280.948
{PS,DS}	185.718	335.878	300.614
{PI, PS(A), DS}	183.580	343.955	291.921
{1_Rho, PS(A), DS}	177.359	330.569	289.786
{PS(A), DS}	212.519	357.866	325.665
{PI,PS(B),DS}	190.710	293.219	274.662†
{1_Rho, PS(B), DS}	183.396	292.267†	279.307
{PS(B), DS}	212.337	323.678	314.623
{PI,DS}	198.578	308.381	293.219
{1_Rho, DS}	209.232	319.002	296.356
{DS}	274.929	384.076	376.167

Note: PI - Perceptual Independence, PS - Perceptual Separability, DS - Decisional Separability, †- Best fitting model

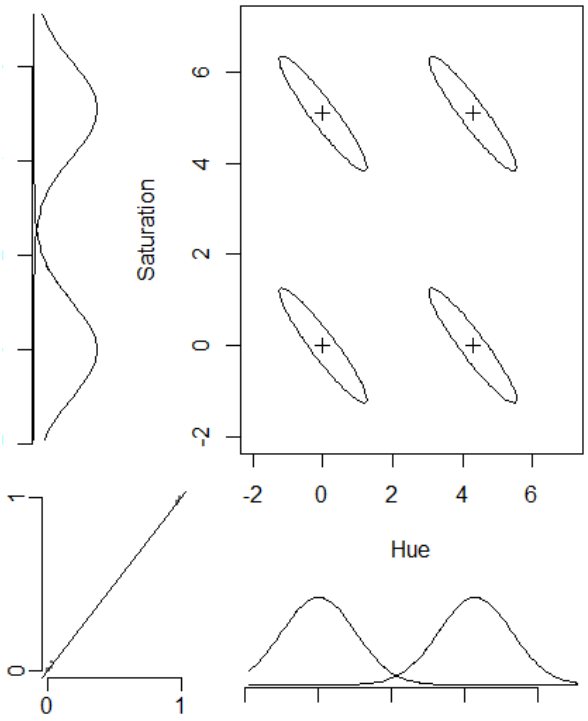


Figure 31. Best GRT Model Fit - {1_Rho, PS, DS} of Hue and Saturation for Participant 1.

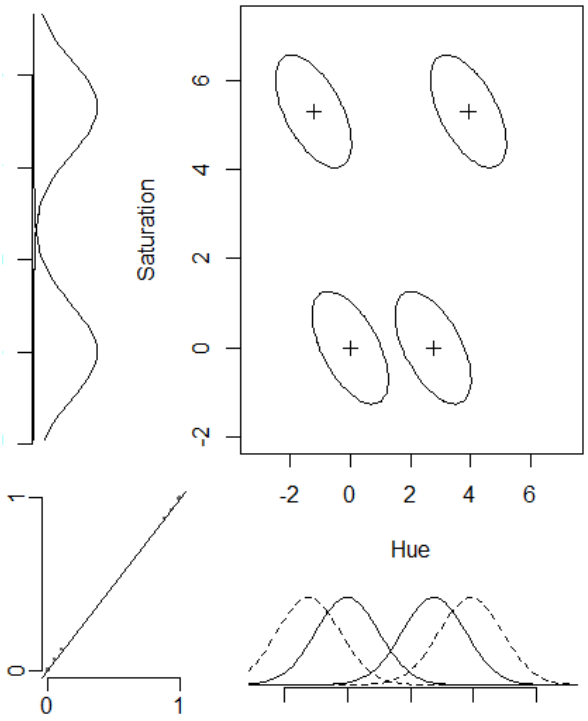


Figure 32. Best GRT Model Fit - {1_Rho, PS(B), DS} of Hue and Saturation for Participant 2.

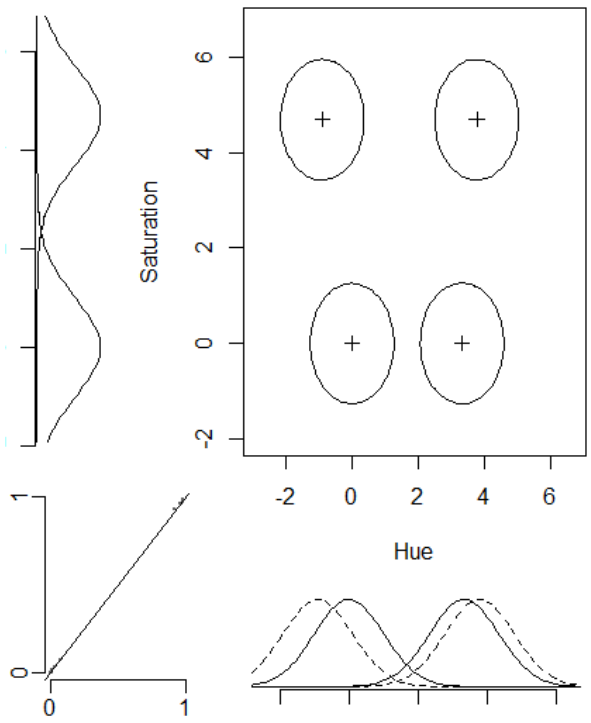


Figure 33. Best GRT Model Fit - {PI, PS(B), DS} of Hue and Saturation for Participant 3.

For Participant 1 the best fitting model was {1_Rho, PS, DS} indicating that the dimensions are perceptually separable and that there is decisional separability, however perceptual independence is violated. This is also evident in Figure 31 by the square orientation of the paired z-score means and the fact the $f(x, y)$s are slanted. For Participant 2 the best fitting model was {1_Rho, PS(B), DS} indicating that the dimensions have asymmetrical perceptual integrality on the hue dimension and are decisionally separable, however like Participant 1 perceptual independence has been violated. The asymmetrical perceptual integrality on A is evident in the fact that the bottom pair of $f(x, y)$s are closer together along only the x-axis than the top pair of $f(x, y)$s. The saturation mean(y-coordinate) for the low saturation $f(x, y)$s functions (bottom pair) remain the same and the saturation mean (y-coordinate) for the high saturation $f(x, y)$s functions (top pair) also remain the same. This indicates that Participant 2 has higher sensitivity in judging hue

at high saturation when compared to low saturation. In other words the level of saturation affects the perception of hue, but the level of hue does not affect the perception of saturation.

For Participant 3 the best model fitting was {PI, PS(B), DS} which indicates that the dimensions are perceptually independent, decisionally separable and have asymmetric perceptual integrality on hue. The perceptual independence is evident by the circular $f(x, y)$s in Figure 33 and the pattern of orientations of the $f(x, y)$s is the same as that for Participant 2 indicating that the participant achieved higher sensitivity in judging hue at higher saturation values.

Means and standard deviations for both correct and error frequencies as well as reaction times for the filtering, control and redundancy tasks for all participants are provided in Appendix B.

Discussion

The intention of this study was twofold, i) to gather data to use in the next set of studies for the determination of weights and ii) to verify that for these ranges of the shape and hue dimensions used in combined stimulus C_1 that perceptual separability observed in previous research holds and that the ranges of saturation and hue used in combined stimulus C_2 that perceptual integrality holds as in previous studies. While previous research indicates how the individual dimensions interact in terms of perceptual separability, the specific ranges of stimulus magnitude may differ from those used in the present study. The results for complex stimulus C_1 showed, for all three participants, that hue and shape were perceptually separable, a finding that follows the results from Garner (1977). The results for two of the participants for complex stimulus C_2 showed that hue and saturation had asymmetric perceptually integrality, that the perception of the level of hue was dependent on the level of saturation, but the perception of the level of saturation was independent of the perception of the level of hue. The perceptual

integrality of hue and saturation is consistent with the results reported by Burns and Shepp (1988). For one of the three participants complete perceptual separability was found for hue and saturation. This result could be due to the particular range of hue and saturation selected for this experiment and it is possible that differing values along the dimension continua would show different results regarding the separability of the two dimensions. However, given the other two participants found the dimensions to be perceptually integral at the extreme dimension values, the two dimensions are probably perceptually integral along these ranges of the dimensions.

For two of the three participants for both set of stimuli perceptual independence was violated, however this may be an artifact of the sparse confusion matrix. By populating the sparse matrices by adding a value of 1 to each cell, the best model fit becomes the same model in terms of perceptual and decisional separability but with perceptual independence no longer violated. While this is not a suggestion to apply this technique when analyzing sparse confusion matrices, it does reveal that if the data was less sparse even by a small amount, perceptual independence would likely not have been violated. To avoid this issue in future work, using less extreme values along each dimensions continuum should increase confusion since the stimuli are closer in physical value and therefore should have a lower degree of discernibility. For example, running the same complete identification task with the shape dimension having the low value be a dodecagon (twelve sided polygon) rather than a hexagon and the high value being a icosagon (twenty sided polygon) rather than thirty-six sided polygon should increase the confusion between the high and low values since the dodecagon has less of a discrepancy in the empty physical space that would be filled by a circle when compared to the hexagon and the icosagon while appearing more like a circle has discernible sides more like a polygon with fewer sides.

CHAPTER 4: WEIGHT DETERMINATION (STUDY 2A)

Method

In order to determine the appropriate weights for each combined mapping function, the data from the identification experiment was used to fit a MFSDT analysis to the MSDT results. The first step involved collapsing the MSDT data into hit (H), miss (M), false alarm (FA) and correct rejection (CR) rates in order to allow comparison to the MFSDT outcomes. By treating each potential pair of comparisons separately a pattern of outcomes emerges from the identification matrix, as seen in Figure 34.

	Response			
Category	1	2	3	4
1	CR/CR/CR	FA	FA	FA
2	M	H/CR/CR	FA	FA
3	M	M	H/H/CR	FA
4	M	M	M	H/H/H

Figure 34. Outcomes based on all possible pairs of categories and responses.

The diagonal values are each repeated three times because each cell is used in three comparisons. For example the category 2-response 2 is a hit when paired with category 1, but it is a correct rejection when paired with category 3 or with category 4 (Thus, H/CR/CR). From this table the frequencies for false alarms and misses is simple to calculate by summing the appropriate values in the table. That is, the total number of false alarms equals the sum of the frequencies from the cells labeled FA in Figure 34 and total number of misses equals the sum of the frequencies from the cells labeled M in Figure 34. Hits and correct rejections are calculated by awarding partial credit of the frequencies along the diagonal in Figure 34. Thus, total hits is calculated by adding the frequency of cell (4,4) with 2/3*the frequency of cell (3,3) and 1/3*the frequency of cell (2,2) from Figure 34 and total correct rejections is calculated by adding the frequency of cell (1,1) with 2/3*the frequency of cell (2,2) and 1/3*the frequency of cell (3,3). Category 1 was

defined as the low/low value of the combined stimulus and category 4 as the high/high value of the combined stimulus. The high/low pairing with the smallest distance from the low/low pair as determined by the grtools best model plot from Study 1b was set as category 2 and the high/low pairing with the farthest distance from the low/low pair was set as category 3. This was done for each of the three participant's identification data.

The second step for fit comparisons required running a MFSDT analysis on the data from Study 1B's identification task. This involves creating a response mapping function as well as a signal mapping function. The signal mapping function followed the proposed form of MFSDT combined mapping functions in conjunction with the mapping functions determined by Study 1A. So the combined signal function was equal to a weight multiplied by the resulting value from the hue a passed into the hue mapping function added to 1 minus the same weight multiplied either to the result from the shape or saturation mapping functions depending on which complex stimulus was being analyzed. The response mapping function was determined by Table 22. This response mapping functions results in the best fit for each weight.

Table 22. *Response mapping function for weight determination*

Hue Value	Shape/Saturation Value	r
Low	Low	0
High	Low	weight
Low	High	1-weight
High	High	1

This step is completed ten times, as it involved computing the MFSDT analysis with the weight set equal to 0.1 and increasing the weight by 0.1 until a maximum weight of 0.9.

The final step was to compute a chi-square test of model fit on observed MSDT outcomes and the expected outcomes created by the MFSDT analysis.

Results

To collapse the MSDT confusion matrices into a form usable for comparison to MFSDT the center z-score means of the best fitting model determined in Study 1b were used to determine the ordering of the four complex stimuli. Table 23 shows the coordinates of each of the four shape and hue stimuli for each participant.

Table 23. *Z-score coordinates of each of the four Shape and Hue stimuli used in Study 1b for each of the three participants.*

	Blue Hexagon Z-Score		Blue Circle Z-Score		Purple Hex. Z-Score		Purple Circle Z-Score	
Participant	X	Y	X	Y	X	Y	X	Y
1	0.0	0.0	0.0	5.00	3.99	0.0	5.00	3.99
2	0.0	0.0	0.0	5.19	4.61	0.0	5.19	4.61
3	0.0	0.0	0.0	4.25	4.08	0.0	4.25	4.08

Since Purple Hexagon's coordinates are closer to the Blue Hexagon coordinates than Blue Circle's coordinates, the ordering for all three participants is (1) Blue Hexagon, (2) Purple Hexagon, (3) Blue Circle and (4) Purple Circle.

Table 24 shows the z-score coordinates of each of the four hue and saturation stimuli for each participant.

Table 24. *Z-score coordinates of each of the four Shape and Hue stimuli used in Study 1b for each of the three participants.*

	Dull Blue Z-Score		Vivid Blue Z-Score		Dull Purple Z-Score		Vivid Purple Z-Score	
Participant	X	Y	X	Y	X	Y	X	Y
1	0.0	0.0	0.0	5.05	4.37	0.0	4.37	5.05
2	0.0	0.0	-1.21	5.25	2.70	0.0	3.85	5.25
3	0.0	0.0	-0.95	4.64	3.45	0.0	3.81	4.64

Since for all three participants Dull Purple's coordinates are closer to Dull Blue's coordinates compared to the distance of Vivid Blue's coordinates from Dull Blue's coordinates, the ordering is (1) Dull Blue, (2) Dull Purple, (3) Vivid Blue, (4) Vivid Purple.

Four separate chi squares were calculated for each weight. One chi-square was calculated for each participant and involved comparing the hit, miss, false alarm and correct rejection rates for each individual participant. Each of these three chi-squares has a degree of freedom equal to one. The fourth chi-square compared observed and expected outcome rates over all three participants. The overall chi-square has a degree of freedom equal to two. Table 25 shows the chi-squares for each participant and the total chi-square for each of the weights tested for complex stimulus C_1, shape and hue. Since the weights on hue and shape can be defined in terms of only hue (w_{shape} = 1 - w_{hue}) only the weight on hue is listed.

Table 25. *Chi squares for individual participants and total chi-square for all hue weights tested for shape and hue.*

w_{hue}	Participant 1	Participant 2	Participant 3	Total
0.1	4.585	2.663	0.899	8.147
0.2	3.599	2.007	0.834	6.440
0.3	2.820	1.413	1.091	5.324
0.4	2.233	0.897	1.747	4.877
0.5	1.629	0.485	2.666	4.780
0.6	1.599	0.298	4.256	6.153
0.7	1.536	0.169	6.135	7.840
0.8	1.635	0.111	8.68	10.426
0.9	1.892	0.138	12.110	14.139

The best fitting hue weight for Participant 1 is w_{hue} = 0.7, the best fitting hue weight for Participant 2 is w_{hue} = 0.8 and the best fitting hue weight for Participant 3 is w_{hue} = 0.2. Across all participants the best fitting hue weight is w_{hue} = 0.5.

Table 26 shows the chi-squares for each participant and the total chi-square for each of the weights tested for complex stimulus C_2, hue and saturation. Again since the weighs can be defined in terms of only hue ($w_{saturation}$ = 1 - w_{hue}) only the weight on hue is listed.

Table 26. *Chi squares for individual participants and total chi-square for all hue weights tested for hue and saturation.*

w_{hue}	Participant 1	Participant 2	Participant 3	Total
0.1	4.350	30.716	10.395	45.461
0.2	3.384	21.366	7.899	32.650
0.3	2.508	14.833	5.835	23.176
0.4	1.735	10.151	4.162	16.048
0.5	1.094	6.746	2.865	10.705
0.6	0.742	4.078	1.955	6.776
0.7	0.494	2.261	1.378	4.132
0.8	0.367	1.086	1.624	3.077
0.9	0.393	0.515	2.363	3.271

The besting fitting hue weight for Participant 1 is w_{hue} = 0.8, the best fitting hue weight for Participant 2 is w_{hue} = 0.9, and the best fitting hue weight for Participant 3 is w_{hue} = 0.7. The best fitting hue weight over all participants is w_{hue} = 0.8.

Discussion

For shape and hue the weights for individual participants vary a great deal from w_{hue} = 0.2 to w_{hue} = 0.8. Given the variation at the individual level using the total chi-square fit statistic for weight determination seems appropriate. Setting w_{hue} = 0.5 makes sense given the grtools model plots in Study 1b. For all participants the four stimuli create a rectangular pattern with similar distances between the sides, making the rectangle more like a square. A square pattern in the model plots would seem to indicate equal ability to tell the difference between the individual dimensions, which in turn would suggest that a weight which treats the two individual dimensions equally is appropriate. Besides having the lowest chi-square for the total at w_{hue} = 0.5, all the individual participant chi-squares are below the chi-square value (χ^2 = 3.84, df = 1)

indicating that the two models are different at a statistically significant level (α = 0.05). The combined mapping function is therefore determined to be

$$S(a_{shape}, a_{hue}) = 0.5 * s_{shape}(a_{shape}) + 0.5 * s_{hue}(a_{hue}) \quad (10)$$

Figure 35 plots the resulting combined mapping function.

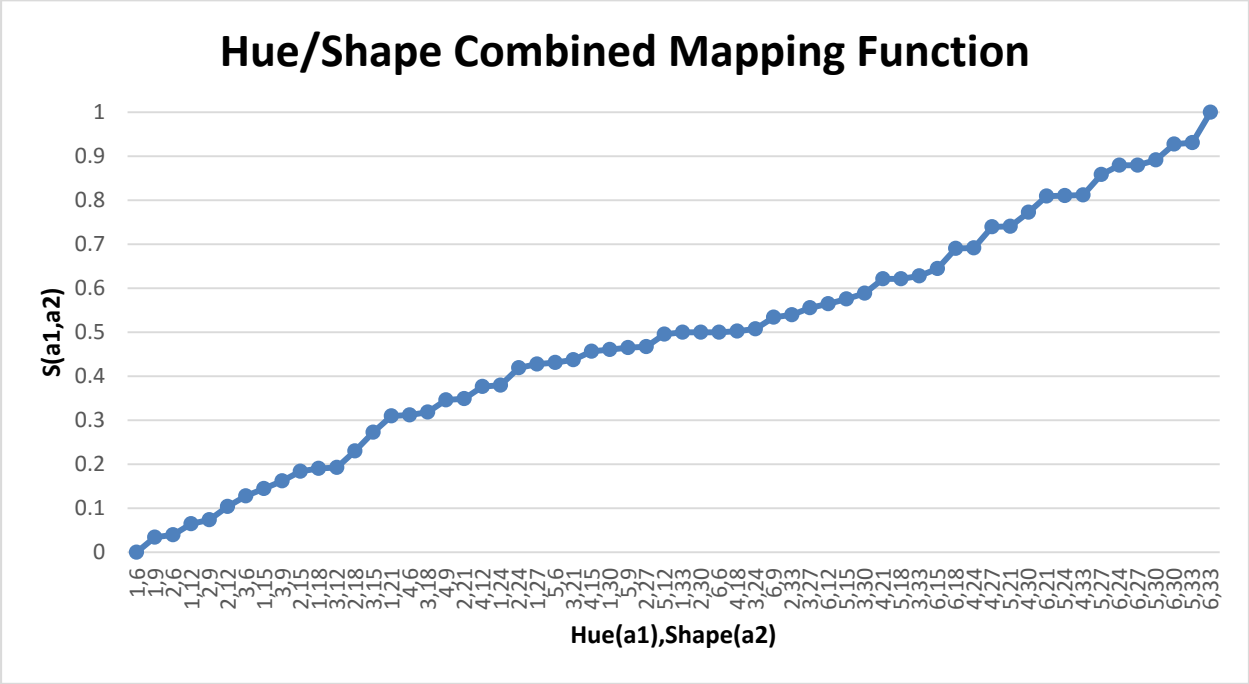


Figure 35. Combined mapping function for hue and shape.

For hue and saturation all of the weights for the individual participants are all high, ranging from w_{hue} = 0.7 to w_{hue} = 0.9. However, the fit for Participant 2 does not minimize in the range between 0.1 and 0.9, in fact the smallest chi-square statistic found for Participant 2 is when w_{hue} = 1 (χ^2 = 0.151). This is likely due the fact that difference between the false alarm outcomes appear to be driving how well the models fit, and the largest value in false alarms can only be created in this data via the hue mapping function. Excluding Participant 2 from the total chi-square value since it can be described via a unidimensional mapping function, the best fitting hue weight decreases to $w_{hue} = 0.7$ (adjusted χ^2 = 1.872) from $w_{hue} = 0.8$ (adjusted χ^2 =

1.991). Even including Participant 2, the individual chi-square fit statistics are all below the critical value (χ^2 = 3.84, df = 1) indicating that the two models are equivalent at a statistically significant level (α = 0.05). The combined mapping function is therefore determined to be

$$S(a_{saturation}, a_{hue}) = 0.3 * s_{saturation}(a_{saturation}) + 0.7 * s_{hue}(a_{hue}). \quad (11)$$

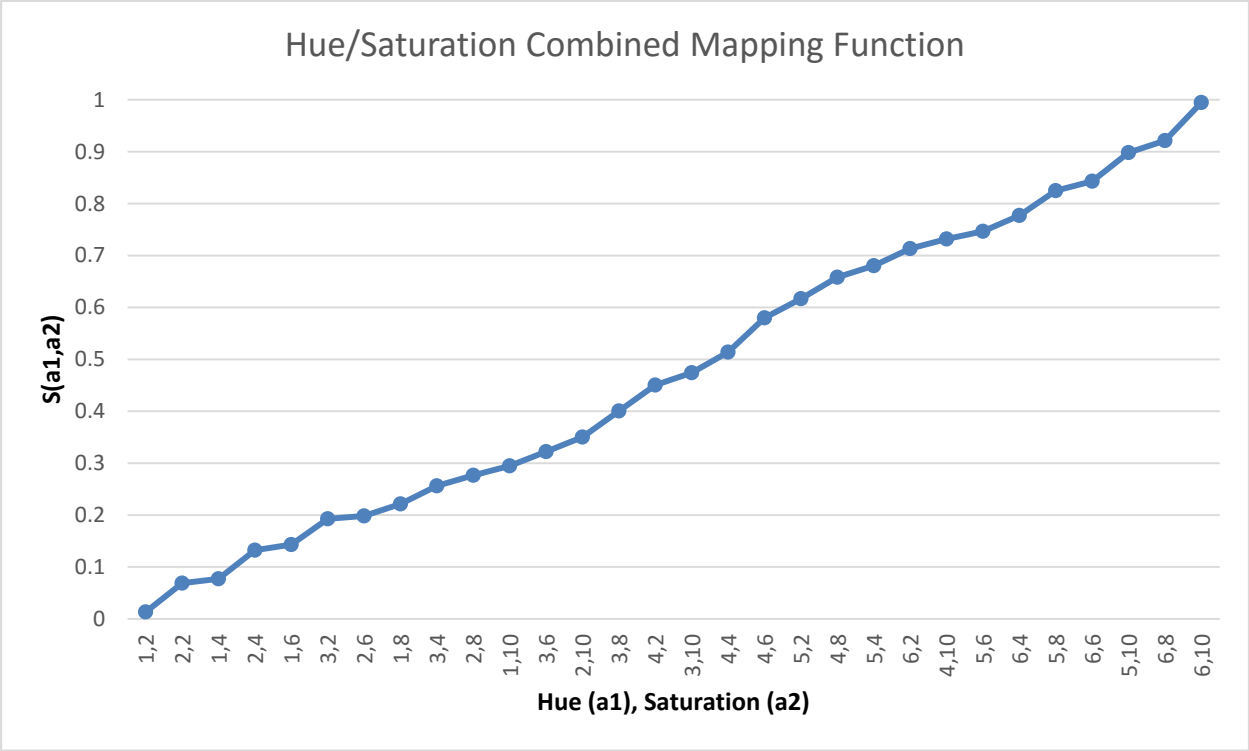


Figure 36. Combined mapping function for hue and saturation.

Setting the hue weight higher than the saturation weight would indicate that the level hue affects the perception of the combined stimulus and thereby the signal membership more than the level of saturation. The grtools plots for the asymmetric perceptually integral cases (Participants 2 and 3), show that there is greater difference in the perception of the hue at high saturation levels, than at low levels of saturation. This can be described in terms of the equation if one were to assume that hue has a greater effect on the perception of the combined stimulus and the level of the saturation adds to the signal membership (or increases the perception of the hue). At low

levels of saturation there is just the base difference in perception which is explained by only hue, and as you increase in saturation the signal qualities of the combined stimuli is amplified.

Future work should consider collecting data from multiple complete identification tasks along the continuum of values of the dimension rather than employing the extreme categories. This should allow for less sparse confusion matrices when there is high discernibility between the extreme values of the dimension. Additionally, since FSDT is intended to be used when there is more confusion about the nature of the stimulus, i.e., the stimulus dimension has properties that are both signal and noise -like, the extreme values are the least representative of the fuzzy nature of the stimulus, so model fitting to these values may result in inappropriate or less than ideal weighting values.

CHAPTER 5: MONTE CARLO ROC ANALYSIS (STUDY 2B)

Method

Szalma and O'Connell (2011) demonstrated the viability of using a Monte Carlo simulation of a perceptual detection task and to generate FSDT data that can be submitted to ROC analysis to test whether the statistical assumptions of SDT extend to FSDT. This methodology was adapted for the present work to test whether the selected mapping function for each complex stimulus can be successfully used to generate an FSDT analysis which satisfies the statistical assumptions of traditional SDT. This procedure was applied to simulate both the C_1 and C_2 complex stimuli.

Data Generation. The data generation is a three step process using multiple generated random variates. In the first step, random pairs of values were generated, simulating the presentation of a trial to a participant. Each value was a discrete integer value sampled from the dimension appropriate range. For hue the range was between one and six. For saturation the range was between 2 and 10, with only the even values capable of being sampled. For shape, the range was between 6 and 36. Since Green & Swets (1966) recommend a large number (e.g. greater than 500) of trials for deriving stable SDT estimates, 1000 random pairs of combined stimulus values were generated. The second step involved simulations of variation in the observer's response. This was introduced by simulating two forms errors of omission. The first form of error attempted to simulate the misperception of the physical stimulus, by creating a perceived value of the stimulus based on the generated value. This involved passing the generated values into a function which used the probabilities shown in Tables 27-29 to determine what each component dimensions was perceived as by the "observer" on a given trial. These

probabilities were chosen arbitrarily, as the goal was simply to introduce perceptual error however how much error did not matter.

Table 27. *Probability of perceived values for generated hue values*

	Perceived Value					
Generated Value	1	2	3	4	5	6
1	0.66	0.19	0.08	0.04	0.02	0.01
2	0.19	0.54	0.19	0.05	0.02	0.01
3	0.08	0.19	0.45	0.19	0.05	0.04
4	0.04	0.05	0.19	0.45	0.19	0.08
5	0.01	0.02	0.05	0.19	0.54	0.19
6	0.01	0.02	0.04	0.08	0.19	0.66

Table 28. *Probability of perceived values for generated shape values*

	Perceived Value									
GV	6- 8	9- 11	12-14	15-17	18-20	21-23	24-26	27-29	30-32	33+
6-8	0.6650	0.1700	0.0700	0.0300	0.0250	0.0200	0.0100	0.0090	0.0008	0.0002
9-11	0.1700	0.4600	0.1700	0.1000	0.0500	0.0200	0.0172	0.0100	0.0020	0.0008
12-14	0.0700	0.1700	0.441	0.1700	0.0700	0.0260	0.0200	0.0140	0.0100	0.0090
15-17	0.0300	0.1000	0.1700	0.3528	0.1700	0.1000	0.0300	0.0200	0.0172	0.0100
18-20	0.0250	0.0500	0.0700	0.1700	0.3490	0.1700	0.01	0.0260	0.0200	0.0200
21-23	0.0200	0.0200	0.0260	0.1000	0.1700	0.3490	0.1700	0.0700	0.0500	0.0250
24-26	0.0100	0.0172	0.0200	0.0300	0.1000	0.1700	0.3528	0.1700	0.1000	0.0300
27-29	0.0090	0.0100	0.0140	0.0200	0.0260	0.0700	0.1700	0.4410	0.1700	0.0700
30-32	0.0008	0.0020	0.0100	0.0172	0.0200	0.0500	0.1000	0.1700	0.4600	0.1700
33+	0.0002	0.0008	0.0090	0.0100	0.0200	0.0250	0.0300	0.0700	0.1700	0.6650

Note. GV = Generated Value

Table 29. *Probability of perceived values for generated saturation values.*

	Perceived Value				
Generated Value	1	2	3	4	5
1	0.72	0.20	0.05	0.02	0.01
2	0.20	0.56	0.20	0.02	0.02
3	0.05	0.20	0.50	0.20	0.05
4	0.02	0.02	0.20	0.56	0.20
5	0.01	0.02	0.05	0.20	0.72

The second form of error, which was simulated was errors due to the observer missing the stimulus presentation, by introducing a 1% probability that the perceived value was 0 (no stimulus observed). The third step and final step takes the paired perceived values and transforms them into a response value. A continuous response function was used based on the ordering of paired stimuli given the weights derived in Study 2a and where in the ordering a particular stimulus pair occurred. For example, for hue and shape, the lowest possible saturation, 2, and the lowest possible hue, 1, creates the low end point and thus assigned a response value of zero and the highest possible saturation, 10, and the highest possible hue, 6, creates the high end point and thus assigned a response value of one. Given $w_{hue} = 0.7$ a saturation of 2 and a hue (number of positions away from low point / most positions away from low point). Figure 37 shows the base response mapping for hue and shape and Figure 38 shows the base response mapping function for hue and saturation.

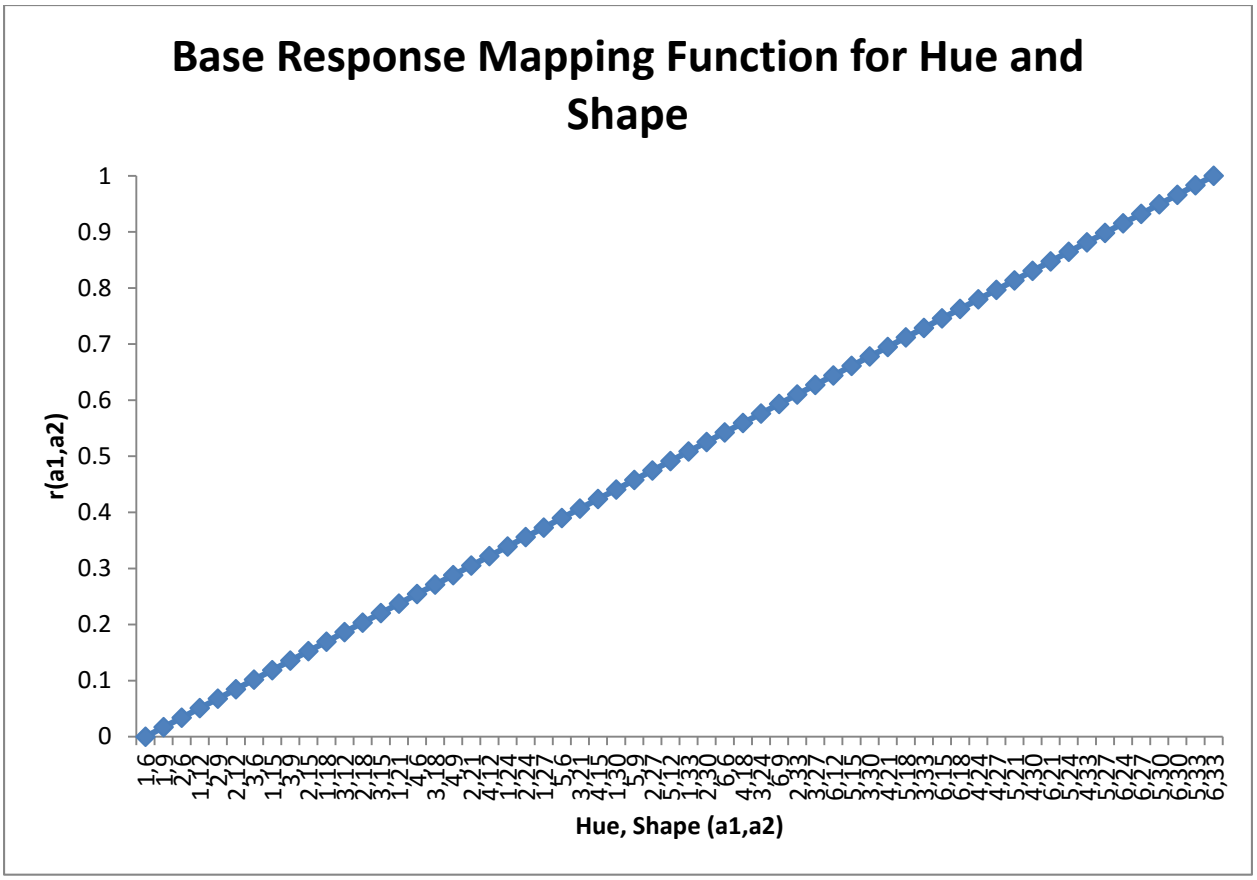


Figure 37. Base response mapping function used in Monte Carlo ROC simulations for hue and shape.

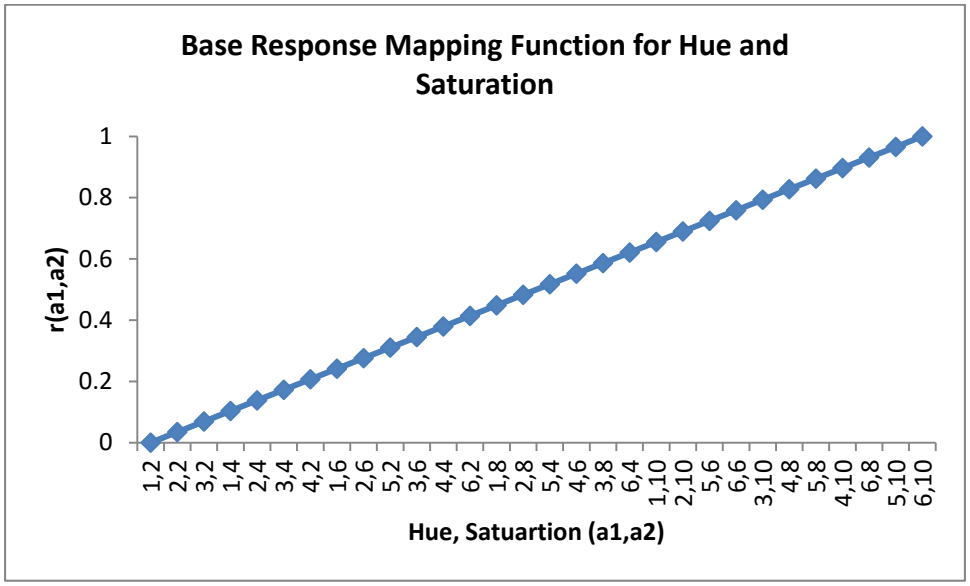


Figure 38. Base response mapping function used in Monte Carlo ROC simulations for hue and saturation.

Since ROC analysis requires the use of a minimum of three criterion values for Yes/No data, three responses are created for each generated pair of stimuli, corresponding to an unbiased criterion, a more lenient criterion, and a more conservative criterion. To achieve this the unbiased criterion is set equal to the base response mapping function, the lenient criterion is set to the base response mapping function plus 0.04, and the conservative criterion is set to the base response mapping function minus 0.04. Figure 39 and Figure 40 plot the three response mapping functions together for hue and shape, and hue and saturation, respectively.

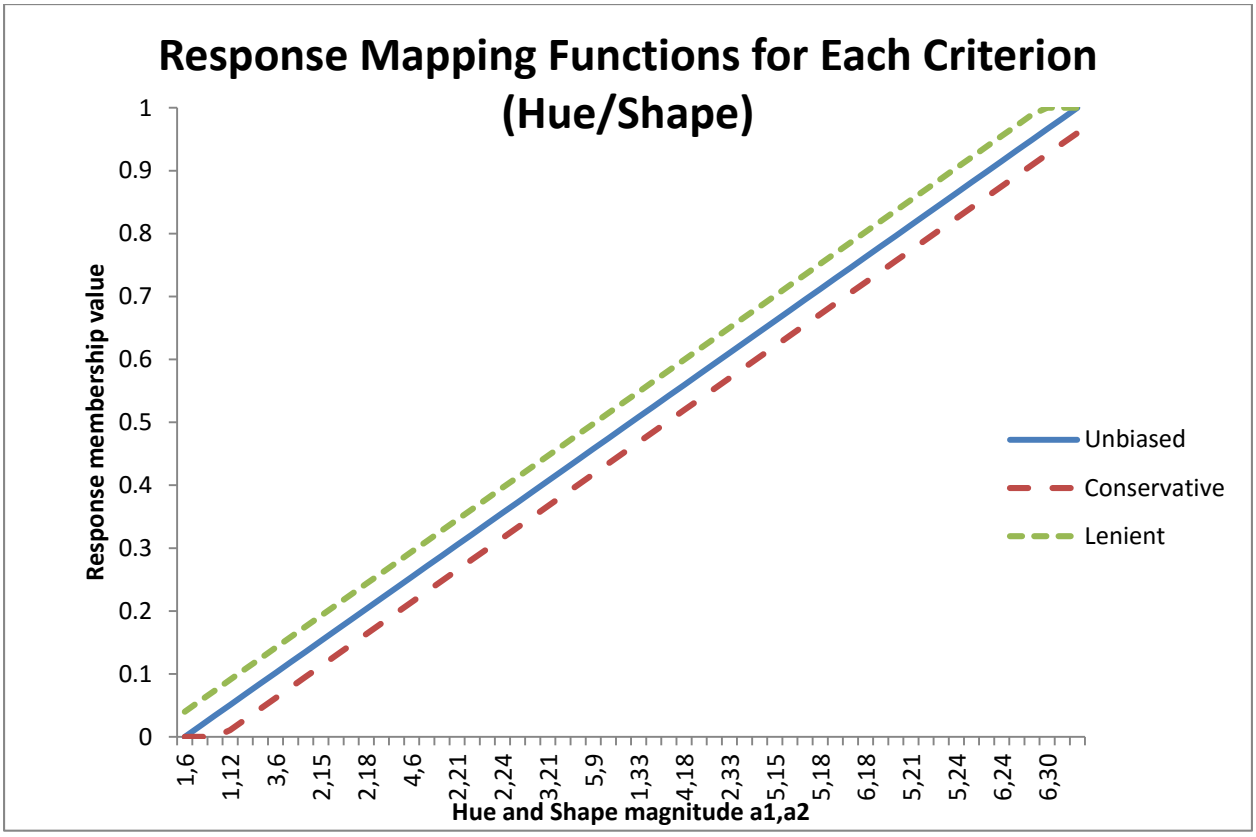


Figure 39. Three criterion response mapping functions used in Monte Carlo ROC stimulations for hue and shape. Note not all stimulus pairing are labeled on the x-axis in order to de-clutter the plot. The order of pairings is the same as Figure 37.

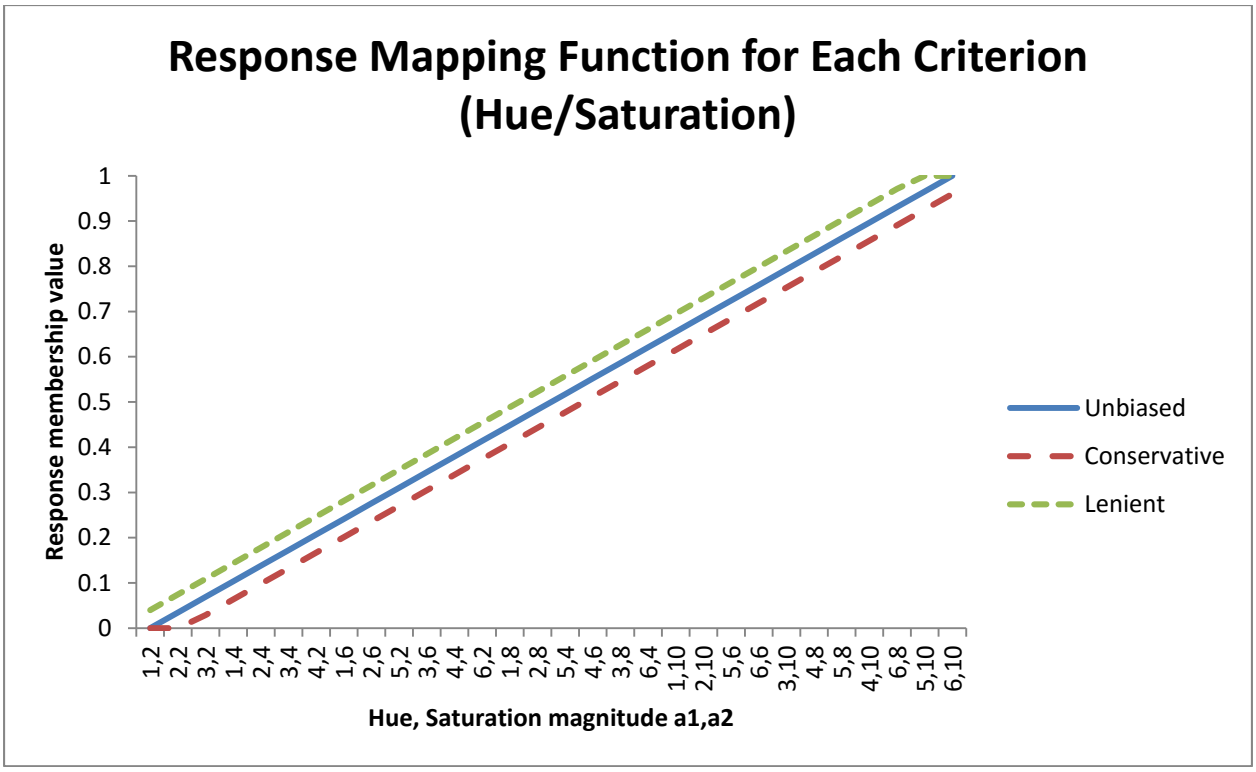


Figure 40. Three criterion response mapping functions used in Monte Carlo ROC stimulations for hue and saturation.

To add variation in the responses a normal random variate with a mean of zero and a standard deviation of 0.0025 was added to the response mapping function value for the perceived pair. Each criterion was associated with its own random normal variate, rather than the same random normal variate being added to all three response mapping function values. At the end of the data generation procedure there are for each trial a generated complex stimuli (i.e., pair of generated values based on dimension appropriate ranges) which represents the true state of the world, and three response membership values corresponding to three different criterion settings.

Generation of Signal Rates and Frequencies. The fuzzy signal membership for each generated pair values for each trial was determined based on the mapping functions derived in Study 2b. Using the signal membership and the response membership in conjunction with the mixed implication functions from Parasuraman, Masalonis and Hancock (2000), fuzzy hit, miss, false alarm and correct rejection outcome membership were calculated for each trial. From these

outcome membership values an average frequency and rates, equations 5-8, were computed over the 1000 trials for each response criterion.

This procedure was repeated over four thousand simulations and the average of the four different outcome frequencies and rates was computed. These simulations were computed for both for complex stimulus C1 and once for complex stimulus C2.

Results

Complex Stimulus C1: Shape and Hue. The simulation outputs the average frequency and rates for the four thousand simulations. The mean frequencies and rates hits, false alarms, misses and correct rejections for each of the response criterion conditions are shown in Tables 30 and 31 respectively for the hue and shape condition.

Table 30. *The average frequency for the four possible outcomes for each of the three response criterion simulated for hue and shape.*

	Hits	Miss	False Alarms	Correct Rejections
Lenient	429.50	72.86	113.04	374.63
Unbiased	412.10	90.26	92.56	395.11
Conservative	391.72	110.63	74.43	413.23

Table 31. *The average rates (proportion) for the four possible outcomes for each of the three response criterion simulated for hue and shape.*

	Hit Rate	Miss Rate	False Alarm Rate	Correct Rejection Rate
Lenient	0.855	0.145	0.227	0.752
Unbiased	0.820	0.180	0.190	0.794
Conservative	0.780	0.220	0.150	0.830

Wicken's FitRoc (2002) was used for the ROC analysis. FitRoc calculated the z-score form ROCs by using maximum likelihood estimation to estimate the slope and intercept with the option to assume equal variance or not. FitRoc also tests the goodness of fit ($\chi 2$) of the resulting

linear equation and generates estimates for the perceptual sensitivity (A_z) and response biases (β_{log}). Since FitRoc only allows integers to be entered into the software the frequencies from Table 30 were rounded to the nearest whole number. Table 32 shows the FitRoc results for complex stimulus C_1, hue and shape.

Table 32. *Displays goodness of fit, sensitivity, and criterion bias calculated for hue and shape.*

Stimulus	χ^2	A(z)	A	B	Conservative β_{log}	Unbiased β_{log}	Lenient β_{log}
C_1	0.004	0.898	1.794	1.00	-0.231	0.038	0.290

For complex stimulus C1, hue and shape, the predicted linear ROC was obtained, as shown in Figure 41. The predicted linear ROC's goodness of fit statistic ($\chi^2 = 0.004$, df = 2) was below the critical chi-square value ($\chi^2 = 5.99$, df = 2, $\alpha = 0.05$) indicating that the predicted linear ROC function fir the observed data. The slope of the linear ROC fit to this data was equal to one indicating the equal variance assumption was met.

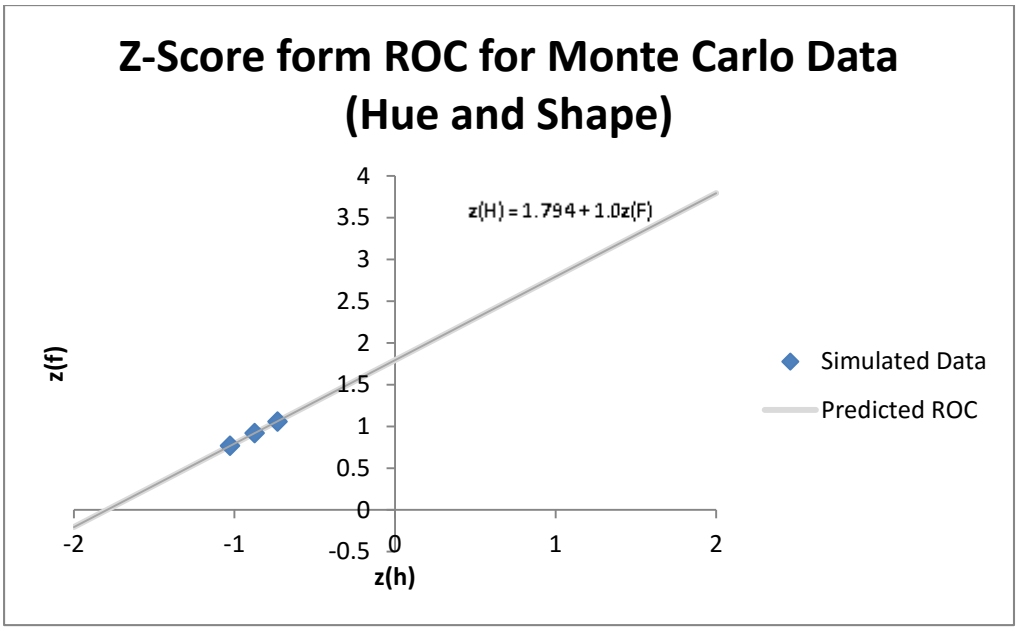


Figure 41. Z-score form linear ROC based on Monte Carlo simulated data for Hue and Shape.

Complex Stimulus C2 - Hue and Saturation. The mean frequencies and rates hits, false alarms, misses and correct rejections for each of the response criterion conditions are shown in Tables 33 and 34 respectively for the hue and shape condition.

Table 33. *The average frequency for the four possible outcomes for each of the three response criterion simulated for hue and saturation.*

	Hits	Miss	False Alarms	Correct Rejections
Lenient	400.30	64.42	132.32	392.98
Unbiased	385.93	78.79	108.59	416.70
Conservative	368.43	96.30	88.01	437.29

Table 34. *The average rates (proportion) for the four possible outcomes for each of the three response criterion simulated for hue and saturation.*

	Hit Rate	Miss Rate	False Alarm Rate	Correct Rejection Rate
Lenient	0.861	0.138	0.247	0.734
Unbiased	0.830	0.170	0.203	0.779
Conservative	0.793	0.207	0.164	0.817

Table 35. *Displays goodness of fit, sensitivity, and criterion bias calculated for hue and saturation.*

Stimulus	χ^2	A(z)	A	B	Conservative β_{log}	Unbiased β_{log}	Lenient β_{log}
C_2	0.030	0.895	1.770	1.00	-0.130	0.124	0.375

For complex stimulus C_2, hue and saturation, the predicted linear z-score form ROC was obtained, as shown in Figure 42. The linear z-score form ROC's goodness of fit statistic (χ^2 = 0.030, df = 2) was less than the critical chi-square value (χ^2 = 5.99, df = 2, α = 0.05) indicating that the predicted linear ROC fit the observed data. The slope of the linear z-score form ROC was equal to one which indicates that the equal variance assumption was met.

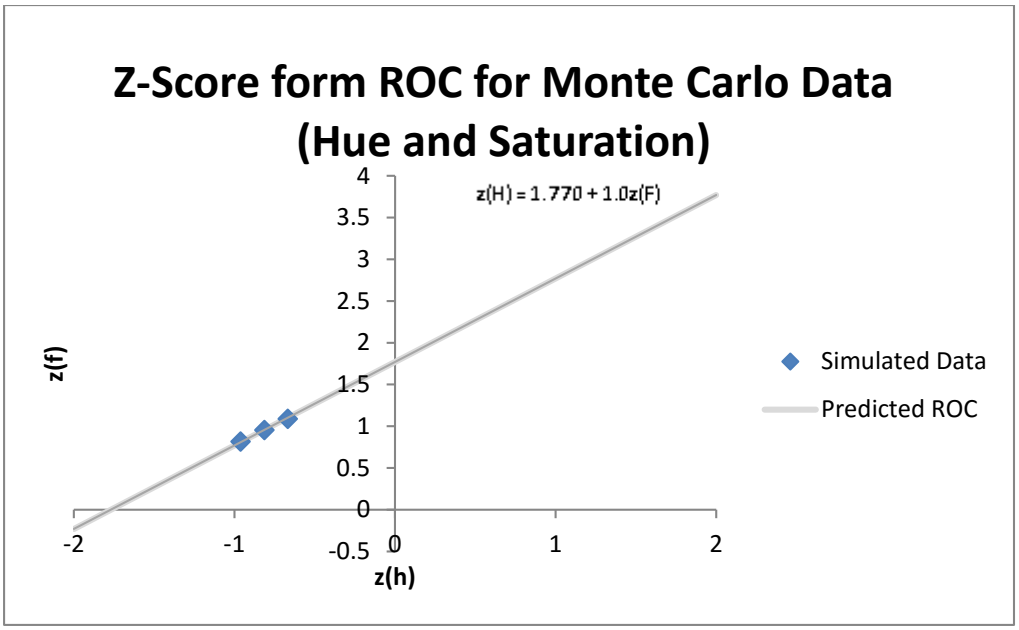


Figure 42. Z-score form linear ROC based on Monte Carlo simulated data for Hue and Saturation.

Discussion

Multidimensional Fuzzy Signal Detection Theory (MFSDT) suffers from the same problem that regular FSDT does with regard to simulation. Whereas both MSDT and traditional SDT allows for simulation based on their respective decision spaces, the decision space for MFSDT and FSDT has not been clearly established (Szalma & Hancock, 2013). Part of the reason the structure of the decision space is ambiguous is there is nothing to indicate in the traditional SDT decision space the fuzzy aspect of the dimensions. While this may be resolved with future research, until it is resolved it leaves the problem as to how to simulate FSDT and MFSDT models. Szalma and O'Connell (2011) suggested simulating the task rather through the use of random variates instead of simulating the decision space in order to create FSDT data that could be evaluated by ROC analysis. While the simulated data may not be exact model of the signal detection process, it can approximate one way the stimulus may be perceived and decisions made. Study 2b adapted the method from Szalma and O'Connell (2011) and modified

it for use in simulating MFSDT data. The main difference in the method aside from changes required to simulate multidimensional stimuli rather than unidimensional stimuli, was having to change the simulation to using discrete dimension values rather than continuous dimension values and through the use of a continuous rather than binary response mapping function.

The discrete dimension values was accomplished through the use of probability mass functions (Tables 27- 29) rather than probability density functions. Additionally with the discrete dimension values a generated data pool was not required as the full universal set of potential data points could easily be defined. As a result the generated pairs were sampled from the full universe rather than a large subset of the universe.

Using a continuous mapping function rather than a binary mapping function simplified the process of creating differing criterion. In the binary mapping function case, three random normal variates were used to represent the different decision criteria on any given trial and then the response membership value was determined based on whether the perceived value was greater or less than the decision criteria. In the continuous case a simple manipulation of the response mapping was used to create additional criteria and each criterion's response membership value was based on what their respective response mapping functions evaluated to given the perceived value with variation added via a small normal variate.

Like Szalma and O'Connell (2011), the generated data from the Monte Carlo simulation was submitted to ROC analysis in order to determine whether the assumptions of SDT were violated by the signal detection variant, in Szalma and O'Connell's case FSDT and for Study 2b case MFSDT. The results for the perceptually separable stimulus C_1, hue and shape, and the perceptually integral stimulus C_2, hue and saturation, showed that in each case a linear z-score form ROC fit the observed data. A linear z-score form ROC indicates that the normality

assumption was not violated by the data. Both complex stimulus C_1 and complex stimulus C_2 data fit a linear ROC model with a slope equal to one, indicating that the equal variance assumption was not violated. Since both the normality and equal variance assumptions were met for both perceptually separable and perceptually integral stimuli there is evidence in favor of the viability of MFSDT. Because both the normality and equal variance assumptions are met, the use of traditional SDT performance measures such as d' and β is possible for MFSDT data.

One thing of note is that for both complex stimulus C_1 and complex stimulus C_2 a large degree of clustering of the three criterion points is observed. This clustering is consistent with the results of previous empirical FSDT research (Murphey et. al, 2004; Szalma et. al, 2006, Szalma & Hancock, 2013) and simulation based FSDT research (Szalma and O'Connell, 2011). The reason for the clustering remains unclear, but it is likely that, because it occurs for FSDT, the reason for the clustering in MFSDT is an artifact the former model rather than due to a multivariate mapping function. It is possible that less clustering may have observed if the differing response mapping function conditions were developed to be more extreme in their differences.

CHAPTER 6: EXPERIMENTAL TESTING AND ROC ANALYSIS (STUDY 3)

Method

Participants. One male participant age 30 and one female participant age 29 volunteered for this study. Each participant had normal or corrected 20/20 vision and normal color vision. Both participants were acquaintances of the author.

Materials. Two sets of complex stimuli were used for this study, the previously defined perceptually separable complex stimulus C_1 and the perceptually integral complex stimulus C_2. Each complex stimulus set consisted of the fully factorial combination of the levels of each individual dimension. For C_1, the shape dimension consisted of 31 levels of polygons ranging from, 6 sided to 36 sided, and the hue dimension has 6 levels of hue, 2.5BP to 5P, combining the two dimensions resulted in 186 unique polygon/hue stimuli. For C_2, the saturation dimension consisted of 5 saturation levels, 2 chroma to 10 chroma (intervals of 2), and the hue dimension consisted of 6 levels of hue, 2.5 BP to 5P, combining the two dimensions resulted in 30 unique saturation/hue stimuli.

Procedure. The detection task required participants to monitor either a series of colored polygons (complex stimulus C_1) or a series of colored squares (complex stimulus C_2) at the center of a color-calibrated LCD screen for 300ms every four seconds. At the beginning of each session the participants were shown examples of a low signal and a high signal and were instructed to drag the indicator on a horizontal scroll bar to the position along a continuum between the low signal and the high signal they determined the presented image belonged. For C_1, the low signal was a blue hexagon and the high signal was a purple circle. For C_2, the low signal was a dull blue square and the high signal was a vivid purple square.

Each participant completed six sessions, each of which lasted 1.0 to 1.5 hours in duration. Each session consisted of a practice session of 50 trials and a detection task session of 700 trials. In order to combat vigilance decrements, five minute breaks occurred every 15 minutes during the session.

The six session comprised of three sessions with different response bias condition for complex stimulus C_1 and three sessions with differing response bias conditions for complex stimulus C_2. Response bias was manipulated by varying the probability distribution of the stimuli. For the unbiased conditions for both complex stimuli, each stimuli had an equal probability of being selected for presentation. For the lenient conditions stimuli determined via the mapping function to have a higher degree of similarity to the high signal had a higher probability of being selected for presentation than stimuli with a low degree of similarity (see Figures 43 and 44).

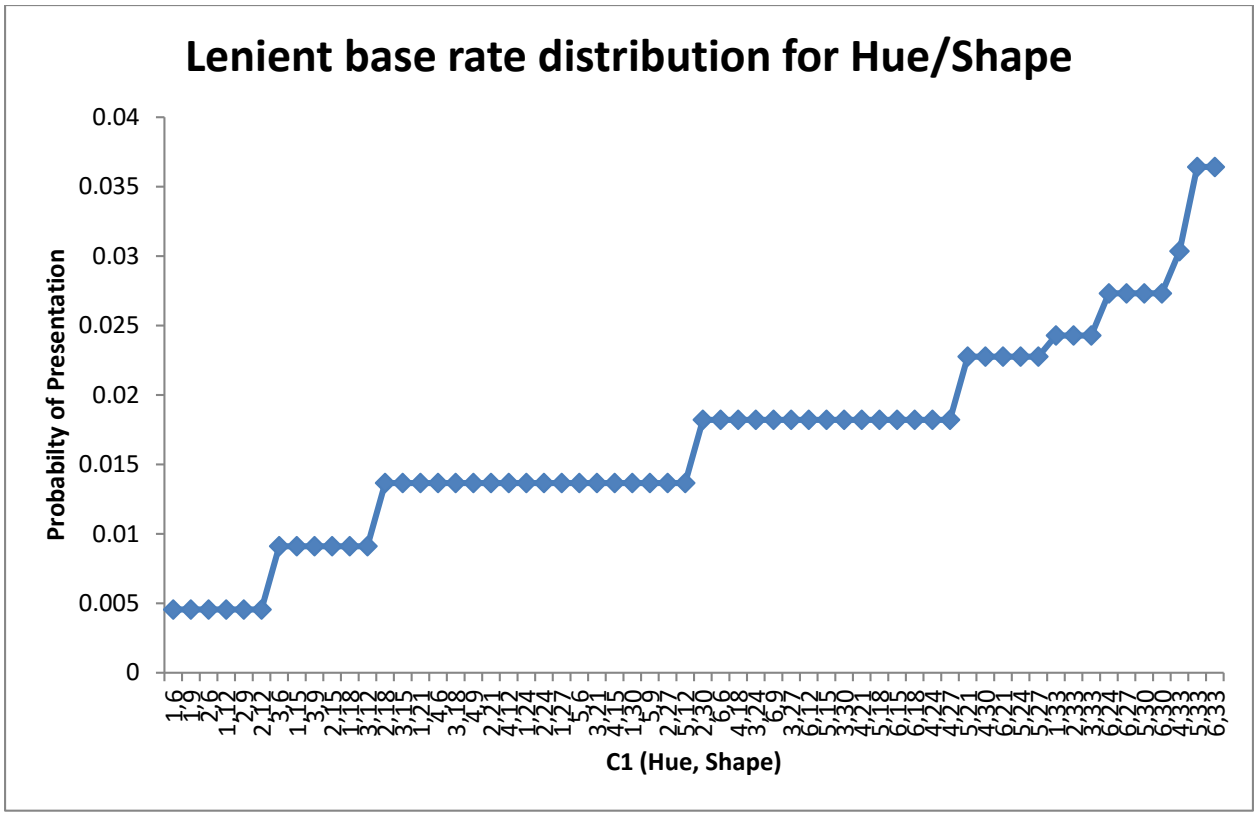


Figure 43. Signal base rate distribution for complex stimuli C_1 (Hue/Shape) for lenient condition. For shape, all polygon sides of the same defined set used the same probability.

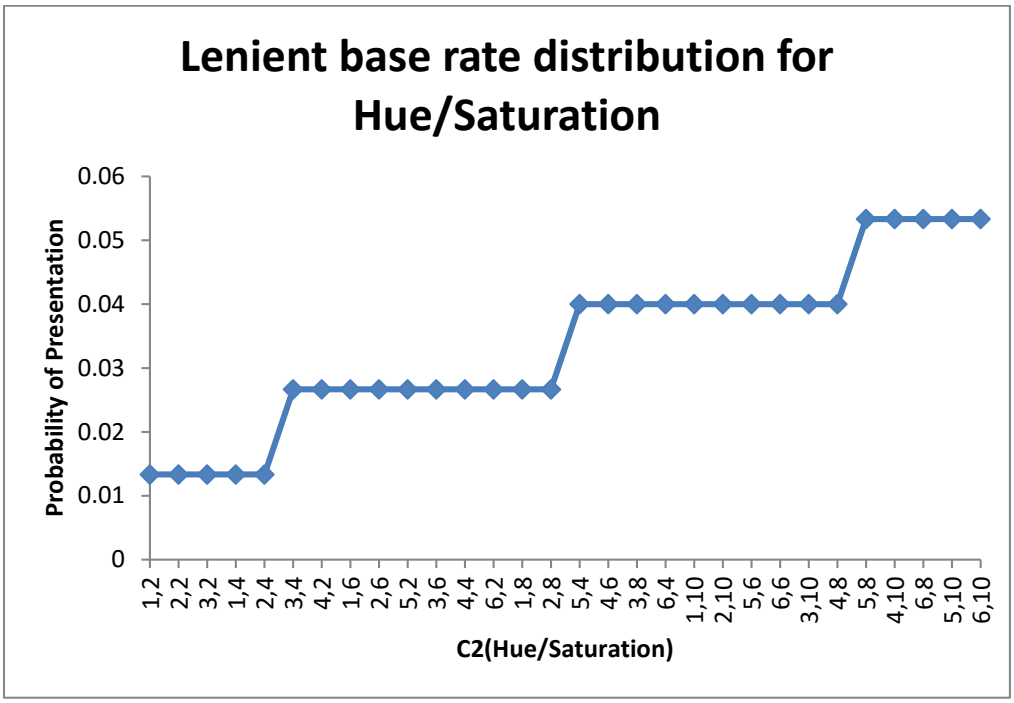


Figure 44. Signal base rate distribution for complex stimulus C_2 (Hue/Saturation) for lenient condition.

For the conservative conditions stimuli determined via the mapping functions to have a higher degree of similarity to the low signal had a higher probability of being selected for presentation than stimuli with a higher degree of similarity (see Figures 45 and 46).

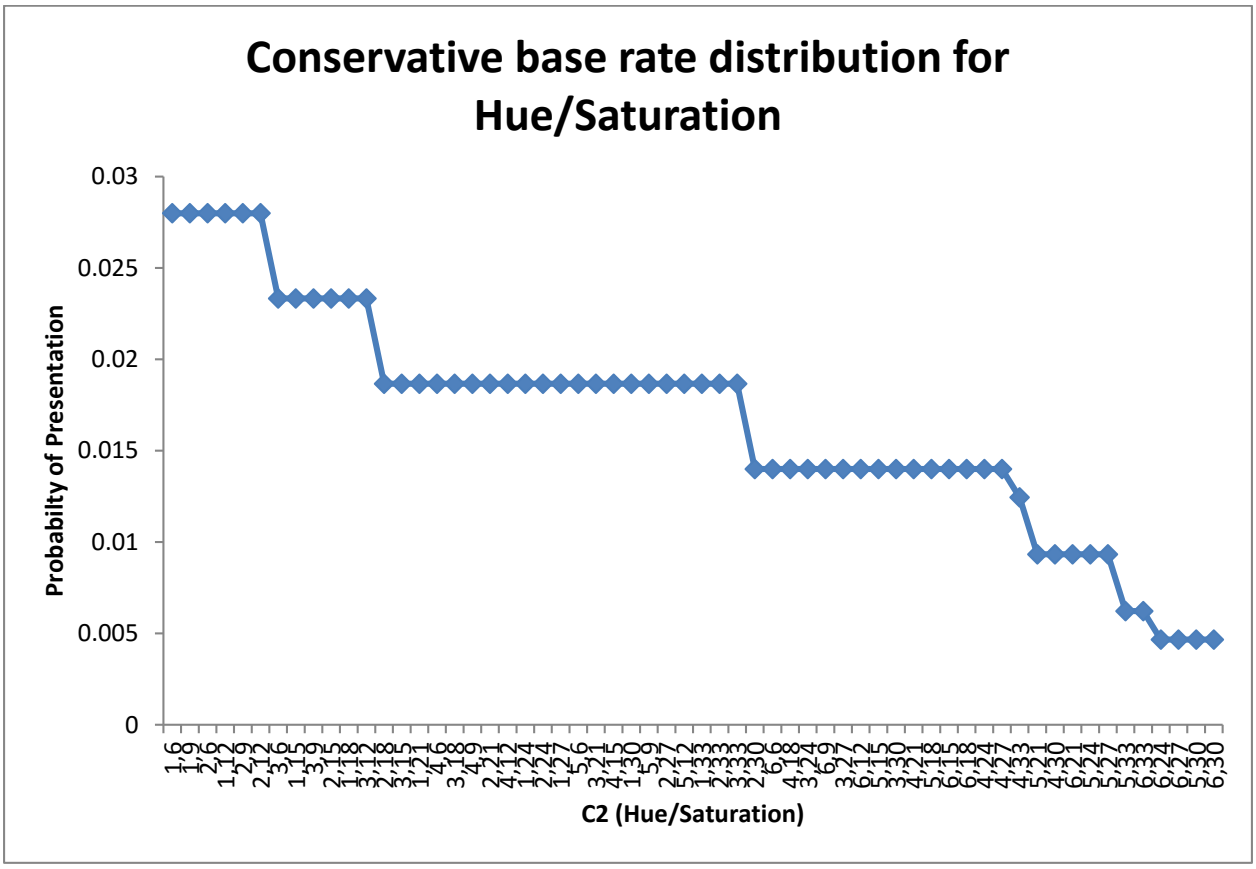


Figure 45. Signal base rate distribution for complex stimuli C_1 (Hue/Shape) for conservative condition. For shape, all polygon sides of the same defined set used the same probability.

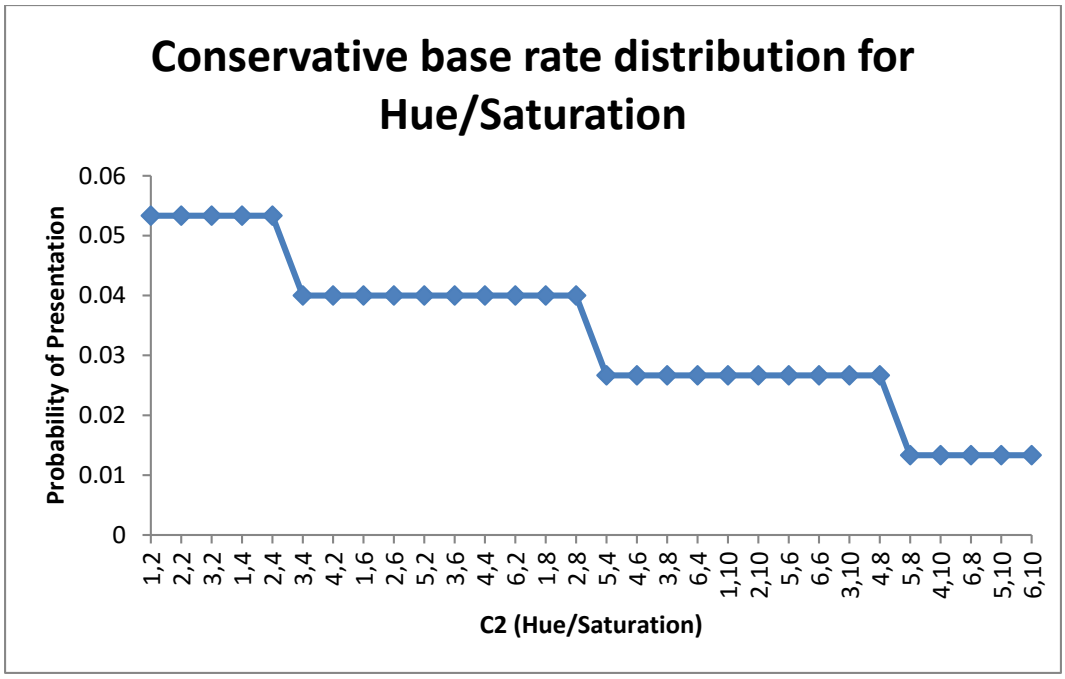


Figure 46. Signal base rate distribution for complex stimulus C_2 (Hue/Saturation) for conservative condition.

Results

The data from Study 3 were analyzed using MFSDT with the combined signal mapping functions determined by Study 2a, as well as by the unidimensional FSDT signal mapping functions determined by Study 1a. The response mapping function used by all three methods of analysis was determined by the distance the horizontal scroll bar was moved by the observer as a proportion of the total length of the scroll bar. For example, if the slider was moved exactly halfway across the sliding bar the response membership would equal 0.5.

Complex Stimulus C1 - Hue and Shape. For participant 1 the frequencies and rates of hits, false alarms, misses and correct rejections for complex stimulus C_1 are shown in Table 36 and 37, respectively, for MFSDT analysis, FSDT analysis with the unidimensional hue mapping function, and FSDT analysis with the unidimensional shape mapping function.

Table 36. *The four outcome frequencies for hue and shape data for Participant 1 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	405.88	26.34	47.69	245.09
Unbiased	351.17	15.89	69.51	288.44
Conservative	275.79	13.74	68.02	367.45
FSDT - Hue Function				
Lenient	372.5	70.55	81.06	200.88
Unbiased	304.29	45.93	116.38	258.39
Conservative	229.58	42.94	114.23	338.26
FSDT - Shape Function				
Lenient	339.74	85.79	113.83	185.64
Unbiased	299.85	84.03	120.83	220.30
Conservative	224.13	82.41	119.68	298.78

Table 37. *The four outcome rates for hue and shape data for Participant 1 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	0.939	0.061	0.163	0.837
Unbiased	0.957	0.043	0.194	0.805
Conservative	0.952	0.048	0.156	0.843
FSDT - Hue Function				
Lenient	0.841	0.159	0.288	0.712
Unbiased	0.869	0.131	0.311	0.689
Conservative	0.842	0.158	0.252	0.748
FSDT - Shape Function				
Lenient	0.798	0.201	0.380	0.620
Unbiased	0.781	0.219	0.354	0.646
Conservative	0.731	0.269	0.286	0.714

Table 38 reports the goodness of fit (χ^2) , estimate of perceptual sensitivity (A_z), response bias (β_{ln}), and estimates of the intercept (a) and slope (b) for the z-score form of the ROC analysis for

MFSDT, FSDT hue mapping function, and FSDT shape mapping function. As all three methods of analysis fit the equal-variance model, only the parameters of the equal variance model are reported.

Table 38. *Displays goodness of fit, sensitivity, and criterion bias calculated for all three methods of analysis for Participant 1 for hue and shape.*

Method	χ^2	A(z)	A	B	Conservative β_{ln}	Unbiased β_{ln}	Lenient β_{ln}
MFSDT	0.543	0.966	2.586	1.00	0.790	1.112	0.749
FSDT - Hue	5.175	0.884	1.688	1.00	0.417	0.540	0.355
FSDT-Shape	0.073	0.796	1.171	1.00	0.031	0.237	0.318

For all three methods of analysis on C_1 conditions for Participant 1, a predicted linear ROC was obtained (see Figure 47), indicating that the data for both MFSDT analysis and the two FSDT analysis was consistent with the assumption of normality. Additionally, as all three slopes of the z-score form ROC are equal to one, the data conforms to the equal variance assumption for all three methods of analysis. The perceptual sensitivity for the MFSDT analysis ($A_z = 0.966$) was higher than both the FSDT Hue mapping function analysis ($A_z = 0.884$) and the FSDT shape mapping function analysis ($A_z = 0.796$).

For the MFSDT and FSDT hue mapping function ROCs for Participant 1 the unbiased condition ($\beta_{ln\text{-}MFSDT} = 1.112$, $\beta_{ln\text{-}FSDT\text{-}Hue} = 0.540$) was more lenient than the lenient condition ($\beta_{ln\text{-}MFSDT} = 0.749$, $\beta_{ln\text{-}FSDT\text{-}Hue} = 0.355$). However the unbiased condition for FSDT shape mapping function ROC ($\beta_{ln\text{-}FSDT\text{-}Shape} = 0.237$) for Participant 1 is more conservative than the lenient condition ($\beta_{ln\text{-}FSDT\text{-}Shape} = 0.318$). For all three method of analysis the conservative condition was the most conservative response bias.

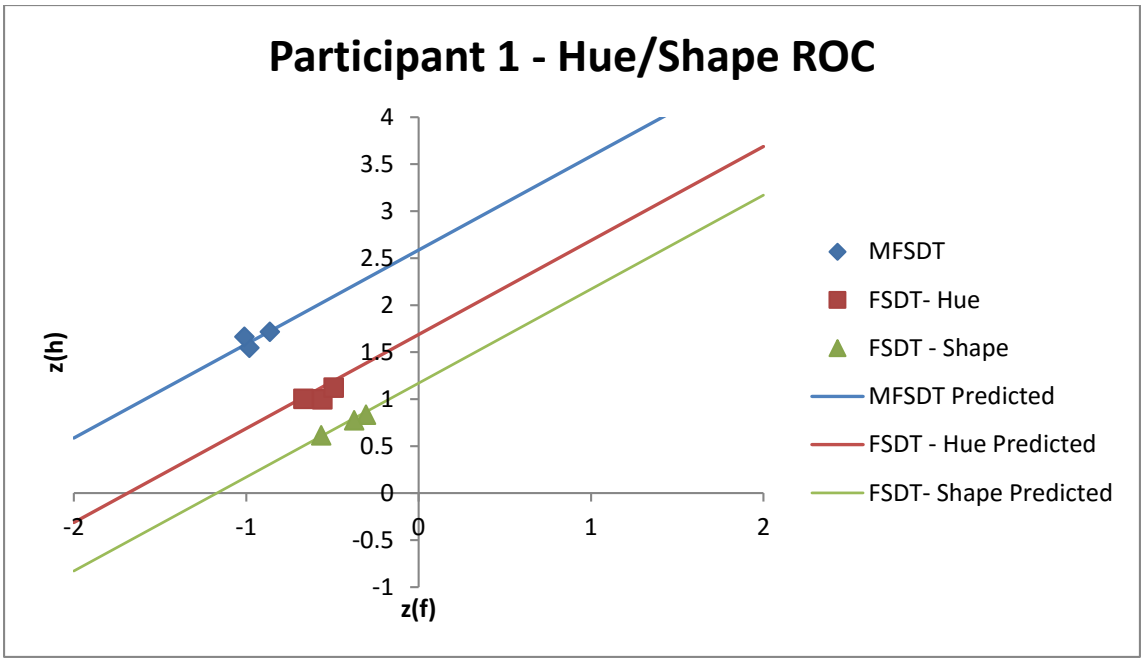


Figure 47. Linear Z-score form ROCs for MFSDT, FSDT - Hue, FSDT - Shape analysis on hue and shape data for Participant 1

For participant 2 the frequencies and rates of hits, false alarms, misses and correct rejections for complex stimulus C_1 are shown in Table 39 and 40, respectively, for all three methods of analysis.

Table 39. The four outcome frequencies for hue and shape data for Participant 2 calculated by each of the three models model

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	417.44	15.63	91.23	200.70
Unbiased	354.88	3.97	133.17	232.99
Conservative	286.90	4.11	123.53	310.46
FSDT - Hue Function				
Lenient	365.50	47.74	143.17	168.60
Unbiased	311.91	29.41	176.14	207.54
Conservative	233.64	37.78	176.78	276.80
FSDT - Shape Function				
Lenient	384.01	68.89	124.65	147.45
Unbiased	326.69	49.68	161.36	187.27
Conservative	252.47	58.14	157.95	256.44

Table 40. *The four outcome rates for hue and shape data for Participant 2 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	0.964	0.036	0.312	0.688
Unbiased	0.989	0.011	0.364	0.636
Conservative	0.986	0.014	0.285	0.715
FSDT - Hue Function				
Lenient	0.884	0.116	0.459	0.541
Unbiased	0.914	0.086	0.459	0.541
Conservative	0.861	0.139	0.390	0.610
FSDT - Shape Function				
Lenient	0.848	0.152	0.458	0.542
Unbiased	0.868	0.132	0.463	0.537
Conservative	0.813	0.187	0.381	0.619

Table 41 reports the goodness of fit (χ^2) , estimate of perceptual sensitivity (A_z), response bias (β_{ln}), and estimates of the intercept (a) and slope (b) for the z-score form of the ROC analysis for

MFSDT, FSDT hue mapping function, and FSDT shape mapping function. As all three methods of analysis fit the equal-variance model, only the parameters of the equal variance model are reported.

Table 41. *Displays goodness of fit, sensitivity, and criterion bias calculated for all three methods of analysis for Participant 2 for hue and shape.*

Method	χ^2	A(z)	A	B	Conservative β_{ln}	Unbiased β_{ln}	Lenient β_{ln}
MFSDT	4.919	0.961	2.501	1.00	1.789	2.294	1.731
FSDT - Hue	1.275	0.835	1.375	1.00	0.555	0.845	0.756
FSDT-Shape	0.319	0.797	1.175	1.00	0.343	0.596	0.542

For all three methods of analysis of C_1 for Participant 2, a predicted linear ROC was obtained (see Figure 48), indicating that the data for both MFSDT analysis and FSDT analyses was consistent with the assumption of normality. Additionally, as all three slopes are equal to one, the data conforms to the equal variance assumption for all three methods of analysis. As with Participant 1, the perceptual sensitivity for Participant 2 for the MFSDT analysis ($A_z = 0.961$) was higher than both the FSDT Hue mapping function analysis ($A_z = 0.835$) and the FSDT saturation mapping function analysis ($A_z = 0.797$).

For all three methods of analysis, Participant 2's response bias in the unbiased condition ($\beta_{ln\text{-}MFSDT} = 2.294$, $\beta_{ln\text{-}FSDT\text{-}Hue} = 0.845$; $\beta_{ln\text{-}FSDT\text{-}Shape} = 0.596$) was more lenient than the lenient condition ($\beta_{ln\text{-}MFSDT} = 1.731$, $\beta_{ln\text{-}FSDT\text{-}Hue} = 0.756$; $\beta_{ln\text{-}FSDT\text{-}Shape} = 0.542$). This differs from Participant 1, whose FSDT shape mapping function ROC showed the unbiased condition was more conservative than the lenient condition.

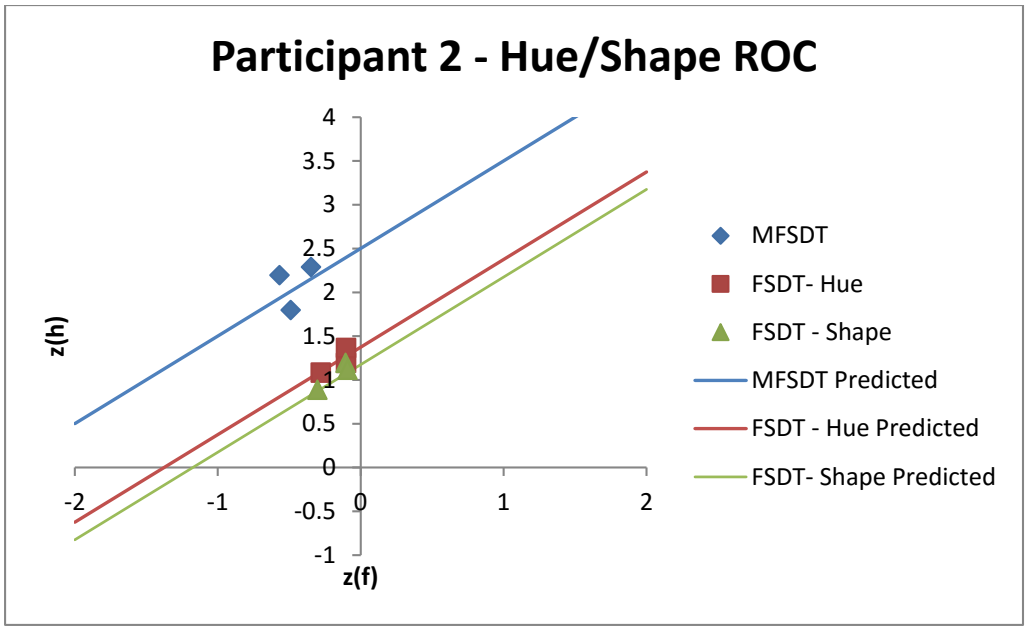


Figure 48. Linear Z-score form ROCs for MFSDT, FSDT - Hue, FSDT - Shape analysis on hue and shape data for Participant 2.

Complex Stimulus C2 - Hue and Saturation. For Participant 1 the frequencies and rates of hits, false alarms, misses and correct rejections for complex stimulus C_1 are shown in Table 42 and 43, respectively, for MFSDT analysis, FSDT analysis with the unidimensional hue mapping function, and FSDT analysis with the unidimensional shape mapping function.

Table 42. *The four outcome frequencies for hue and saturation data for Participant 1 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	262.65	127.00	21.76	313.59
Unbiased	265.59	83.54	26.01	349.86
Conservative	195.27	94.20	21.79	413.73
FSDT - Hue				
Lenient	223.64	154.29	60.78	286.29
Unbiased	229.21	114.52	62.39	318.87
Conservative	155.19	128.94	61.88	378.99
FSDT - Saturation				
Lenient	264.06	152.93	20.36	287.66
Unbiased	254.53	111.52	41.40	321.88
Conservative	188.57	113.36	28.50	394.58

Table 43. *The four outcome rates for hue and saturation data for Participant 1 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	0.674	0.324	0.065	0.935
Unbiased	0.761	0.239	0.069	0.931
Conservative	0.675	0.325	0.050	0.950
FSDT - Hue				
Lenient	0.592	0.408	0.175	0.825
Unbiased	0.667	0.333	0.164	0.836
Conservative	0.546	0.454	0.140	0.860
FSDT - Saturation				
Lenient	0.633	0.367	0.066	0.934
Unbiased	0.692	0.308	0.114	0.886
Conservative	0.625	0.375	0.067	0.933

Table 44 reports the goodness of fit (χ^2), estimate of perceptual sensitivity (A_z), response bias (the natural logarithm of β: β_{ln}), and estimates of the intercept (a) and slope (b) for the z-

score form of the ROC for MFSDT analysis, FSDT analysis using only the unidimensional hue mapping function, and the FSDT analysis using only the unidimensional saturation mapping function for Participant 1. As all three methods of analysis fit the equal variance model, only the parameters of the equal variance model are reported.

Table 44. *Displays goodness of fit, sensitivity, and criterion bias calculated for all three methods of analysis for hue and saturation for Participant 1.*

Method	χ^2	A(z)	A	B	Conservative β_{ln}	Unbiased β_{ln}	Lenient β_{ln}
MFSDT	1.674	0.930	2.084	1.00	-1.235	-0.779	-1.161
FSDT - Hue	3.359	0.813	1.257	1.00	-0.605	-0.341	-0.450
FSDT-Saturation	0.890	0.897	1.790	1.00	-1.052	-0.655	-1.022

For all three methods of analysis, the predicted linear ROC was obtained (see Figure 49), indicating that the data for both MFSDT analysis and FSDT analysis was consistent with the assumption of normality. Additionally, as all three slopes are equal to one, the data conforms to the equal variance assumption for all three methods of analysis. The perceptual sensitivity for the MFSDT analysis ($A_z = 0.93$) was higher than both the FSDT Hue mapping function analysis ($A_z = 0.813$) and the FSDT saturation mapping function analysis ($A_z = 0.897$).

For all three methods of analysis the Unbiased condition ($\beta_{ln\text{-}MFSDT} = -0.779$, $\beta_{ln\text{-}FSDT\text{-}Hue} = -0.341$; $\beta_{ln\text{-}FSDT\text{-}Saturation} = -0.655$) was more lenient than the lenient condition ($\beta_{ln\text{-}MFSDT} = -1.161$, $\beta_{ln\text{-}FSDT\text{-}Hue} = -0.450$; $\beta_{ln\text{-}FSDT\text{-}Saturation} = -1.022$).

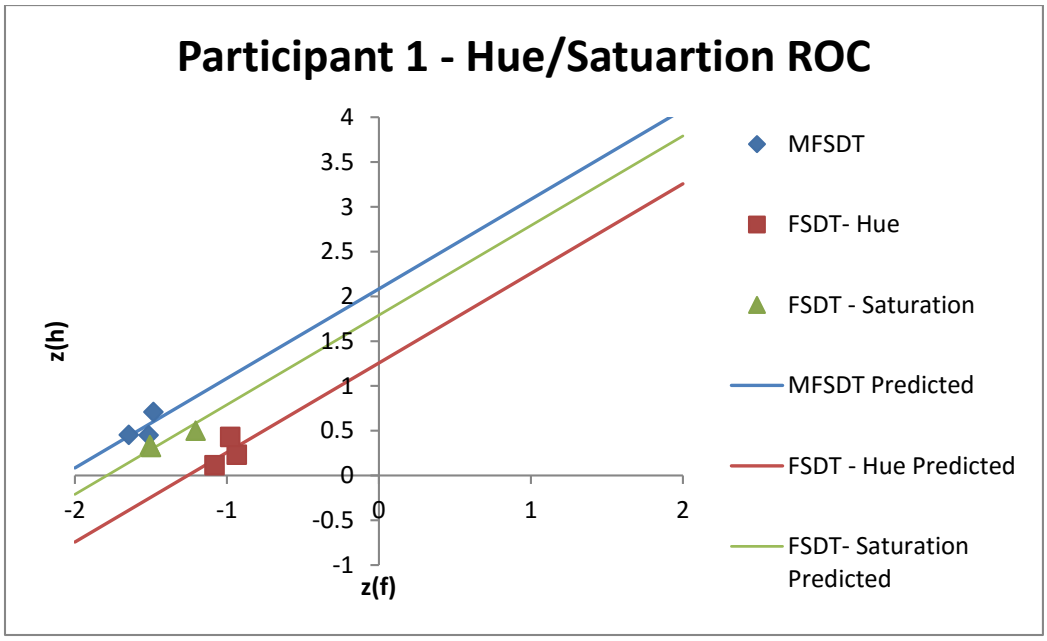


Figure 49. Linear Z-score form ROCs for MFSDT, FSDT - Hue, FSDT - Saturation analysis on hue and saturation data for Participant 1.

For Participant 2 the frequencies and rates of hits, false alarms, misses and correct rejections for complex stimulus C_2 are shown in Table 45 and 46, respectively, for all three methods of analysis.

Table 45. *The four outcome frequencies for hue and saturation data for Participant 2 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	363.08	53.47	45.94	262.52
Unbiased	278.85	61.21	60.82	324.12
Conservative	223.38	64.65	61.77	375.0
FSDT - Hue				
Lenient	320.09	89.60	88.92	226.39
Unbiased	233.33	95.45	106.34	289.88
Conservative	179.18	104.56	105.97	335.30
FSDT - Saturation				
Lenient	346.64	85.88	62.37	230.10
Unbiased	277.92	88.47	61.75	296.85
Conservative	228.28	74.93	60.75	366.20

Table 46. *The four outcome rates for hue and saturation data for Participant 2 calculated by each of the three models model*

	Hits	Misses	False Alarms	Correct Rejections
MFSDT				
Lenient	0.872	0.128	0.149	0.851
Unbiased	0.820	0.180	0.158	0.842
Conservative	0.776	0.224	0.141	0.859
FSDT - Hue				
Lenient	0.781	0.219	0.282	0.718
Unbiased	0.710	0.290	0.268	0.732
Conservative	0.632	0.368	0.240	0.760
FSDT - Saturation				
Lenient	0.801	0.199	0.213	0.787
Unbiased	0.759	0.241	0.172	0.828
Conservative	0.753	0.247	0.142	0.858

Table 47 reports the goodness of fit (χ^2), estimate of perceptual sensitivity (A_z), response bias (the natural logarithm of β: β_{ln}), and estimates of the intercept (a) and slope (b) for the z-

score form of the ROC for MFSDT analysis, FSDT analysis using only the unidimensional hue mapping function, and the FSDT analysis using only the unidimensional saturation mapping function for Participant 2. As all three methods of analysis fit the equal variance model, only the parameters of the equal variance model are reported.

Table 47. *Displays goodness of fit, sensitivity, and criterion bias calculated for all three methods of analysis for hue and saturation for Participant 2.*

Method	χ^2	A(z)	A	B	Conservative β_{ln}	Unbiased β_{ln}	Lenient β_{ln}
MFSDT	5.098	0.918	1.970	1.00	-0.300	-0.080	0.118
FSDT - Hue	4.757	0.799	1.186	1.00	-0.208	-0.038	0.126
FSDT-Saturation	0.575	0.883	1.680	1.00	-0.329	-0.203	0.035

As with Participant 1, all three methods of analysis for Participant 2 resulted in a predicted linear ROC was obtained (see Figure 50), indicating that the data for both MFSDT analysis and FSDT analysis was consistent with the assumption of normality. Additionally, as all three slopes are equal to one, the data conforms to the equal variance assumption for all three methods of analysis. The perceptual sensitivity for the MFSDT analysis ($A_z = 0.918$) was higher than both the FSDT Hue mapping function analysis ($A_z = 0.799$) and the FSDT saturation mapping function analysis ($A_z = 0.883$).

For all three methods of analysis of C_2 for Participant 2, the response bias of the conservative condition ($\beta_{ln\text{-}MFSDT} = -0.300$, $\beta_{ln\text{-}FSDT\text{-}Hue} = -0.208$; $\beta_{ln\text{-}FSDT\text{-}Saturation} = -0.329$) was more conservative than the response bias for the unbiased condition ($\beta_{ln\text{-}MFSDT} = -0.080$, $\beta_{ln\text{-}FSDT\text{-}Hue} = -0.038$; $\beta_{ln\text{-}FSDT\text{-}Saturation} = -0.203$) which in turn was more conservative than the response bias for the lenient condition ($\beta_{ln\text{-}MFSDT} = 0.118$, $\beta_{ln\text{-}FSDT\text{-}Hue} = 0.126$; $\beta_{ln\text{-}FSDT\text{-}Saturation} = 0.035$). This differs from Participant 1 whose response bias in the unbiased condition was more lenient than their response bias in the lenient condition.

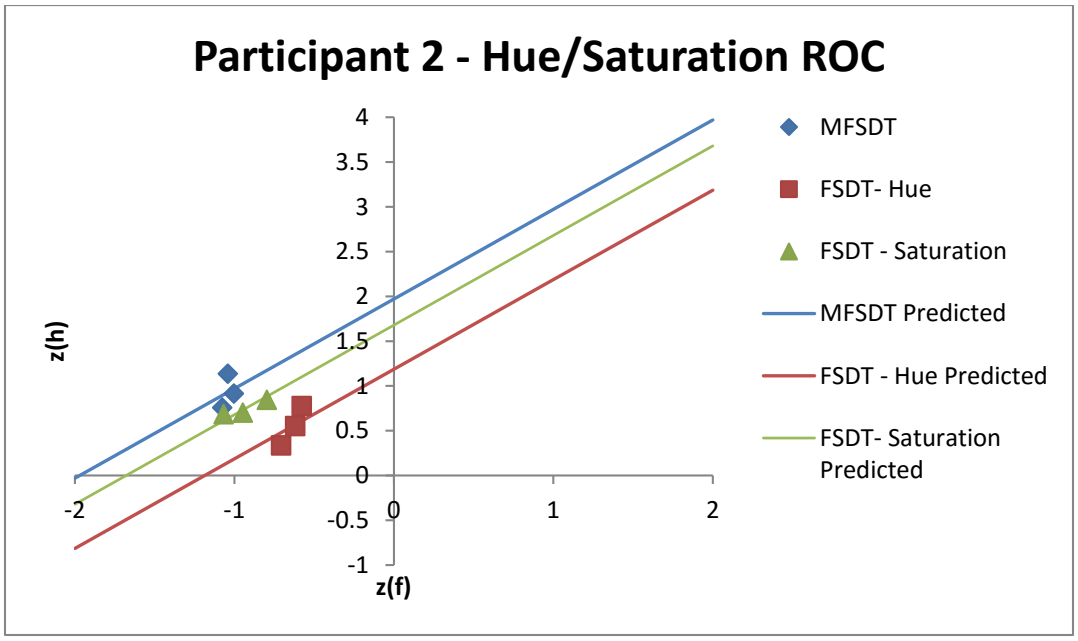


Figure 50. Linear Z-score form ROCs for MFSDT, FSDT - Hue, FSDT - Saturation analysis on hue and saturation data for Participant 2

Discussion

The purpose of Study 3 was to empirically test the viability of MFSDT. If a linear z-score form ROC with a slope of one was found, this would be evidence that MFSDT met the normality and equal variance assumptions required to use traditional SDT performance measures, such as d' and β. The results revealed that the statistical assumptions of SDT appear to extend to MFSDT. For both participants and both types of complex stimuli, perceptually separable and perceptually integral, MFSDT analysis yielded a linear z-score form ROC with a slope of one. This is consistent with results from previous unidimensional FSDT studies regarding the normality and equal variance assumptions of SDT when using fuzzy analysis and indicates the transformation of the mapping function from unidimensional to multidimensional does not violate these assumptions.

While participants were asked to actively consider both dimensions of the stimulus when determining the stimulus signal-ness this does not preclude running a unidimensional FSDT

analysis on the data. FSDT analysis can be applied to data where the stimulus to be detected has identifiable component dimensions, it but only uses the mapping function for one of these dimensions in the analysis. This may result in a loss of information regarding the observer's performance, but such information loss is likely to occur when a multidimensional stimulus is evaluated using a unidimensional SDT model. By analyzing the data from Study 3 with MFSDT, FSDT using the hue mapping function, and FSDT using the shape or saturation mapping function depending on the complex stimulus, we are able to compare the two methods with the same response data. The results indicated an increase in sensitivity with MFSDT analysis when compared to either of the FSDT analysis for both participants and with both perceptually separable and perceptually integral stimulus dimensions. These empirical results are consistent with the artificial example described in the introduction to this work as well as the comparison of MSDT to SDT with a brightness/loudness example from Macmillan and Creelman (2005).

As with the simulated ROC in Study 2, a high degree of clustering of points on the ROC plots was observed for the differing bias conditions. As stated previously this clustering effect has been observed in multiple FSDT studies both empirical and simulated (e.g. Szalma & Hancock, 2013) and is likely a function of FSDT rather than the multidimensional aspect of MFSDT. However, with MFSDT analysis the difference between β_{ln} between conditions was typically larger than with either FSDT analysis, resulting in less clustering.

An additional item of note was that in some cases the bias manipulation did not result with the appropriate bias ordering. That is, in some cases what was to be a lenient manipulation yielded with response biases that were more conservative than an unbiased manipulation. This result of having bias manipulations not correspond to the expected response biases has been

observed in other FSDT studies (e.g., Szalma et al., 2006), though they used a different bias manipulation technique, payoff matrixes, rather than signal base rate manipulation to effect bias in the present investigation. The reordering of the biases may be due to the high degree of clustering between the separate bias conditions, and the reordering may be due to statistical error since the differences between the response biases are small. For example, Participant 2's response bias in the unbiased condition for complex stimulus C_1 when analyzed by FSDT shape mapping function was $\beta_{ln} = 0.596$ and the lenient condition was $\beta_{ln} = 0.542$, only a 0.054 difference in β_{ln}.

CHAPTER 7: DISCUSSION

An underlying goal of any statistical model is to model the true state of the world as closely as possible. While Signal Detection Theory is an extremely useful tool for analyzing information about detection tasks and describing observer's performance, it fails in capturing some aspects of real world detection. There have been proposed extensions of SDT which seek to account for information that SDT fails to capture, such as Multidimensional Signal Detection Theory and the potential multidimensional nature of signals, and Fuzzy Signal Detection Theory and the potential ambiguous nature of the signal and/or response. Multidimensional Fuzzy Signal Detection Theory attempts to integrate both MSDT and FSDT in order to capture more information about detection tasks where there are multiple identifiable dimensions to an ambiguous signal or response. MFSDT retains the fuzzy set theory and mixed implication function procedure from FSDT in order to maintain the ability to describe fuzzy stimuli while modifying FSDT mapping functions to allow for signals or responses to be defined by multiple component dimensions. The proposed form of MFSDT mapping functions was as shown in Equation 9.

$$S = \sum_{i=1}^{n} w_i \, s_i(a_i) + \mathrm{C}$$
$$w_1 + \ldots + w_n = 1$$

Where each component dimension had its own mapping function with which the combined mapping function scaled individual component mapping functions and summed all the scaled components. The purpose of the present sequence of studies was to test the viability of this proposed multidimensional change to FSDT. From MSDT it is known that the individual component dimensions may either be perceptually separable, where the level of one dimension has no effect on the perception of the level of the other dimensions, or perceptually integral, where the level of one dimension has an effect on the perception of the other dimension. The

manipulations in each of the three studies were applied to two difference kinds of stimulus combinations, one with a perceptually separable complex stimuli, hue and shape, and one with a perceptually integral complex stimuli, hue and saturation. This was done to test the validity of MFSDT and whether it is applicable to both types of multidimensional stimuli.

Mapping Functions

In order to use FSDT and consequentially MFSDT, information on the signal dimensions is needed in order to drive the development of mapping functions. As noted by Parasuraman et al. (2000) mapping functions may be derived through regulations, theory or empirical observations. This should also be true for MFSDT, however given the lack of regulations or theory to drive the definition of mapping functions for hue, saturation, and shape, empirical observations were needed. Pair-wise comparisons and Thurstone scaling were used to develop the individual mapping functions. A clear result of applying these methods is that the development of valid mapping functions requires serious consideration of the methods applied, as empirical observations may result in drastically different functions depending on the way in which the data are collected and analyzed.

The pair-wise comparison method appears to be an effective approach to obtaining relational information for stimuli to be used in a mapping function where the stimuli has a small number of potential levels. However when the stimuli, such as shape in this case, has a large number of potential combinations of stimulus levels, either a potentially prohibitively large sample size is needed to gather relational information about the full factorial combined pairs, or adjustments to the method are required, such as grouping stimuli into sets and comparing these sets against each other rather than comparing the specific stimulus levels. These adjustments may influence the form of the mapping function, as was the case with shape in the present

investigation, where unique signal membership values could not be empirically assigned to individual stimulus levels, and instead all members of a given set was assigned the same signal membership value.

The use of Thurstone scaling as a method of analysis for developing mapping functions was also successful but with caveats. For both the hue and the adjusted shape pair-wise data, applying the scaling was successful with no indication of problems with applying a normalization process to the results. For saturation however when the Thurstone scaling was applied the resulting mapping function was influenced by the forced normalization processes in a way that made the function non-monotonic. When the function was calculated through the frequency of target similarity judgments without a normalization adjustment, the function was increasing and monotonic, i.e., each increase in magnitude resulted in an increase in signal membership value. However, when Thurstone scaling and thus normalization was applied the ordinality of stimulus levels were violated, i.e., there were cases in which signal membership value decreased compared to signals with a lower physical magnitude. While there is no theoretical requirement that a signal membership value must increase as signal magnitude increases, saturation has been studied in both psychophysics (e.g. Gorden, Abramov, & Chan, 1994) and MSDT research (e.g. Garner and Felfoldy, 1977). Indeed, hue, saturation and shape were selected for the present investigation because the properties of these dimensions have been well established by previous research. Hence, for saturation there is no logical reason why such a loss of ordinality should occur, indicating that for this dimension applying a normalization procedure was not appropriate.

As of now there has been no research with regard to the effects of systematic manipulation of mapping functions on FSDT outcomes, and while that question is outside the

scope of this research, these results do indicate that method and analysis for empirical observations does have an effect on the shape of the resulting mapping function. That is *the form of a mapping function may be influenced by the method used to derive it*. As such it is advisable to consider whether or not the empirical method one uses to develop your mapping function is appropriate for the perceptual dimension of interest.

Determining Scale Factor

One way to establish a multidimensional mapping function is to derive a weighted combination of the mapping function for each individual component dimension. Established MSDT methodologies, specifically a complete identification task, a redundancy and control tasks and a filtering task, can be employed to determine these weights or scale factors. Aside from providing a way to determine the scale factors methods were used to verify the results of previous research that hue and shape are perceptually separable and hue and saturation are perceptual integral. Most of the information collected from the MSDT tasks was, in the end, unnecessary for determining scale factors. The best approach to developing the scale factors empirically was to adjust the scale factors on the complex mapping function and fit the function outcomes to the confusion matrixes that resulted from the complete identification task. Hence, this method was not without its problems. First, the confusion matrix from the complete identification experiment had to be transformed into the four SDT outcomes. While many of the cells on the confusion matrix were identifiable as belonging to only one outcome, there are cells which could be defined as more than one outcome. The multiple outcomes of a single cell was accounted for however the result was to fit to the modified MSDT results rather than fit to MSDT results.

A second problem was one of extremes, extreme performance and extreme stimuli. For all three participants the confusion matrix from the complete identification experiment was sparse, that is to say the participants did not commit many errors. For example, no participants ever responded to a purple circle by identifying it as a blue hexagon. With any form of SDT, the closer to perfect performance the participants are the more difficult it is to fit the model. This may be due to the fact the end points of each individual dimension range were used for the complete identification task and there was therefore very little confusion in identifying whether a stimuli is at the lowest level or at the highest level. For instance, participants experienced little difficulty in determining that a presented shape was either a hexagon or a circle. Using the end points for the complete identification made the FSDT analysis less than ideal. According to our mapping functions, any of the three stimuli types either have a low signal membership value between 0.0 and 0.05 or a high signal membership value between 0.95 and 1.0. With membership values that close to the end points of signal membership values, the outcome membership values would in many cases mimic the crisp case where one outcome membership would receive a membership value of one, and the other outcomes would have a membership value of zero. Essentially with any given event only one outcome would be increase in membership value, rather than membership credit being given to multiple of the four potential outcomes. As the goal was to try to fit a multidimensional mapping function to the MSDT data, this method may have been hindered by acting like the crisp case rather than the fuzzy one, since when the values are that close to one or zero, the combined mapping function value was determined almost exclusively by one of the individual component mapping functions, instead of both component dimension mapping function contributing to the combined signal membership.

A potential way of solving some of these problems would be to choose non-endpoint levels in the range of stimuli, so that there is more uncertainty as to the level of stimulus in the complete identification task. This should result in a less sparse confusion matrix, as well as allowing FSDT to benefit from assigning membership values to potentially more than one outcomes on a given trial. A second way could be as simple running only one of the MSDT tasks as oppose to the five tasks run in this study, so that a larger N could be achieved, allowing for techniques such as factor analysis or multiple regression to be used rather than relying on trying to fit FSDT and MSDT together.

MFSDT and ROC Analysis

Since MFSDT is an extension of FSDT there is nothing in the traditional SDT decision space which indicates the fuzzy aspect of the theory and therefore the structure of the theoretical decision space is unclear (Szalma & Hancock, 2013). FSDT research has shown that information regarding the tenability of FSDT can be gained through the use of ROCs, specifically determining whether the normality and equal variance assumptions of SDT are being met (Murphy, et al., 2004; Szalma, et al. 2006; Szalma & O'Connell, 2011; Szalma & Hancock, 2013). Since ROC analysis was successfully applied to FSDT, MFSDT tenability was tested using the ROC approach as well. One ROC was computed from data derived from a simulated decision task involving both of the complex stimuli. The results from the simulated ROC indicated that MFSDT was viable for both perceptually separable and perceptually integral complex stimuli and that it did not violate the traditional SDT assumptions of normality or equal variance. The second ROC analysis used both complex stimuli in an empirical decision making task, and the results were consistent with those obtained with the simulated data. Thus, the

evidence indicates that MFSDT is viable and that combining the two models does not result in violations of the normality and equal variance assumptions.

MFSDT and Applied Research

The operational definitions of signal or response categories in operational contexts are often more complex than in laboratory settings. In the laboratory, information regarding the signal and response is fully defined; the experimenters know the true state of the world, as they often define it. However, in the real world, the true state of the world is often unknown or unclear, in that the definition of what constitutes an instance of the category 'signal' is sometimes not easily specified as mutually exclusive categories. It was this fact of uncertainty that motivated the development of FDST. FDST is ideally suited to capture ambiguity of the signal category itself that is often present in signal detection tasks in operational environments. MFSDT is also ideally suited to capture that ambiguity, and to extend the model to cases in which the stimulus categories cannot or should not be represented as a unidimensional variable. However, use of MFSDT requires a level of understanding about the signal or response that FSDT does not necessitate. In order to derive weights one must understand the components of that comprise a complex signal. MFSDT in applied research could be used as an additional technique for identifying how component dimensions affect human performance in decision and detection tasks.

Conclusions and Future Work

One area unique to MFSDT that needs further examination is how to develop the scaling factors (weights) for the multidimensional mapping function. While the use of MSDT to gather information about the individual component dimensions was effective, there are potentially many other ways both empirically and theoretically to drive the determination of scale factors.

However most of the issues to be resolved for MFSDT are the same as what needs to be done for FSDT.

Mapping functions in general need further exploration, both in how best to develop them since method and analysis have significant effects on the shape of the functions, and how differences in the form of mapping functions affect the fuzzy outcomes. In addition, to date work with MFSDT and FSDT has focused solely on the signal mapping functions. The form of response mapping function has not been investigated. The mixed implication functions are determined not only by the signal membership value of an event but also the response membership value, making the response mapping function just as important as the signal mapping function in the determination of fuzzy outcomes.

The meaning of response bias needs to be explored further with regards to FSDT and thus MFSDT. Research on both methods has shown high degree of clustering of ROC points with both empirical and simulated data indicating the manipulation of response bias have a small effect on the position of the points in ROC space. The cause for this clustering is unknown, but may be related to difficulty setting a fuzzy criterion or potentially in the empirical cases to be a factor of experimental design. It is possible that traditional SDT methodology for bias manipulation such as base rate manipulation and payoff matrices are not appropriate for experiments using fuzzy sets. The fact that biases often do not conform to their expected manipulated conditions may be evidence of the inappropriate nature of these methods. Alternatively, the clustering may be due to compression resulting from the mathematical and statistical procedures.

Despite the theoretical and methodological issues that need to be resolved in the development of mapping functions and the determination of scale factors, the results from both

the simulated ROC and the empirical ROC in the present set of studies indicate that a multidimensional fuzzy mapping function and thus MFSDT may be a viable tool for analysis of signal detection tasks. Additionally, given the evidence of normality and equal variance assumptions being met for data derived using these methods, the use of SDT performance metrics such as d' and β is appropriate for MFSDT. MFSDT offers a potentially useful methodological approach to modeling performance for complex signal detection tasks

APPENDIX A: THURSTONE SCALING FOR INDIVIDUAL SHAPE LEVELS

Table A- 1. *Average Z-score, Adjusted Z-Score and Proportion of highest z-score for each level of shape (number of sides).*

Number of Sides of Polygon	**Averaged Z-score**	**Adjusted Z-score**	**Proportion of highest z-score**
6	-4.05	0.371	0.046
7	-4.385	0.039	0.005
8	-4.424	0.00	0.00
9	-3.776	0.648	0.081
10	-3.557	0.867	0.108
11	-3.677	0.748	0.093
12	-3.050	1.374	0.171
13	-3.43	0.994	0.124
14	-2.567	1.858	0.232
15	-2.396	2.028	0.253
16	-2.502	1.922	0.24
17	-1.736	2.688	0.335
18	-1.613	2.811	0.351
19	-1.971	2.454	0.306
20	-0.510	3.915	0.489
21	-0.690	3.735	0.466
22	0.215	4.64	0.579
23	-0.627	3.797	0.474
24	0.662	5.087	0.635
25	0.756	5.18	0.464
26	0.928	5.352	0.668
27	1.465	5.890	0.735
28	1.687	6.112	0.763
29	2.329	6.754	0.843
30	2.350	6.775	0.846
31	3.166	7.59	0.947
32	1.290	5.714	0.713
33	3.404	7.829	0.977
34	2.548	6.973	0.870
35	3.588	8.013	1
36	1.559	5.983	0.747

APPENDIX B: RESULTS OF FILTERING, CONTROL AND REDUNDANCY TASKS

Table B - 1. *Means and Standard Deviations for Correct and Error Response Times and Percent Correct by Condition for Participant 1 and the Hue - Shape condition*

	Hue Condition					Shape Condition				
	Correct			Error		Correct			Error	
Condition	M	SD	%	M	SD	M	SD	%	M	SD
Filtering	607.9	319.9	98.2	565.2	194.2	580.2	286.2	96.3	414.4	191.5
Control 1	581.1	262.5	98.3	777.0	491.1	536.8	228.9	98.1	271.4	80.9
Control 2	546.4	270.1	98.1	365.0	177.5	544.3	243.0	97.8	547.6	200.3
Redundancy +	565.3	309.6	99.2	930.3	494.2	520.3	307.2	98.9	354.5	173.8
Redundancy -	579.3	259.4	98.1	781.9	357.5	572.2	259.0	98.3	449.3	148.2

Table B - 2. *Means and Standard Deviations for Correct and Error Response Times and Percent Correct by Condition for Participant 2 and the Hue - Shape condition*

	Hue Condition					Shape Condition				
	Correct			Error		Correct			Error	
Condition	M	SD	%	M	SD	M	SD	%	M	SD
Filtering	447.3	234.5	98.3	329.4	137.5	436.2	203.2	97.6	241.4	197.5
Control 1	459.1	263.9	98.9	1087.5	711.6	442.7	237.0	99.1	623.0	522.2
Control 2	422.5	209.5	99.2	270.0	85.6	433.2	211.4	98.6	318.0	268.4
Redundancy +	463.1	252.4	99.4	1154.0	789.9	420.8	277.6	98.1	233.6	137.3
Redundancy -	438.4	224.1	97.8	807.0	475.7	413.0	189.5	99.7	358.0	0.0*

* - no standard deviation due to only a single data point.

Table B - 3. *Means and Standard Deviations for Correct and Error Response Times and Percent Correct by Condition for Participant 3 and the Hue - Shape condition*

	Hue Condition					Shape Condition				
	Correct			Error		Correct			Error	
Condition	M	SD	%	M	SD	M	SD	%	M	SD
Filtering	280.2	145.1	97.9	124.3	58.9	309.1	151.2	94.6	124.9	68.4
Control 1	259.3	118.9	98.6	102.6	28.1	319.1	147.2	95.8	92.1	40.6
Control 2	291.6	159.4	97.2	107.4	59.2	309.2	170.6	93.9	137.1	77.1
Redundancy -	277.5	136.8	98.0	106.7	27.6	256.0	107.1	99.4	132.0	55.2
Redundancy +	277.9	135.9	97.5	154.0	150.4	292.4	158.6	98.3	199.7	234.2

Table B - 4. *Means and Standard Deviations for Correct and Error Response Times and Percent Correct by Condition for Participant 1 and the Hue - Saturation Condition.*

	Hue Condition					Saturation Condition				
	Correct			Error		Correct			Error	
Condition	M	SD	%	M	SD	M	SD	%	M	SD
Filtering	686.6	342.2	95.7	667.4	482.0	575.4	248.5	97.6	614.6	340.7
Control 1	572.3	245.8	98.3	418.3	129.9	532.0	271.1	98.1	427.6	244.1
Control 2	553.0	227.7	98.6	648.4	480.4	533.5	252.2	97.5	415.6	154.0
Redundancy -	524.5	235.7	98.3	342.7	116.7	543.2	226.3	98.6	333.4	130.0
Redundancy +	597.6	389.3	96.9	589.4	236.4	555.3	318.5	96.9	320.1	105.0

Table B - 5. *Means and Standard Deviations for Correct and Error Response Times and Percent Correct by Condition for Participant 2 and the Hue - Saturation Condition.*

	Hue Condition					Saturation Condition				
	Correct			Error		Correct			Error	
Condition	M	SD	%	M	SD	M	SD	%	M	SD
Filtering	527.9	739.6	95.4	540.6	362.4	441.1	224.4	97.9	536.3	494.5
Control 1	488.3	273.7	95.2	597.9	246.0	415.8	195.4	99.2	259.3	35.8
Control 2	404.3	198.3	98.9	299.8	144.2	417.8	212.7	99.4	202.5	44.5
Redundancy -	418.1	173.0	99.4	382.0	99.0	423.3	215.4	99.4	460.0	55.2
Redundancy +	483.4	351.8	98.3	415.7	205.8	405.8	232.8	98.1	287.1	129.2

Table B - 6. *Means and Standard Deviations for Correct and Error Response Times and Percent Correct by Condition for Participant 3 and the Hue - Saturation Condition.*

	Hue Condition					Saturation Condition				
	Correct			Error		Correct			Error	
Condition	M	SD	%	M	SD	M	SD	%	M	SD
Filtering	350.4	172.9	95.9	290.2	296.0	306.9	174.9	91.5	215.2	146.8
Control 1	321.5	136.5	98.0	115.6	32.2	278.4	146.5	96.9	108.6	48.9
Control 2	272.1	177.1	97.8	116.6	52.1	292.0	151.5	96.4	109.9	32.0
Redundancy -	290.0	198.0	97.2	180.6	232.3	287.5	168.2	97.5	133.2	46.2
Redundancy +	284.6	144.7	98.6	77.6	46.5	259.7	141.3	97.2	144.6	78.9

REFERENCES

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

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

Ashby, F.G., & Maddox, T.D. (1994). A response time theory of separability and integrality in speeded classification. *Journal of Mathematical Psychology, 38,* 428-466.

Ashby, F.G., & Maddox, T.D. (1991). A response time theory of perceptual independence. In J.P. Doignon & J.C. Falmange, (Eds.), *Mathematical Psychology: Current developments*, 389-413. Springer-Verlag.

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

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

Bonnel, A., Mottron, L., Peretz, I., Trudel, M., Gallun, E., & Bonnel A-M., (2003). Enhanced pitch sensitivity in individuals with autism: A signal detection analysis. *Journal of Cognitive Neuroscience, 15(2),* 226-235.

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

CIE (2004). CIE 15-2004. *Colorimetry*, CIE Central Bureau: Vienna, Austria.

Garner, W.R., (1977). The effect of absolute size on separability of the dimensions of size and brightness. *Bulletin of the Psychonomics Society, 9,* 380-382.

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

Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Gorden, J., Abramov, I., & Chan, H. (1994). Describing color appearance: Hue and saturation scaling. *Perception & psychophysics, 56(1),* 27-41.

Guilford, J.P. (1928). The method of paired comparisons as a psychometric method. *Psychological Review, 35(6),* 494-506.

Hancock, P.A., Masalonis, A.J., & Parasuraman, R. (2000). On the theory of fuzzy signal detection: Theoretical and practical considerations. *Theoretical Issues in Ergonomic Science, 1,* 207-230.

Hofer, F. & Schwaninger, A. (2005). Using threat image projection data for assessment of individual screener performance, *WIT Transactions on the Built Environment, 82,* 417-426.

Ishihara, S. (1993). *Ishihara's tests for colour-blindness*. Tokyo, Japan: Kanehara & Co.

Macmillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide, 2nd ed*. Mahwah, NJ: Erlbaum.

Maddox, T.D., (1992). Perceptual and decisional separability. In F.G. Ashby (Ed.) *Multidimensional Models of Perception and Cognition,* 147-180. Erlbaum.

Maddox, T.D., & 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.

Masalonis, *A.J.,* & Parasuraman, R. (2003) Fuzzy signal detection theory: Analysis of human and machine performance in air traffic control, and analytic considerations. *Ergonomics, 46,* 1045-1074.

MacAdam, D.L. (1935). Theory of the maximum visual efficiency of colored materials. *Journal of Optical Society of America, 25*, 249-252.

Morgan, J.F., Oron-Gilad, T., Hancock, P.A., & Szalma, J.L. (2005). *Perception of morphed objects: Generating psychophysical functions for the evolution of a shape*. Paper presented at the 2005 American Psychological Association Division 21/19 & Human Factors and Ergonomics Society Potomac Chapter Annual Symposium on Applied Experimental Research, George Mason University, Fairfax, VA.

Murphy, L., Szalma, J.L., & Hancock, P.A. (2004). Comparison of fuzzy signal detection and traditional signal detection theory: Analysis of duration discrimination of brief light flashes. *Proceedings of the Human Factors and Ergonomics Society, 48,* 2494-2498.

Newhall, S.M., Judd, D.B., & Nickerson, D. (1943) Final report of the OSA subcommittee on the spacing of Munsell Colors. *Journal of the Optical Society of America, 33,* 385-418.

Nickerson, D. (1976). History of the Munsell color system, company, and foundation. *Color Research and Application, 1(1),* 7-10.

Peterson, W.W., & Birdsall, T.G., (1953). *The theory of signal detectability* (*Technical Report No. 13*).University of Michigan: Electronic Defense Group,

Parasuraman, R., Hancock, P. A., & Olofinboba, O. (1997). Alarm effectiveness in driver-centred collision-warning systems. *Ergonomics*, *40*(3), 390-399.

Parasuraman, R., Masalonis, A.J., & Hancock, P.A. (2000). Fuzzy signal detection theory: Basic postulates and formulas for analyzing human and machine performance. *Human Factors, 42*, 636-659.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics, 30*, 286-297.

Posner, M.I. (1964). Information reduction in the analysis of sequential tasks. *Psychological Review, 71*, 491-504.

See J.E., Warm, J.S., Dember, W.N., & Howe, S.R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors, 39(1),* 14-29.

Siewert, B., Bly, B.M., Schlaug, G., Darby, D.G., Thangaraj, V., Warach, S., Edelman, R.R. (1996). Comparison of the BOLD- and EPISTAR- techniques for brain imaging by using signal detection theory. *Magnetic Resonance in Medicine, 36 (2),* 249-255.

Sorkin, R.D., & Woods, D. (1985). Systems with human monitors: A signal detection analysis. Human-Computer Interaction, 1, 49-75.

Soto, F.A., & Zheng, E. (2015). grtools: General recognition theory tools for the analysis of perceptual independence. R package version 0.1.2.

Stafford, S.C., Szalma, J.L., Hancock, P.A., & Mouloua, M. (2003). Application of fuzzy signal detection theory to vigilance: The effect of criterion shifts. *Proceedings of the Human Factors and Ergonomics Society*, *47*, 1678-1682.

Swets, J.A. (1979). ROC analysis applied to the evaluation of medical imaging techniques. *Investigative Radiology, 14*, 109-121.

Swets, J.A., & Pickett, R. M. (1982). *Evaluation of diagnostic systems: methods from signal detection theory.* Academic Press, New York, USA.

Szalma, J.L., & Hancock, P.A. (2013). A signal improvement to signal detection analysis: Fuzzy SDT on the ROCs. *Journal of Experimental Psychology: Human Perception and Performance.*

Szalma, J.L., & O'Connell, M. (2011). Fuzzy signal detection theory: A Monte Carlo investigation. *Proceedings of the Human Factors and Ergonomics Society, 55,* 1366-1369.

Szalma, J.L., Oron-Gilad, T., Saxton, B., & Hancock, P.A. (2006). Application of fuzzy signal detection theory to the discrimination of morphed tank images. *Proceedings of the Human Factors and Ergonomics Society, 50,* 1716-1720.

Tanner, W.P., & Swets, J.A. (1954). A decision making theory of visual detection. *Psychological Review, 61,* 401-409.

Thomas, E.A., & Hogue, A. (1976). Apparent weight of evidence, decision criteria, and confidence ratings in juror decision making. *Psychological Review, 83(6),* 442-465.

Travis, D. (1991). Effective color displays: Theories and practice. London: Academic Press.

Wickens, T.D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.