Exploring Strategies for Melanoma Detection Utilizing Discrimination Training

Jonathan R. Miller

Submitted to the graduate degree program in Applied Behavioral Science and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

_____

Derek D. Reed, Ph.D. Chairperson

_____

Florence DiGennaro Reed, Ph.D. Committee Member

_____

Kimberly Engelman, Ph.D. Committee Member

_____

David P. Jarmolowicz, Ph.D. Committee Member

_____

Kathryn J. Saunders, Ph.D. Committee Member

Date Proposed: __5/2/13__

Date Defended: __7/18/13_

The Dissertation Committee for Jonathan R. Miller certifies that this is the approved version of the following dissertation:

Exploring Strategies for Melanoma Detection Utilizing Discrimination Training

_____

Derek D. Reed, Ph.D. Chairperson

Date approved: _____

Abstract

Melanoma is the deadliest form of skin cancer. Although melanoma is amenable to visual identification by those who might suffer from the disease, no consensus exists on a single strategy for promoting early detection. To date, the application of behavioral principles has been all but absent from the literature attempting to address this issue. The conceptually systematic knowledgebase on learning and behavior offered by behavior analysis has the potential to contribute substantially toward developing strategies for early detection of melanoma. In particular, generalization is a ubiquitous behavioral process with an extensive literature from which to draw. As such, the purpose of the current series of studies is to employ a use-inspired translational approach to explore strategies for promoting early detection by capitalizing on behavior analytic research regarding the processes of generalization and discrimination (i.e., peak shift). The purpose of the experiments was to (1) use discrimination training to establish generalization and postdiscrimination gradients with moles as stimuli, (2) determine the effects of parametric manipulations of training on postdiscrimination gradients, and (3) evaluate training with multiple discriminative stimuli. Results from Study 1 indicated that discrimination training produced gradient shifts as compared to a control group trained only with the S+. Results from Study 2 indicated that training with an S- more distinct from the S+ produced gradient shifts, but that S- stimuli more similar to the S+ did not. Results from Study 3 indicated that training with two S- stimuli from one extreme of the stimulus array produced relatively weak shifts in postdiscrimination gradients, but that training with an S- at both extremes of the array was effective in producing highly consistent response patterns. Theoretical implications and future directions toward more clinically relevant studies are discussed.

Acknowledgements

I would like thank my advisor, Dr. Derek Reed, for his encouragement and guidance while working on this project, as well as providing a highly reinforcing environment in which to work and learn. I would also like to thank Dr. Thomas Critchfield for his help in developing the materials and insightful comments. Additionally, I would like to thank the research assistants who aided in conducting this research. Finally, I would like to thank the former and current members of my committee for their contributions.

Table of Contents

Exploring Strategies for Melanoma Detection Utilizing Discrimination Training

Cutaneous melanoma accounts for approximately 75% of skin-cancer-related deaths each year and is considered the deadliest form of skin cancer (Karakousis & Czerniecki, 2011). According to the National Cancer Institute's Surveillance Epidemiology and End-Results (SEER) data, an estimated 76,690 new cases of melanoma will be diagnosed and 9,480 deaths will occur in United States during 2013. Epidemiological data from this report indicate that there has been an increasing trend in diagnosis rates over the past several decades—a pattern unlikely to decline in the foreseeable future. The state of Kansas fares worse in these statistics, exceeding the national average in both incidence and mortality rates from 2005 through 2009 (Howlader et al., 2011).

Despite being the most lethal form of skin cancer, melanoma is particularly amenable to treatment if detected early (Kasparian, McLoone, & Meiser, 2009). Consequently, early detection is of utmost concern in mitigating negative outcomes of this disease. Extensive research has been conducted in attempts to identify population characteristics that are predictive of developing melanoma. Phenotypic factors that have been consistently associated with melanoma incidence include (a) being male, (b) aged over 65 years old, (c) having light skin, numerous moles, and freckles, and (d) having personal or family history of melanoma (Watson et al., 2011). Higher socioeconomic status (SES) and education level have also been positively correlated with melanoma incidence—a peculiar finding given that SES and education level are often correlated with lower incidence of disease (Little & Eide, 2012; Singh et al., 2011). The positive correlation has been attributed to the generally higher proportion of phenotypic factors (a-c, above) among higher SES individuals and that individuals of higher SES tend to have more leisure time, which is often spent in contexts that increase UV exposure. Behavioral histories

involving frequency and severity of sunburns and use of tanning beds are also statistically related to developing melanoma (Epstein, Gilchrest, Eller, Geller, & Yaar, 1999; Little & Eide, 2012). These risk factors have helped focus investigations into causal variables for the disease, as well as inform the screening practices of medical professionals (e.g., Geller, Miller, Swetter, Demierre, & Gilchrest, 2006).

Early Detection and Skin Self-Examination (SSE)

Unfortunately, no consensus currently exists on best practices for early detection (Goodson & Grossman, 2008). The primary method for initial identification of melanoma is via visual inspection, which is then followed by skin biopsy to allow for histological examination. In 1985, Friedman, Rigel, and Kopf developed a mnemonic device to help educate practitioners and laypeople regarding visual characteristics of melanoma that are related to malignance, referred to as the ABCDs: asymmetry, border irregularity, color variegation (uneven pigmentation), and diameter greater than 6 mm. Friedman and colleagues also presented a body chart for mapping moles and a 10-step task analysis of how individuals could perform self-examinations. These authors subsequently advocated for self-examinations as useful, inexpensive means for identifying melanoma, especially when used in conjunction with medical examinations by practitioners. This recommendation is widely accepted, with self-examinations being a relatively commonplace strategy for detecting skin cancer. Miller et al. (1996) conducted a nationwide survey and found that 46% of respondents reported engaging in some form of skin self-examination (although, the thoroughness of these SSEs are unknown; cf. Weinstock et al., 2004).

Unfortunately, while self-examinations are growing in popular use and recommended by both the American Academy of Dermatology and the American Cancer Society, the utility of

self-examinations is unclear. The U.S. Preventive Services Task Force report on screening for skin cancer (2009) stated that there is currently insufficient evidence to determine the efficacy of SSE. Only one study has been conducted to date showing that skin examinations result in lower mortality rates (Berwick, Begg, Fine, Roush, & Barnhill, 1996). The dearth of support could potentially be due to a related body of literature indicating that melanomas detected by patients during self-examinations tend to be more advanced than those detected by both specialized and non-specialized practitioners (Brady et al., 2000; Epstein, Lange, Gruber, Mofid, & Koch, 1999; Lamerson, Eaton, Sax, & Kashani-Sabet, 2012; McPherson et al., 2006; Schwartz et al., 2002).

It has been suggested that the tendency for self-examinations to detect tumors at later stages of progression as compared to clinical examinations is cause for placing greater emphasis on ensuring more frequent examinations by physicians (De Giorgi et al., 2012; Geller et al., 2006). This may not be an ideal solution due to the results of several investigations that have indicated that non-specialized practitioners both report feeling undertrained in conducting skin examinations (Moore et al., 2006) and frequently fail to identify potential cases of melanoma (e.g., Cassileth et al., 1986; Robinson & McGaghie, 1996). In an experimental preparation, Brochez, Verhaeghe, Bleyen, and Naeyaert (2001) evaluated general practitioners' (n = 16) and dermatologists' (n = 60) ability to discriminate melanoma. All participants were presented 13 images of pigmented lesions (moles, birthmarks, etc.) with accompanying brief patient histories and asked to select a diagnosis for each image from among seven diagnoses, or simply indicate whether the lesion was suspicious or benign. General practitioners accurately diagnosed skin lesions only 49% of opportunities, with malignant melanomas being missed 28% of the time; whereas dermatologists accurately diagnosed 85% of skin lesions, missing only 9% of melanomas.

Despite the unfavorable results for general practitioners of Brochez et al. (2001), other research has been more encouraging. Argenziano and colleagues (2012) conducted an archival analysis spanning 10 years and found that of moles removed by non-specialized practitioners, the frequency and proportion of moles that were malignant melanoma increased over time, indicating greater accuracy in correctly identifying moles needing to be excised. The authors attributed the increase in accuracy to recently developed technologies, such as dermatoscopy and digital monitoring, which are now available when conducting skin examinations. Additionally, the study found that specialists (i.e., dermatologists) have been, and continue to be, adept in identifying melanoma. Although promising, these data only included removed moles and did not account for those missed. Continuing technological advancements will likely enhance detection during clinical examinations, but the benefits may be limited by other factors such as limited access to health care in general or to specialists in particular.

Even with the technological advances available to medical professionals, patient self-examination remains the most cost-effective primary strategy for initial detection of melanoma. Given that the majority of melanomas are detected by non-practitioners, such as the patients, their spouses, or other laypeople (Epstein, Lange, et al., 1999; Koh et al., 1992; McGuire, Secrest, Andrulonis, & Ferris, 2011; McPherson et al., 2006; Weinstock et al., 2004), it seems that strategies for enhancing self-detection remain a worthwhile endeavor. Although the initial strategies for skin self-examination were published over 25 years ago, only a handful of investigations have sought to evaluate the effects of skin examination training on melanoma detection or ways to improve self-examinations. A review by Hamidi, Peng, and Cockburn (2010) identified only seven papers between 1987 and 2007 that evaluated the accuracy of SSE. Of these studies, four focused on the patient simply counting the number of moles on their body.

These studies ranged in levels of agreement by the patients with either experimenters or physicians, ranging from fairly high agreement (79% of the sample within ±3 moles; Lawson et al., 1994) to rather poor agreement (K = 0.14 for total body count of moles; Buettner & Garbe, 2000).

Other studies have examined patients' ability to detect changes in moles during SSEs. The evolution, or change over time, of moles has been considered as an additional criterion by which to judge malignance. As such, Friedman and colleagues' mnemonic has been modified with the addition of an "E" for "evolution," making it the ABCDEs of melanoma (Abbasi et al., 2004; Thomas et al., 1998). One study examining the ability of patients to detect changes in moles was conducted by Muhn, From, and Glied (2000). The authors conducted an experiment in which participants were asked to perform skin self-examinations. Participants were recruited from a cancer clinic and only those who were at high-risk of developing melanoma, and reported both being taught how to conduct SSEs and consistently performing them over the previous year were included. The experimenters modified existing moles on the participants' backs using cosmetic eyebrow pencil (color matched to the mole) and asked participants to perform SSEs. Specifically, experimenters increased mole size (average 5 to 7 mm in diameter) by either 2 or 4 mm, or left them unchanged (traced with a noncolored HB pencil). The experimenters found that only 58% of participants correctly identified a change of 2 mm, 75% correctly identified a change of 4 mm, and 62% correctly identified that a mole did not change in size. Consequently, the authors suggested that SSEs may not be a particularly sensitive method for clients to identify changes in moles, but they may still be useful for detecting other signs of malignance.

Other studies have sought to improve identification during SSE through educational interventions targeting specific aspects of SSE. Mickler, Rodrigue, and Lescano (1999)

compared three strategies for increasing knowledge about skin cancers, identifying pictures of skin cancer, and conducting SSEs. Participants experienced either a brochure, a video, one-on-one training by a nurse practitioner, or a wait-list control. The results indicated that all three interventions improved performance in each area as compared to the control group, with knowledge improved most by the video and brochure, identification improved most by the one-to-one teaching by a nurse, and SSE improved most by the brochure. Although the average correct identification was approximately 78% across the interventions (approximately 69% in control group), the implications of these results for melanoma detection is unclear as the authors did not indicate the proportion of pigmented lesions that were melanoma or the accuracy in identifying melanoma in particular.

Targeting identification. A number of studies have sought to evaluate specific aspects of identifying melanoma and other skin cancers through visual detection. Miles and Meehan (1995) conducted a series of studies to determine individuals' ability to respond appropriately regarding the need for medical attention both with and without training. Participants were shown pictures of melanoma, dysplastic nevi, seborrheic keratosis, and benign nevi and asked to respond as to whether they would "do nothing", "keep an eye on it", "show someone else", "show doctor next visit", or "see doctor immediately." The experimenters evaluated whether length of viewing the pictures (0.5, 1.5, or 5.0 s) and the size of the pictures influenced participants' responses to the lesions and found no evidence that viewing time or size influenced performance. Results for responding with appropriate urgency indicated that participants correctly responded to melanoma ("see doctor immediately") only 36% of the time. The experimenters then evaluated the individual effects of pictorial examples and written descriptions across two viewing times (0.5 and 5.0 s). Pictorial examples consisted of a single picture of each

type of lesion accompanied by its name and the appropriate response. Written descriptions—presented to participants and read aloud by experimenters—consisted of detailed descriptions of each lesion type and the ABCD criteria for melanoma. Results indicated that there were interactions between lesion type and instruction type for correct responding in that performance was improved by pictorial examples for seborrheic keratosis (not seeking medical attention) and melanoma (seeking immediate medical attention), and by written descriptions for benign nevi (not seeking medical attention) and dysplastic nevi (show someone/doctor). Additionally, the longer viewing time only improved performance for melanoma, which occurred across both instruction types, and the highest level of correct identification was 60% (pictorial example, 5-s viewing).

Borland, Mee, and Meehan (1997) extended the work of Miles and Meehan (1995) by evaluating the effects of the number of examples during training, as well as the individual and combined effects of pictorial and written information on responses of need for treatment (i.e., immediate need, need, monitor, do nothing). The results indicated no significant difference between the number of examples shown (4, 8, or 16) and correct responses of need for treatment for early melanoma, dysplastic nevi, seborrheic keratoses, or benign nevi. The results also indicated that presenting pictorial and written information improved correct responding for melanoma, pictorial information alone improved correct responding for seborrheic keratosis, while written information alone did not improve responding in any case.

Hanrahan, Hersey, Menzies, Watson, and D'Este (1997) used computer-simulated lesions to determine participants' ability to detect changes in moles. The effects of age (younger than 30 versus older than 45) and an educational pamphlet were assessed in a 4 by 4 group design. Participants saw pictures of original moles and then immediately observed pictures in which

those moles were either changed or unchanged and asked to identify whether there was a change. This assessment was repeated one week later, with half of the participants in each age group having received an educational pamphlet prior to the second assessment. The participants were subsequently tested at delays of 29 and 60 days without viewing the originals to assess their ability to identify change over time. The results indicated no difference in ability to identify change in moles as a function of age, found that the educational pamphlet resulted in a response bias toward indicating that the moles had changed (that is, increased responding to both original and changed moles as being changed), and found poor ability to detect changes at both delays of 29 and 60 days. The authors suggested that the poor ability to detect changes over time is a barrier to patient self-identification of symptom evolution and that future efforts may be best served by investigating programs that occur under temporally proximate conditions, such as comparing pictures of non-symptomatic moles to existing moles.

Brooks, Predebon, and Van der Zwan (2001) investigated the effects of phrasing on skin cancer identification. In a series of studies, each using a pre-/posttest design, the experimenters asked participants to judge pictures of pigmented lesions according to whether or not they were (a) skin cancer, (b) dangerous, and (c) distinctive. The effects of these three phrasings were each evaluated in conjunction with an educational brochure on skin cancer (experimental group) or acne (control group). Results for all experimental groups suggested that the skin cancer brochure actually decreased correct responding to the pictures; however, the negative effects of the brochure were somewhat mitigated when participants responded based on distinctiveness. Although these data do not suggest an effective intervention, they are useful in drawing attention to the potential for phrasing to influence individuals' responses to interventions and also in

highlighting the fact that simply providing information about melanoma will not necessarily enhance correct identification.

Girardi et al. (2006) also examined the effects of several strategies on melanoma detection. Using a group design, the experimenters evaluated three methods for presenting information to participants on their ability to identify melanoma correctly. The ABCD intervention described the ABCD criteria, presented a single picture of melanoma, and indicated visually how each criterion applied to the example. The "cognitive" intervention presented nine sets of three pictures of nevi, each set with common morphological features. The first picture was of a regular benign nevus with the descriptor "good", the second (adjacent) was an irregular benign nevus with the descriptor "also good", and the third (adjacent, but spaced farther away) was a malignant melanoma with the descriptor "danger." The first and second pictures were grouped under text stating "these moles are benign" and the third was alone under text stating "these ones are cancers." The "cognitive + explanation" intervention used the same pictures in the "cognitive" intervention but grouped the benign regular nevi, the benign irregular nevi, and the melanoma separately. Each group of pictures had the same descriptors as in the "cognitive" intervention. All of the information from ABCD description was also included. Additionally, the study employed a "no intervention" control group. During the study, the participants were shown pictures of lesions and asked to indicate whether each was melanoma. Participants were tested prior to experiencing the intervention, one week after a 10-min exposure to the intervention, and after being exposed to a stressing statement ("Knowing that you mustn't miss a cancer which could kill you or one of your relatives, review thoroughly these pictures to identify which ones could be cancers") immediately prior to the test. The results indicated that the ABCD group showed an increase in sensitivity but decreases in specificity and overall accuracy

following intervention (38.2%), the "cognitive" and "cognitive + explanation" groups showed no changes in sensitivity and increases in specificity and accuracy (53.9% and 48.9%, respectively). Following the stressing statement, sensitivity increased and specificity decreased for all groups, reflecting an overall bias for indicating melanoma. Accuracy did not decrease significantly for the ABCD group (35.8%), but did for the "cognitive" (49.6%) and "cognitive + explanation" (approx. 45%) groups. These findings suggest that visual comparisons can be beneficial in identifying melanoma and that information in the ABCD intervention was not helpful and possibly counterproductive toward this end. However, this approach is still in great need for refinement, given that accuracy following the stressing statement (an approximation of realistic circumstances) remained below 50%.

A Behavioral Model of Melanoma Detection

Further exploration of methods to enhance early detection is clearly warranted. The lack of clear gains in self-detection in the existing research may be due to a failure to design procedures capitalizing on the learning and behavioral processes involved detecting melanoma. Incorporating principles and technologies from behavior analysis to this problem is a novel approach that may prove beneficial.

To date, the application of behavioral principles has been all but absent from the literature attempting to address early identification of melanoma; however, behavioral principles have already been applied with great success to self-detection of breast cancer (Pennypacker et al., 1982; Saunders, Pilgrim, & Pennypacker, 1986). For melanoma detection, the process involves a complex chain of behaviors that can be addressed through behavioral interventions. The steps of identification include (1) contacting information that the individual is at risk of having or developing melanoma, (2) detecting a melanoma visually, and (3) seeking medical

attention to confirm the need for intervention. These general steps can be further broken down into more refined behaviors (e.g., steps to seeking medical attention; de Nooijer, Lechner, & de Vries, 2001) which may vary somewhat across people and contexts. However, regardless of the particular nuances, a behavioral framework applied to the underlying processes may serve to elucidate novel ways for improving the behaviors occurring at each step. The information of Step 1 can be considered a discriminative stimulus signaling to the individual that engaging in a skin-examination may be negatively reinforced by the detection and subsequent treatment of skin cancer. Knowing that one is at risk for developing melanoma could simultaneously function as an establishing operation, evoking behavior that might eliminate the grave consequences of undiagnosed melanoma. In Step 2, the individual discriminates whether moles exhibit symptoms of malignant melanoma (e.g., the ABCDs). The presence of the stimulus features of a mole leading to identification of melanoma in Step 2 serves as a discriminative stimulus for Step 3: going to a medical professional to receive the appropriate level of intervention. Although this initial interpretation is clearly oversimplified, it points to specific behavioral processes that have received a great deal of attention in the behavioral literature, namely those involved in stimulus discrimination. The factors influencing the ability of information to serve as discriminative stimuli in Step 1 are presumably wide-ranging. It seems the process of discriminating features of moles that are potentially malignant (Step 2) is clearer in focus, and therefore more immediate gains may be made by improving performance at this step (especially given the estimated 35-46% of individual who already engage in some form of SSE; Miller et al., 1996; Weinstock et al. 1999). For Step 3, the medical literature has not recognized a major deficit in patients seeking medical attention after having identified melanoma, and so the need for intervention at this stage

does not appear warranted (Richard et al., 2000; Temoshok, Diclemente, Sweet, Blois, & Sagebiel, 1984).

Interventions aimed at enhancing discrimination of potentially malignant moles will necessarily rely on aspects of generalization, as it would be impossible to train the exact presentation of all possible variations of melanoma. Following this logic, understanding and applying the behavioral research on discrimination and generalization of stimuli to the detection of melanoma may prove to be a productive venture.

Generalization Gradients

Stimulus control is a fundamental aspect of behavior. The extent to which an organism has experienced more reinforcing events following responding in the presence of a stimulus (or set of stimulus features) relative to its absence will directly affect the likelihood of that response occurring in the presence of that stimulus in the future (given similar establishing operations for the reinforcing event). This history allows a stimulus or stimulus complex to exert control of a response through its association with differential consequences. Discriminated responding (or discrimination) refers to the relation between response probability and the level of the discriminative stimulus. Although there is some debate as to whether this is the case, it appears that discrimination and generalization are inverse properties of responding (Honig & Urcuioli, 1981), describing opposing ends of a continuum of stimulus control. To the extent that responding occurs only in the presence of a particular stimulus, the response is discriminated; to the extent that it occurs in the absence of that stimulus, the response is generalized.

Spence (1937) hypothesized that generalization could be described in terms of the spread of reinforcement effects to stimuli sharing similar features, forming a gradient peaking at the stimulus associated with reinforcement. Guttman and Kalish (1956) conducted the first study to

demonstrate empirically the existence of generalization gradients. The experimenters reinforced key pecks to a specific wavelength of light (S+), and examined rates of pecking to several additional untrained wavelengths. Four groups of pigeons were trained to respond to different initial wavelengths (530, 550, 580, and 600 nm) under intermittent schedules of reinforcement. Following training, pigeons were exposed to generalization tests in which 11 wavelengths (ranging from 470 to 640 nm) were presented in 12 series of trials under conditions of extinction. For two groups, the S+ was among the test stimuli and for the other two groups the S+ was not. Mean response rates during generalization test trials of the different wavelengths revealed gradients with peaks at or near the value of the S+ for each group and nearly symmetrical decelerating slopes to either side of the peaks (see Figure 1).

The importance of generalization as a behavioral process is immense. The spread of reinforcement effects to novel stimuli has likely been a critical factor in evolution, as organisms with a capacity to respond to similar but novel conditions were likely to be more prepared to respond effectively in their environments. As such, the experimental demonstration of generalization through the procedures designed by Guttman and Kalish (1956) was groundbreaking in that it allowed researchers to begin to explore this process. The results of the experiment served to inspire a multitude of investigations aimed at understanding the conditions influencing stimulus generalization. Several review papers have provided detailed summaries of the extensive literature that has progressed in this area (e.g., Ghirlanda & Enquist, 2003; Honig & Urcuioli, 1981). Honig and Urcuioli (1981) noted that the generalization gradient is the standard metric for assessing how different independent variables affect generalization across studies. Specifically, the authors identified area, height, slope, and form of gradients as the major components (described below) for which such comparisons can be made. Given that

generalization gradients are measures of stimulus control, these characteristics of gradients should translate to measurements of various aspects of stimulus control. Area is determined by the distribution of responses across a range of stimuli. Height (i.e., peak) is the maximum level of responding observed, indicating the greatest level of stimulus control. Slope displays the relation (i.e., change) between responding at two levels of the stimulus complex assessed and is considered the most sensitive measure of stimulus control (Honig & Urcuioli). Form is a more general description that accounts for multiple qualities such as whether the gradient is symmetric or monotonic, its location on the continuum, and variability of the gradient. Ghirlanda & Enquist (2003) found that the form of generalization gradients tend to be best described by a normal (Gaussian) curve with the peak at S+. Thus, an additional means of analyzing generalization gradients is calculation of the area under the curve at each side of the S+.

Postdiscrimination Gradients and Peak Shift

A behavioral phenomenon related to generalization gradients that has garnered a great deal of attention is the peak shift. A peak shift occurs when, after experiencing discrimination training in which responding in the presence of one stimulus (S+) is reinforced and responding in the presence of a second stimulus (S-) results in extinction (or less reinforcement), the highest levels of responding do not occur at the S+ but are displaced in the direction opposite of the S- (Honig & Urcuioli, 1981). The gradients obtained using these procedures are sometimes referred to as postdiscrimination gradients.

Peak shift was first observed in a study by Hanson (1959) that used similar procedures to those of Guttman and Kalish (1956). Five groups of pigeons were initially trained to a single S+ (550 nm), after which four of the groups experienced additional discrimination training with a specific S- stimulus (555, 560, 570, or 590 nm). During discrimination training, S+ and S-

stimuli were presented successively in semi-random order. Following training, all groups were exposed to generalization test trials comprised of 13 stimuli. The major findings of the study were that peak response rates were displaced *away* from the S+ *opposite* the S- for all four test groups, the degree of displacement was inversely related to the distance between the S+ and S- (smaller distances resulting in greater displacement), and peak response rates were notably higher in the test groups as compared to the control group. The results of this study (and the peak shift phenomenon itself) are particularly interesting because they run somewhat counter to traditional behavioral accounts of stimulus control. That is, one would assume that behavior would be most likely to occur in the presence of a stimulus that had been paired directly with reinforcement as opposed to a stimulus that had not. This paradox has led to numerous investigations to understand the process by which peak shift occurs.

Since the initial study by Hanson (1959), an assortment of variables has been shown to influence the conditions under which peak shift will occur and the characteristics of obtained postdiscrimination gradients. Three general categories of factors include the features of the stimulus dimension(s) being evaluated, the methods used during discrimination training, and the procedures used to test generalization.

Stimulus dimension. Stimuli assessed in generalization paradigms are typically those that occur on a continuous dimension such as light wavelength (e.g., Hanson, 1959), auditory frequency (e.g., Galizio & Baron, 1979), object weight (MacKinnon, 1972), degree of line-tilt (e.g., Bloomfield, 1967), rate of an event such as light flickers (e.g., Sloane, 1964) or movement (Dickinson & Hedges, 1986). Stimulus intensity, such as light brightness (Newlin, Rodgers, & Thomas, 1979) or auditory intensity (Pierrel & Sherman, 1960), has also been examined. Organisms' sensory capabilities for detecting different values in these various dimensions dictate

the psychophysical properties of the stimuli being evaluated (Blough, 1965). Sensitivity to changes in stimuli in terms of just-noticeable difference (JND) can influence an organism's response to stimuli along a continuum and, therefore, influence the generalization gradient for a set of stimuli. Specifically, when the difference between two points on a continuum is less discriminable due to deficits in sensory receptors, the slope will be less steep than between two points for which the difference is more discriminable due to higher concentrations of sensory receptors, thereby changing the shape of the gradient.

Ghirlanda and Enquist (2003) proposed that the effects of a particular stimulus dimension on an organism can be separated into two general categories: rearrangement and intensity. They defined "rearrangement" effects as when the changes along a stimulus dimension result in an approximately equal number of the same type of sensory receptors being activated by the stimuli. For example, stimuli on a line-tilt continuum activate the same number of visual receptors in the eye but the specific receptors change with each stimulus. The authors presented evidence that generalization gradients along rearrangement dimensions tend to be symmetrical and can be reliably biased (i.e., display peak shift) through discrimination training. Ghirlanda and Enquist defined "intensity" effects as when the changes along the stimulus dimension result in changes in the amount of receptor activity. Increases in receptor activity could be in the form of changes in rate of activity of a specific receptor or an increase in the total number of receptors. For example, a loud tone would result in a higher level of activity than a softer tone in receptors capable of being stimulated by that sound frequency. Research has found that when the stimulus dimension of interest is the level of intensity, there tends to be a bias toward responding to more intense stimuli on the continuum (Thomas, Mood, Morrison, & Wiertelak, 1991). Such findings have led researchers to debate the role of adaptation (or habituation) on responding during

generalization tests (Newlin et al., 1979; Thomas, Strub, & Dickson, 1974; White & Thomas, 1979). Monotonic gradients, in which responding continues to accelerate (or decelerate) from one end of the continuum to the other, are commonly seen with intensity dimensions. Additionally, Ghirlanda and Enquist noted that some stimulus dimensions exhibit effects of both rearrangement and intensity. Complex sound and complex light dimensions were identified as being subject to strong effects characteristic of both. Of particular note, object size also was indicated as a dimension that is subject to large effects from both categories; however, results of these gradients generally show a peak. Similarly, object shape was indicated as being subject to both; however, relatively stronger effects were noted from rearrangement as compared to intensity.

Beyond psychophysical aspects of single-dimension stimuli, researchers have investigated generalization gradients with stimuli comprised of multiple dimensions. Such investigations have examined a number of stimulus complexes, including two dimensions in the same sensory modality (e.g., light wavelength and intensity; Furrow & LoLordo, 1975) and two or more unrelated dimensions (e.g., tones presented simultaneously with light wavelengths; Blough, 1969). Stimulus complexes of two separate dimensions (i.e., compound stimuli) have often been used to investigate the role of both inhibition and attention on generalization. Findings from this type of research have shown that some stimulus dimensions are more apt to gain control of responding than others. For example, color has been shown to exert stimulus control more readily than line tilt (Farthing, 1972). Additionally, Miles and Jenkins (1973) demonstrated with compound auditory tone and light intensity stimuli (tone present vs. absent, 5 light intensities) that the more salient feature will come to control responding. When the stimulus dimension being altered is multifaceted and these dimensions vary in concert,

generalization and peak shift will typically show similar effects to that of a simple single-dimension continuum (Spetch, Cheng, & Clifford, 2004).

Discrimination training. The second category of variables influencing peak shift involves the methods used to train the organism regarding the conditions in which responding will produce reinforcement. In the initial study on peak shift, Hanson (1959) evaluated several S- values differing in distance from the S+. The results indicated a relation between the degree of separation of the S+ and S-, with greater shifts away from S+ occurring with smaller disparities between the S+ and S- values. This effect has been replicated across a number of different studies (for a review see Ghirlanda & Enquist, 2003). Additionally, Hanson found that discrimination training requires more time as the difference between S+ and S- approaches the JND threshold.

A key factor in the method of discrimination training is the alternation of the stimuli. Akins, Gouvier, and Lyons (1981) found that the alternation of S+ and S- is an essential part of discrimination training, in that peak shift will not occur if organisms are trained to respond in the presence of the S+ at one time and trained not to respond to S- at a separate time. Additionally, procedures that utilize errorless training in which the S- is slowly faded into the preparation such that the organism does not contact extinction for responding to the S-, as well as those using massed extinction to the S-, have also failed to produce peak shift (e.g., Honig, Thomas, & Guttman, 1959; Terrace, 1964). Studies have demonstrated that in certain discrimination preparations the S- can function as an aversive stimulus (Rilling, Askew, Ahlskog, & Kramer, 1969). It has been hypothesized that it is the aversive aspects of the S- that lead to behavioral contrast, which may be an underlying process for producing peak shift (Grusec, 1968; Terrace, 1968). Behavioral contrast is a phenomenon in which the response rate to an S+ increases with

the introduction of an S- to which the response rate decreases (Reynolds, 1961). However, research has indicated that behavioral contrast is neither necessary nor sufficient to produce peak shift (Ellis, 1970), but that behavioral contrast is an instance of inhibitory stimulus control, and that any procedure which establishes inhibitory control by the S- will result in peak shift (Honig & Urcuioli, 1981).

Procedures in which stimuli are alternated during training have effectively produced shifted postdiscrimination gradients and can be considered a sequential or successive training preparation in that each training stimulus is presented in isolation. Another method that involves simultaneous presentation of the stimuli. In this procedure, the organism is presented two response options during training and only responses to the S+ are reinforced. This preparation has had limited success in producing peak shift in nonhuman animals (Honig, 1962), and it is has been suggested that the features of the S- do not gain inhibitory control due to overshadowing (Honig & Urcuioli, 1981). However, empirical work on transposition typically employs this procedure (Lazareva, 2012). Transposition refers to responding in the context of stimulus pairs in which after training the organism responds to the relative aspects of the training stimuli as opposed to the absolute features. For example, after being trained that the larger of a pair of circles produces reinforcement, the organism will select the larger circle even if the smaller circle was the original S+. Current procedural differences in assessing transposition and peak shift preclude determining whether they are both products of the same behavioral process.

Another factor includes the amount of reinforcement that responding produces during training. Guttman (1959) demonstrated that having a richer schedule of reinforcement (VI 1-min) associated with the S+ and a leaner schedule (VI 5-min) associated with the S- also produced peak shift with pigeons' responding to spectral stimuli. This effect has been replicated

across several studies (Terrace, 1966b, 1968; Wheatley & Thomas, 1974) and is attributed to behavioral contrast.

The amount of discrimination training also influences peak shift in some cases; however, the relation is complicated. Terrace (1966a) found that prolonged training reduced peak shift. Conversely, Hearst (1971) and Dukhayyil and Lyons (1973) did not find extended training to reduce peak shift. The differences in results have been attributed to procedural differences in that Terrace used a preparation in which the S- prevented access to the S+ (as opposed to fixed durations of both S+ and S- in the later studies), which presumably resulted in an initial aversive quality of the S- that dissipated over the course of training.

Although peak shift has most often been investigated with a single S+ and S-, studies have frequently examined the effects of multiple S+ or S- stimuli (e.g., Hanson, 1961). Galizio (1985) investigated the effects of various arrangements of multiple S+ and S- stimuli on postdiscrimination gradients of auditory stimuli with humans, as well as the effects of amount of training, across a series of studies. In this study, the researcher trained participants to respond during trials in which the S+ was present and not to respond during trials in which it was absent. During training, responding during S+ trials produced the experimenter feedback "correct" and responding during S- trials produced the feedback "incorrect." In the first study, the preparation entailed exposing eight groups of 12 participants to various trainings. Four groups experienced a single S- (located one or two steps above or below the S+) and two groups experienced two S-'s (both located either above or below the S+). Training of the control group entailed the S+ only and another trained with two S-'s (one above and the other below the S+). Following training, all participants encountered the same generalization test in which six blocks of the seven test stimuli were presented without feedback. The results indicated that training with two S-'s

required more time than with a single S-, single S- training only produced area shifts, and stronger shifts (true peak shift) occurred when training entailed two S- stimuli on one side of the S+. In order to rule out that stronger shifts with the two S- groups were due to more training, the experimenters held training trials constant at 12 or 42 trials in the second study. Twelve groups of 12 participants experienced one of the six experimental training procedures from the first study with one of the two fixed trial amounts. Results indicated that amount of training did not influence the amount of peak shift. In the final study, the experimenters examined the effects of training with two S+ stimuli to evaluate whether adaptation level or the Spence model accounted for the stronger shifts produced by two S- training. Groups of six participants were trained either with an S+ one step above or below the middle array value and no S-, with a single S+ one step above or below and S- at the middle array value, or S+ both one step above and below the middle S-. Generalization tests included an array of nine stimuli. Results indicated the three-stimulus training again required more time, a central tendency for one of the S+ groups occurred, area shift only occurred following single S- training, and greater peak shift was again observed with the three-stimulus training.

Generalization test. The third category of variables influencing generalization and peak shift are specific to the assessment of the effect through generalization tests. Generalization tests can vary in numerous ways including the number of stimuli in the test array, the distribution of the test values in the array along the stimulus dimension, and the length of testing, among others. In Hanson's (1959) study, the generalization test included 13 test stimuli presented in randomized sequences across 10 blocks (130 30-s stimulus presentations). Generalization testing took place under conditions of extinction.

Some of the earliest experiments on peak shift investigated the effect of various distributions of test stimuli in generalization tests on postdiscrimination gradients. Research with nonhuman animals has established that the number and spacing of the test stimuli has minimal effect on the shape of the gradient (Thomas, 1993). For example, Friedman (1963) examined pigeons' wavelength generalization using 11 test stimuli spaced in units of 5, 10, or 20 nm and found that the gradients were essentially identical.

Generalization tests are frequently conducted under conditions of extinction; however, steady-state procedures, in which the organism continues to receive reinforcement during the generalization test, have also been used (Blough, 1975). Under extinction, decrements in responding to the test stimuli usually occur over the course of the generalization test. As such, the length of the generalization test can influence whether the peak shift is captured.

Generalization and Peak Shift with Humans

The vast majority of research on generalization and peak shift employed nonhuman animals; however, a number of studies have sought to determine the extent to which these findings are directly applicable to human behavior.

In assessing generalization with nonhuman animals, experimental preparations frequently involve extended training in the presence of S+ and S- stimuli to achieve steady-state levels of a free operant response. Rates of the free operant response observed at the test stimuli during generalization tests produce the shape of the gradient. However, given humans' ability to rapidly learn the types of discriminations typically investigated in these studies, the experimental preparation is often modified. Rather than using a free-operant preparation, experimenters often employ a discrete trial format, either in the form of a "go/no-go" response or a two-choice response, such as "yes/no" or "left/right" (Thomas, Lusky, & Morrison, 1992).

Although much of the research can be extended to humans, several findings from the nonhuman literature have failed to be replicated with human subjects or have been reproduced inconsistently. Many of the effects have indicated complex interactions between variables in the experimental preparation and the target stimulus dimension. Contrary to what has been observed with nonhumans, there appears to be a dynamic interaction between the S+/S- separation on the resulting postdiscrimination gradient of humans (Thomas et al., 1991; Thomas, Svinicki, & Vogt, 1973). For instance, Doll and Thomas (1967) observed greater peak shift in postdiscrimination gradients with humans when the S- was 20 nm from the S+ as compared with an S- that was only 10 nm away. Another discrepancy repeatedly demonstrated is that human generalization can be influenced by both the range of stimuli included in the generalization test and the distribution of stimuli within the test array (Thomas, 1993). The range influences responding through a phenomenon referred to as the central tendency effect, in which responding tends to regress toward the central value of the array over the course of the generalization test (Thomas & Jones, 1962; Thomas et al., 1973). Additionally, responding can be biased through overrepresentation of a single test value or group of test values in the array (Thomas et al., 1992).

Despite findings in the nonhuman literature are compatible with the excitation-inhibition gradient-interaction model derived from Spence's (1937) theory of discrimination, the aforementioned discrepancies present a particular challenge. To account for these, an alternative theory known as the adaptation-level (AL) model has been suggested (Thomas, 1993). This theory posits that as the subject encounters stimuli during training, the average stimulus value experienced by the subject (the adaptation level) functions as a reference against which each upcoming stimulus value is compared. During the course of the generalization test, the subject constantly adjusts their AL as they experience the test stimuli. Assuming each stimulus value in

the array is represented equally, the average stimulus value is also the middle value, and therefore the AL serves to increase responding to central values in the array as the test progresses. According to AL theory, the criterion developed during discrimination training is AL + X (or AL – X), where AL is the average of the S+ and S- stimulus values and X is the distance from the AL to the S+. Although AL model seems to provide an adequate explanation for some response patterns not predicted by the gradient-interaction model (e.g., effects of S+/S- distance), it does not sufficiently account for other factors (e.g., effects of training with multiple S+ or S- stimuli). AL and the central tendency effect appear to be more pronounced when the stimulus dimension is on a simple intensity continuum such as brightness (Verbeek, Spetch, Cheng, & Clifford, 2006).

Verbal behavior and categorization. A clear difference between humans and nonhumans is the capacity for complex verbal behavior. The tendency to identify common features of stimuli and subsequently categorize and label these abstractions (e.g., naming colors) is a behavioral repertoire ubiquitous among typically developing humans. Although many of the underlying behavioral processes of generalization are the same across humans and other species (Ghirlanda & Enquist, 2003), verbal behavior can substantially alter the ways in which generalization occur. In particular, categorizing and labeling can limit or alter the extent to which generalization or peak shift occur (e.g., Kalish, 1958; McLaren, Bennett, Guttman-Nahir, Kim, & Mackintosh, 1995). In an early experiment on peak shift with humans, Landau (1968) examined generalization on a line-tilt continuum with both children and adults. During training, the S+ was either vertical (90º) or tilted (120º) with various S- conditions across groups (no line, 60º, 90º, 120º, 150º). Although the procedures were limited in several respects, the findings were still noteworthy: training failed to produce peak shift following any of the training

arrangements. Additionally, clear dissimilarities in response patterns emerged between adults and children when the S+ was a 120º line and the S- was a 60º line in that the children did not respond to any of the stimuli, whereas the adults responded to the S+ and not the S-. Landau proposed that verbal behavior and categorization were responsible for this effect, as well as the lack of peak shift in general. It should be noted that although verbal behavior facilitates categorization, it is not essential. Nonhuman animals (e.g., monkeys, pigeons) can also engage in stimulus categorization and similar effects on generalization are seen (Wakita, 2004), but humans are substantially more likely to do so. Additionally, some stimuli are more readily categorized than others. It appears that rearrangement-based stimulus dimensions (e.g., line-tilt, color) are more easily labeled than intensity-based stimulus dimensions (e.g., brightness), and such simple dimensions are more easily labeled than complex dimensions (e.g., morphed faces). However, experience with the stimuli also influences generalization, typically with greater amounts of experience decreasing generalization (Ghirlanda & Enquist, 2003).

Notwithstanding the research described above, categorizing does not always occur and research has found that many of the discrepancies between human and nonhuman generalization disappear with stimuli that are less amenable to categorization. Through using stimuli that were presumably difficult to categorize, Baron (1973) found no evidence of categorization or AL effects on generalization of pure tones with humans. More recently, Verbeek and colleagues (2006) found that complex stimuli (morphed faces) were less susceptible to these effects as well. Additionally, Livesey and McLaren (2009) found that strategies specifically designed to make categorization more difficult were effective in mitigating categorization, albeit temporarily, with color generalization. The results showed an initial peak shift in responding to hues on a continuum from yellow-green to green-blue, but responding became monotonic over the course

of testing. Livesey and McLaren attributed this response pattern to participants' developing a verbal rule or heuristic over the course of the generalization test and concluded that when the features of the generalization test or the stimuli themselves preclude generating a response strategy, human generalization is governed by the same processes that have been clearly observed in nonhumans.

Generalization, Peak Shift, and Melanoma

To date, the majority of experiments have examined generalization with stimuli selected for their suitability for laboratory study rather than their relevance to society, per se. However, researchers have begun studying biological stimuli that specific to humans. Some of the first investigations of human peak shift with biologically-relevant stimuli examined the phenomenon using facial features (Lewis & Johnston, 1999; Spetch et al., 2004). Derenne (2010) examined peak shift of bilateral face symmetry. In another study, Derenne, Breitstein, and Cicha (2008) examined peak shift of female waist-to-hip ratio. The results of these studies confirmed that peak shift was applicable to biological stimuli, and may therefore have played a part in influencing evolution though sexual selection. Stimulus generalization is a ubiquitous aspect of behavior (Ghirlanda & Enquist, 2003). Taken together, the aforementioned studies lay the foundation for further extensions to clinically relevant stimuli such as melanoma.

One study that has attempted to apply concepts of generalization to the issue of melanoma detection was conducted by Dalianis, Critchfield, Howard, Jordan, and Derenne (2011). The main goal of the study was to evaluate participants' ability to detect changes (i.e., "evolution") in two of the symptom criteria for melanoma: asymmetry and border irregularity. Prior to evaluating the evolution of malignance, the authors first created two sets of stimuli using morphing software. For each symptom type, an image of an asymptomatic mole and a clearly

symptomatic mole were used as the two ends of the continuum along which 98 images were generated to reflect the progression. To provide support that the generated stimuli constituted and orderly progression, Dalianis et al. employed procedures similar to those of Guttman and Kalish (1956) to assess generalization gradients for each set of stimuli. The results showed that generalization gradients for both symptom types peaked at the S+ (Stimulus 50) and were approximately symmetrical to either side. The subsequent experiments in the study determined that changes in stimuli across the progression were equally detectable (identifying change after approximately 7 steps) and that identifying change occurred earlier in the stimulus array when viewing stimuli over extended periods (28 days) as opposed to a single, brief session (approximately 2 to 7 min).

This study was an important first step in the translation of existing behavioral research toward the applied goal of promoting early melanoma detection. Unfortunately, the results of the experiments were consistent with previous research examining symptom evolution in that individuals are not particularly adept in identifying change over time. In considering these and previous results it seems that, rather than working against people's natural deficit in evaluating stimuli over time, procedures would most likely benefit from capitalizing on temporally proximate strategies. However, the finding that stimulus evolution can be conceptualized in terms of stimulus generalization is in line with a behavioral model of melanoma detection. The importance of understanding generalization and peak shift becomes apparent when considering that the form of these universal patterns of responding can be reliably altered in predictable ways through discrimination training. That is, knowing the variables that influence peak shift can facilitate the design of interventions that capitalize on generalization, and specifically generalization of responding toward or away from certain stimuli.

Following this logic, further understanding generalization and beginning to explore peak shift as they relate to melanoma detection appear worthwhile. Such investigation could take a number of forms as it develops. An application-focused research program aimed at producing an immediately consumable technology for melanoma detection, while goal-directed toward socially important outcomes, would require a great deal of collaboration with professionals across disciplines (e.g., behavior analysts, dermatologists, public health professionals), may take a long time to produce an effective intervention, and would expose individuals with potentially life-threatening conditions to unknown consequences of the proposed treatments. An alternative approach involves starting with an initial demonstration of the conceptual principles involved in the area of inquiry and progressively integrating considerations of application as the research line develops. One benefit of the second approach is that the information gained regarding the conceptual analysis could be applicable to related interventions in the area. For example, knowledge acquired about how people respond to changes in moles could be applied to both how laypeople are trained to look for moles, as well as how medical professionals are trained, although the actual procedures to accomplish these goals may differ. Another benefit is that participants are not exposed to potentially unwanted or adverse effects of ineffective or countertherapeutic interventions. Conversely, this approach is limited in that it is less applicable to the problem at hand. At this time, using a basic-to-applied model appears to have potential for advancing melanoma detection from a behavioral perspective without exposing participants to any risk.

Purpose

It is the aim of the currently proposed experiments to incorporate the extant literature on generalization and peak shift to explore procedures for enhancing early detection of melanoma.

Through applying a use-inspired approach to basic research (Stokes, 1997), the gains of an entire field of conceptually systematic research may be brought to bear on this important problem. The proposed experiments seek to (1) use discrimination training to establish generalization and postdiscrimination gradients with moles as stimuli, (2) determine the effects of parametric manipulations of training on postdiscrimination gradients, and (3) evaluate training with multiple discriminative stimuli.

General Method

Participants

Undergraduates enrolled in an introductory course on behavior analysis and/or child development served as participants. For participating, students earned 1% of extra credit added to their final grade in the course from which they were recruited.

Settings, Apparatus, and Stimuli

Data were collected in a computer lab (9 m by 6 m) containing approximately 20 Dell Optiplex 360 computers. Dell P1911 17" wide-aspect flat-panel monitors were equipped with screen protectors to minimize the ability of adjacent participants from viewing others' screens. Coby CV185 headphones were plugged into the computers. A computer mouse was placed within reach and the keyboard was absent. Datasheets (see Appendix A) were placed on the table directly in front of the monitor along with a ballpoint pen to record responses.

Stimuli were created using the same morphing software (Morpheus Photo Morpher®) as reported in Dalianis et al. (2011). Two images were loaded into the program and the software used an algorithm to create a series of images that combined features of each original image progressing in equal intervals from one to the next. The first image was an abstraction of a completely benign mole and the second image was an abstraction of a completely malignant

mole, resembling superficial spreading melanoma. The program generated 98 iterations of images, representing the progression from asymptomatic to symptomatic. Test stimuli consisted of the two original images and 13 images at interval steps of 7, for a total of 15 stimuli (see Appendix B). Each stimulus was labeled in terms of its position on the progression ranging from 1 (benign; S1) to 99 (malignant; S99). A noteworthy limitation of these stimuli is that the image morphs and stimulus progressions are linear. That is, the difference between S1 and S8 are assumed to be equivalent to the differences between S64 and S71. The psychophysical differences between stimuli are mostly likely nonlinear. Unfortunately, quantitative technologies to describe the psychophysical scaling of stimulus differences for the sake of the current studies could not be identified.

Procedure

Participants experienced the experimental preparation individually; however, multiple participants (up to 12) completed the study in the computer lab at the same time. An experimenter individually instructed each participant to read the text instructions on the screen, to the wear headphones during the duration of the study, and to record responses on the sheet in front of them. Participants began the study by clicking a button labeled "Start" once they had finished reading the instructions.

Training. Training entailed a successive discrimination arrangement. This preparation is ideal because it likely approximates how individuals judge their own moles; that is, by looking at one mole at a time and determining whether it is or is not of concern. In natural contexts, individuals are not able to look simultaneously at the same mole under different levels of symptomatic progression to determine whether signs of melanoma are present. Furthermore, the

design of the preparation aimed to minimize the length of training in order to accommodate future use in applied settings.

Participants experienced the stimuli via a Microsoft's PowerPoint® program (see Appendix 3). The size of the stimuli ranged from approximately 4 mm by 5 mm (S1) to 8mm by 7 mm (S99) and were presented in the center of the screen on a flesh-colored background (RGB code = 213 R, 172 G, 129 B). The S+ appeared initially for 10 s with text identifying it as the "original" stimulus. Training immediately followed, in which a single stimulus appeared in each trial. After an observation period, the program prompted the participant to record their response (mark in one of two boxes labeled "Y" or "N") and indicate their response by clicking on the corresponding "Yes" or "No" button. Following correct responses to the S+, feedback was presented in that the word "Correct!" appeared in green text on the screen for 2 s and a chime sound played into the headphones. Following incorrect responses to the S+, the word "Incorrect." appeared in red text for 2 s and no sound played. For training conditions including an S-, responding produced the same feedback as with the S+ except that the chime did not play following correct responses to the S-. Experimenters observed participants to ensure they recorded their response on the datasheet before clicking to receive feedback. If a participant appeared to record following a response, the experimenter provided corrective instructions. Each training trial was followed by a black screen on which a text prompt to continue by clicking on the text appeared after 5 s. The number of training trials differed across studies. In order to be included in the study, participants were required to demonstrate a minimum performance in the last half of training trials, which varied slightly due to differing numbers of trials across studies. The last half of training trials contained equal proportions of the stimulus values included in the

training. Following training, participants were informed that they would no longer be using the computer mouse and to place it to the side. This was followed by a 60-s break.

Generalization test. During the generalization test, no feedback appeared and the program did not allow the participant to click the response corresponding with the answer they recorded. All 15 test stimuli appeared individually in trials lasting a specified number of seconds, which varied across studies, across 9 trial blocks. After each trial, a black screen appeared for 5 s before automatically progressing to the next trial. Within each block, all 15 stimuli were arranged in a randomized order and, therefore, differed across blocks; however, this order was fixed for all participants. Two additional 60-s breaks occurred during the generalization test, one after the third block and the second after the sixth block.

Dependent Variables

Given the exploratory nature of this investigation, multiple dependent measures were analyzed to evaluate the effects of the independent variables. Common metrics for evaluating stimulus control identified by Honig & Urcuioli (1981) include area, height (peak), slope, and form of gradients. Additionally, the mean response within gradients has been another common metric (e.g., Galizio & Baron, 1979). As gradient mean provides the most global measure of the generalization gradient, this was used as the primary dependent measure for assessing stimulus control on the dimension of mole malignance. The means of individual gradients were calculated by multiplying the number of "yes" responses to each stimulus by the stimulus value, summing the products, and dividing by the total number of responses. In addition to providing a summary metric of stimulus control, this calculation transforms the binary measure of the "yes/no" response into a continuous measure amenable to ANOVA comparisons, which were used to compare effects across groups.

Additionally, gradient peak and area are highly amenable to quantitative analysis for gauging overall stimulus control and, therefore, were included as supplementary dependent measures. The researcher calculated peaks of individual gradients by identifying the stimulus value at which the greatest number of "yes" responses was observed. If the greatest number of responses was observed at more than one stimulus value, the values were averaged to obtain the peak. To obtain a measure of the relative distribution of "yes" responses to stimuli displaying less malignance, the researcher analyzed the area of each gradient in terms of its distribution across stimulus values equal to or lesser than the S+ (S1 through S50) as compared to those equal to or greater than the S+ (S50 through S99). This analysis evaluated the overall amount of responding to stimuli earlier in the progression of malignance as compared to later in the progression with respect to the S+. The area of an individual gradient for S1-S50 was calculated by entering the frequency of "yes" responses for each of the eight stimuli S1 through S50 into GraphPad Prism® software and conducting an "Area under curve" analysis. The area was calculated separately for S50-S99 using the same steps.

Slope provides a measure of change in stimulus control between two points on a gradient and indicates differences in discriminability from the S+. The researcher calculated slopes between each adjacent pair of stimulus values for each group. Slopes were analyzed across all adjacent points on the gradients to assess the effects of the independent variables on discriminability of stimuli within the array.

In addition to traditional means of analyzing generalization gradients, signal detection theory (SDT; Green & Swets, 1966) provides an alternative approach to interpreting the discriminability of stimuli. In SDT, events are categorized according to two dichotomous dimensions: the presence or absence of a target stimulus and the presence or absence of a target

response. A responses in the presence of a stimulus is called a "hit," a response in the absence of a stimulus is a "false alarm," the absence of a response in the presence of the stimulus is a "miss," and the absence of a response in the absence of the stimulus is a "correct rejection." The researcher used this approach to determine if additional information could be garnered. Four indices of discriminability were calculated based on the frequencies of hits, false alarms, misses, and correct rejections. These include accuracy (hits + correct rejections / total opportunities), sensitivity (hits / hits + misses), specificity (correct rejections / correct rejections + misses), and precision (hits / hits + false alarms).

As the experimental design used was a group design, the researcher examined the effects of the independent variables on the dependent variables using inferential statistics.

Study 1: Establishing Peak Shift

As this is the first known attempt to produce peak shift with moles varying along the stimulus dimension of degree of symptomology, there were two purposes of Study 1. First, similar to Dalianis et al. (2011), this study sought to determine whether generalization across the array of stimuli would approximate curves obtained with other stimuli (i.e., peak responding to the S+, symmetrical deceleration in responding to stimuli above and below the S+ on the continuum). Second, this study sought to determine the effects of discrimination training on postdiscrimination gradients with these stimuli and whether training with S- stimuli above or below the S+ would produce differential effects on the obtained gradients.

Participants

Participants included 24 undergraduate students (19 female, 5 male) between the ages of 18 and 23 year (M = 19.8, SD = 1.1).

Procedures

The researcher quasi-randomly assigned participants to one of three groups (n = 8 participants, each) that differed only in regards to the training condition. Each training condition consisted of 20 trials and generalization tests (identical across groups) consisted of 135 trials. Stimulus 50 (S50), the stimulus halfway in the progression from benign to malignant, served as the S+ for all groups. In the control group, participants experienced training in which all 20 training trials entailed the S+. In the two experimental groups, training consisted of both the S+ and an S- presented in semi-random order (no more than 3 consecutive trials of one stimulus type) in a sequential arrangement. For one experimental group the S- was S1; for the other, the S- was S99. Minimum performance criteria for being included in the study included 1) correct responding on at least 9 of the last 10 training trials, 2) at least one affirmative response during generalization trials, and 3) a response recorded for each trial throughout the experiment.

During training and test trials, the stimulus appeared on the screen throughout the trial. After 3 s a text prompt to record a response appeared and a 1-s tone played in an attempt to promote participants attending to the trial. During training, the buttons corresponding to recorded responses appeared after an additional 3 s. During the generalization test, trials automatically advanced after 11 s.

Results

The frequency of affirmative responses—that is, the participant recorded "Yes" in response to the question "Is this the same as the original?"—at each stimulus value during the generalization test was averaged across all participants in the group to produce a gradient. The average gradient for all three groups are depicted in Figure 2. The average responding to stimuli during the generalization test for the S+ 50 group show a generalization gradient that peaks at the

S+ and decelerates toward each extreme of the stimulus array. These data indicate a bias for responding to stimuli to the right of S50 (i.e., presenting with greater malignance). Average postdiscrimination gradients for both the S- 1 and S- 99 groups indicate that responding was shifted away from the S+, each in the direction opposite the S- for that group.

Individual data for each participant are depicted in Figures 3, 4, and 5 for the control, S- 1, and S- 99 groups, respectively. The gradients generally displayed the same effect demonstrated by the average for the group, although some variability occurred.

Figure 6 depicts the average gradient means per trial blocks during the generalization test for each group. Average gradient means were 54.25, 64.10, and 46.20 for the control, S- 1, and S- 99 groups, respectively. A two-way ANOVA of the three groups indicated a main effect of group level ($F_{2,187} = 69.35$, $p < 0.01$), but not trial block ($F_{8,187} = 0.6245$, $p = 0.76$). The trial block by group interaction was not statistically significant ($F_{16,187} = 0.1528$, $p > 0.99$). Exploring the main effect of group more fully, a Fisher's LSD posttest revealed that all group means significantly differed at $p < 0.01$, such that each training produced statistically different effects on the participants' responding during the generalization test.

The average stimulus values of peaks were 53.06, 63.13, and 45.63 for the control, S- 1, and S- 99 groups, respectively. The researcher analyzed differences in peak responding between groups using one-tailed nonparametric t-tests (Mann-Whitney *U*-test). Analysis entailed a one-tailed test because the difference in peaks was predicted to occur in specific directions (peaks at higher stimulus values for the S- 1 group and at lower stimulus values for the S- 99 group as compared to the control group). Stimulus values at which peak responding occurred differed significantly when comparing the S- 1 (*Mdn* = 58.75 [57.00, 69.25]) and control (*Mdn* = 51.75

[50.00, 57.00]) groups (U = 14.50, p = 0.03), the S- 99 (*Mdn* = 44.75 [40.38, 49.13]) and control groups (U = 11.50, p = 0.01), and the S- 1 and S- 99 groups (U = 4.50, $p < 0.01$).

The average areas for each group are depicted in Table 1. Separate one-tailed Mann-Whitney *U* tests compared the area for each group with respect to the less malignant and more malignant halves of the stimulus values (see Table 1). For comparisons of area to the left of the S+ (S1-S50), *U* tests yielded significant differences between the control (*Mdn* = 70.00 [42.88, 100.6]) and S- 1 (*Mdn* = 14.00 [7.875, 31.50]) groups (U = 9.00, $p < 0.01$) and the S- 1 and S- 99 (*Mdn* = 96.25 [62.13, 167.1]) groups (U = 3.50, $p < 0.01$). For the comparison of the control and S- 99 groups, the difference approached significance (U = 19.00, p = 0.09). For comparisons of area to the right of the S+ (S50-S99), *U* tests yielded significant differences between the control (*Mdn* = 147.0 [84.88, 155.8]) and S- 99 (*Mdn* = 56.00 [13.13, 109.4]) groups (U = 7.00, $p <$ 0.01) and the S- 1 (*Mdn* = 183.8 [138.3, 230.1]) and S- 99 groups (U = 5.00, $p < 0.01$). For the comparison of the control and S- 1 groups, the difference approached significance (U = 16.00, p = 0.05). These results indicate that the effect of the each of the discrimination training groups was primarily to decrease responding to the half of the stimulus array containing the S- as compared to the control group, but not necessarily to increase responding away from the S-.

The differences in slopes between adjacent pairs of stimuli were compared across groups using a two-way ANOVA (see Table 2). The test of the three groups indicated a main effect of the stimulus pair ($F_{13,294}$ = 10.90, $p < 0.01$), but not group ($F_{2,294}$ = 0.10, p = 0.91). The stimulus pair by group interaction was statistically significant ($F_{26,294}$ = 3.64, $p < 0.01$). Exploring this interaction more fully, a Fisher's LSD posttest revealed that significant differences in slopes were obtained across a variety of adjacent stimuli depending on the groups compared, indicating that the trainings differentially affected only a subset of the stimuli in the array. Significant

differences were observed between the adjacent stimulus pairs ranging from S36-43 to S64-S71. The greatest effect on discriminability occurred between S50-S57, in that all comparisons revealed statistically significant differences in slope.

Lastly, values for each SDT index (accuracy, specificity, sensitivity, and precision) were calculated for each group and compared using a Kruskal-Wallis *H* tests (see Table 3). The *H* test of accuracy indicated no significant difference between groups ($H_2 = 3.17$, $p = 0.20$). The *H* test of sensitivity indicated a significant difference between groups ($H_2 = 6.76$, $p = 0.03$). A Dunn's multiple comparisons posttest revealed that significance was a result of the difference between the control (*Mdn* = 0.89 [0.69, 1.0]) and S- 1 (*Mdn* = 0.33 [0.56, 0.78]). The *H* test of specificity also indicated a significant difference between groups ($H_2 = 7.12$, $p = 0.03$). A Dunn's multiple comparisons posttest revealed that significance was again a result of the difference between the control (*Mdn* = 0.99 [0.97, 1.0]) and S- 1 (*Mdn* = 0.94 [0.91, 0.98]). The *H* test of precision indicated that the difference between groups approached significance ($H_2 = 4.70$, $p = 0.10$). Taken together, these results indicate that the SDT indices did not capture the changes in responding observed through the behavioral measures. However, the statistical differences obtained for sensitivity and specificity between the control and S- 1 groups are interesting that the lower sensitivity of the S- 1 group indicated that these participants were significantly less sensitive to the presence of the S+ and were significantly more likely to respond in the absence of the S+.

Overall, these data indicate that the current preparation is a viable means for obtaining gradient shifts and changes in discriminability of stimuli resembling cutaneous melanoma along the complex dimension of malignance.

Monte Carlo Simulation

A limitation of the experimental preparation was that the primary dependent measure—participants' marking "yes" or "no" on the datasheet—was essentially a permanent product measure. Although this method of data collection is common in research published in medical journals on melanoma detection (e.g., Robinson & Turrisi, 2006), behavioral research often employs either automated data collection (often the case in human operant research) or data collection by an independent second observer to increase the confidence that the measures reflect the actual occurrence of the behavior under the conditions of the experiment. The researcher conducted two Monte Carlo simulations in order to increase confidence that the responses were not random, and therefore, more likely to be a result of the experimental preparation.

To determine the probability of obtaining results similar to those observed in the generalization test of Study 1 if responding was random, 10,000 datasets were generated in which the frequency of responding to each of the 15 stimuli in the array was randomly assigned a frequency from 0 to 9 (thereby reflecting the constraints of the generalization test). The simulated data were then compared to criteria based on the range of observed responding at each stimulus value for the control group participants in Study 1. For example, no participants in the control group responded to S8 during the generalization test; therefore, the criterion range of responding for this stimulus was 0. As another example, the range of responding to S50 was between 3 and 9 (inclusive) across the control group; therefore, the criterion range of S50 for the simulated data was a frequency between 3 and 9 (inclusive). In these two cases, one or more responses to S8 would not meet the criterion for that stimulus, nor would less than three responses meet the criterion for S50. The number of criteria met for each of the 10,000 simulated datasets was then determined. The results of this simulation are depicted in Table 4.

The results indicate that a maximum of eight criteria were met when responding was completely random (probability of 0.0001) and, therefore, increases confidence that the participants were not responding at random during the generalization test.

A second simulation was conducted to further increase confidence that the responses occurred under the assumed experimental conditions. In this simulation, 10,766 datasets were generated using a similar method to the previous simulation; however, datasets were constrained to the range of the total number of "yes" responses observed in the control group of Study 1. That is, the minimum number of responses recorded across all trials in the generalization test by a participant in the control group was 23 and the maximum number of recorded responses was 36—a range of 14. A total of 769 datasets were generated for each of the 14 total response values in the observed range in order to produce approximately 10,000 datasets. In each dataset, the total constrained responses were distributed randomly among the 15 stimuli of the array. The number of criteria met for each of the 10,766 simulated datasets was then determined. The results of this simulation are depicted in Table 4. The results indicate that a maximum of 12 criteria were met when the frequency of responding was constrained to that observed for the control group of Study 1 completely random (probability of 0.00009) and, therefore, further increases confidence that the participants were not responding at random during the generalization test.

Study 2: Parametric Evaluation of S+/S- Distance

Previous literature indicates that the distance between S+ and S- stimuli used in discrimination training can influence the degree of peak shift obtained in postdiscrimination gradients. Research with nonhumans consistently demonstrates that smaller distances produce greater shifts, but results are mixed among these types of studies with humans in that larger

distances sometimes produce larger gradient shifts (Derenne, 2006). This study sought to determine the effects of S+/S- distance on postdiscrimination gradients in the current preparation, as such information could eventually prove useful in designing interventions for melanoma self-detection. This parametric evaluation assessed several S- stimuli regarding their effects on generalization; however, all S- stimuli displayed greater malignance than the S+ (i.e., greater than S50) in an attempt to shift the postdiscrimination gradients in potentially beneficial direction for early detection.

Participants

An a priori power analysis using G*Power 3.1.5 (see Faul, Erdfelder, Lang, & Buchner, 2007) suggested a minimum total sample size of 16 to detect an effect of this size in a repeated measures ANOVA investigating between-factors effects with 4 groups and nine repeated measures. This study employed a total sample size of 28 undergraduate students (24 female, 4 male) between the ages of 18 and 25 years (M = 20.8, SD = 1.7). Seven additional participants did not meet the minimum performance criteria and were excluded from the study.

Procedures

The experimental preparation was identical to Study 1 with the following exceptions. Experimental sessions contained a maximum of nine participants per single session. The researcher quasi-randomly assigned participants to one of four groups (n = 7 participants, each) that differed only in regards to the training condition. Training conditions consisted of a control group trained with S+ 50 and experimental groups trained with S+ 50 and S- of 64, 78, or 92. These S- values correspond to distances from the S+ of 2, 4, and 6 stimulus progression steps, respectively, within the stimulus array. Pilot data indicated that participants had difficulty mastering the discrimination when the S- was only 1 stimulus value away from the S+ (S- 57)

during the brief training provided in the current preparation. The minimum performance criteria for being included in the study were the same as in Study 1. Of the seven participants who did not meet the performance criteria, six failed the training criterion (5 experienced S- 64 training, 1 experienced S- 92 training). One participant from the S- 78 training group did not respond affirmatively to any generalization trials.

Several procedural modifications reduced the overall length of the experimental session. During training and test trials, the text prompt to record and the 1-s tone were presented 1 s after the trial began. During training, the buttons corresponding to the participant's recorded response appeared after 5 s. During the generalization test, trials automatically advanced after 5 s. This change was made to promote participants attending to stimuli throughout the generalization test. During Study 1, participants generally recorded responses within the first 1-2 s and often spent the remaining time engaging in behavior such as drawing or closing their eyes, which may have interfered with attending to subsequent trials. Previous research has indicated that differences in brief observation time have minimal effects on accuracy when recording the malignance of moles (Miles & Meehan, 1995) and these trial lengths are similar to those used in other studies evaluating generalization gradients with humans (e.g., Livesey & McLaren, 2009).

Results

Average gradients for participants in Study 2 are displayed in Figure 7. Average responding for participants in the control group replicated those seen in Study 1, with peak responding observed at the S+ and a bias for responding to stimuli more malignant than the S+. The average gradients of both the S- 64 and the S- 78 groups did not differ from the control group; however, the average gradient for the S- 92 group differed from all groups, with responding shifted away from the S+ in the direction opposite the S-.

Individual data for each participant are depicted in Figures 8, 9, 10, and 11 for the control, S- 64, S- 78, and S- 92 groups, respectively. The gradients generally displayed the same effect demonstrated by the average for the group, although some variability was observed.

The average gradient means per trial blocks during the generalization test for each group are depicted in Figure 12. The graph indicates that mean responding for all discrimination training groups were initially shifted away from the S+ in the direction opposite the S- (toward stimuli earlier in the malignance progression). This effect did not maintain for the S- 64 and S- 78 groups, but became more pronounced for the S- 92 group. Average gradient means and standard error were 57.20 (0.87), 54.25 (1.02), 54.78 (1.31), and 46.00 (1.02) for the control, S- 64, S- 78, and S- 92 groups, respectively. The researcher intended to conduct a two-way repeated measures ANOVA but could not due to participants not always responding affirmatively to at least one stimulus in each trial block across the generalization test. Rather than interpolating gradient means for these trial blocks, the researcher opted to conduct a two-way ANOVA without repeated measures. The two-way ANOVA of the four groups indicated a main effect of group level ($F_{3,209} = 13.43$, $p < 0.01$), but not trial block ($F_{8,209} = 1.22$, $p = 0.29$). The trial block by group interaction was not statistically significant ($F_{24,209} = 0.45$, $p = 0.99$). Exploring the main effect of group more fully, a Fisher's LSD posttest revealed that significance was a result of the difference between S- 92 and each other group, such that the S- 92 training group featured a mean significantly left of the three other groups (as described above). A more thorough evaluation of the interaction found statistically significant differences across trial blocks 2 and 5-9 of the generalization test when comparing the S- 92 and control group, across trial blocks 6-9 when comparing the S- 92 and S- 64 groups, and across trial blocks 7-9 when comparing the S- 92 and S- 78 groups. These results indicate that the effect of the S- 92 training

became more pronounced over the course of the generalization test and that responding by participants in the S- 64 and S- 78 trainings became more similar to those of the control group over the course of the generalization test.

The average stimulus values of peaks were 56.00, 54.00, 54.50, and 44.50 for the control, S- 64, S- 78, and S- 92 groups, respectively. Differences in peak responding between experimental and control groups were analyzed using one-tailed Mann-Whitney *U* tests. No statistically significant differences were observed when comparing the control to the S- 64 (U = 17.50, p = 0.18) or the S- 78 (U = 22.50, p = 0.42) groups; however, the difference between the control (*Mdn* = 57.00 [50.00, 57.00]) and S- 92 (*Mdn* = 43.00 [36.00, 53.50]) groups was statistically significant (U = 9.00, p = 0.03). A Kruskal-Wallis test indicated no statistical differences in peak responding between any of the groups that experienced discrimination training (H_2 = 5.81, p = 0.12).

The average areas for each group are depicted in Table 5. Individual one-tailed Mann-Whitney *U* tests compared each discrimination training group to the control group to determine whether differences existed in area for each half of the stimulus array (see Table 5). The *U* tests indicated statistically significant differences only for the comparison of area to the left of S50 for the S- 92 group (U = 8.00, p = 0.02). No other comparisons were statistically significant. These results suggest the S- 92 group primarily affected responding by increasing responding to stimuli less malignant than the S+ and not by reducing the bias for responding to more malignant stimuli. Krsukal-Wallis *H* tests compared the areas for the discrimination training groups and found no significant differences in area to the left of S50 (H_2 = 4.47, p = 0.11) or in area to the right of S50 (H_2 = 1.87, p = 0.41), indicating that the discrimination training groups overlapped

in their effects on responding. This is not entirely surprising given the parametric nature of this study.

The differences in slopes between adjacent pairs of stimuli were compared across groups using a two-way ANOVA (see Table 6). The test of the four groups indicated a main effect of the stimulus pair ($F_{13,336} = 13.41$, $p < 0.01$), but not group ($F_{2,336} = 0.0009$, $p = 1.00$). The stimulus pair by group interaction was statistically significant ($F_{26,336} = 1.50$, $p = 0.03$). Exploring this interaction more fully, a Fisher's LSD posttest revealed that significant differences in slopes were obtained across a variety of adjacent stimuli depending on the groups compared, indicating that the trainings differentially affected only a subset of the stimuli in the array. Significant differences were observed between the adjacent stimulus pairs ranging from S36-43 to S57-S64. Comparisons between S- 92 and each other group revealed statistically significant differences in slope between S43-50, indicating this was where the S- 92 exerted the greatest effect on discriminability.

Analyses of SDT indices were conducted as described in Study 1. Kruskal-Wallis tests of accuracy ($H_3 = 2.75$, $p = 0.43$), sensitivity ($H_3 = 2.19$, $p = 0.53$), specificity ($H_3 = 2.29$, $p = 0.51$), and precision ($H_3 = 2.01$, $p = 0.57$) indicated no statistically significant differences for any of the indices for any of the comparisons between groups (see Table 7).

Taken together these results indicate that the group with the S- that was furthest away from the S+ produced the strongest and most consistent shifts. Although less pronounced, the effects of training with the S- at 64 and 78 appeared to differ from the control group in some of the analyses; however, differences only approached statistical significance.

Study 3: Examining Effects of Multiple S- Stimuli

In an effort to explore additional means of altering discrimination to promote early detection of melanoma, Study 3 evaluated the effects of discrimination training that employed two S- stimuli. Galizio and Baron (1985) found that this preparation was effective in producing larger shifts in postdiscrimination gradients with adult humans as compared to training with a single S-. The results of Study 2, in which the largest shift occurred when the S- was furthest away from the S+, suggest that under the current experimental preparation participants are likely to respond to relative characteristics of S+/S- disparities. The use of two S- stimuli may facilitate categorization of malignant features, resulting in stronger peak shift toward less malignant stimuli. Understanding the effects of multiple training preparations will likely prove important in ensuring clinically beneficial changes in generalization with moles.

Participants

An *a priori* power analysis using G*Power 3.1.5 (see Faul et al., 2007) suggested a minimum total sample size of 15 to detect an effect of this size in a repeated measures ANOVA investigating between-factors effects with 3 groups and nine repeated measures. This study employed a total sample size of 21 undergraduate students (16 female, 5 male) between the ages of 18 and 23 (M = 20.3, SD = 1.3). Two additional participants were excluded from the study.

Procedures

The experimental preparation was similar to Study 2. The researcher quasi-randomly assigned participants to one of three training groups. Each training condition consisted of 30 training trials and generalization tests consisted of 135 trials. One control group experienced training only with the S+ (+50) to determine the effects of increased training trials on generalization. A second control group experienced training with the S+ of S50 and two S- of S1

and S99 (-1/+50/-99) to determine the effects of training two S- stimuli on the postdiscrimination gradient. The experimental group experienced trained with S+ 50, S- 78, and S- 99 (+50/-78/-99) to determine the effects of training two S- stimuli on one extreme of the stimulus array on the postdiscrimination gradient. The minimum performance criteria for being included in the study were the same as in two previous studies except that participants needed to respond correctly to 13 of the last 15 training trials. Of the two participants who did not meet the performance criteria, one failed to respond affirmatively to any generalization test trials (+50 training group) and one failed to respond to every trial throughout the experiment (+50/-78/-99 training group).

Results

Average gradients for participants in Study 3 are displayed in Figure 13. Average responding for participants in the +50 group replicated those of both previous studies, with peak responding occurring at the S+ and a bias for responding to stimuli more malignant than the S+; however, there were no clear differences when comparing the average gradients from any of the groups from Study 3.

Individual data for each participant are depicted in Figures 14, 15, and 16 for the control, +50, -1/+50/-99, and +50/-78/-99 groups, respectively. The gradients generally displayed the same effect demonstrated by the average for the group, although some variability was observed.

The average gradient means per trial blocks during the generalization test for each group are depicted in Figure 17. The graph indicates that mean responding for the +50 group replicated the bias observed in the previous studies and that mean responding for the -1/+50/-99 group occurred more consistently at the S+ value. Additionally, the graph indicates that mean responding for the +50/-78/-99 group initially shifted responding away from the S+ in the direction opposite the S- (toward stimuli earlier in the malignance progression); however, this

effect did not maintain. The average gradient means were 53.59, 51.43, and 48.96 for the +50, -1/+50/-99, and +50/-78/-99 groups, respectively. The researcher intended to conduct a two-way repeated measures ANOVA but could not due to participants not always responding affirmatively to at least one stimulus in each trial block across the generalization test. Rather than interpolating gradient means for these trial blocks, the researcher opted to conduct a two-way ANOVA without repeated measures. The two-way ANOVA of the three groups indicated a main effect of group level ($F_{2,161} = 6.503$, $p < 0.01$), but not trial block ($F_{2,161} = 1.41$, $p = 0.19$). The trial block by group interaction was not statistically significant ($F_{2,161} = 0.66$, $p = 0.83$). Exploring the main effect of group more fully, a Fisher's LSD posttest revealed that significance was a result of the difference between the +50 ($M = 53.52$, $SEM = 0.56$) and +50/-78/-99 ($M = 49.04$, $SEM = 1.09$) groups, such that the +50/-78/-99 training group featured a mean significantly left of the +50 group (as described above). A more thorough evaluation of the interaction found statistically significant differences across the first three trial blocks of the generalization test, indicating that the effect of the +50/-78/-99 was fleeting.

The average stimulus values of peaks were 52.50, 50.00, 54.50, and 48.00 for the +50, -1/+50/-99, and +50/-78/-99 groups, respectively. A two-tailed *U* test compared differences in peak responding between the two control groups because there was no predicted direction of difference, whereas one-tailed tests compared the experimental group to the control groups due to the predicted direction of change. No statistically significant differences were observed when comparing the +50 to the -1/+50/-99 group ($U = 15.00$, $p = 0.18$) or the -1/+50/-99 to the +50/-78/-99 group ($U = 18.00$, $p = 0.20$). The comparison of differences in peak responding for the +50 ($Mdn = 50.00$ [50.00, 53.50]) and the +50/-78/-99 ($Mdn = 50.00$ [39.50, 50.00]) group approached statistical significance ($U = 12.00$, $p = 0.05$).

The average areas for each group are depicted in Table 8. Two-tailed *U* tests compared differences in area with respect to the less malignant and more malignant halves of the stimulus values for the +50 and -1/+50/-99 groups and one-tailed *U* tests compared differences between the +50/-78/-99 group and each of the control groups using (see Table 8). Statistically significant differences were only obtained for the comparison of area to the left of S50 for the +50 (*Mdn* = 70.00 [38.50, 73.50]) and -1/+50/-99 (*Mdn* = 101.5 [80.50, 136.5]) groups (*U* = 5.50, $p = 0.01$), suggesting a relatively increased propensity for participants in the -1/+50/-99 to respond to stimuli less malignant than the S+. No other comparisons were statistically significant, although the difference between areas to the left of S50 approached significance for the comparison of the +50 and +50-78/-99 (*Mdn* = 80.50 [56.00, 143.5]) groups (*U* = 12.50, $p = 0.07$).

To determine differential effects of training on slope, a two-way ANOVA compared differences in slopes between adjacent pairs of stimuli across groups (see Table 9). The test of the four groups indicated a main effect of the stimulus pair ($F_{13,336} = 20.30$, $p < 0.01$), but not group ($F_{2,252} = 0.0$, $p > 0.99$). The stimulus pair by group interaction approached statistical significance ($F_{26,252} = 1.49$, $p = 0.07$). Exploring the main effect more fully, a Fisher's LSD posttest revealed that significant differences in slopes were obtained across a variety of adjacent stimuli depending on the groups compared, indicating that the trainings differentially affected only a subset of the stimuli in the array. Significant differences were observed between the adjacent stimulus pairs ranging from S43-50 to S71-S78. The greatest effect on discriminability occurred between S50-S57, in that all comparisons revealed statistically significant differences in slope. The only significant differences in slope for the +50 and +50/-78/-99 comparisons were between S43-S50; however, comparisons between -1/+50/-99 and +50/-78/-99 indicated

significant differences for slopes between S43-S50, S64-S71, and S71-S78. These results suggest that each of the training preparations differentially affected discriminability in unique ways across the array of melanoma progression.

Analyses of SDT indices were conducted as described in Study 1. Kruskal-Wallis tests of accuracy ($H_2 = 1.91$, $p = 0.38$), sensitivity ($H_2 = 3.58$, $p = 0.15$), specificity ($H_2 = 3.34$, $p = 0.19$), and precision ($H_2 = 1.96$, $p = 0.38$) indicated no statistically significant differences for any of the indices for any of the comparisons between groups (see Table 10).

Overall, these results indicate that training with two S- stimuli to one side of the S+ was not effective in producing a substantial gradient shift. Training with an S- to either side of the S+ was effective at producing consistent response patterns across participants.

Discussion

This series of investigations provides an initial foray into the effects of discrimination training on melanoma detection from a use-inspired basic behavioral approach. Capitalizing on a wealth of behavioral literature on generalization and peak shift to explore methods for understanding melanoma identification provides a unique joining of two largely independent areas of study. The experiments sought to (1) use discrimination training to establish generalization and postdiscrimination gradients with moles as stimuli, (2) determine the effects of parametric manipulations of training on postdiscrimination gradients, and (3) evaluate training with multiple discriminative stimuli. The results of Study 1 indicated that it was possible to obtain orderly generalization gradients using moles along the dimension of malignance and that discrimination training could shift responding along the stimulus array. The results of Study 2 indicated that, under the current preparation, shifted postdiscrimination gradients occurred only when the discrepancy between the S+ and S- stimuli was relatively large. The results of Study 3

indicated that discrimination training with S- stimuli at both extremes of the stimulus continuum was effective in producing highly consistent responding across the generalization test, whereas training with two S- stimuli on one side on the continuum of the S+ initially produced a shift in responding away from the S+ but this effect did not maintain throughout the generalization test.

In addition to the primary measure of gradient mean, comparisons of several supplementary measures provided further analysis of the differential effects of the training conditions. The analyses of peak responding were highly consistent with the primary measure and so did not afford a great deal more insight regarding the effects across groups. Logically, this makes sense in that the peak is also a measure of central tendency (i.e., mode). One interesting relation consistently observed between the peak and gradient mean was that the peak was always at a lower value on the continuum of malignance than the gradient mean. This, again, makes sense in that this relation is a product of the biased responding observed during the generalization test.

Unlike comparisons of peak, analyses of area to either side of the S+ offered a more sensitive account of the effects of discrimination training. Of particular interest, across all three studies the area of the gradients between S50 and S99 was nearly twice that of the area between S1 and S50 for the control groups, providing a direct measure of the bias observed during the generalization test. Additionally, mean area between S1 and S50 was larger than the area between S50 and S99 only for the S- 99, S- 92, and +50/-78/-99 groups, suggesting that these training conditions were effective at changing the distribution of responses as compared to the control groups (however, note that the comparison of +50 and +50/-78/-99 only approached statistical significance).

The between-group analyses of slope revealed some interesting consistencies across the experiments. In particular, a large proportion of the slopes did not differ significantly across the experiments, indicating that the effects of the trainings were typically restricted to a few stimuli within each group (often for the slopes near the center of the gradients). Interestingly, whenever a statistically significant difference in gradients between a test and control group was obtained on another measure, a statistically significant difference also was obtained for the S50-S57 slope. This suggests that a change in the discriminability between the S+ (S50) and S57 was an important factor in shifting the gradient.

The use of nearly identical procedures over the course of experiments allows for additional comparisons of the results as a whole. Of particular note is the consistent bias in responding toward stimuli displaying malignance greater than the S+ for the control groups across all three experiments. Such a consistent bias may engender questions about the validity of the stimuli given their complex nature and development via computer software. However, as previously noted, biased responding during generalization tests is a common finding in the literature (see Ghirlanda & Enquist, 2003). Rather than being a confound, the presence of the response bias may provide insight with respect to the nature of the stimuli used in the current investigations and potentially to progressions of malignant melanoma in general. Specifically, progressing malignance likely involves changes in the morphology of a mole along dimensions of both intensity and rearrangement. The ABCDs originally described by Friedman, Rigel, and Kopf (1985) can be broken down into these perceptual categories outlined in the review by Ghirlanda & Enquist (2003). Specifically, both asymmetry and border irregularity are properties of the more general dimension of object shape, color variegation (under natural conditions) is comprised of both hue and color intensity dimension, and diameter is a descriptor of size.

According to the results of their review, object shape dimensions are consistent with rearrangement-type gradients (symmetrical and peaked), as is hue. Color intensity tends to produce intensity-type gradients (asymmetrical or monotonic) and size dimensions tend to produce gradients displaying a combination of rearrangement and intensity dimensions (asymmetrical and peaked). Based on the results of Dalianis et al. (2012) that directly assessed generalization based on the asymmetry and border irregularity of moles, these two dimensions appear to produce gradients consistent with rearrangement due to their observed peaks and symmetry. Gradients obtained across all three of the present experiments appear consistent with the rearrangement/intensity explanation offered by Ghirlanda and Enquist in that the multiple dimensions of object shape, size, and color interacted in such a way that participants' responses produced peaked, asymmetrical (biased) gradients. A limitation exists in that the extent to which a particular dimension exerted a greater degree of control with respect to the obtained gradients is unknown. For example, generalization may have been strictly a result of stimulus size, limiting the degree to which the results are reflective of a progression of malignance. It should be noted, however, that other stimulus arrays utilizing complex, biologically relevant stimuli have displayed similar effects. In particular, Derenne's (2008, 2010) studies examining female waist-to-hip ratio and facial symmetry also found biases toward one end of the stimulus array, specifically, the end of the array associated with evolutionarily advantageous preferences. That is, there was a bias toward larger waist-to-hip ratios, which are indicators of fertility and ability to successfully birth children, and toward symmetrical faces, which are indicators fewer genetic disorders.

Although the extent to which the current findings are representative of what would be seen in naturally occurring progressions of malignance is unknown at this juncture, they are

immediately informative with respect to the application of existing theories to this line of inquiry. It is apparent that the adaption-level model more readily accounts for the present results than does the gradient-interaction model. In Study 1, the S- stimulus at either extreme of the continuum of test stimuli produced gradient shifts. In Study 2, the S- more similar to the S+ did not produce gradient shifts, but the more distinct S- did. In Study 3, the tendency to obtain shifted gradients with distal S- stimuli (i.e., S92, S99) observed in the previous studies decreased when the training also included a more proximal S-, suggestive of an averaging of the S- features. Each of these findings is consistent with AL model predictions. However, contrary to the AL model, there was no clear central tendency in responding over the course of the generalization test in any of the experiments. Participants' responding occasionally exhibited trends, but these were not consistently toward the central value of the array. Furthermore, the group level analyses of mean responding displayed varying patterns across the experiments, with no trend observed in Study 1, a trend away from the central value in Study 2, and a slight trend toward the central value in Study 3.

That the current results only partly fit the AL model adds further ambiguity to the existing theoretical debate regarding human generalization. Continued exploration of the applicability of the AL model to generalization with complex stimuli could be done in such a way as to expand both theoretical and practical knowledge. In particular, the AL model may need to be restricted to the conditions under which the test stimulus shares only a certain proportion of its properties with the S+. For instance, when the test stimulus shares a small proportion of the features of the S+, the discrimination may occur rapidly, but when the test stimulus shares a larger proportion of the properties, the discrimination may take more time. It may be the case that the rule for responding based on average experience is applied only with the

second stimulus in this example—a subtlety not currently proposed by the AL model. Researchers could evaluate this through measuring response latencies during the generalization test. Relating this to the present experiments, a response to S92 by a participant having experienced training with S- 99 may occur more quickly than that same participant's response to S64.

Additionally, it would be important to assess the degree to which the immediately preceding test stimulus influenced responding on the subsequent test stimulus. To evaluate this, researchers could use a fixed order of test stimuli, with a target stimulus preceded by stimuli of varying, predetermined distances to allow for analysis of interactions. This becomes relevant to identification of melanoma when one considers the relative distribution of malignant moles among benign moles, either in a generalization test or (more importantly) on a person. Having numerous moles is a risk factor for melanoma and, presumably, the vast majority of moles on a person are not malignant. If the preceding benign mole influences the response to the subsequent malignant mole in a reliable way, information about this additional determinant of stimulus control would be valuable in the further development of behavioral techniques for promoting early identification.

A noteworthy consideration when discussing the merits of current theories of generalization is whether generalization gradients do, in fact, describe the actual behavior of organisms. An argument against the use of generalization gradients as an explanatory device was posited by the quantal interpretation of stimulus control (Bickel & Etzel, 1985). This interpretation holds that generalization gradients are an artifact the methods for training and testing the target stimulus-response relation(s) because responding is integral (quantal) in nature and the averaging of unitary stimulus-response relations produces gradients that inappropriately

imply continuous relations. Difficulties in accounting for observed responding in generalization gradients with a single theory may be because they are products of the procedures and not the behavioral process under investigation. However, the interpretation largely criticized studies utilizing response rates as the primary dependent measure, as well as studies inferring some level of induction (response generalization) occurring concomitantly with changes in the level of the stimulus dimension(s) of interest. The use of discrete-trial procedures for examining generalization and peak shift eliminates this criticism to some extent. Additionally, Bickel and Etzel noted that peak shift itself is a product of multiple stimulus-response relations in that discrimination training often involves explicitly training a stimulus-response relation in the presence of the S+ and implicitly training a stimulus-response relation (the absence of the S+ response plus whatever behavior takes its place) in the presence of the S-. By explicitly training a second stimulus-response relation for the S-, researchers can describe peak shift more adequately via its quantal components. The authors also suggested that, although the quantal interpretation allows for analysis at a more molecular level than the traditional (or descriptive) interpretation that relies on generalization gradients, the descriptive interpretation offers a more molar level analysis that can lead to effective scientific action as well.

In further considering other approaches to analyzing generalization, SDT offers another alternative. As such, data from each of the studies were analyzed according to SDT to determine whether this model would provide additional understanding of the obtained results. The SDT indices of accuracy, sensitivity, specificity, and precision are percentages of responding in the presence of versus absence of a target event and are intended to provide measures of an individual's criteria for responding to a signal and their sensitivity to changes in the signal. With respect to the current series of studies, all SDT indices provided little additional information for

interpreting participants' generalization. There were no statistically significant differences for any of the indices when comparing across groups, except when comparing the control and S- 1 groups of Study 1. As a result, little can be said about the utility of this model for melanoma detection. Perhaps an intervention that is more powerful or that includes greater number of participants is required for SDT analyses to detect effects.

As only one other study to date has attempted to apply behavioral strategies to the issue of melanoma detection (i.e., Dalianis et al., 2011), this line of work is clearly in its nascence. Numerous extensions of the present findings are immediately apparent for further exploring how the conceptual systems of behavior analysis apply to the specific issue of melanoma detection, for correcting existing procedural limitations and evaluating alternative preparations, and particularly for advancing a technology designed for application.

A major limitation of the current studies existed in the recording methods (participants recording their own responses). However, both the orderliness of the data and the Monte Carlo simulations speak to the improbability of obtaining these results if participants were responding irrespective of the stimuli presented during the experiments. Future investigations warrant automated data collection in order to eliminate such concerns and provide additional measures of interest (e.g., response latency).

In addition to increasing confidence in the data, utilizing a more advanced, customizable computer program as the interface for the participants would allow for the examination of more complex independent variables, such as training equivalence classes/relational frames along with or in place of the mole discriminations. The benefit of using training that involves verbal behavior (e.g., rules) is that it can produce long-lasting changes in behavior. Along this line of inquiry, researchers could examine the effects of verbal behavior on mole discrimination by

providing information about the stimuli. In the aforementioned study by Girardi et al. (2006), stressing statements about the lethal outcomes of cancer biased participants' ability to detect cancer. As such, simply informing participants that the stimuli include melanoma may alter their responding, and therefore is a critical variable to isolate under experimental conditions prior to examining more clinically oriented manipulations.

A second limitation, specifically related to the clinical utility of the experiments, was that the stimulus array was created without consulting an expert dermatologist to assess its face validity. Given the basic nature of the investigations, it was acceptable to proceed without such consultation, as this series of studies examined the conceptual application of generalization gradients to melanoma detection. Additionally, approaching an expert medical professional in the absence of any preliminary research seems unlikely to be successful. However, prior to conducting further investigations toward a clinically relevant technology, it is important that an expert in skin cancer be consulted with respect to the stimulus array, as well as determining the stimulus values along the array at which it would be ideal for patients to be able to identify the progression of malignance. This is would increase the potential for a technology based on the procedures of these studies to produce beneficial outcomes for participants.

A related limitation regarding the generality of the results is that the stimulus array was restricted to the progression of a single mole from benign to malignant. As malignant moles differ in appearance, the progression from benign to malignant would necessarily contain different proportions of the rearrangement and intensity dimensions evaluated in this series of experiments. Future investigations should develop and assess numerous progressions based on actual cancerous lesions as their malignance progresses. However, as it would be unethical to track the progression of malignance in a human without intervening, using standard medical

diagnostic criteria (e.g., the American Joint Committee on Cancer staging guidelines) to produce abstractions at each level of malignance would possibly be a fruitful alternative.

As a whole, this series of experiments provides a first attempt to understand how discrimination training might help promote visual detection of melanoma. The attempt to apply behavior analytic methods and conceptual systems to this issue suggests this to be a fertile area for future research. Although a technology ready for consumption by the public is obviously not at hand, a line of research stemming from these use-inspired basic experiments may bring this goal to fruition.

References

Abbasi, N. R., Shaw, H. M., Rigel, D. S., Friedman, R. J., McCarthy, W. H., Osman, I., . . . Polsky, D. (2004). Early diagnosis of cutaneous melanoma. *Journal of the American Medical Association, 292*, 2771-2776.

Akins, F. R., Gouvier, W. D., & Lyons, J. (1981). The effect of S+ and S− alternation on stimulus control and the peak shift. *The Journal of General Psychology, 105*, 101-114.

Argenziano, G., Cerroni, L., Zalaudek, I., Staibano, S., Hofmann-Wellenhof, R., Arpaia, N., . . . Bandelloni, R. (2012). Accuracy in melanoma detection: A 10-year multicenter survey. *Journal of the American Academy of Dermatology*, *67*, 54-59.e51.

Baron, A. (1973). Postdiscrimination gradients of human subjects on a tone continuum. *Journal of Experimental Psychology, 101*, 337-342.

Berwick, M., Begg, C. B., Fine, J. A., Roush, G. C., & Barnhill, R. L. (1996). Screening for cutaneous melanoma by skin self-examination. *Journal of the National Cancer Institute, 88*, 17-23.

Bickel, W. K., & Etzel, B. C. (1985). The quantal nature of controlling stimulus-response relations as measured in tests of stimulus generalization. *Journal of the Experimental Analysis of Behavior, 44*, 245-270.

Bloomfield, T. M. (1967). A peak shift on a line-tilt continuum. *Journal of the Experimental Analysis of Behavior, 10*, 361-366.

Blough, D. S. (1965). Definition and measurement in generalization research. In D. I. Mostofsky (Ed.), *Stimulus generalization*. Stanford: Stanford University Press.

Blough, D. S. (1969). Attention shifts in a maintained discrimination. *Science, 166*, 125-126.

Blough, D. S. (1975). Steady state data and a quantitative model of operant generalization and discrimination. *Journal of Experimental Psychology: Animal Behavior Processes, 104*, 3-21.

Borland, R., Mee, V., & Meehan, J. W. (1997). Effects of photographs and written descriptors on melanoma detection. *Health Education Research, 12*, 375-384.

Brady, M. S., Oliveria, S. A., Christos, P. J., Berwick, M., Coit, D. G., Katz, J., & Halpern, A. C. (2000). Patterns of detection in patients with cutaneous melanoma. *Cancer, 89*, 342-347.

Brochez, L., Verhaeghe, E., Bleyen, L., & Naeyaert, J.-M. (2001). Diagnostic ability of general practitioners and dermatologists in discriminating pigmented skin lesions. *Journal of the American Academy of Dermatology, 44*, 979-986.

Brooks, A., Predebon, J., & Van der Zwan, R. (2001). Perceptual strategies to improve skin cancer discriminations in naive observers. *Public Health, 115*, 139-145.

Buettner, P. G., & Garbe, C. (2000). Agreement between self-assessment of melanocytic nevi by patients and dermatologic examination. *American Journal of Epidemiology, 151*, 72-77.

Cassileth, B. R., Clark Jr, W. H., Lusk, E. J., Frederick, B. E., Thompson, C. J., & Walsh, W. P. (1986). How well do physicians recognize melanoma and other problem lesions? *Journal of the American Academy of Dermatology, 14*, 555-560.

Dalianis, E. A., Critchfield, T. S., Howard, N. L., Jordan, J. S., & Derenne, A. (2011). Using morphed images to study visual detection of cutaneous melanoma symptom evolution. *The Psychological Record, 61*, 341-362.

De Giorgi, V., Grazzini, M., Rossari, S., Gori, A., Papi, F., Scarfi, F., . . . Gandini, S. (2012). Is skin self-examination for cutaneous melanoma detection still adequate? A retrospective study. *Dermatology, 225*, 31-36.

Derenne, A. (2006). Effects of S+ and S- separation on gradient shifts in humans. *Journal of General Psychology, 133,* 163-173.

Derenne, A. (2010). Shifts in postdiscrimination gradients within a stimulus dimension based on bilateral facial symmetry. *Journal of the Experimental Analysis of Behavior, 93,* 485-494.

Derenne, A., Breitstein, M., & Cicha, R. J. (2008). Shifts in postdiscrimination gradients within a stimulus dimension based on female waist-to-hip ratios. *The Psychological Record, 58,* 51-60.

Dickinson, J., & Hedges, D. (1986). Adaptation level as an explanation of the peak shift in generalization with movement stimuli. *Journal of Motor Behavior, 18,* 101-110.

Doll, T. J., & Thomas, D. R. (1967). Effects of discrimination training on stimulus generalization for human subjects. *Journal of Experimental Psychology, 75,* 508-512.

Dukhayyil, A., & Lyons, J. E. (1973). The effect of overtraining on behavioral contrast and the peak-shift. *Journal of the Experimental Analysis of Behavior, 20,* 253-263.

Ellis, W. R. (1970). Role of stimulus sequences in stimulus discrimination and stimulus generalization. *Journal of Experimental Psychology, 83,* 155-163.

Epstein, D. S., Lange, J. R., Gruber, S. B., Mofid, M., & Koch, S. E. (1999). Is physician detection associated with thinner melanomas? *Journal of the American Medical Association, 281,* 640-643.

Epstein, F. H., Gilchrest, B. A., Eller, M. S., Geller, A. C., & Yaar, M. (1999). The pathogenesis of melanoma induced by ultraviolet radiation. *New England Journal of Medicine, 340,* 1341-1348.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Friedman, H. (1963). Wavelength generalization as a function of spacing of test stimuli. *Journal of Experimental Psychology, 65*, 334-338.

Friedman, R. J., Rigel, D. S., & Kopf, A. W. (1985). Early detection of malignant melanoma: The role of physician examination and self-examination of the skin. *CA: A Cancer Journal for Clinicians, 35*, 130-151.

Furrow, D. R., & LoLordo, V. M. (1975). Stimulus control in a discrimination based on a distinctive feature. *Journal of the Experimental Analysis of Behavior, 23*, 217-222.

Galizio, M. (1985). Human peak shift: Analysis of the effects of three-stimulus discrimination training. *Learning and Motivation, 16*, 478-494.

Galizio, M., & Baron, A. (1979). Human postdiscrimination gradients: The effects of three-stimulus discrimination training. *Animal Learning & Behavior, 7*, 53-56.

Geller, A. C., Miller, D. R., Swetter, S. M., Demierre, M.-F., & Gilchrest, B. A. (2006). A call for the development and implementation of a targeted national melanoma screening program. *Archives of Dermatology, 142*, 504-507.

Ghirlanda, S., & Enquist, M. (2003). A century of generalization. *Animal Behaviour, 66*, 15-36.

Girardi, S., Gaudy, C., Gouvernet, J., Teston, J., Richard, M. A., & Grob, J. J. (2006). Superiority of a cognitive education with photographs over ABCD criteria in the education of the general population to the early detection of melanoma: a randomized study. *International Journal of Cancer, 118*, 2276-2280.

Goodson, A. G., & Grossman, D. (2008). Strategies for early melanoma detection: Approaches to the patient with nevi. *Journal of the American Academy of Dermatology, 60,* 719-735.

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

Grusec, T. (1968). The peak shift in stimulus generalization: Equivalent effects of errors and noncontingent shock. *Journal of the Experimental Analysis of Behavior, 11*, 239-249.

Guttman, N. (1959). Generalization gradients around stimuli associated with different reinforcement schedules. *Journal of Experimental Psychology, 58*, 335-340.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51*, 79-88.

Hamidi, R., Peng, D., & Cockburn, M. (2010). Efficacy of skin self-examination for the early detection of melanoma. *International Journal of Dermatology, 49*, 126-134.

Hanrahan, P. F., Hersey, P., Menzies, S. W., Watson, A. B., & D'Este, C. A. (1997). Examination of the ability of people to identify early changes of melanoma in computer-altered pigmented skin lesions. *Archives of Dermatology, 133*, 301-311.

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology, 58*, 321-334.

Hanson, H. M. (1961). Stimulus generalization following three-stimulus discrimination training. *Journal of Comparative and Physiological Psychology, 54*, 181-185.

Hearst, E. (1971). Contrast and stimulus generalization following prolonged discrimination training. *Journal of the Experimental Analysis of Behavior, 15*, 355-363.

Honig, W. K. (1962). Prediction of preference, transposition, and transposition-reversal from the generalization gradient. *Journal of Experimental Psychology, 64*, 239-248.

Honig, W. K., Thomas, D. R., & Guttman, N. (1959). Differential effects of continuous extinction and discrimination training on the generalization gradient. *Journal of Experimental Psychology, 58*, 145-152.

Honig, W. K., & Urcuioli, P. J. (1981). The legacy of Guttman and Kalish (1956): 25 years of research on stimulus generalization. *Journal of the Experimental Analysis of Behavior, 36*, 405-445.

Howlader, N., Noone, A., Krapcho, M., Neyman, N., Aminou, R., Waldron, W., . . . Tatalovich, Z. (2011). SEER Cancer Statistics Review, 1975–2008. National Cancer Institute; Bethesda, MD: 2011.

Kalish, H. I. (1958). The relationship between discriminability and generalization: A revaluation. *Journal of Experimental Psychology, 55*, 637-644.

Karakousis, G. C., & Czerniecki, B. J. (2011). Diagnosis of Melanoma. *PET Clinics, 6*, 1-8.

Kasparian, N. A., McLoone, J. K., & Meiser, B. (2009). Skin cancer-related prevention and screening behaviors: A review of the literature. *Journal of Behavioral Medicine, 32*, 406-428.

Koh, H. K., Miller, D. R., Geller, A. C., Clapp, R. W., Mercer, M. B., & Lew, R. A. (1992). Who discovers melanoma?: Patterns from a population-based survey. *Journal of the American Academy of Dermatology, 26*, 914-919.

Lamerson, C. L., Eaton, K., Sax, J. L., & Kashani-Sabet, M. (2012). Comparing melanoma invasiveness in dermatologist-versus patient-detected lesions: A retrospective chart review. *Journal of Skin Cancer, 2012*, 1-5.

Landau, J. S. (1968). Postdiscrimination generalization in human subjects of two different ages. *Journal of Experimental Psychology, 76*, 656-663.

Lawson, D. D., Moore II, D. H., Schneider, J. S., & Sagebiel, R. W. (1994). Nevus counting as a risk factor for melanoma: comparison of self-count with count by physician. *Journal of the American Academy of Dermatology, 31*, 438-444.

Lazareva, O. F. (2012). Relational learning in a context of transposition: A review. *Journal of the Experimental Analysis of Behavior, 97*, 231-248.

Lewis, M., & Johnston, R. (1999). Are caricatures special? Evidence of peak shift in face recognition. *The European Journal of Cognitive Psychology, 11*, 105-117.

Little, E. G., & Eide, M. J. (2012). Update on the current state of melanoma incidence. *Dermatologic Clinics, 30*, 355-361.

Livesey, E. J., & McLaren, I. P. L. (2009). Discrimination and generalization along a simple dimension: Peak shift and rule-governed responding. *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 554-565.

MacKinnon, M. M. (1972). Adaptation-level theory, anchor theory, and the peak shift phenomenon. *Journal of Motor Behavior, 4*, 1-12.

McGuire, S. T., Secrest, A. M., Andrulonis, R., & Ferris, L. K. (2011). Surveillance of patients for early detection of melanoma: patterns in dermatologist vs patient discovery. *Archives of Dermatology, 147*, 673-678.

McLaren, I., Bennett, C., Guttman-Nahir, T., Kim, K., & Mackintosh, N. (1995). Prototype effects and peak shift in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 662-673.

McPherson, M., Elwood, M., English, D. R., Baade, P. D., Youl, P. H., & Aitken, J. F. (2006). Presentation and detection of invasive melanoma in a high-risk population. *Journal of the American Academy of Dermatology, 54*, 783-792.

Mickler, T. J., Rodrigue, J. R., & Lescano, C. M. (1999). A comparison of three methods of teaching skin self-examinations. *Journal of Clinical Psychology in Medical Settings, 6,* 273-286.

Miles, C., & Jenkins, H. (1973). Overshadowing in operant conditioning as a function of discriminability. *Learning and Motivation, 4,* 11-27.

Miles, F., & Meehan, J. W. (1995). Visual discrimination of pigmented skin lesions. *Health Psychology, 14,* 171-177.

Miller, D. R., Geller, A. C., Wyatt, S. W., Halpern, A., Howell, J., Cockerell, C., . . . Rosenthal, L. (1996). Melanoma awareness and self-examination practices: results of a United States survey. *Journal of the American Academy of Dermatology, 34,* 962-970.

Moore, M. M., Geller, A. C., Zhang, Z., Hayes, B. B., Bergstrom, K., Graves, J. E., . . . Miller, D. R. (2006). Skin cancer examination teaching in US medical education. *Archives of Dermatology, 142,* 439.

Muhn, C. Y., From, L., & Glied, M. (2000). Detection of artificial changes in mole size by skin self-examination. *Journal of the American Academy of Dermatology, 42,* 754-759.

Newlin, R. J., Rodgers, J. P., & Thomas, D. R. (1979). Two determinants of the peak shift in human voluntary stimulus generalization. *Perception & Psychophysics, 25,* 478-486.

Pennypacker, H., Bloom, H., Criswell, E. L., Neelakantan, P., Goldstein, M. K., & Stein, G. H. (1982). Toward an effective technology of instruction in breast self-examination. *International Journal of Mental Health, 11,* 98-116.

Pierrel, R., & Sherman, J. G. (1960). Generalization of auditory intensity following discrimination training. *Journal of the Experimental Analysis of Behavior, 3,* 313-322.

Reynolds, G. S. (1961). Behavioral contrast. *Journal of the Experimental Analysis of Behavior, 4*, 57-71.

Richard, M. A., Grob, J. J., Avril, M. F., Delaunay, M., Gouvernet, J., Wolkenstein, P., . . . Dalac, S. (2000). Delays in diagnosis and melanoma prognosis (I): The role of patients. *International Journal of Cancer, 89*, 271-279.

Rilling, M., Askew, H. R., Ahlskog, J. E., & Kramer, T. J. (1969). Aversive properties of the negative stimulus in a successive discrimination. *Journal of the Experimental Analysis of Behavior, 12*, 917-932.

Robinson, J. K., & McGaghie, W. C. (1996). Skin cancer detection in a clinical practice examination with standardized patients. *Journal of the American Academy of Dermatology, 34*, 709-711.

Robinson, J. K., & Turrisi, R. (2006). Skills training to learn discrimination of ABCDE criteria by those at risk of developing melanoma. *Archives of Dermatology, 142*, 447.

Saunders, K. J., Pilgrim, C. A., & Pennypacker, H. S. (1986). Increased proficiency of search in breast self-examination. *Cancer, 58*, 2531-2537.

Schwartz, J. L., Wang, T. S., Hamilton, T. A., Lowe, L., Sondak, V. K., & Johnson, T. M. (2002). Thin primary cutaneous melanomas. *Cancer, 95*, 1562-1568.

Singh, S. D., Ajani, U. A., Johnson, C. J., Roland, K. B., Eide, M., Jemal, A., . . . Ekwueme, D. U. (2011). Association of cutaneous melanoma incidence with area-based socioeconomic indicators–United States, 2004-2006. *Journal of the American Academy of Dermatology, 65*, S58-S68.

Spence, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review, 44*, 430-444.

Spetch, M. L., Cheng, K., & Clifford, C. W. (2004). Peak shift but not range effects in recognition of faces. *Learning and Motivation, 35,* 221-241.

Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*: Brookings Inst Press.

Temoshok, L., Diclemente, R. J., Sweet, D. M., Blois, M. S., & Sagebiel, R. W. (1984). Factors related to patient delay in seeking medical attention for cutaneous malignant melanoma. *Cancer, 54,* 3048-3053.

Terrace, H. S. (1964). Wavelength generalization after discrimination learning with and without errors. *Science, 144,* 78-80.

Terrace, H. S. (1966a). Behavioral contrast and the peak shift: Effects of extended discrimination training. *Journal of the Experimental Analysis of Behavior, 9,* 613-617.

Terrace, H. S. (1966b). Discrimination learning and inhibition. *Science, 154,* 1677-1680.

Terrace, H. S. (1968). Discrimination learning, the peak shift, and behavioral contrast. *Journal of the Experimental Analysis of Behavior, 11,* 727-741.

Thomas, D. R. (1993). A model for adaptation-level effects on stimulus generalization. *Psychological Review, 100,* 658-673.

Thomas, D. R., & Jones, C. G. (1962). Stimulus generalization as a function of the frame of reference. *Journal of Experimental Psychology, 64,* 77-80.

Thomas, D. R., Lusky, M., & Morrison, S. (1992). A comparison of generalization functions and frame of reference effects in different training paradigms. *Perception & Psychophysics, 51,* 529-540.

Thomas, D. R., Mood, K., Morrison, S., & Wiertelak, E. (1991). Peak shift revisited: A test of alternative interpretations. *Journal of Experimental Psychology: Animal Behavior Processes, 17*, 130-140.

Thomas, D. R., Strub, H., & Dickson, J. F. (1974). Adaptation level and the central tendency effect in stimulus generalization. *Journal of Experimental Psychology, 103*, 466-474.

Thomas, D. R., Svinicki, M. D., & Vogt, J. (1973). Adaptation level as a factor in human discrimination learning and stimulus generalization. *Journal of Experimental Psychology, 97*, 210-219.

Thomas, L., Tranchand, P., Berard, F., Secchi, T., Colin, C., & Moulin, G. (1998). Semiological value of ABCDE criteria in the diagnosis of cutaneous pigmented tumors. *Dermatology, 197*, 11-17.

Verbeek, E. L., Spetch, M. L., Cheng, K., & Clifford, C. W. (2006). Determinants of range effects in face recognition. *Learning & Behavior, 34*, 229-240.

Wakita, M. (2004). Categorical perception of orientation in monkeys. *Behavioural Processes, 67*, 263-272.

Watson, M., Johnson, C. J., Chen, V. W., Thomas, C. C., Weir, H. K., Sherman, R., . . . Saraiya, M. (2011). Melanoma surveillance in the United States: overview of methods. *Journal of the American Academy of Dermatology, 65*, S6-S16.

Weinstock, M. A., Martin, R. A., Risica, P. M., Berwick, M., Lasater, T., Rakowski, W., . . . Dubé, C. E. (1999). Thorough skin examination for the early detection of melanoma. *American Journal of Preventive Medicine, 17*, 169-175.

Weinstock, M. A., Risica, P. M., Martin, R. A., Rakowski, W., Smith, K. J., Berwick, M., . . . Lasater, T. (2004). Reliability of assessment and circumstances of performance of

thorough skin self-examination for the early detection of melanoma in the Check-It-Out Project. *Preventive Medicine, 38,* 761-765.

Wheatley, K. L., & Thomas, D. R. (1974). Relative and absolute density of reinforcement as factors influencing the peak shift. *Journal of the Experimental Analysis of Behavior, 22,* 409-418.

White, K. G., & Thomas, D. R. (1979). Postdiscrimination stimulus generalization in humans: An extension of Galizio and Baron. *Learning & Behavior, 7,* 564-565.

Wolff, T., Tai, E., & Miller, T. (2009). Screening for skin cancer: An update of the evidence for the US Preventive Services Task Force. *Annals of Internal Medicine, 150,* 194-198.

Tables

Table 1

Area Means for Each Group in Study 1; One-Tailed U *Tests were Used for Comparing Groups*

	Area		Comparison			
	Left of S50	Right of S50	Left of S50		Right of S50	
Group	*M* (*SEM*)	*M* (*SEM*)	S- 1	S- 99	S- 1	S- 99
Control	71.31 (13.26)	132.6 (14.89)	0.0089**	0.0943	0.0514	0.0050**
S- 1	23.19 (9.446)	180.7 (23.86)		0.0016**		0.0027**
S- 99	117.7 (23.64)	62.56 (17.39)				

* $p < .05$, ** $p < .01$

Table 2

Slope of Adjacent Stimuli for Study 1; a Two-Way ANOVA was Used for Comparisons Between Groups

Adjacent Stimuli	Comparison		
	Control vs. S- 1	Control vs. S- 99	S- 1 vs. S- 99
1-8	0.8873	0.8873	0.7769
8-15	> 0.9999	0.8873	0.8873
15-22	> 0.9999	0.3956	0.3956
22-29	> 0.9999	0.2576	0.2576
29-36	0.0663	0.6709	0.024*
36-43	0.0481*	0.6709	0.0165*
43-50	0.571	< 0.0001**	0.0003**
50-57	0.002**	0.002**	< 0.0001**
57-64	0.1573	> 0.9999	0.1573
64-71	0.0031**	0.0005**	0.571
71-78	0.8873	0.571	0.4789
78-85	0.6709	0.7769	0.4789
85-92	0.3956	0.8873	0.3218
92-99	0.8873	> 0.9999	0.8873

* $p < .05$, ** $p < .01$

Table 3

Signal Detection Theory Indices of Discriminability for Each Group in Study 1; a Kruskal-Wallis Test was Used for Comparisons Between Groups

Index	Control	S- 1	S- 99	Control vs. S- 1	Control vs. S- 99	S- 1 vs. S- 99
Accuracy	0.78	0.72	0.79	ns	ns	ns
Sensitivity	0.82	0.39	0.68	*	ns	ns
Specificity	0.98	0.94	0.97	*	ns	ns
Precision	0.25	0.14	0.26	ns	ns	ns

* $p < .05$

Table 4

Monte Carlo Simulations to Estimate Probability of Random Responding Based on Observed Ranges of "Yes" Responses in Study 1 when Numbers of "Yes" Responses were Unconstrained and Constrained to Actual Totals Observed

			Unconstrained (n = 10,000)	Constrained (n = 10,766)
Stimulus	Criterion Frequency of Responding	No. of Criteria Met	Probability	Probability
		0	0	0
1	0	1	1	1
8	0	2	0.9903	1
15	0	3	0.8472	0.99981
22	0	4	0.3815	0.99350
29	0	5	0.0320	0.93414
36	0 to 7	6	0.0022	0.73398
43	0 to 9	7	0.0001	0.43294
50	3 to 9	8	0	0.17750
57	4 to 9	9	0	0.05424
64	1 to 9	10	0	0.01096
71	0 to 5	11	0	0.00149
78	0 to 2	12	0	0.00009
85	0 to 1	13	0	0
92	0	14	0	0
99	0	15	0	0

Table 5

Area Means for Each Group in Study 2; One-Tailed U *Tests were Used for Comparisons Between Discrimination Training Groups and the Control Group; a Kruskal-Wallis Test was Used for Comparisons Between Discrimination Training Groups*

	Area		Comparison					
	S1 - S50	S50 - S99	S1 - S50			S50 - S99		
Group	*M* (*SEM*)	*M* (*SEM*)	S- 64	S- 78	S- 92	S- 64	S- 78	S- 92
Control	52.5 (10.360)	125.5 (13.038)	0.1520	0.4883	0.0184*	0.3692	0.3167	0.0787
S- 64	63.3 (12.076)	105.0 (24.055)		> 0.9999	0.3767		> 0.9999	0.9034
S- 78	68.5 (22.924)	113.5 (9.504)			0.1277			0.5886
S- 92	131.0 (26.763)	77.5 (36.256)						

* $p < .05$

Table 6

Slope of Adjacent Stimuli for Study 2; a Two-Way ANOVA was Used for Comparisons Between Groups

	Comparison					
Adjacent Stimuli	Control S- 64	Control S- 78	Control S- 92	S- 64 vs. S- 78	S- 64 vs. S- 92	S- 78 vs. S- 92
1-8	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999	> 0.9999
8-15	> 0.9999	> 0.9999	0.7553	> 0.9999	0.7553	0.7553
15-22	> 0.9999	0.8762	0.213	0.8762	0.213	0.2757
22-29	0.8762	0.5331	0.1198	0.6402	0.1613	0.35
29-36	0.35	0.7553	0.0621	0.5331	0.35	0.1198
36-43	0.5331	0.7553	0.0434*	0.7553	0.1613	0.0871
43-50	0.5331	0.1613	< 0.0001**	0.436	0.0002**	0.0033**
50-57	0.35	0.35	0.0434*	0.0621	0.2757	0.0033**
57-64	0.2757	0.5331	0.0297*	0.6402	0.2757	0.1198
64-71	0.35	0.0621	0.7553	0.35	0.5331	0.1198
71-78	0.8762	> 0.9999	0.5331	0.8762	0.436	0.5331
78-85	0.436	0.8762	0.436	0.5331	> 0.9999	0.5331
85-92	0.5331	0.5331	0.5331	> 0.9999	> 0.9999	> 0.9999
92-99	0.7553	0.8762	0.7553	0.6402	> 0.9999	0.6402

* $p < .05$, ** $p < 0.01$

Signal Detection Theory Indices of Discriminability for Each Group in Study 2; a Kruskal-Wallis Test was Used for Comparisons Between Groups

Index	Control	S- 64	S- 78	S- 92	Control vs. S- 64	Control vs. S-78	Control vs. S- 92	S- 64 vs. S- 78	S- 64 vs. S-92	S- 78 vs. S-92
Accuracy	0.84	0.85	0.83	0.79	ns	ns	ns	ns	ns	ns
Sensitivity	0.75	0.73	0.68	0.60	ns	ns	ns	ns	ns	ns
Specificity	0.98	0.98	0.97	0.97	ns	ns	ns	ns	ns	ns
Precision	0.26	0.29	0.24	0.20	ns	ns	ns	ns	ns	ns

Table 8

Area Means for Each Group in Study 3; a Two-Tailed U *Tests was Used for the Comparison Between the Control Groups; One-Tailed* U *Tests were Used for Comparisons Between the +50/-78/-99 Group and the Control Groups*

	Area		Comparison			
	S1 - S50	S50 - S99	S1 - S50		S50 - S99	
Group	*M* (*SEM*)	*M* (*SEM*)	-1/+50/-99	+50/-78/-99	-1/+50/-99	+50/-78/-99
+50	60.50 (8.799)	109.5 (19.18)	0.0128*	0.0691	0.9289	0.3016
-1/+50/-99	96.50 (8.431)	110.5 (11.68)		0.4889		0.3447
+50/-78/-99	102.0 (17.84)	94.00 (27.43)				

* $p < .05$

Table 9

Slope of Adjacent Stimuli for Study 3; a Two-Way ANOVA was Used for Comparisons Between Groups

Adjacent Stimulus Pair	Comparison		
	+50 vs. -1/+50/-99	+50 vs. +50/-78/-99	-1/+50/-99 vs. +50/-78/-99
1-8	> 0.9999	> 0.9999	> 0.9999
8-15	> 0.9999	> 0.9999	> 0.9999
15-22	0.7352	> 0.9999	0.7352
22-29	0.8657	0.0637	0.0917
29-36	0.0663	0.6709	> 0.9999
36-43	0.8657	0.0917	0.1288
43-50	0.3981	0.0026**	0.0287*
50-57	0.0432*	0.2371	0.3981
57-64	0.3107	0.0637	0.3981
64-71	0.3107	0.3107	0.0432*
71-78	0.3107	0.3107	0.0432*
78-85	0.8657	0.7352	0.612
85-92	0.8657	> 0.9999	0.8657
92-99	> 0.9999	> 0.9999	> 0.9999

* $p < .05$, ** $p < .01$

Table 10

Signal Detection Theory Indices of Discriminability for Each Group in Study 3; a Kruskal-Wallis Test was Used for Comparisons Between Groups

Index	+50	-1/+50/-99	+50/-78/-99	+50 vs. -1/+50/-99	+50 vs. +50/-78/-99	-1/+50/-99 vs. +50/-78/-99
Accuracy	0.78	0.72	0.79	ns	ns	ns
Sensitivity	0.82	0.39	0.68	ns	ns	ns
Specificity	0.98	0.94	0.97	ns	ns	ns
Precision	0.25	0.14	0.26	ns	ns	ns

Figures

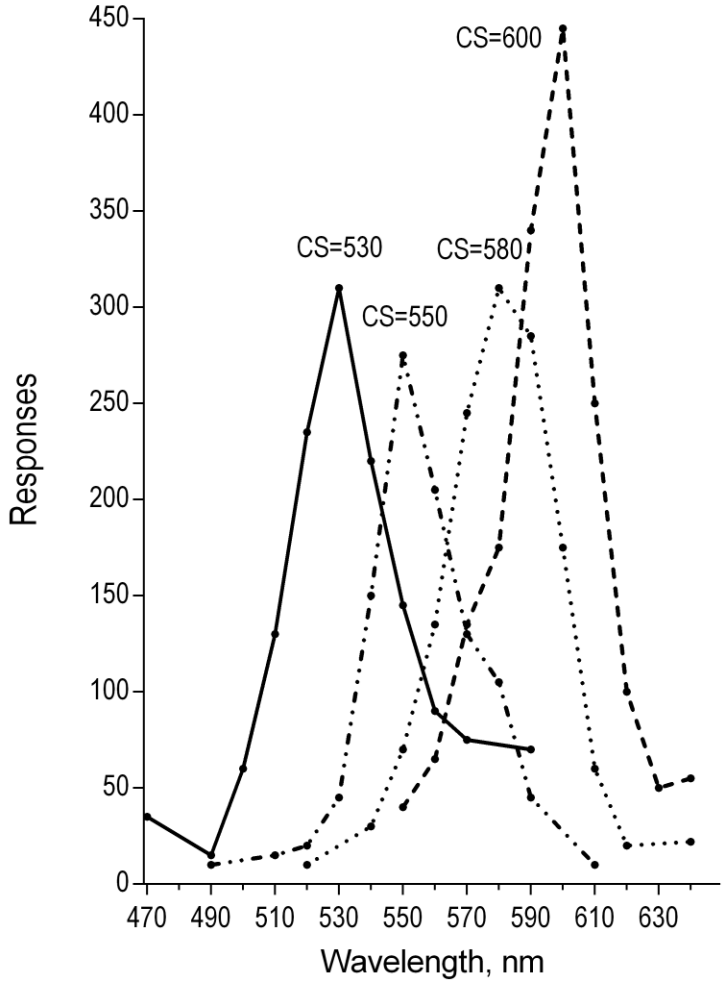

Figure 1. Mean generalization gradient, first test (adapted from Guttman & Kalish, 1956).

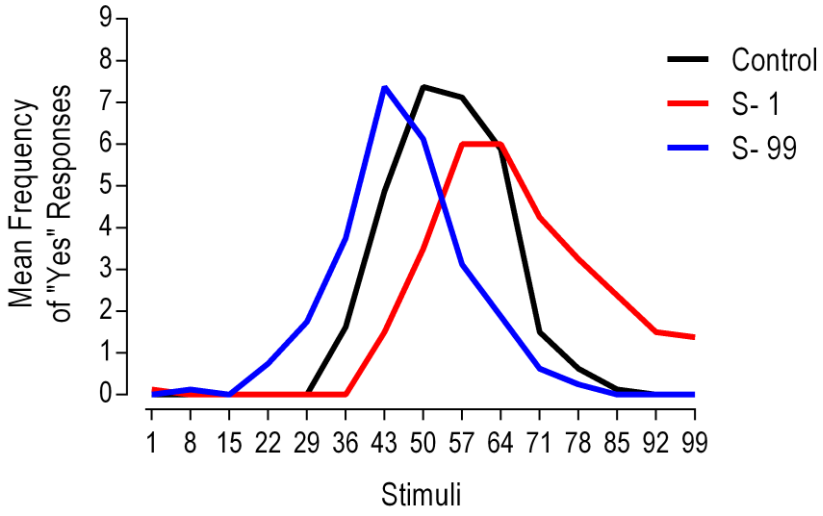

Figure 2. Average gradients for Control, S- 99, and S- 1 groups for Study 1. Stimuli are scaled to the x-axis, mean frequency of "yes" responses are scaled to the y-axis.

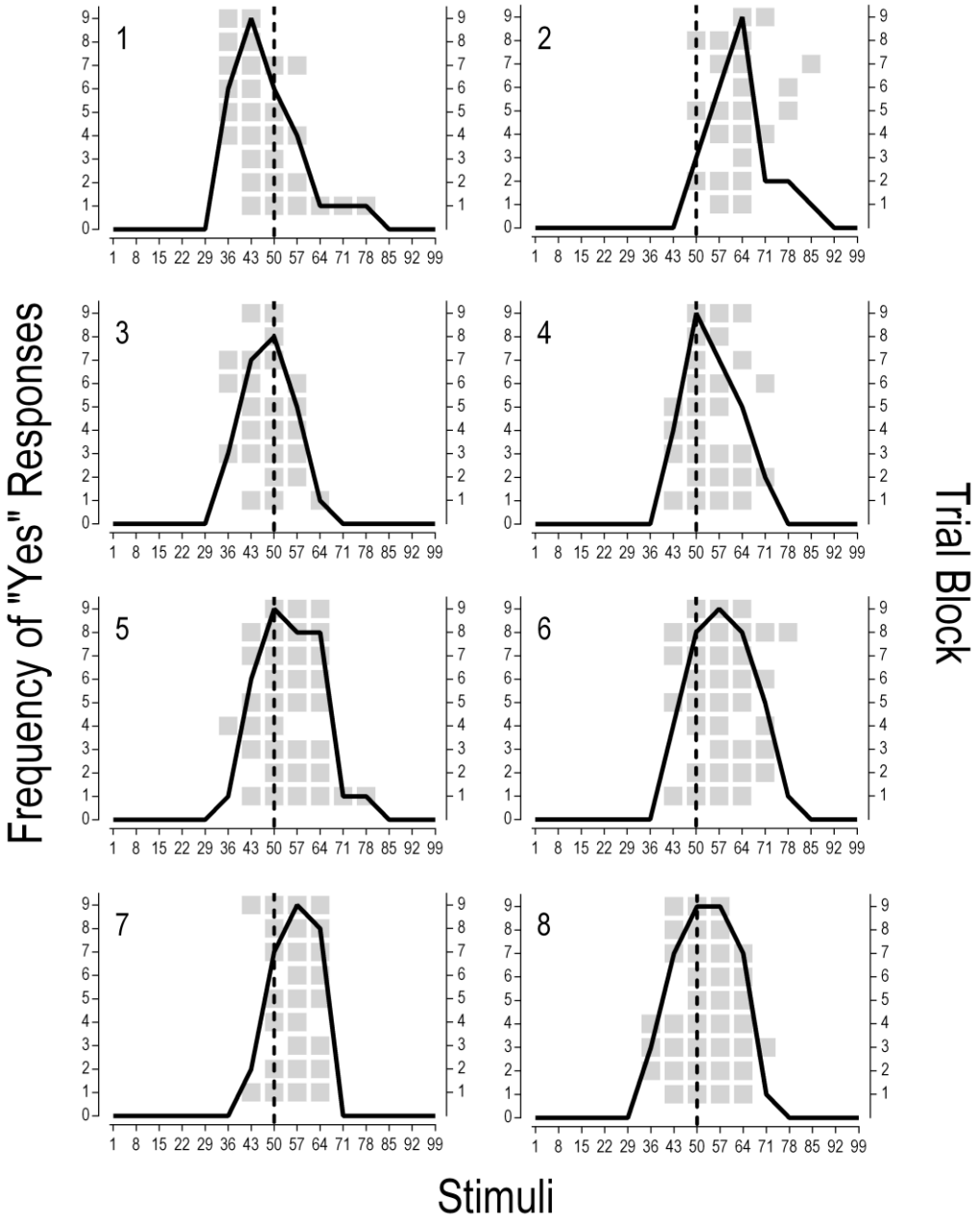

Figure 3. Individual generalization gradients for the Control group for Study 1. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

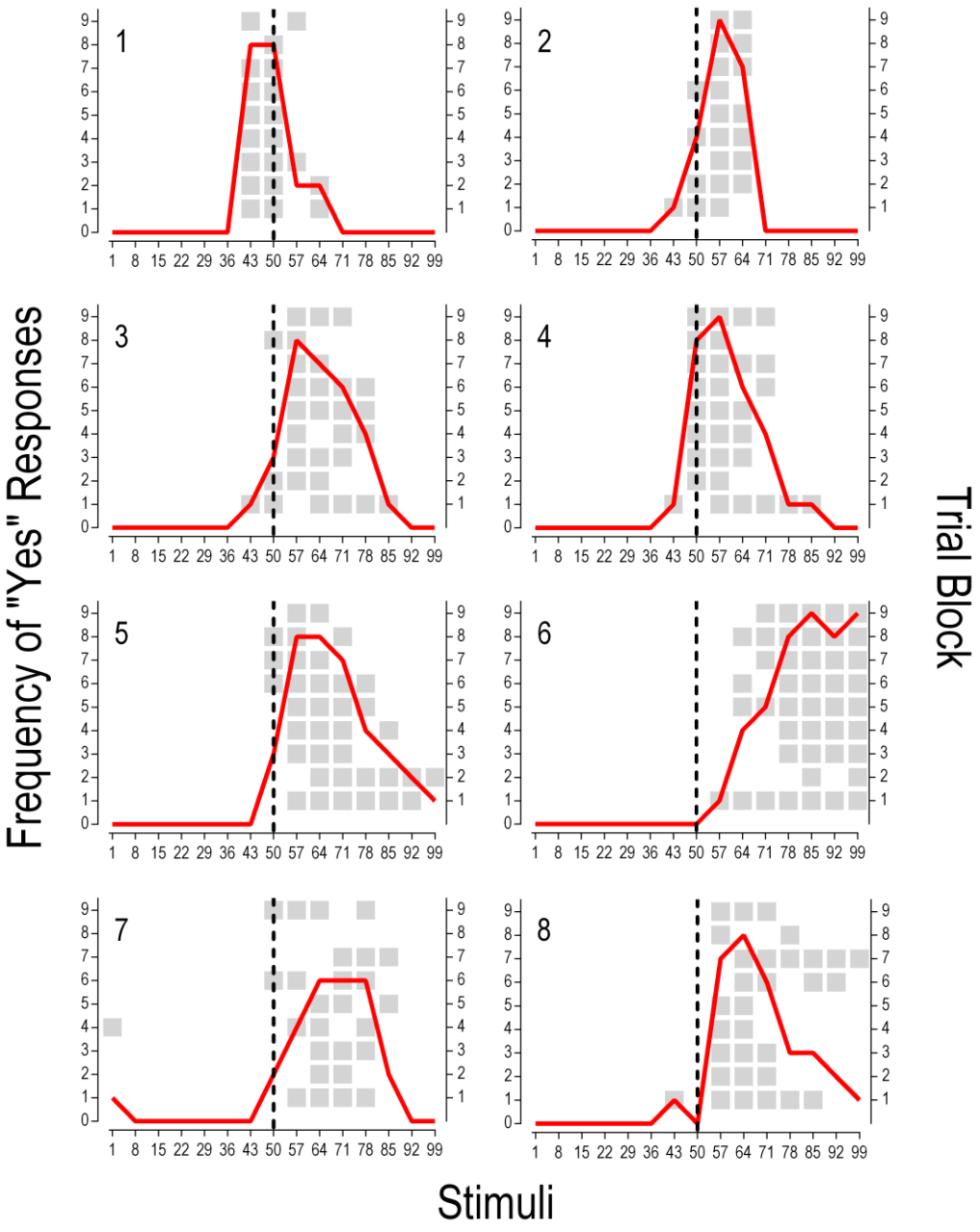

Figure 4. Individual generalization gradients for the S- 1 group for Study 1. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

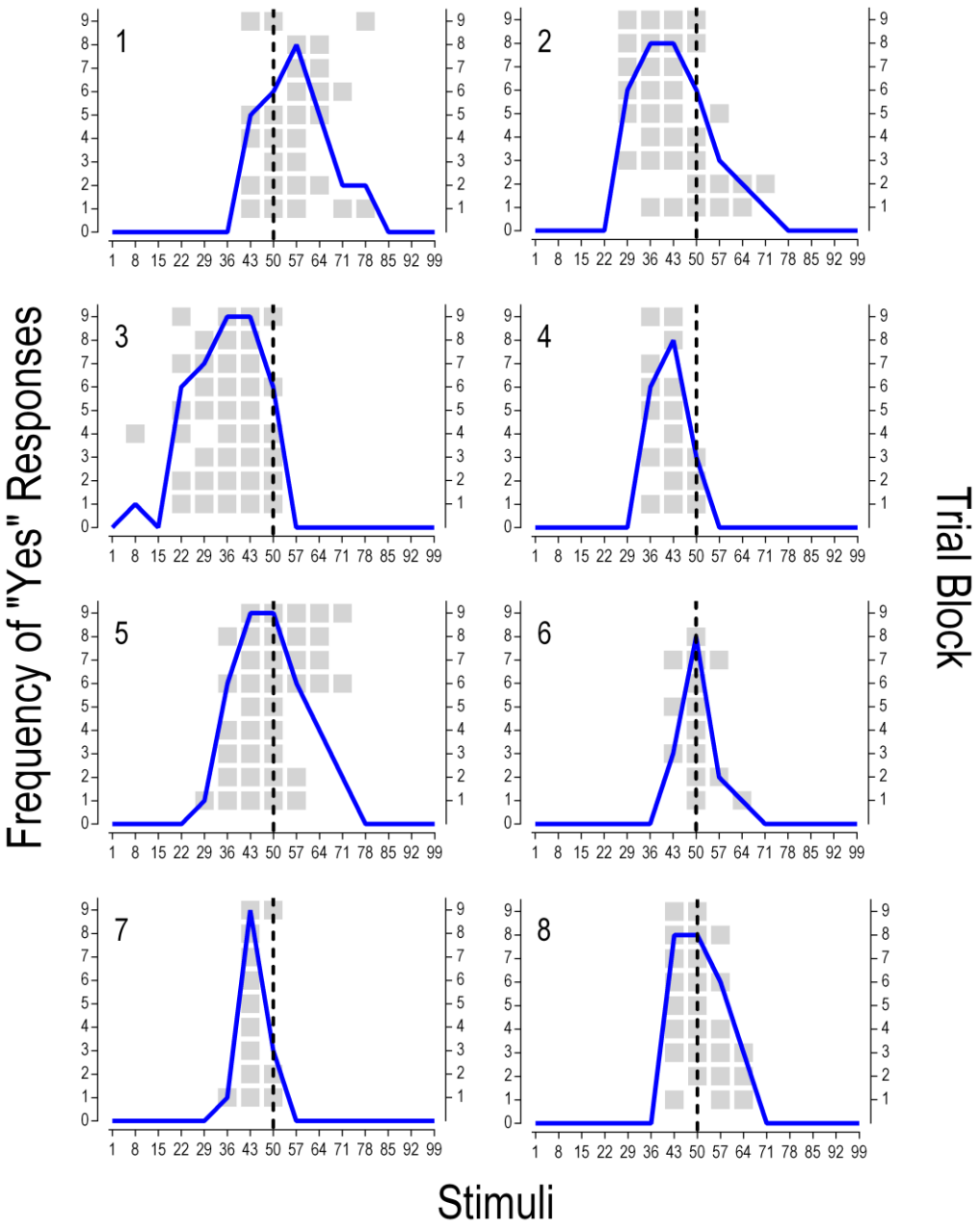

Figure 5. Individual generalization gradients for the S- 99 group for Study 1. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

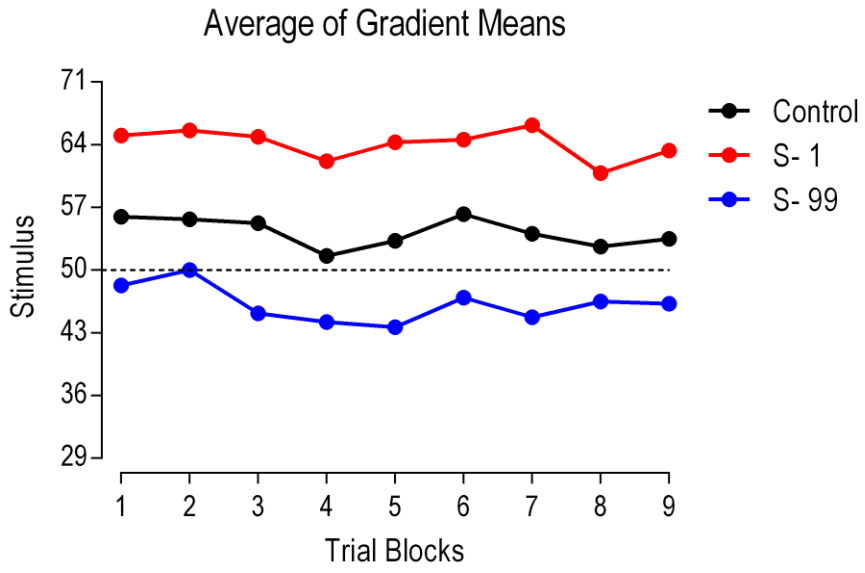

Figure 6. Average gradient means across trial blocks during the generalization test for Study 1. Trial blocks are scaled to the x-axis; stimulus value is scaled to the y-axis.

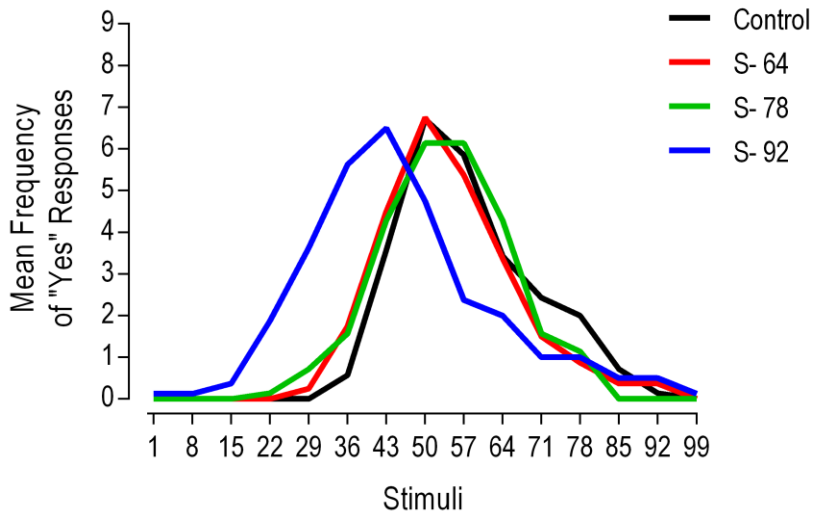

Figure 7. Average gradients for the control, S- 64, S- 78, and S- 92 groups for Study 2. Stimuli are scaled to the x-axis, mean frequency of "yes" responses are scaled to the y-axis.

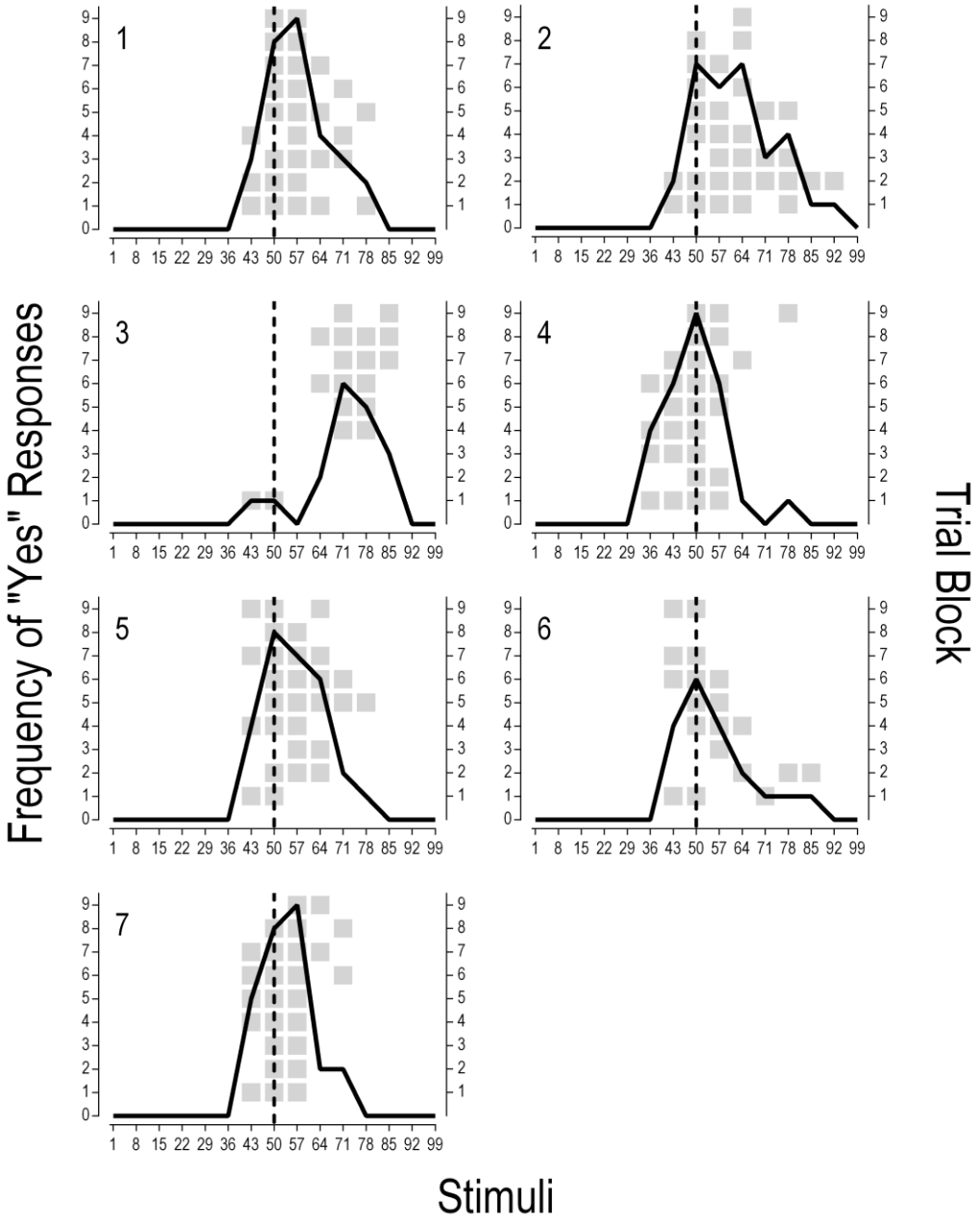

Figure 8. Individual generalization gradients for the control group for Study 2. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

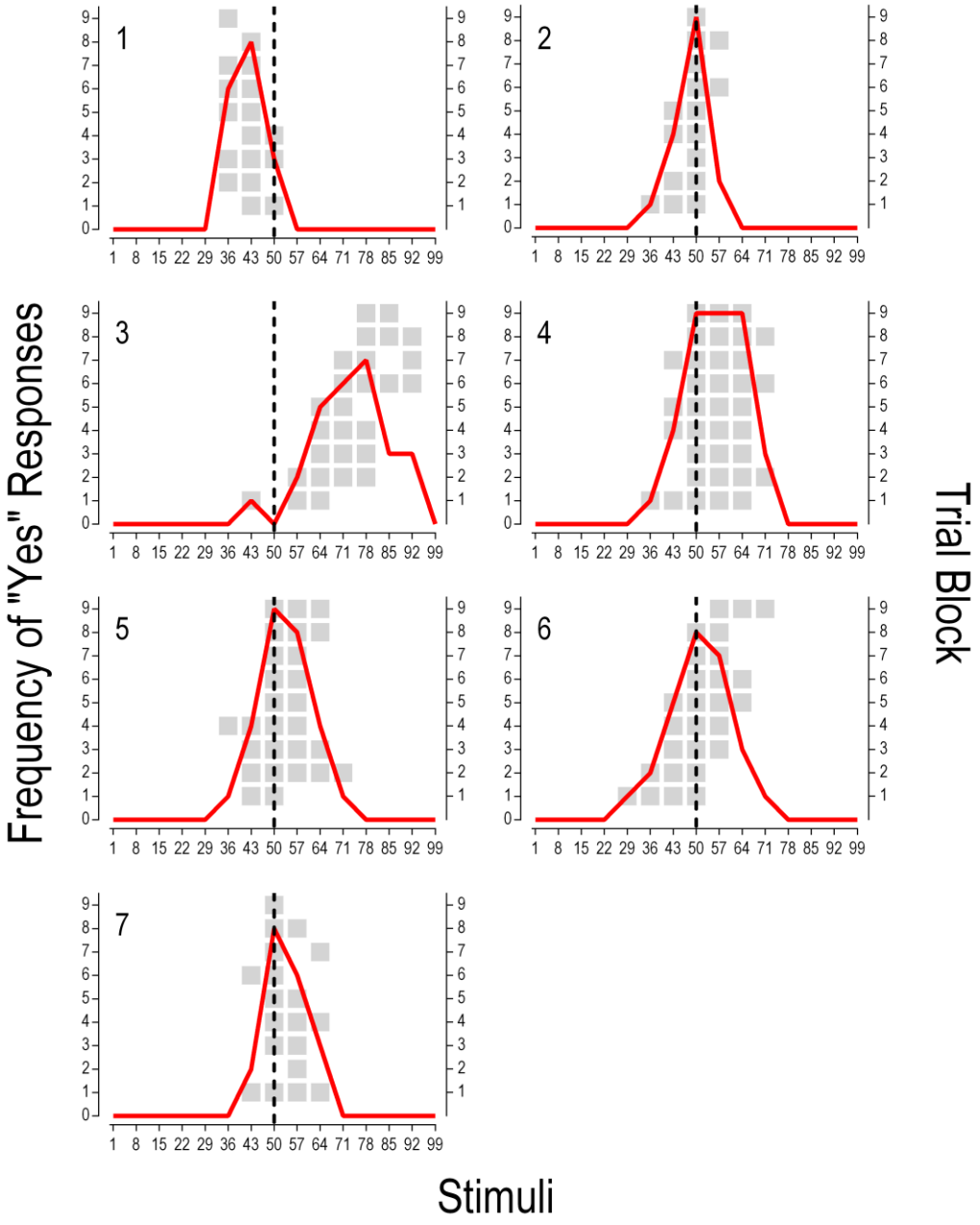

Figure 9. Individual generalization gradients for the S- 64 group for Study 2. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

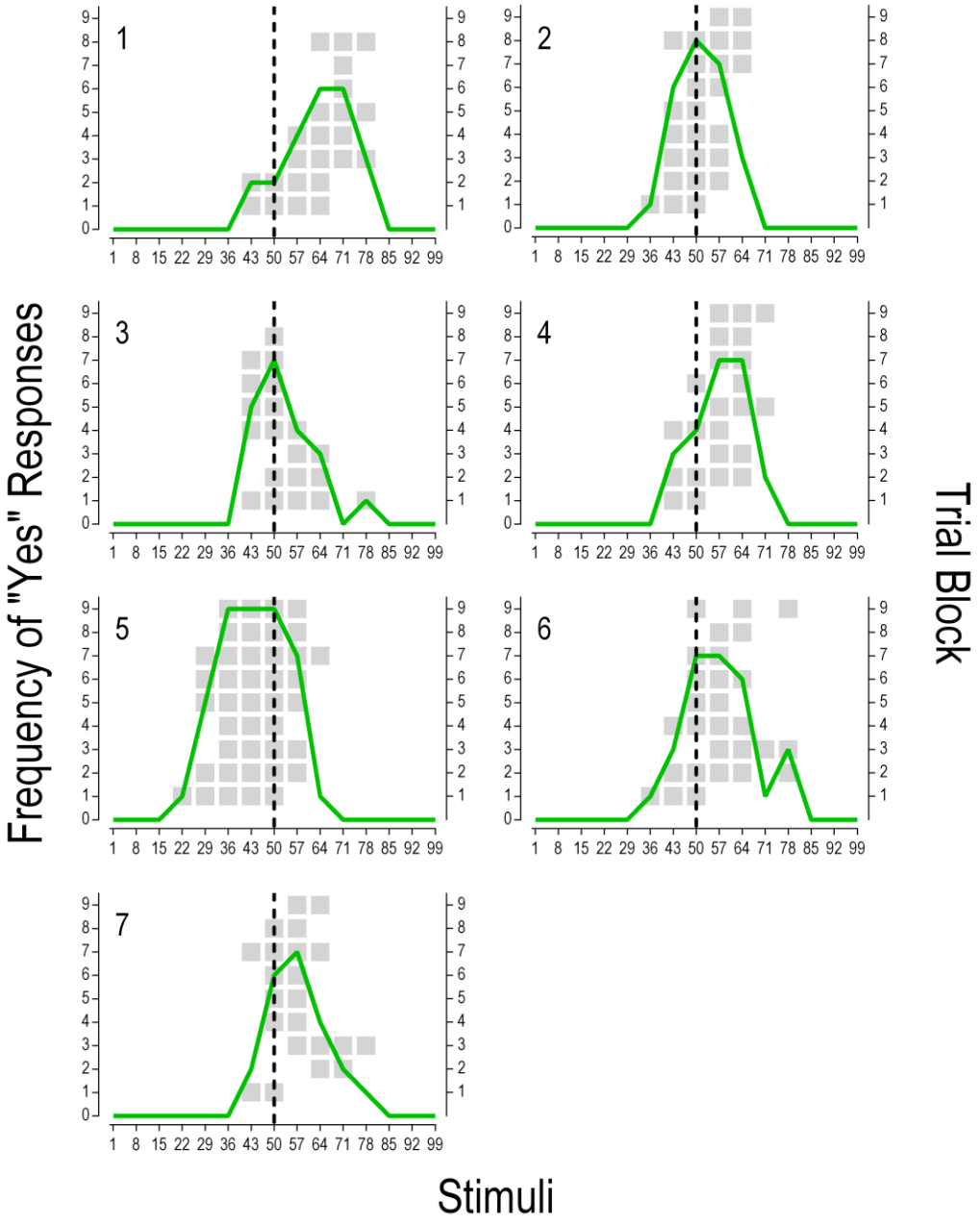

Figure 10. Individual generalization gradients for the S- 78 group for Study 2. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

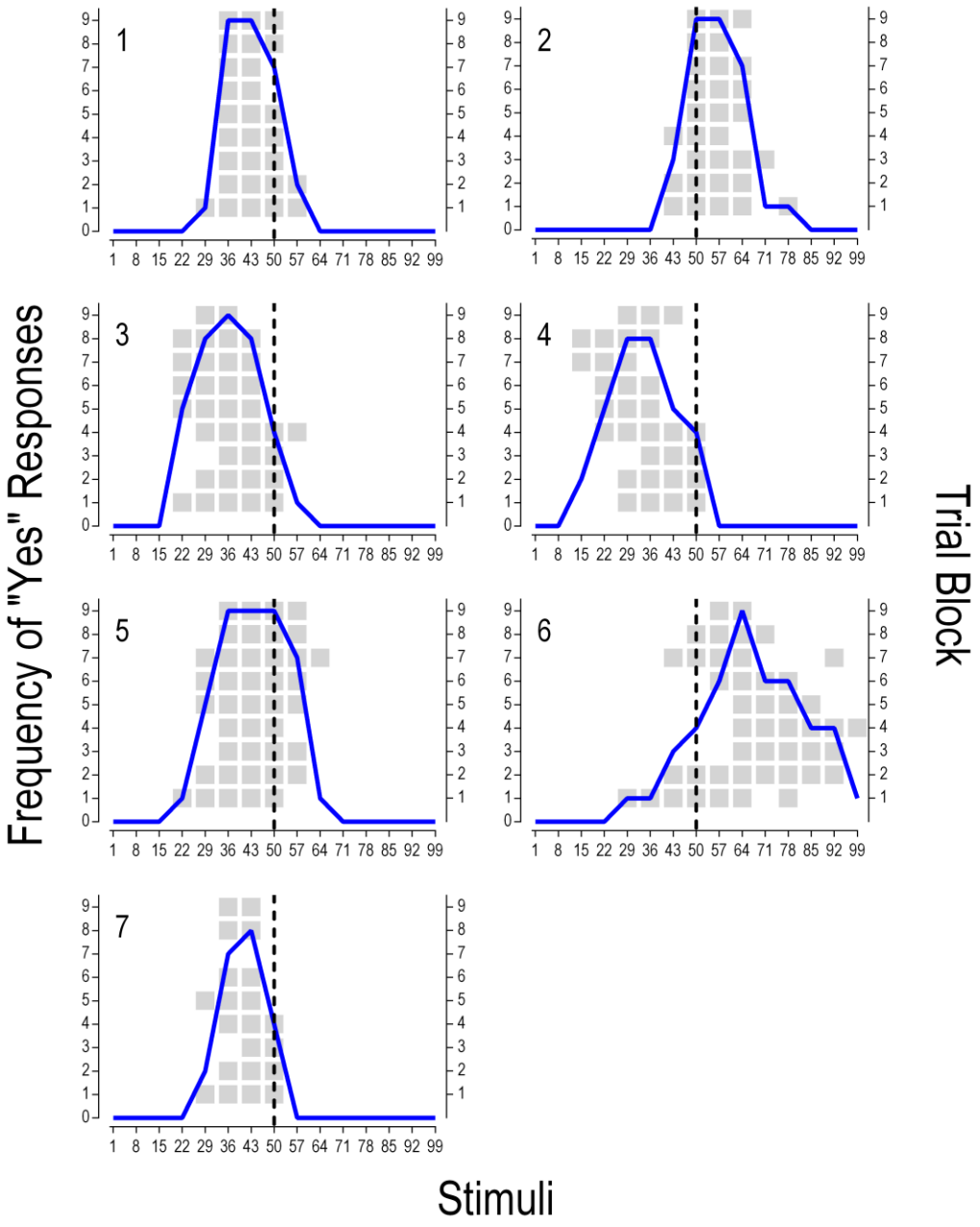

Figure 11. Individual generalization gradients for the S- 92 group for Study 2. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

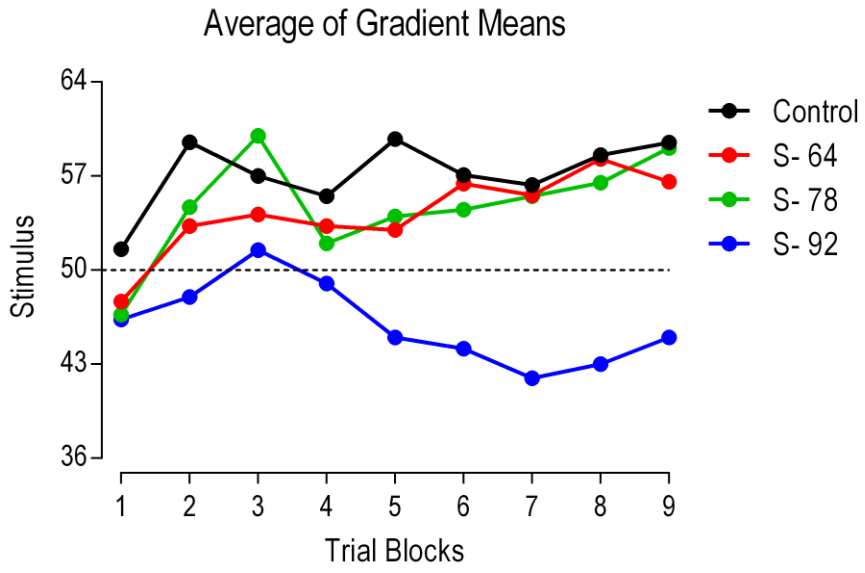

Figure 12. Average gradient means across trial blocks during the generalization test for Study 2. Trial blocks are scaled to the x-axis; stimulus value is scaled to the y-axis.

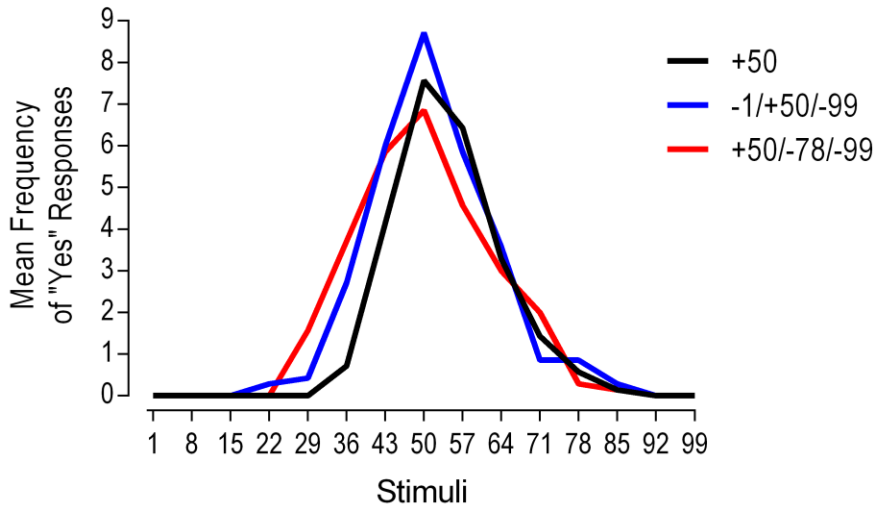

Figure 13. Average gradients for +50, -1/+50/-99, and +50/-78/-99 groups for Study 3. Stimuli are scaled to the x-axis, mean frequency of "yes" responses are scaled to the y-axis.

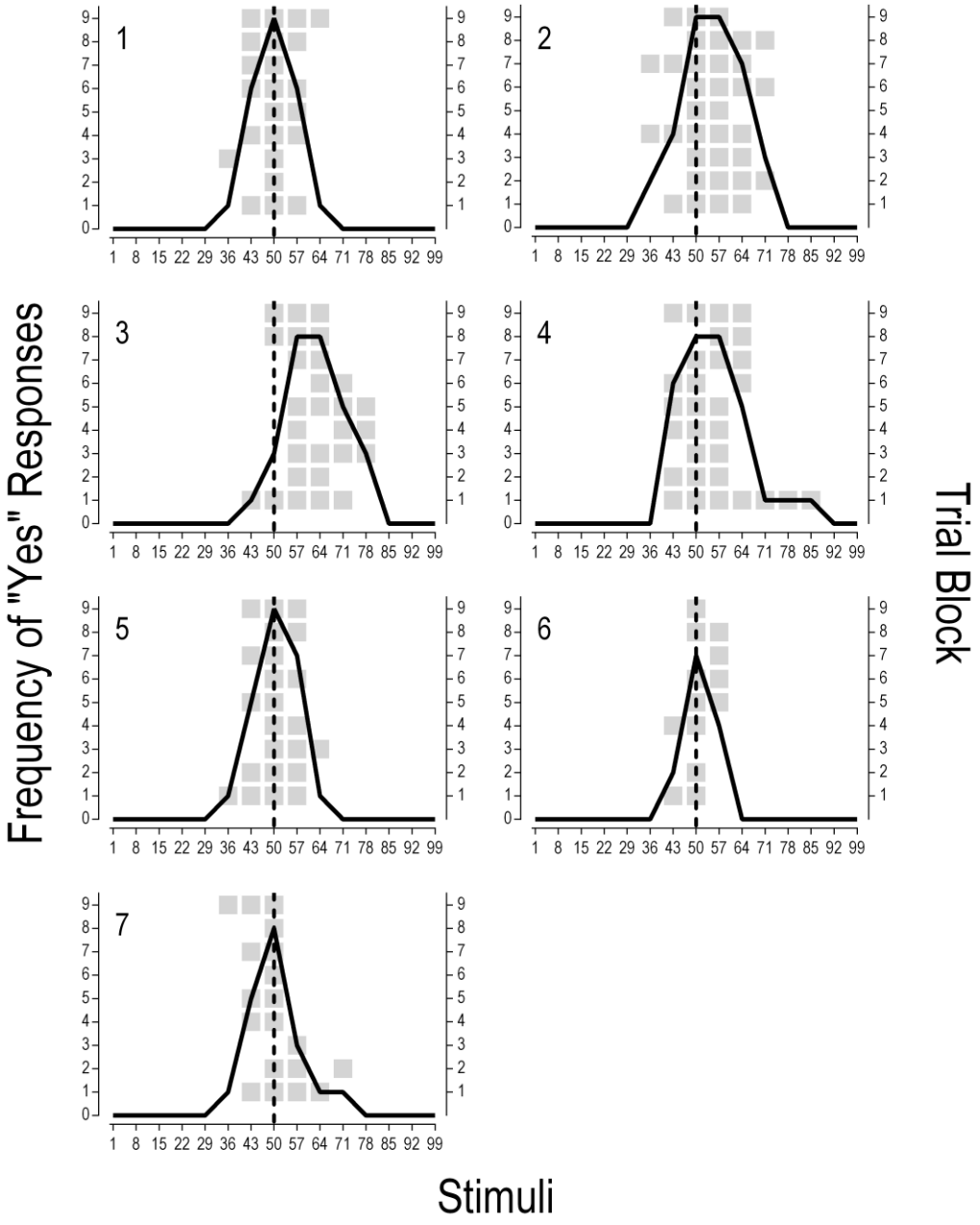

Figure 14. Individual generalization gradients for the +50 group for Study 3. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

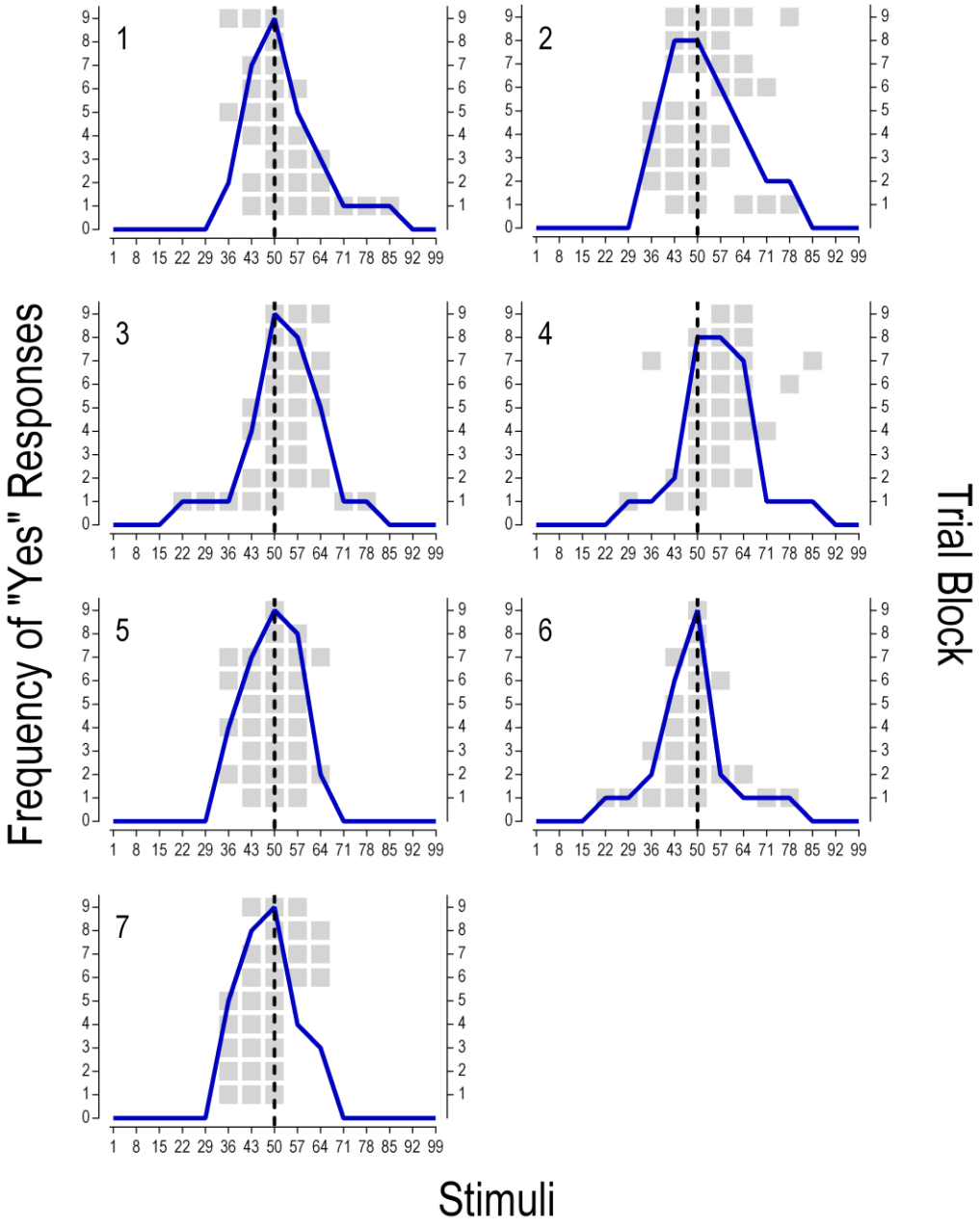

Figure 15. Individual generalization gradients for the -1/+50/-99 group for Study 3. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

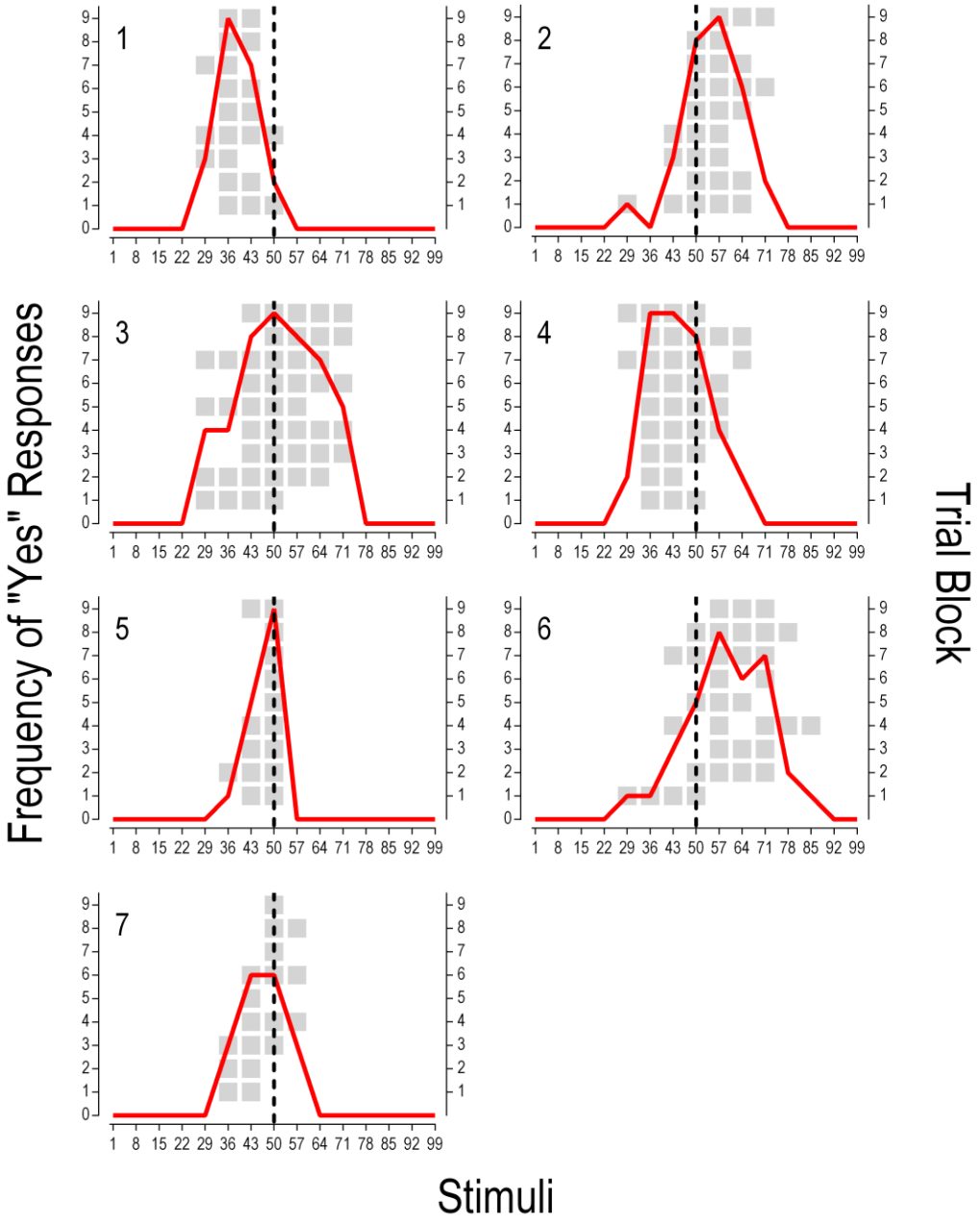

Figure 16. Individual generalization gradients for the +50/-78/-99 group for Study 3. Data paths are scaled to the left y-axis. Each gray block depicts a "yes" response to a stimulus within a 15-stimulus trial block, scaled to the right y-axis.

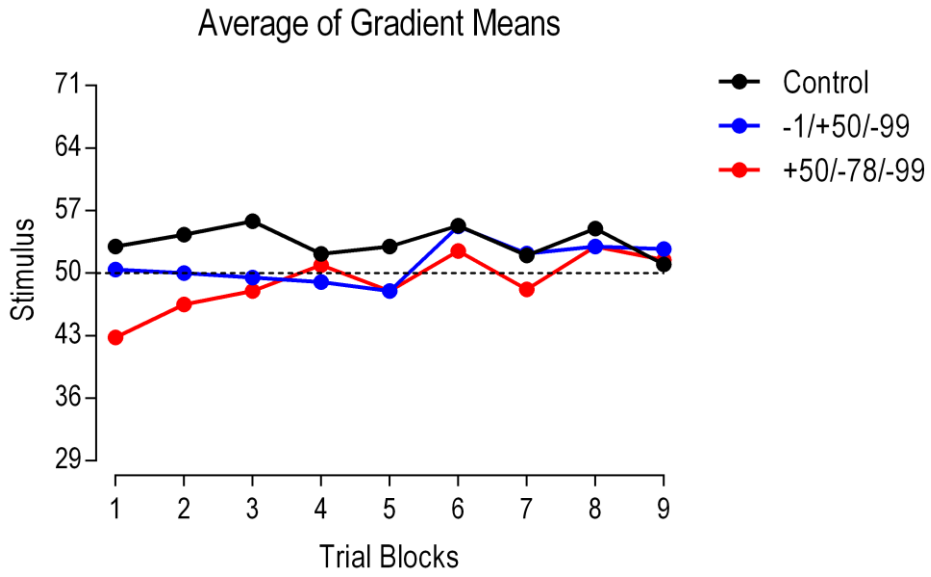

Figure 17. Average gradient means across trial blocks during the generalization test for Study 3. Trial blocks are scaled to the x-axis; stimulus value is scaled to the y-axis.

Appendix A: Data Sheets

Studies 1 & 2: 20-trial training data sheet (1 of 2)

Trial #	Is this the original?		Trial #	Is this the original?	
1	Y ☐	N ☐	46	Y ☐	N ☐
2	Y ☐	N ☐	47	Y ☐	N ☐
3	Y ☐	N ☐	48	Y ☐	N ☐
4	Y ☐	N ☐	49	Y ☐	N ☐
5	Y ☐	N ☐	50	Y ☐	N ☐
6	Y ☐	N ☐	51	Y ☐	N ☐
7	Y ☐	N ☐	52	Y ☐	N ☐
8	Y ☐	N ☐	53	Y ☐	N ☐
9	Y ☐	N ☐	54	Y ☐	N ☐
10	Y ☐	N ☐	55	Y ☐	N ☐
11	Y ☐	N ☐	56	Y ☐	N ☐
12	Y ☐	N ☐	57	Y ☐	N ☐
13	Y ☐	N ☐	58	Y ☐	N ☐
14	Y ☐	N ☐	59	Y ☐	N ☐
15	Y ☐	N ☐	60	Y ☐	N ☐
16	Y ☐	N ☐	61	Y ☐	N ☐
17	Y ☐	N ☐	62	Y ☐	N ☐
18	Y ☐	N ☐	63	Y ☐	N ☐
19	Y ☐	N ☐	64	Y ☐	N ☐
20	Y ☐	N ☐	65	Y ☐	N ☐
	60-sec Break			60-sec Break	
21	Y ☐	N ☐	66	Y ☐	N ☐
22	Y ☐	N ☐	67	Y ☐	N ☐
23	Y ☐	N ☐	68	Y ☐	N ☐
24	Y ☐	N ☐	69	Y ☐	N ☐
25	Y ☐	N ☐	70	Y ☐	N ☐
26	Y ☐	N ☐	71	Y ☐	N ☐
27	Y ☐	N ☐	72	Y ☐	N ☐
28	Y ☐	N ☐	73	Y ☐	N ☐
29	Y ☐	N ☐	74	Y ☐	N ☐
30	Y ☐	N ☐	75	Y ☐	N ☐
31	Y ☐	N ☐	76	Y ☐	N ☐
32	Y ☐	N ☐	77	Y ☐	N ☐
33	Y ☐	N ☐	78	Y ☐	N ☐
34	Y ☐	N ☐	79	Y ☐	N ☐
35	Y ☐	N ☐	80	Y ☐	N ☐
36	Y ☐	N ☐	81	Y ☐	N ☐
37	Y ☐	N ☐	82	Y ☐	N ☐
38	Y ☐	N ☐	83	Y ☐	N ☐
39	Y ☐	N ☐	84	Y ☐	N ☐
40	Y ☐	N ☐	85	Y ☐	N ☐
41	Y ☐	N ☐	86	Y ☐	N ☐
42	Y ☐	N ☐	87	Y ☐	N ☐
43	Y ☐	N ☐	88	Y ☐	N ☐
44	Y ☐	N ☐	89	Y ☐	N ☐
45	Y ☐	N ☐	90	Y ☐	N ☐

Studies 1 & 2: 20-trial training data sheet (2 of 2)

Trial #	Is this the original?		Trial #	Is this the original?	
91	Y ☐	N ☐	136	Y ☐	N ☐
92	Y ☐	N ☐	137	Y ☐	N ☐
93	Y ☐	N ☐	138	Y ☐	N ☐
94	Y ☐	N ☐	139	Y ☐	N ☐
95	Y ☐	N ☐	140	Y ☐	N ☐
96	Y ☐	N ☐	141	Y ☐	N ☐
97	Y ☐	N ☐	142	Y ☐	N ☐
98	Y ☐	N ☐	143	Y ☐	N ☐
99	Y ☐	N ☐	144	Y ☐	N ☐
100	Y ☐	N ☐	145	Y ☐	N ☐
101	Y ☐	N ☐	146	Y ☐	N ☐
102	Y ☐	N ☐	147	Y ☐	N ☐
103	Y ☐	N ☐	148	Y ☐	N ☐
104	Y ☐	N ☐	149	Y ☐	N ☐
105	Y ☐	N ☐	150	Y ☐	N ☐
106	Y ☐	N ☐	151	Y ☐	N ☐
107	Y ☐	N ☐	152	Y ☐	N ☐
108	Y ☐	N ☐	153	Y ☐	N ☐
109	Y ☐	N ☐	154	Y ☐	N ☐
110	Y ☐	N ☐	155	Y ☐	N ☐
	60-sec Break				
111	Y ☐	N ☐			
112	Y ☐	N ☐			
113	Y ☐	N ☐			
114	Y ☐	N ☐			
115	Y ☐	N ☐			
116	Y ☐	N ☐			
117	Y ☐	N ☐			
118	Y ☐	N ☐			
119	Y ☐	N ☐			
120	Y ☐	N ☐			
121	Y ☐	N ☐			
122	Y ☐	N ☐			
123	Y ☐	N ☐			
124	Y ☐	N ☐			
125	Y ☐	N ☐			
126	Y ☐	N ☐			
127	Y ☐	N ☐			
128	Y ☐	N ☐			
129	Y ☐	N ☐			
130	Y ☐	N ☐			
131	Y ☐	N ☐			
132	Y ☐	N ☐			
133	Y ☐	N ☐			
134	Y ☐	N ☐			
135	Y ☐	N ☐			

Study 3: 30-trial training datasheet (1 of 2)

Trial #	Is this the original?		Trial #	Is this the original?	
1	Y ☐	N ☐	46	Y ☐	N ☐
2	Y ☐	N ☐	47	Y ☐	N ☐
3	Y ☐	N ☐	48	Y ☐	N ☐
4	Y ☐	N ☐	49	Y ☐	N ☐
5	Y ☐	N ☐	50	Y ☐	N ☐
6	Y ☐	N ☐	51	Y ☐	N ☐
7	Y ☐	N ☐	52	Y ☐	N ☐
8	Y ☐	N ☐	53	Y ☐	N ☐
9	Y ☐	N ☐	54	Y ☐	N ☐
10	Y ☐	N ☐	55	Y ☐	N ☐
11	Y ☐	N ☐	56	Y ☐	N ☐
12	Y ☐	N ☐	57	Y ☐	N ☐
13	Y ☐	N ☐	58	Y ☐	N ☐
14	Y ☐	N ☐	59	Y ☐	N ☐
15	Y ☐	N ☐	60	Y ☐	N ☐
16	Y ☐	N ☐	61	Y ☐	N ☐
17	Y ☐	N ☐	62	Y ☐	N ☐
18	Y ☐	N ☐	63	Y ☐	N ☐
19	Y ☐	N ☐	64	Y ☐	N ☐
20	Y ☐	N ☐	65	Y ☐	N ☐
21	Y ☐	N ☐	66	Y ☐	N ☐
22	Y ☐	N ☐	67	Y ☐	N ☐
23	Y ☐	N ☐	68	Y ☐	N ☐
24	Y ☐	N ☐	69	Y ☐	N ☐
25	Y ☐	N ☐	70	Y ☐	N ☐
26	Y ☐	N ☐	71	Y ☐	N ☐
27	Y ☐	N ☐	72	Y ☐	N ☐
28	Y ☐	N ☐	73	Y ☐	N ☐
29	Y ☐	N ☐	74	Y ☐	N ☐
30	Y ☐	N ☐	75	Y ☐	N ☐
	60-sec Break			60-sec Break	
31	Y ☐	N ☐	76	Y ☐	N ☐
32	Y ☐	N ☐	77	Y ☐	N ☐
33	Y ☐	N ☐	78	Y ☐	N ☐
34	Y ☐	N ☐	79	Y ☐	N ☐
35	Y ☐	N ☐	80	Y ☐	N ☐
36	Y ☐	N ☐	81	Y ☐	N ☐
37	Y ☐	N ☐	82	Y ☐	N ☐
38	Y ☐	N ☐	83	Y ☐	N ☐
39	Y ☐	N ☐	84	Y ☐	N ☐
40	Y ☐	N ☐	85	Y ☐	N ☐
41	Y ☐	N ☐	86	Y ☐	N ☐
42	Y ☐	N ☐	87	Y ☐	N ☐
43	Y ☐	N ☐	88	Y ☐	N ☐
44	Y ☐	N ☐	89	Y ☐	N ☐
45	Y ☐	N ☐	90	Y ☐	N ☐

Study 3: 30-trial training datasheet (2 of 2)

Trial #	Is this the original?		Trial #	Is this the original?	
91	Y ☐	N ☐	136	Y ☐	N ☐
92	Y ☐	N ☐	137	Y ☐	N ☐
93	Y ☐	N ☐	138	Y ☐	N ☐
94	Y ☐	N ☐	139	Y ☐	N ☐
95	Y ☐	N ☐	140	Y ☐	N ☐
96	Y ☐	N ☐	141	Y ☐	N ☐
97	Y ☐	N ☐	142	Y ☐	N ☐
98	Y ☐	N ☐	143	Y ☐	N ☐
99	Y ☐	N ☐	144	Y ☐	N ☐
100	Y ☐	N ☐	145	Y ☐	N ☐
101	Y ☐	N ☐	146	Y ☐	N ☐
102	Y ☐	N ☐	147	Y ☐	N ☐
103	Y ☐	N ☐	148	Y ☐	N ☐
104	Y ☐	N ☐	149	Y ☐	N ☐
105	Y ☐	N ☐	150	Y ☐	N ☐
106	Y ☐	N ☐	151	Y ☐	N ☐
107	Y ☐	N ☐	152	Y ☐	N ☐
108	Y ☐	N ☐	153	Y ☐	N ☐
109	Y ☐	N ☐	154	Y ☐	N ☐
110	Y ☐	N ☐	155	Y ☐	N ☐
111	Y ☐	N ☐	156	Y ☐	N ☐
112	Y ☐	N ☐	157	Y ☐	N ☐
113	Y ☐	N ☐	158	Y ☐	N ☐
114	Y ☐	N ☐	159	Y ☐	N ☐
115	Y ☐	N ☐	160	Y ☐	N ☐
116	Y ☐	N ☐	161	Y ☐	N ☐
117	Y ☐	N ☐	162	Y ☐	N ☐
118	Y ☐	N ☐	163	Y ☐	N ☐
119	Y ☐	N ☐	164	Y ☐	N ☐
120	Y ☐	N ☐	165	Y ☐	N ☐
	60-sec Break				
121	Y ☐	N ☐			
122	Y ☐	N ☐			
123	Y ☐	N ☐			
124	Y ☐	N ☐			
125	Y ☐	N ☐			
126	Y ☐	N ☐			
127	Y ☐	N ☐			
128	Y ☐	N ☐			
129	Y ☐	N ☐			
130	Y ☐	N ☐			
131	Y ☐	N ☐			
132	Y ☐	N ☐			
133	Y ☐	N ☐			
134	Y ☐	N ☐			
135	Y ☐	N ☐			

Appendix B: Stimuli

Appendix C: Instructions & Trial Format