December 2022

Stimulus Discrimination Training And Its Effects On Generalization Gradients

Christopher Veenstra

STIMULUS DISCRIMINATION TRAINING AND ITS EFFECTS ON GENERALIZATION GRADIENTS

Christopher Colley Veenstra
Bachelor of Science, University of North Dakota, 2020
Bachelor of Art, University of North Dakota, 2020

A thesis

submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of requirements

for the degree of

Master of Science

Grand Forks, North Dakota
December 2022

Name: Christopher C. Veenstra

Degree: Master of Science

This document, submitted in partial fulfillment of the requirements for the degree from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

DocuSigned by:
Adam Derenne
2144173C63FD4E6...

Adam Derenne

DocuSigned by:
Joelle Ruthig
DFECC2324D58448...

Joelle Ruthig

DocuSigned by:
Jeffrey Weatherly
0F653400B3F6454...

Jeffrey Weatherly

This document is being submitted by the appointed advisory committee as having met all the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

DocuSigned by:
Chris Nelson
2E0AF088C733403...

Chris Nelson
Dean of the School of Graduate Studies

9/7/2022

Date

PERMISSION

Title	Stimulus Discrimination Training and its Effects on Generalization Gradients
Department	Psychology
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in his absence, by the Chairperson of the department or the dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Christopher C. Veenstra
August 31, 2022

Contents

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Adam Derenne, for all of his guidance during this work. The insights and experience shared have been invaluable, and I continue to find our conversations both exciting and encouraging.

I would also like to thank the faculty and my fellow students for welcoming me into an environment that is so fun, supportive, and challenging.

Last, I would like to thank my parents for their willingness to listen to my bizarre ramblings about stimulus generalization, David Haeselin for his help in beginning this journey in the first place, and Dutch for his persistent reminders to go outside from time to time.

Abstract

Adaptation-level theory has been used to explain how the size and shape of generalization gradients depend on the procedure used to study generalization. However, most of the empirical support for this theory has come from studies using a relatively simple stimulus dimension (e.g., lights that vary in brightness). In addition, much of the research has focused on one specific prediction of adaptation-level theory: that responding during a generalization test should shift towards intermediate values within the tested portion of the stimulus dimension. The present research used a more complex set of stimuli. Specifically, the stimulus dimension was based on bar length, but participants (192 undergraduates) assessed bar length indirectly by estimating the combined length of six parallel bars that were nonidentical in length. The research additionally examined adaptation-level theory predictions for how gradients are affected by the amount of discrimination training that participants receive, and the relative difficulty of the discrimination (S+ and S- were relatively similar in one condition and relatively dissimilar in another). Data collection occurred online.

The research yielded orderly generalization gradients, but not the outcomes predicted by adaptation-level theory. Instead, the generalization tests showed such effects as a progressive decrease in the area under the generalization gradient, greater area under the gradient when the range of tested stimuli was wider (with this difference emerging more rapidly than adaptation-level theory would predict), and pronounced individual differences in the basic form of the generalization gradient. These results suggest the need for additional theories of stimulus generalization.

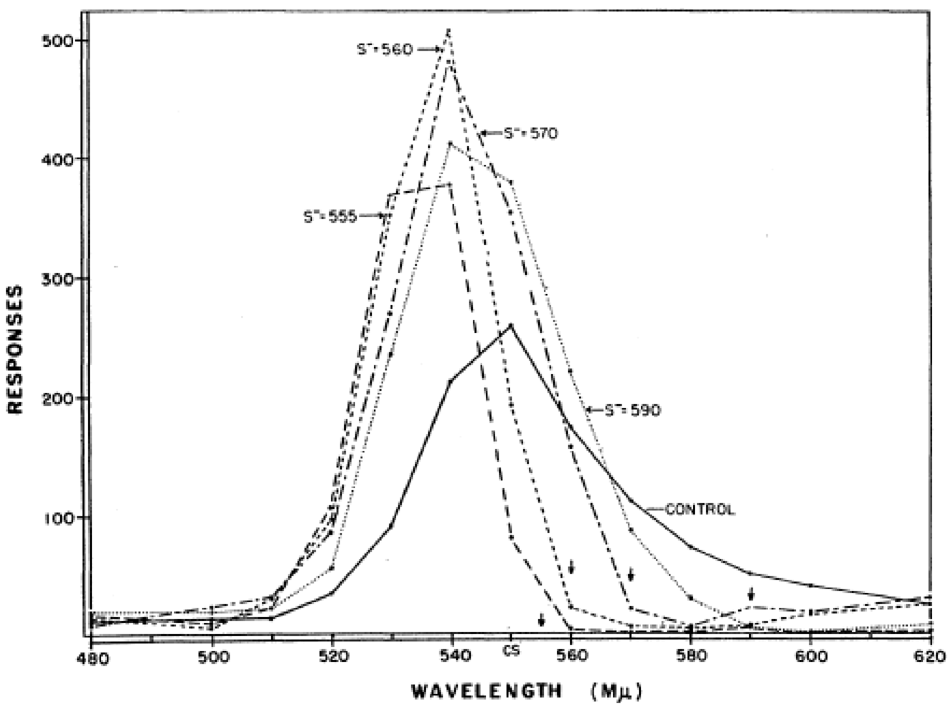
Stimulus Discrimination Training and its Effects on Generalization Gradients

An early development within behaviorism was the ability to quantify and systematically study stimulus generalization. In summary, it was found that organisms that had been trained to provide a particular behavioral response when presented with a specific stimulus (S+) would continue to make the same response to a variety of other related stimuli. The probability of a response could then be found across a range of stimuli, to determine the amount of generalization that was occurring (Guttman & Kalish, 1956). The likelihood of this response was found to be related to the degree of similarity between the S+ and the presented stimulus. Specifically, response rates were found to be highest in the presence of the S+, with a decrease in response rates as the stimuli became increasingly dissimilar to the S+. This resulted in what would be termed a "generalization gradient." Visualized graphically, summated responses formed a peak that was centered over the S+. In the case of Guttman and Kalish's study, pigeons that had been trained to peck a key when a light of a particular wavelength was presented would also peck the key when exposed to shorter or longer wavelengths of light. Gradients were found to be generally symmetrical; responses attenuated at similar rates as a function of stimulus dissimilarity regardless of the direction of the stimulus change.

Peak Shift

The seemingly straight-forward relationship between responding and stimulus similarity to the S+ would become complicated by research into the effects of discrimination training on generalization gradients. Specifically, Hanson (1957, 1959) found that the introduction of an unreinforced stimulus (S-), which exists along the same continuum as the S+, during training could affect the position and characteristics of the generalization gradient. These changes can be described as a higher peak response rate, a skewing of the gradient away from the S-, and a shift in the location of the response mode (Hanson, 1957, 1959). This final change would be termed a "peak shift", and is the subject of this study. In effect, it was found that pigeons that had undergone discrimination training were more likely to respond to a novel stimulus than they were to respond to the S+, which often experienced a decrease in response rates. The results of this experiment can be seen in Figure 1.

Figure 1
Generalization Gradients in Trained Pigeons (Hanson, 1959)



Note. This graph shows five generalization gradients for pigeons that received either single-stimulus reinforcement training or discrimination training. The S+ (labeled "CS") was a light with a wavelength of 550 Mμ; four groups received discrimination training (S-: 590, 570, 560, or 555), while one group was only trained using the S+ (termed "control group").

Peak Shift in Human Participants

Peak shift has since been replicated with a wide range of species, conditions, and stimulus types (Galizio & Baron, 1979; Hearst, 1962; Hebert et al., 1974; Honig & Slivka, 1964;

Thomas & Decapito, 1962). However, a number of differences have emerged between the response patterns of humans and animals. These discrepancies may be due to differences in experimental methodologies or the influence of different processes. Relating to the latter, while peak shift in animals is generally considered to be the result of associative learning processes (McLaren & Mackintosh, 2002), evidence suggests that peak shift in humans is at least partially the result of cognitive processes (Lee et al., 2018; Thomas & Jones, 1962). Four of the findings that are unique to humans, and are difficult to explain using any single associative theory, are the central tendency effect in stimulus generalization, instances in which peak shift occurs towards S-, a more pronounced peak shift after discrimination training with dissimilar stimuli, and increased stimulus generalization resulting from increased ranges of test stimuli. These effects are discussed below.

The central tendency effect describes the propensity for generalization gradients to progressively shift toward intermediate stimulus values regardless of their proximity to S+ (Bizo & McMahon, 2007; Hebert et al., 1972; Helson & Avant, 1967; Thomas et al., 1973). The result is that generalization gradients that are initially distant from the center of the stimulus range will shift toward that center, with that shift being a function of the number of testing trials that are undergone. For example, should the peak of a generalization gradient initially occur at Stimulus 2, on a range from Stimulus 1 to 9, that peak will progressively move toward Stimulus 5. This form of peak shift is also noteworthy for the fact that it can occur following conditions of single-stimulus reinforcement training, meaning that this type of peak shift is not contingent on discrimination training.

Per the second of these human-specific phenomena, in cases where the S- is between the S+ and the center of the continuum, human peak shift has been observed to occur in the direction of the S- rather than away from it (Bizo & McMahon, 2007; Thomas et al., 1991). In these cases, participants were often more likely to respond to the S-, which had received no reinforcement during discrimination training, than they were to respond to the S+, which had been reinforced. For example, should one train participants to respond to Stimulus 2, but not Stimulus 4, and then test those participants using a stimulus continuum ranging from 1 to 9, then the generalization gradients would likely be shifted toward the S-, rather than away from it.

Next is the negative relationship between stimulus similarity during discrimination training and the magnitude of peak shift. In animals, the generalization gradient is found to shift away from the S-, with the magnitude of the shift being inversely related to the distance between the S+ and S- (Hanson, 1957, 1959). In effect, the more similar the S- is to the S+, the greater the peak shift. This same relationship has been observed in humans (Baron, 1973); however the reverse has also been observed in humans, with peak shift increasing as S+/- similarity decreases, meaning that in these instances, peak shift is negatively correlated with S+/- similarity (Doll & Thomas, 1967; Thomas et al., 1973).

The last of these phenomena that will be discussed is the gradient-expansion effect, or the tendency for human participants to show greater generalization when a wider range of stimulus values are presented during the generalization test (Derenne, 2019; Hansen et al., 1974; Thomas & Bistey, 1964; Thomas & Hiss, 1963; Verbeek et al., 2006). This means that participants who are tested using a relatively narrow stimulus range can be expected to show relatively little stimulus generalization. Conversely, participants who are tested using a wider stimulus range, despite having undergone the same response training, will show greater generalization.

Adaptation-Level Theory

Adaptation-level (AL) theory (Helson, 1947; Thomas, 1993; Thomas & Jones, 1962) is a cognitive account of peak shift that may be able to explain the findings that are unique to studies with human participants, and is the primary concern of this thesis. AL theory postulates that a mental representation of averageness (referred to as an "adaptation level"), which exists between the S+ and S-, is created within the mind of the participant. This AL then acts as a point of reference for the participant's decision-making criteria. For participants trained with an S+ and S-, the adaptation level would be a point intermediate to the two stimuli. The learned relation of the S+ to the adaptation level can be expressed as S+ = AL + X, where X is the distance between the adaptation level and S+. Because the AL exists as an average of experienced stimuli, it remains plastic, and can be affected by new stimuli. As such, the exposure to the new range of stimuli involved in the generalization test causes this AL reference point to shift toward the new stimulus mean, which in turn shifts the generalization gradient.

To illustrate, imagine a set of five stimuli on a stimulus dimension that are numbered 1 through 5. The developed adaptation level will be equal to the average value of stimuli presented up to a particular point in time. For example, if participants were trained to respond to Stimulus 3 (S+ in this case) and no other stimulus were presented, then the adaptation level would be equal to 3. If participants were trained with an S+ of 3 and an S- of 5, and both stimuli were presented equally often, then the adaptation level would be equal to 4 (the average of Stimulus 3 and Stimulus 5). The theory additionally proposes that participants will learn the relation of S+ to the adaptation level. If the S+ is 3 and the AL is 4, then participants will learn that the S+ is one unit less than the adaptation level. This learned rule becomes important during the generalization test for two reasons. First, reinforcement (usually affirming feedback in research with human participants) is withheld during the generalization test, so that participants must rely on this previously constructed rule when responding. Second, because the adaptation level is considered to be the average of the presented stimuli, it may change during the generalization test. A change in the AL, which is used as the point of reference, will thus cause responding to shift away from the S+. Within this example, if Stimuli 1, 2, 3, 4, and 5 all appear equally often during the generalization test, then the adaptation level will eventually move from 4 to 3. Because participants are still using the rule “respond to the stimulus that is one unit less than the adaptation level”, they will respond more often to Stimulus 2 than Stimulus 3.

Evidence for Adaptation-Level Theory

AL theory’s greatest successes have been in providing plausible explanations for the previously mentioned human-specific phenomena, though some evidence exists that it may also pertain to animal behavior under some circumstances (Malone et al., 2004). First, AL theory predicts the observed negative relationship between stimulus similarity and peak shift. Under AL theory, the more dissimilar the S+ and S- are, the more offset the average of those two stimuli will be from the S+. For example, should the S+ be Stimulus 3 while the S- is Stimulus 4, then the AL would be equal to 3.5 (the average of the presented stimuli). However, if the S+ were Stimulus 3 and the S- were Stimulus 5, then the AL would be equal to 4. During the generalization test, when the average presented stimulus is Stimulus 3, the gradient produced by AL 3.5 would shift .5 units to the left while the gradient produced by AL 4 would shift 1 unit to the left.

AL theory can also explain how it is possible for peak shift to occur in the direction of the S-. Should the S+ be offset from the mean of the testing stimulus continuum, with the S- being closer to the mean of the testing continuum, then AL theory would predict a peak shift toward the S-. For example, if the S+ were Stimulus 2 and the S- were Stimulus 4 (producing an AL of 3), while the testing continuum ranged from 1 to 9 (leading to a new stimulus mean of 5), then the gradient would be predicted to peak at Stimulus 4 (the S-). Indeed, this was observed by Newlin et al. (1979).

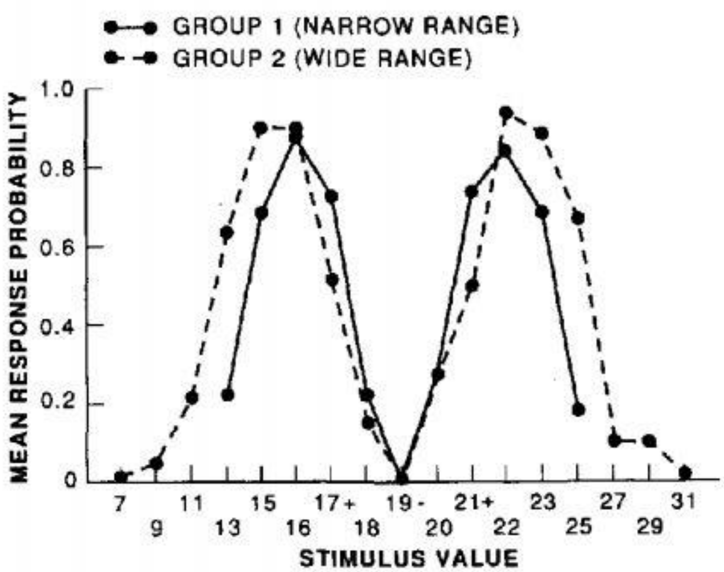
AL Theory would also be used to account for range effects, or the effects of the stimulus-range used in the generalization test upon the generalization gradient. The first of these is the central tendency, or the observed trend for generalization gradients to move toward the center of the presented stimulus-range (Thomas & Jones, 1962). As explained by AL theory, the adaptation level shifts over the course of the testing phase toward the new average of the presented stimuli. If all stimuli are presented with equal frequency, then this average would be at the center of the stimulus-range. This shift of the AL would cause generalization gradients that were initially distant from the center of the range to move toward the center.

However, under the AL theory, the central tendency effect would be limited under several important conditions. First, should the center of the stimulus-range be between S+ and the AL, the movement of the AL would cause the gradient to move away from the center. Second, should the S+ be between the center of the stimulus-range and the AL, the gradient would first move toward the center of the range and then continue to move past it. Last, should the AL be between the S+ and the center of the stimulus-range, AL theory would predict that generalization gradients would move toward the center of the range, but not reach it. The reason for this is that the movement of the gradient should stop once the AL (which is offset from the response peak) reaches the range mean.

The presence of greater peak shift when a wider range of stimuli is presented during testing has also been described as a range effect (Thomas et al., 1991). Specifically, Thomas et al. cited a study by Galizio and Baron (1979) in which discrimination training was conducted with a single S- (placed at the center of the testing continuum) and two S+s (one on each side of the S-, equidistant). Galizio and Baron claimed that because the testing range was symmetrical in regard to the S-/S+s, there should be no movement of the AL from training to testing, and thus no peak shift. However, the result of this experiment showed a double peak shift, in which the two peaks of the bimodal generalization gradient were offset away from the S-. While Galizio and Baron suggested this result was inconsistent with AL theory, Thomas et al. (1991) suggested that this too can be accounted for should one consider the AL decision criteria to be relative rather than absolute. In effect, if the "X" in the decisional rule "respond to AL +/- X" were affected by the range of stimuli (rather than being constant), then the double peak shift would be explained. Indeed, Thomas et al. showed greater double peak shift and wider gradients when participants were tested using a wider range than a narrower range, supporting the hypothesis that this double peak shift was a range effect. This result can be seen in Figure 2. However, in this case analyses were not performed to determine if this shift was progressive (a characteristic that is necessary to AL shift).

While it may be argued that this effect on the decisional X is a range effect, AL theory would still not explain the gradient-expansion effect, in which response gradients become wider when tested using a wider range of stimuli. However, should the AL be conceptualized as a range in itself (rather than a point), which is also subject to range effects, then gradient expansion might be accounted for. For example, during training involving an S- that has a value of 1 and an S+ that has a value of 5, participants may learn that S+ = 2-4 + 2. Here, the AL exists as a range from 2 to 4, while X is 2 (the average distance from the AL to the S+). The range of this AL could then increase once the generalization test begins, and a wider range of stimuli are experienced. While this effect might also be explained by conceptualizing the X as a range, describing the AL this way provides the additional benefit of explaining gradient expansion following single-stimulus reinforcement training, in which no S- is presented (thus, the decisional rule is simply S+ = AL).

Figure 2
Amplified Double Peak Shift Resulting from Greater Stimulus Range (Thomas et al., 1991)



Note. *This graph shows two generalization gradients, both of which were created using discrimination training between one S- (Stimulus 19) and two S+s (Stimuli 17 and 21). However, participants in the two conditions experienced different stimulus ranges during generalization testing, which accounts for the differing amounts of peak and area shift.*

Unresolved Questions

An unresolved question about peak shift is why the results of experiments with human participants sometimes conform to the results of animal experiments (Baron, 1973), suggesting control by associative processes, and sometimes instead appear to show the effects of other processes, such as control of responding by the adaptation level (Doll & Thomas, 1967; Thomas et al., 1973). It may be that certain details of the method determine which outcome is observed.

As such, the primary purpose of the present research is to examine some of the conditions that may determine whether AL-like effects occur.

One possible explanation is that it is the amount of training with the task that is responsible for the discrepancies. It has been observed that the degree of peak shift is affected by the number of training trials, with fewer trials leading to greater AL-like shift (negative correlation between S+/- similarity and peak shift) and more trials leading to lower AL-like shift (positive correlation between S+/- similarity and peak shift) (Newlin et al., 1979; Thomas et al. 1973; Wisniewski et al., 2010). However, the amount of training has rarely been manipulated, and the effects on peak shift in humans remains unclear.

Two ways that the amount of training may be relevant are through perceptual anchoring or the adaptation level decisional rule. Regarding anchoring, per AL theory, peak shift occurs when the AL changes over the course of the generalization test. It is possible that the amount of training affects how susceptible the AL is to this change. Specifically, extensive training may "anchor" the AL in one place, so that little change in the AL, and thus minimal AL-like effects, occur. This is an intuitive explanation; because the AL exists as the average of experienced stimuli, exposure to more stimulus-presentations during training would dilute the effect of later stimuli upon that average.

Relating to the decisional rule, with little training, the rule may be inexact. Instead of learning that the S+ is a particular distance from the AL, it is possible that participants are only able to observe that the S+ lies in a particular direction from the AL. This would lead to wide gradients, as greater generalization would occur toward the end of the stimulus-range that is away from the S-. However, with greater training, participants would become better able to assess the exact distance between the S+ and AL, and gradients would narrow toward the S-.

If extended training makes the adaptation level resistant to change, through anchoring, then participants who are extensively trained should show the poorest correspondences with the predictions of adaptation-level theory. However, if extended training improves participants' ability to discern X and adhere to the rule S+ = AL + X then participants who are extensively trained should show the greatest correspondences with the predictions of adaptation-level theory.

One way to help determine how greater training affects correspondence with adaptation-level theory is to closely examine whether, and how, responding changes during a generalization test. The previously mentioned "anchoring", through which training seems to moderate such effects as central tendency, has not been adequately examined in the context of the number of testing trials underwent. Rather, response data is generally aggregated across the generalization test. As such, little is known about how anchoring interacts with changes in the generalization gradient, which occur over the course of the testing phase.

This question is particularly relevant due to AL theory's reliance on progressive change, meaning that the effect of AL shift increases relative to the number of testing trials underwent (Thomas et al., 1973). More specifically, at the beginning of the generalization test, the adaptation level that controls responding is the adaptation level that was present at the end of training. This means that, should AL shift be the only operating source of peak shift, responding at the beginning of the generalization test should be relatively accurate, with the response peak aligning with S+. However, as the stimuli selected for the generalization test cause the adaptation level to change, then responding should become progressively less accurate during the generalization test as the new adaptation level becomes established. Whether this actually occurs is not well known. Only a subset of studies concerned with adaptation-level theory report such an analysis, and the topic is rarely examined elsewhere in the peak shift literature.

The central tendency effect and gradient-expansion effect have also been shown to be moderated by training factors in several noteworthy ways. In a study by Verbeek et al. (2006), it was shown that these effects (referred to under the umbrella term “range effects”) are decreased by increasing stimulus complexity (i.e., line angles compared to face morphing), but that this could be countered by increasing S+/- similarity. This finding is noteworthy for the fact that the relationship is the reverse of that predicted by AL theory. In this case, increased S+/- similarity caused greater range effects, and was correlated with an increase in training trials that were required to meet the advancement criteria. It is particularly worth noting that gradient expansion has not yet been examined with respect to progressive change across the generalization test.

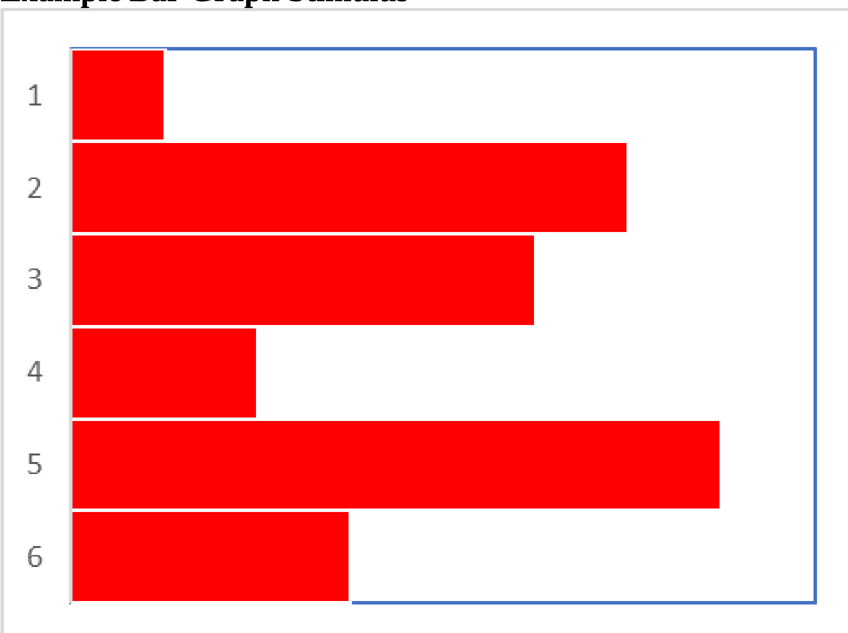
The Current Study

Goals and Approach

The primary goal of the research was to resolve the effect that training has on human participants’ adherence to the predictions of AL theory; specifically in the context of the S+/S- distance effect. A secondary goal of the research was to examine how well performances adhere to AL theory when the stimuli were relatively complex. Thomas consistently utilized relatively simple stimuli such as lights of varying brightness, but Verbeek et al. (2006) and Spetch et al. (2004) suggested that range effects were diminished when a relatively complex stimulus (morphed faces) was used. However, morphed faces differ from previously used stimuli in a number of ways that may be unrelated to complexity (e.g., pre-existing familiarity with discrimination between the stimuli). As such, horizontal bar graphs, with varying average bar lengths, were selected as the stimulus; see Figure 3 for an example stimulus and Appendix A for a detailed description of how these stimuli were constructed.

This stimulus is somewhat novel, although a similar stimulus has been used in recent research (Derenne et al., 2022). While this novelty does complicate predictions, the stimulus offers two important advantages over a stimulus such as human faces. First is improved face validity in comparisons between this study and past research involving the conceptually similar, but less complex, stimulus of lines that vary in length. Second is a low likelihood of pre-existing participant familiarity with this type of task.

Figure 3
Example Bar Graph Stimulus



Note. This graph represents a stimulus score of 30 on a scale from 0 to 60, with a medium standard deviation between the bars (i.e., bar mean = 5.0, bar standard deviation = 2.5).

To this end, two experiments were conducted, each using a 2 (S+/- Similarity) x 2 (Training Amount) x 18 (Testing Iteration) mixed-model design. The first variable, S+/- Similarity, relates to the degree of similarity between the S+ and the S- during discrimination training. Under both conditions, the S+ was the same (Stimulus 4), while the S- was either relatively similar to the S+ (Stimulus 3) or relatively dissimilar to the S+ (Stimulus 1). The condition in which the S- was relatively similar to the S+ will be referred to as the Near S- condition, and the condition in which the S- was dissimilar will be referred to as the Far S- condition. As will later be discussed, the exact values of these stimuli varied between Experiment 1 and 2.

The inclusion of Near S- and Far S- conditions allowed for the examination of how stimulus similarity affects generalization gradients, both alone and in combination with the other variables that were examined. AL theory would predict that peak shift should be greater under the Far S- condition, while associative theories (Boneau & Cole, 1967; Spence, 1937) would predict greater shift under the Near S- condition.

The second variable, Training Amount, relates to the number of training trials that participants completed during discrimination training in order to advance to the testing phase. Participants were randomly assigned to one of two training conditions: 8 training trials or 40 training trials. Should the posited explanation of anchoring be accurate, peak shift under the Far S- conditions would be greatest when fewer training trials were undergone.

For questions regarding range effects, generalization gradients at the beginning (Iteration 1) and end (Iteration 18) of the testing phase were of particular interest as repeated measures. Comparisons between gradients at these different stages of the testing phase allowed for the examination of relationships between the amount of training and progressive effects such as central tendency and gradient expansion.

The primary variables of interest were gradient means, which represent the average stimulus that received an affirmative response, and gradient widths, which describe how many stimuli within the range received affirmative responses. These will be explained further within the Methods section. While not sufficient to provide a complete description of a gradient's characteristics, these variables do provide more information than the gradient mode alone. The gradient mean in particular is more sensitive to changes in the distribution of responses than the mode. Changes in the gradient mean are referred to as "area shift", and represent a displacement of the gradient along the stimulus dimension, which may exist despite the peak remaining in one location.

Hypotheses

Based on previous findings (Thomas et al., 1973; Wisniewski, 2010), which have shown AL-like trends (where peak shift is greater when the S+ and S- are dissimilar) to be most prominent under conditions of less extensive training, while associative-like trends (where peak shift is greater when the S+ and S- are similar) are greater under moderate training:

> Hypothesis 1a: Generalization gradients will differ depending on the similarity of the S- used in training and the amount of training underwent.
>
> Hypothesis 1b: An interaction will occur between similarity of the S- and the amount of training underwent, such that greater training with a dissimilar S- leads to less shift, while greater training with a similar S- leads to increased shift.

As the central tendency effect has been shown to be progressive, increasing with the number of testing trials undergone (Thomas et al., 1973):

> Hypothesis 2a: The difference between the gradient mean and the range mean (the average of all stimuli used in the generalization test) will decrease as the testing phase progresses.

Given observations that AL-like effects (i.e., central tendency and progressive shift) are negatively associated with the amount of discrimination training (Newlin et al., 1979; Thomas et al. 1973):

Hypothesis 2b: The amount of progressive decrease in the difference between the gradient mean and range mean will be greater in the 8-training trial condition than the 40-training trial conditions.

Given observations by Verbeek et al. (2006) of positive control of range effects by S+/- similarity:

Hypothesis 3: Gradient expansion will be greater in Near S- conditions than Far S- conditions.

As discrimination proficiency has been shown to be progressive and inversely related to S+/- similarity (Hanson, 1959):

Hypothesis 4a: Participants will be more accurate at the end of training than the beginning of training.

Hypothesis 4b: Participants within the Far S- condition will be more accurate than participants in the Near S- condition.

Hypothesis 4c: Participants in the 40-training trial condition will be more accurate than participants in the 8-training trial conditions.

General Method

This study involved two experiments that differed in some ways, but were methodologically identical in most regards. As such, the general method will be discussed here, and the unique aspects of each experiment will be discussed in turn.

Participants

This project involved 192 participants (96 participants for each experiment). These participants were UND students, recruited through SONA, and compensated with .5 research credits (based on an expected participation time of 15 to 25 minutes) to be applied to their psychology course. It was required that participants be 18 years of age or older, and had not participated in any similar studies conducted by the author or research advisor. Participants were not asked to provide information regarding visual acuity or demographics.

Apparatus

Data collection occurred through the Qualtrics website, accessed through the participant's computer, tablet, or cellphone. The only requirement for said devices was that they be capable of accessing and interfacing with the Qualtrics website.

Stimulus

Each stimulus consisted of a graph displaying six horizontal red bars, which initiate at the left edge of the graph area and extend to the right (see Figure 3). Each stimulus differed from every other, either in terms of individual or summated bar lengths, so that no two graphs were identical. Along the y-axis, the bars were labeled from 1 (the topmost bar) to 6 (the bottommost bar). No other markings or labels appeared on the graphs.

Each bar was quantified on a relative linear scale from 0 (not visually discernable from the leftmost edge of the graph area) to 10 (spanning the length of the graph area). The absolute length of each bar depended on the specifications of the device that was utilized to view the stimuli.

Each stimulus was designated based on the sum of all six bars that it contained; this sum was referred to as the stimulus's "score". Experiment 1 used seven stimuli; the values of which were 15, 20, 25, 30, 35, 40, and 45. Experiment 2 also used seven stimuli, though in this case the values of which were 0, 10, 20, 30, 40, 50, and 60. For the sake of concision, stimuli will be referred to as "S" followed by the stimulus's score (e.g., stimuli with a score of 20 will be referred to as "S20").

A detailed description of the method by which the stimuli were constructed can be found under Appendix A.

Design

These experiments both followed 2 (Training Type) x 2 (Training Amount) x 18 (Testing Iteration) mixed-model designs.

The Training Type (Near S- and Far S-) relates to the stimuli that were presented to the participant during discrimination training. Under the Near S- condition, participants were asked to discriminate between S30 (S+) and those and a relatively similar S- (S25 in Experiment 1, S20 in Experiment 2); similarly, under the Far S- condition participants discriminated between S30 (S+) and a relatively dissimilar S- (S15 in Experiment 1, S0 in Experiment 2).

Participants' assigned Training Amount (8 or 40 training trials) determined the number of discrimination trials that had to be completed to advance from the training phase to the testing phase.

Any participant that failed to complete the testing phase within 2 hours of initiating the Qualtrics survey was excluded from the data set. Additional participants were recruited to replace these cases.

Procedure

Participant involvement occurred over three phases: initial instructions, discrimination training, and generalization testing.

Initial Instructions

Prior to agreeing to take part in the study, participants were provided with a brief summary of their task and the study's overall purpose. Specifically, they were told that the research examines people's ability to detect differences between similar-appearing bar graphs.

Upon initiation of the Qualtrics survey, participants were provided with an informed consent form that they had to agree to in order to continue with the study. Once this was done, participants were provided with more detailed instructions and information. It was explained that they would be shown a bar graph (S+), referred to as the “target”. They were told that the combined bars of this graph formed this graph’s “score”, and that they should observe this graph closely, as they would have to compare subsequent graphs against it.

Last, participants were informed that they would initially receive feedback after each response as to whether they were correct or incorrect, and that this feedback would cease when they entered the final phase. Once participants acknowledged that they understood these instructions, they were presented with the S+. Upon acknowledgement that they had observed the S+ and were ready to continue, participants advanced to the discrimination training phase.

Discrimination Training Phase

In this phase, participants were first assigned to one of two training types (Far S- or Near S-) as well as one of two training amounts (8 trials or 40 trials). Regardless of assignment, the general process can be described thusly: participants were shown a series of bar graphs one by one, all of which matched the S+ in format but may have differed in overall bar length. Each stimulus was displayed with the text “Does this match the target?” and participants were given the opportunity to respond “Yes” or “No”. In the event of a correct response to a presented S+, the text “CORRECT! This graph has the same overall score as the target.” was displayed; a correct response to a presented S- resulted in the text “Correct! This graph’s overall score is less than the target’s.” This text was bolded and colored green. In the event of an incorrect response to an S+, the text “INCORRECT: this graph’s score matches the target score.” was displayed.

Similarly, an incorrect response to an S- resulted in the text "INCORRECT: this graph's score was below the target score." This text was bolded and colored red.

The graphs that were presented depended on the group to which the participant was assigned. Both groups were shown graphs that matched the S+ in overall bar length (a summated score of 30), as well as an S- that depended on the participant's assigned condition. S+/- presentations were arranged in a predetermined semi-random order, with the only constraint being that neither S+ nor S- were presented in more than 3 consecutive trials.

Participants' Training Amount (8 trials or 40 trials) determined the number of discrimination trials that were completed during the training phase. Once all of the assigned discrimination trials were completed, the participant was informed that feedback would no longer be provided, and that they were advancing to the testing phase.

Generalization Testing Phase

The generalization testing phase consisted of 18 iterations, with each iteration containing one presentation of each of the seven stimuli, for a total of 126 test stimuli presentations. This was done so that each participant would encounter every possible stimulus score once before any repeats were presented, thus ensuring that no score was disproportionately present within a given section of the testing phase. Participants were not given any indication when one iteration was ending and another was beginning.

The stimuli were presented in one of four possible orders, to which participants were randomly assigned. A detailed representation of presentation orders can be found in Table 1.

Once the testing phase was completed, no further data was collected.

Data Preparation

Responses made during discrimination training were used to calculate accuracies toward both the S+ and the S-, as well as an overall accuracy. Initial accuracies (over the first 4 training trials) and final accuracies (over the last 4 training trials) were used to compare the experimental conditions. Responses made during the generalization test were used to construct generalization gradients for each testing iteration, as well as for the entire test.

To quantify gradient characteristics, gradient means and widths were calculated. Gradient means represented the average stimulus that received a response within a given condition or testing iteration. These were calculated by multiplying each stimulus-score by the number of responses that were made to that stimulus. These products were summated and divided by the total number of responses made. In rare cases where a participant made no affirmative responses within a given testing iteration, the gradient mean for this iteration was calculated as the average of the means found in that participant's preceding and following iterations.

The gradient width was calculated for each participant, within each testing iteration, as the number of stimuli that received an affirmative response or were enclosed by two stimuli that received affirmative responses, multiplied by the distance between adjacent stimuli (5 in Experiment 1, 10 in Experiment 2). For example, if a participant in Experiment 1 only responded to S30 within Trial 1, then this would represent a gradient width of 5, while another participant who responded to S30, S35, and S40 would show a gradient width of 15. These widths were used in statistical analyses, and the means of these widths were used in graphical representations. Quantifying widths in this way, rather than simply as the number of stimuli that received an affirmative response, allowed for comparisons between the gradient widths of Experiment 1 and Experiment 2.

Gradient area was the total responses made to all stimuli, divided by the number of stimuli presented. As such, areas existed on a scale from 0 (did not respond to any stimulus) to 1 (responded to every stimulus).

Primary Analyses

Historically, the majority of analyses involving generalization gradients have been based on qualitative comparisons between these gradients. This is largely due to the complexity of generalization gradients, as statistical comparisons of all characteristics would not be feasible. Further, visual analyses have been, and still are, important because 1) behavior analysts commonly make extensive observations of the target behavior, and graphs may communicate observations more effectively than a report of statistics, 2) a tradition of showing all of the data increases transparency, as readers can see exactly what happened and not rely on the researchers' reports that certain differences were or were not significant, and 3) visual analyses emphasize effect sizes, which may be considered to be a more important matter than *p* values.

The exceptions are generally either the gradient mode or mean, which have more often been subject to quantitative analyses. Given the number of gradient characteristics of interest in this study, a similar approach was taken. Gradient means and widths were analyzed quantitatively, while the other characteristics were described quantitatively and compared qualitatively.

Each experiment involved two analyses of primary interest. The first was a two-way ANOVA, using Training Type and Training Amount as independent variables, and response accuracy over the final four training trials as the dependent variable. This was done to allow for assessment as to whether the discrimination training resulted in meaningfully different learning outcomes. The second analysis of primary interest was a repeated measures MANOVA, which used Training Type, Training Amount, and Testing Iteration (the repeated measures component) as independent variables, while gradient means and widths were analyzed as the dependent variables.

Conditions were also compared qualitatively using graphic representations of the generalization gradients. This was done with responses aggregated across the entire generalization test, as well as generalization gradients composited from different sections of the testing phase (i.e., testing iterations 1 and 18).

Additional unplanned analyses were also performed to explore unexpected trends that emerged in the data.

Experiment 1

Method

Experiment 1 followed the general method previously discussed, with two specifications. First, the Far S- was a stimulus with a score of 15, while the Near S- was a stimulus with a score of 25. Second, the generalization test exclusively involved presentations of stimuli with the following scores: 15, 20, 25, 30, 35, 40, and 45.

Results and Discussion

Overview

Results for both the training and testing phases will be discussed. Of particular interest within the training phase was response accuracy and how this changed with trial progression. Assessing accuracy allowed for the determination of whether learning occurred and whether this learning differed notably between conditions.

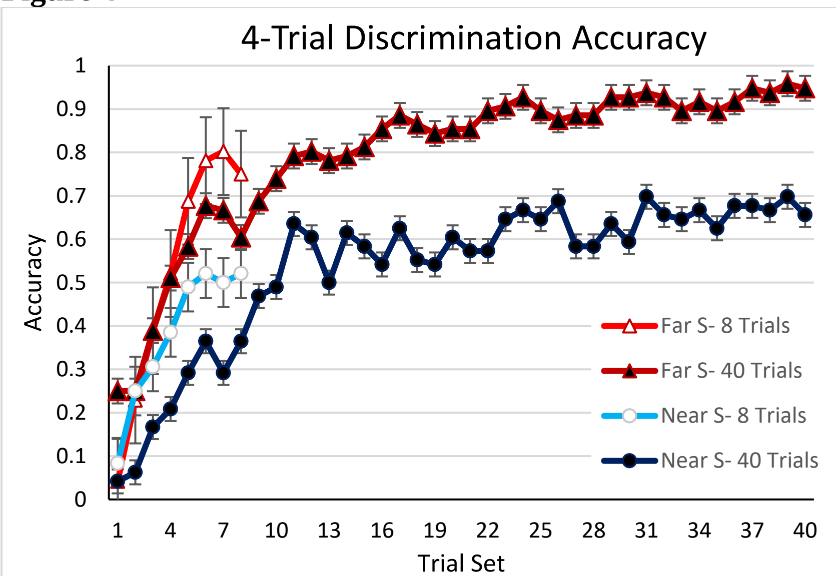
Data from the testing phase was more directly relevant to the hypotheses of this research. The gradient mean and width, while not sufficient to provide a complete picture of the gradients, are adequate for describing the relevant aspects (i.e., gradient location and width along the X-axis). Statistical comparisons of these dependent variables were used to determine whether these key components differed by experimental condition, and visual analyses of the gradients provided supplementary comparisons, some of which were analyzed with additional post-hoc tests.

Training Accuracy

Over the training phase, accuracy was calculated using the running average ratios of correct/incorrect responses within sets of four consecutive trials. Particular attention was paid to initial accuracies (over the first 4 trials) and final accuracies (over the last 4 trials), as these would allow assessment as to how much learning took place during training, as well as participants' discrimination proficiency as they began the generalization test. Both the initial and final trial-sets consisted of two S+ presentations and two S- presentations.

Given the relative complexity of the stimuli, it was expected that participants' initial accuracies would be near, or slightly above, those of random chance (accuracy = .50); however, as displayed in Figure 4, this was not the case. The overall average accuracy (M = .406, SD = .267) was significantly below chance during the first four training trials, $t(95) = -3.438$, $p < .001$. To determine whether this was the result of a particular training condition, two-tailed single sample t-tests were performed to compare the initial accuracies of each condition against an expected accuracy of .50. These revealed that accuracies under the F8 condition (M = .521, SD = .207) were not significantly different from chance, $t(23) = .492$, $p = .627$, and accuracies under the F40 condition (M = .510, SD = .227) were similarly not different from chance, $t(23) = .225$, $p = .824$. However, accuracies under the N8 condition (M = .385, SD = .276) were near to being significantly less than one would expect by chance, $t(23) = -2.037$, $p =.053$, and accuracies under the N40 condition (*M* = .208, *SD* = .241) were significantly below chance $t(23) = -5.935$, $p <$ .001.

Figure 4



Note. Accuracies are presented as running averages of the four most recent training trials. For example, Trial Set 2 represents accuracies in response to trials 1 and 2, while Trial Set 40 represents average accuracies in response to trials 37 – 40.

The initially poor performance in discrimination was somewhat unexpected, as even random guessing would be expected to yield a greater overall accuracy in two of the four conditions. One possible explanation is that participants were attending to aspects of the stimuli that did not adequately represent similarity to the S+. For example, it may be that when participants were initially shown the S+, the shape of the graph or the length of a particular bar were observed, while the overall area was disregarded. If this were the case, it would be expected that few of these participants would correctly make an affirmative response to presentations of the S+ that did not match this irrelevant feature. Under this explanation, one would also expect far more participants to correctly make a negative response to presentations of the S-, which would not match the S+ in terms of overall bar length or general shape. This interpretation is supported by the data presented in Figure 5, which shows accuracies toward the S+ were initially far lower than accuracies toward the S-. However, it is worth noting that the presentation order was the same in all conditions, and that the first S- was not presented until after two presentations of the S+. As such, it may be that some learning had occurred prior to the first S- presentation, which would account for the greater accuracy in responding to the first S-.

Another potential explanation is that the relative complexity of the stimulus-type introduced noise to the perceptual task, which biased participants toward negative responses. As participants were not aware of the specifics of what stimuli would be presented (particularly how many different S-s would be presented, and how similar these would be to the S+), it stands to reason that any perceived difference, no matter how slight, would result in a negative response. While this hypersensitivity to perceived difference would explain the low initial accuracy, without requiring the assumption that many of the participants had failed to understand the instructions of the task, it does not explain why S- accuracies did not initiate near 100%. Again, it may be that some learning had taken place by the time the first S- was presented, so that participants had become more liberal in making affirmative responses. This possibility is supported somewhat by the trends observed in Figure 5, which generally showed that while both S+ and S- accuracies trended upward, they often did so independently. That is, improvements in classifying the S+ did not necessarily correspond with equal improvements in identifying the S-. Rather, the relationship was often weak or slightly negative, suggesting that the development of sensitivity to differences was occurring alongside changing amounts of conservativeness.

While the reason for the low initial accuracy is not clear, it is suggestive that discrimination involving this stimulus set is more challenging than that of traditional stimuli, either due to the novelty of the task or complexity of the stimuli. This is a likely source of other unusual trends that will later be discussed.

Regardless of the reason for the initially poor accuracy, subsequent performance more closely matched expectations. Specifically, accuracies improved at a decreasing rate, with initially rapid improvements that plateaued and approached horizontal asymptotes (Figure 4). The y-values of these asymptotes differed depending on the location of the S-. On average, these final accuracies (M = .719, SD = .306) were greater than chance. To determine whether the final accuracies of each group differed significantly from chance, two-tailed single sample t-tests were performed for each condition using accuracies over the final 4 trials. These showed that the F8 group (M = .750, SD = .346) was significantly greater than chance ($t(23) = 3.542$, $p = .002$), the F40 group (M = .948, SD = .127) was significantly greater than chance ($t(23) = 17.245$, $p < .001$), the N8 group (M = .521, SD = .220) was not significantly more accurate than chance ($t(23) = .464$, $p = .647$), and the N40 group (M = .656, SD = .320) was significantly greater than chance ($t(23) = 2.394$, $p = .025$). While the N8 group was not more accurate than chance at this point, a paired samples t-test showed that this final accuracy was more accurate than the initial performance (M = .385, SD = .276), indicating that learning had occurred, $t(23) = 2.251$, $p = .017$.

While the Near and Far S- conditions reached different maximum accuracies, the number of trials that it took to approach these maxima were similar. In both cases, improvements leveled off around the 11th trial-set (trials 8-11), although modest improvements did continue through the 40th trial. As training ended for the participants assigned to the 8-trial condition before this point, learning was likely still occurring in the majority of these participants.

These trends form traditional “learning curves”, in which improvements are initially rapid, and then plateau. One important note is that the maximum accuracy achieved differs between conditions. Those who underwent 8 training trials achieved a lower final accuracy (*M* = .635) than those who underwent 40 (*M* = .802), and those who were trained using a relatively similar S- achieved a lower accuracy (*M* = .589) than those who were trained with a dissimilar S- (*M* = .849). The difference in accuracies that resulted depending on training amount is useful, as it indicates that learning was still occurring at the time of the 8th trial, while the minimal upward slope at the 40th trial would suggest that little learning was still taking place. As such, should differences in generalization gradients exist purely as a result of how much learning has occurred, or how close the participants were to achieving maximal accuracy, it would be reasonable to expect these differences to be present under the current circumstances. However, as with all procedures conducted using discrimination training, it is possible that an effect may exist at some training amount that was not examined here. That is, it may be that some training amount prior to the 8th trial, between the 8th and the 40th trial, or after the 40th trial would result in effects that these training amounts do not. This possibility is somewhat less likely given the rarity of sub-8 training trial procedures in the literature, as well as the relatively small change in accuracies over the middle and later trials.

To determine whether observed differences between the groups were significant, a two-way ANOVA was conducted using the final accuracies of each group. This revealed significant main effects of both S- Location ($F(1, 92) = 22.727$, $p < .001$) and Training Amount ($F(1, 92) = 9.309$, $p = .003$) on final accuracies. The Far S- condition showed greater accuracy (*M* = .849) than the Near S- condition (*M* = .589); while members of the 40-trial condition showed greater accuracy (*M* = .802) than members of the 8-trial condition (*M* = .635). A significant interaction was not observed ($F(1, 92) = .327$, $p = .569$).

The difference in final accuracies between the F40 (M = .948) and N40 (M = .656) conditions is noteworthy, as it is suggestive that training alone is not sufficient to overcome the difficulty of the discrimination task when the S- is relatively similar to the S+. As with the differences observed between the final accuracies of the 8 and 40-trial conditions, this suggests that the S-s are different enough from each other to affect any characteristics of the gradients that are subject to this variable. Again, it may be that some other S- value would cause effects that these do not, but the accuracies observed do not contradict the hypothesis that these S-s are suited to the task.

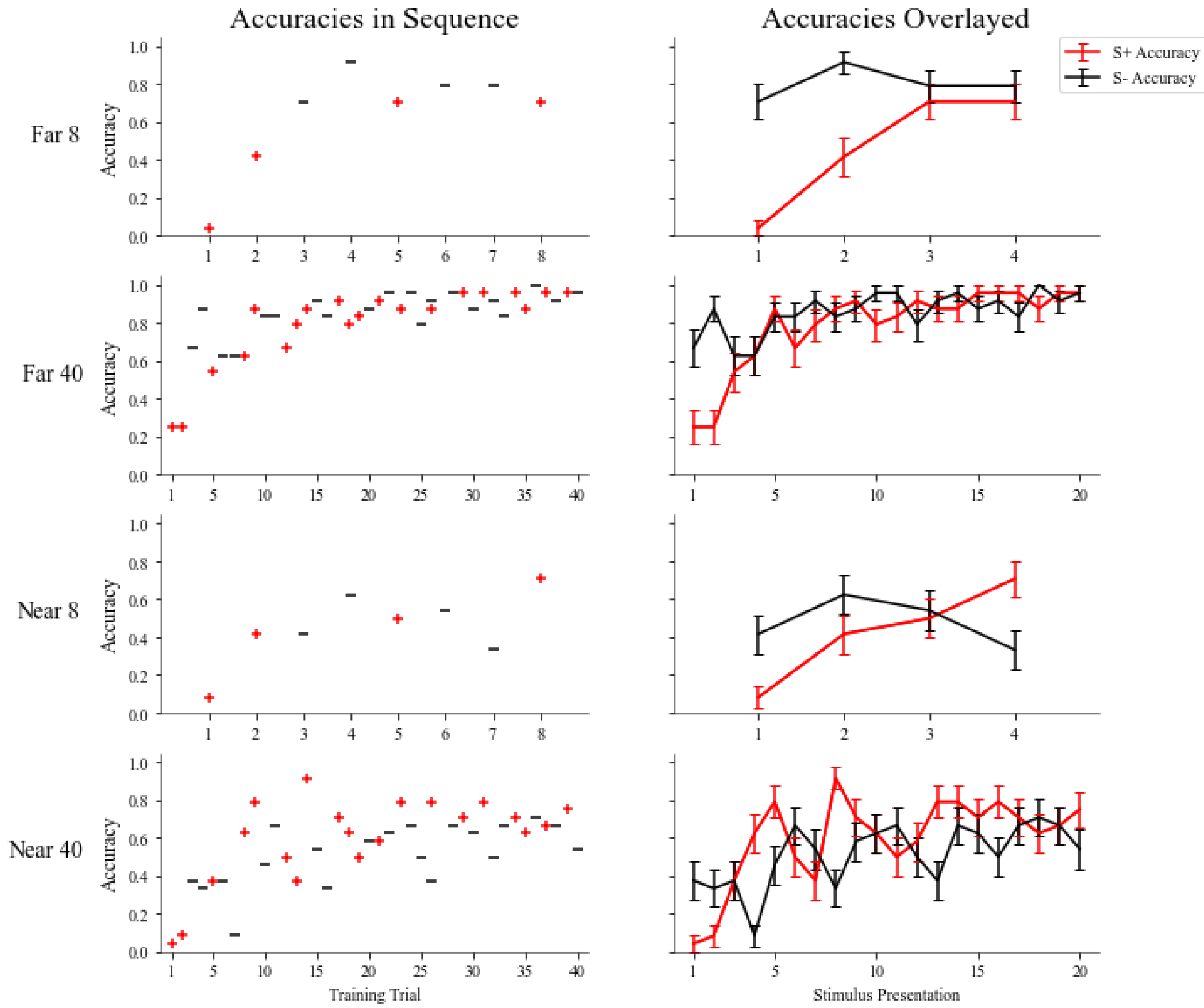
Although improvements in overall accuracies formed relatively typical trends, it is possible that improvements may not have been identical between the S+ and S-. To determine these more specific trends, accuracy was further broken down into S+ accuracy and S- accuracy. These are depicted in Figure 5, which displays the proportion of correct responses at each presentation of the S+ and S-. Across all conditions, S+ accuracy was initially much lower than S- accuracy. While both S+ and S- accuracies showed progressive improvement, accuracy toward the S+ improved more quickly, so that participants were approximately as accurate when responding to the 3rd S+ as they were when responding to the 3rd S-.

Generalization Test

Data collected during the testing phase will be described in both quantitative and qualitative terms. To minimize risk of Type I error, only two aspects of the generalization gradients (i.e., gradient mean and gradient width) were quantified and subjected to statistical tests. As previously discussed, the purpose of these tests was to determine whether central tendency in the area under the generalization gradient (gradient mean) and the amount of generalization (gradient width) varied by experimental condition and across the testing phase. As such, a repeated measures approach was taken to investigate changes in gradients over the course of the generalization test. This allowed for the evaluation of progressive gradient shift, which would be expected under AL theory. Additionally, qualitative visual analyses allowed comparisons of aspects of the gradients that are not described by the mean or width.

Figure 5

S+/- Accuracies in Sequence and Overlayed



Note. *The left set of graphs represents the training sequence of each condition, in which the stimulus presentation order was the same for all participants. The right set of graphs shows the same data, but with accuracies toward successive presentations of the S+ and S- overlayed, so that trends may be more easily compared.*

A repeated measures MANOVA was conducted using training amount, S- similarity, and testing iteration as independent variables, while gradient means and gradient widths were used as dependent variables. No participants represented significant outliers. Assumptions of homogeneity of variance and covariance were met. Gradient means and widths were moderately correlated (r =.359), which was within the acceptable range (.2 to .9).

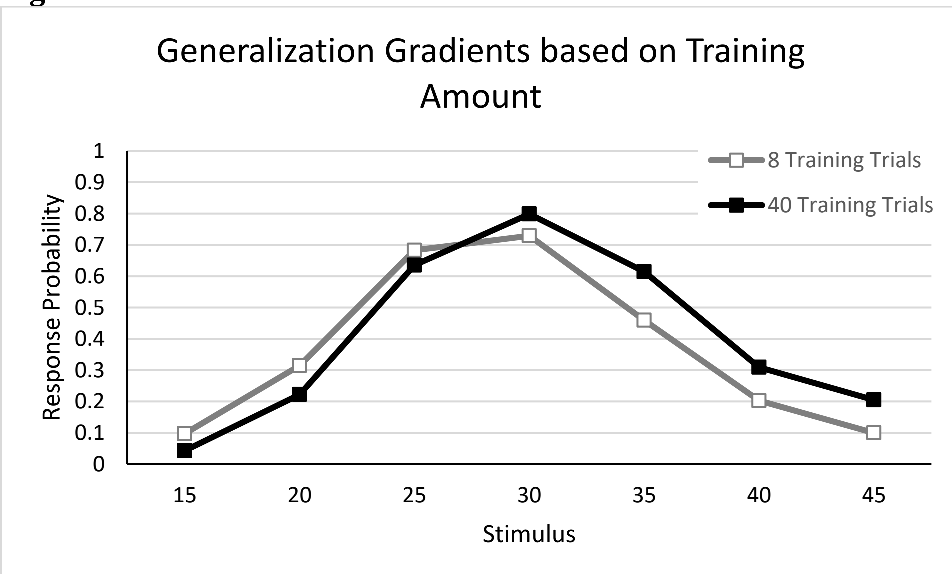
As previously discussed, AL theory predicts that generalization gradients should have been more offset from the S- when: 1) the S- was relatively dissimilar to the S+; 2) participants underwent less training. Under AL theory, the effect of S- similarity is due to the location at which the AL is established, with a more dissimilar S- leading to an AL that is farther from the S+, and thus facilitates greater shift. The effect of greater training is suggested to be either the result of anchoring the AL, which buffers against changes during the generalization test, or the ability of participants to identify a specific "X" value in forming a decisional rule.

Amount of Training

MANOVA results showed that the main effect of Training Amount was significant, F (2, 91) = 3.432, p = .037; Wilk's Λ = 0.930, partial η^2 = .070. As such, homogeneity of variances was tested for both dependent variables using Levene's F test. It was found that these assumptions were satisfied, and follow-up ANOVAs were conducted. These revealed a significant main effect of training amount on the gradient mean, F (1, 92) = 6.935, p = .010, partial η^2 = .070. Under the 8-trial training condition, the mean (M = 28.826) was significantly less than that of the 40-trial training condition (M = 30.566).

As shown in Figure 6, when data was combined with respect to S- location, gradients produced after 40 training trials were slightly right of those produced with only 8 trials. Notably, the lesser training conditions produced gradients that were biased toward the left of the S+, in the direction of the S-. Contrary to expectations, Figure 6 shows that the 8-trial participants were far more likely to respond to S25 than they were to respond to S35. These results do not support AL theory, which would predict that gradients should be more offset away from the S- (either due to an absence of anchoring or broadness of the decision rule) when less training is undergone.

Figure 6



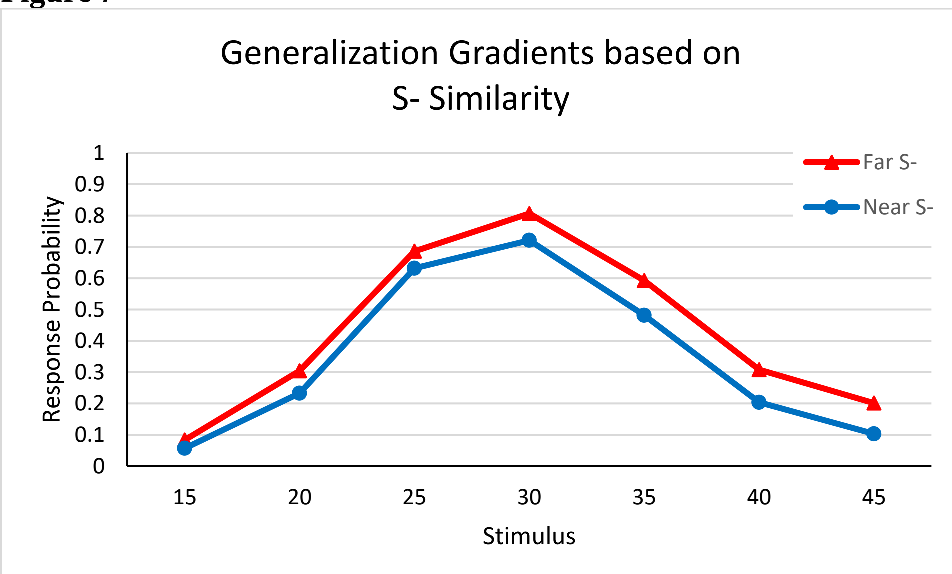
Note. Generalization gradients were constructed by merging the conditions with respect to S-similarity, so that only the effect of training amount is shown.

S- Similarity

The location of the S- during training did not have a significant effect on either gradient means or widths, F (2, 91) = .958, p = .143; Wilk's Λ = 0.930, partial η^2 = .070. Figure 7 shows the generalization gradients with respect to the main effect of S- placement. These gradients did not differ noticeably in mode or mean, however there was a difference in terms of area under the curve, with the Near S- condition producing a gradient that was lower than that of the Far S-. This difference was relatively small at the low end of the stimulus range, and greater at the higher end.

Again, these results do not support AL theory, which predicts that gradients should show greater shift away from the S+ when the S- is relatively dissimilar. While participants were slightly more likely to respond to greater stimuli following training with the Far S- than the Near S-, this difference was minimal.

Figure 7



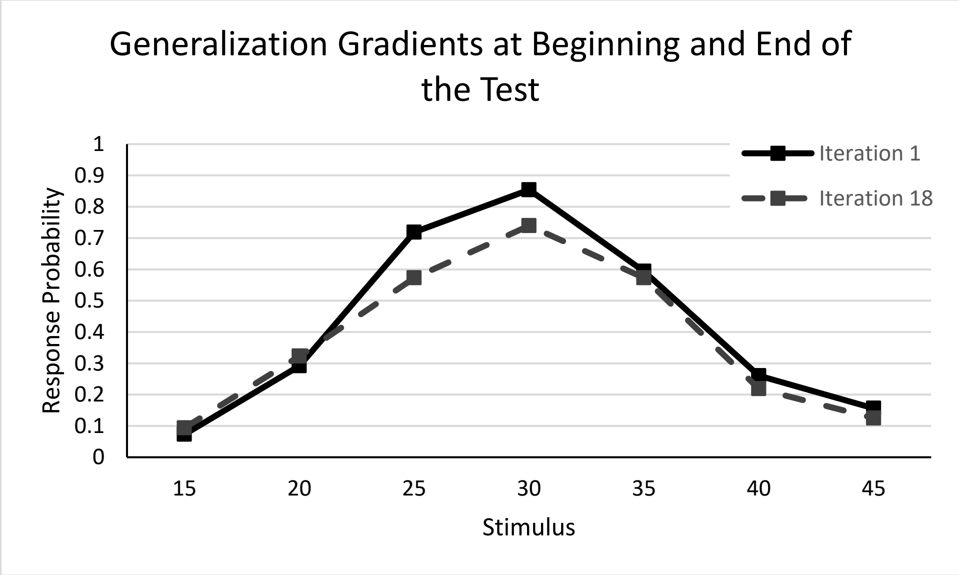
Note. Generalization gradients were constructed by merging the conditions with respect to training amount, so that only the effect of S- similarity is shown.

Testing Iteration

A significant main effect of testing iteration, the within subjects variable, was also observed, F (34, 59) = 1.959, p = .012; Wilk's Λ = 0.470, partial η^2 = .530. The follow-up ANOVA revealed that testing iteration affected gradient widths, F (17, 1564) = 1.633, p = .049, partial η^2 = .017. Overall, gradient widths decreased as the testing phase progressed (Iteration 1 mean width = 15.781, Iteration 18 mean width = 14.115). Using Iteration 1 widths as a baseline, 8 of the 17 following iterations showed gradient widths that were significantly narrower. Only one testing iteration, Iteration 10, had an average gradient width that was wider than that of Iteration 1.

While gradient widths did progressively decrease over the course of the testing phase, this is not clearly visible within the aggregate gradients. Rather, the progressive change appears as a general depression in response probability, particularly near the S+. This can be seen in Figure 8, as well as Figure 13.

Figure 8



Note. Generalization gradients were constructed by merging all conditions to display overall differences in responding between the first and final iterations.

While some extinction may be expected to occur due to the absence of reinforcement during testing, this gradient contraction is the reverse of the expansion that was predicted by AL theory, and is not easily explained using these data alone. Potential explanations for this finding will be discussed in the General Discussion section, with the results of Experiment 2 taken into consideration.

Simple Effects and Interactions

The interactions between training amount and S- similarity (p = .444), training amount and testing iteration (p = .968), S- similarity and testing iteration (p = .251), as well as the three-way interaction (p = .967), were all non-significant. Although the interactions were not statistically significant with respect to gradient means and widths, a number of differences between the gradients were visually apparent.

First, Figure 9 shows that under the Far S- condition the amount of training affected the area more than the peak. Specifically, the F8 condition produced a gradient that was roughly symmetrical in regard to the extremes of the stimulus range, with response probabilities being approximately the same for S15 as S45, while the F40 condition produced a gradient that favored the right side of the stimulus range (i.e., participants responded to S45 29.4% of the time, and responded to S15 3.5% of the time.) In effect, when using a relatively dissimilar S-, greater training increased the probability of responding to extreme high-value stimuli, and decreased the probability of responding to extreme low-value stimuli. This created a right-shifted area. In comparison, the F8 condition created a gradient that was leftward of the F40 gradient, and even somewhat left of the S+ (S30).

Another important distinction is that the N40 (Figure 10) condition led to a gradient that was roughly symmetrical, while both the N8 (Figure 10) and F8 (Figure 11) conditions led to gradients that were shifted slightly leftward (in the direction of the S-). This is difficult to explain, as AL theory would not predict a shift in the direction of the S- when the S+ is at the center of the stimulus range.

Figure 9

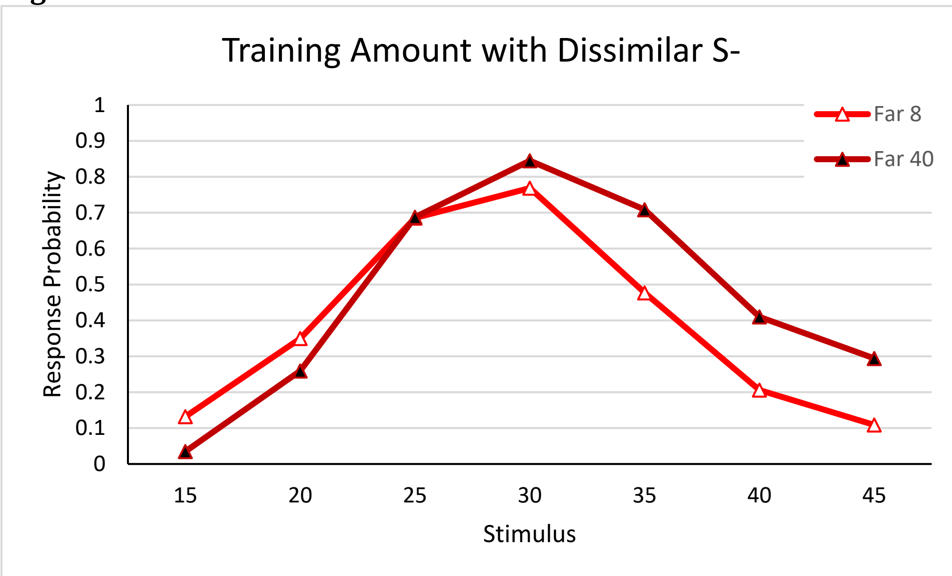
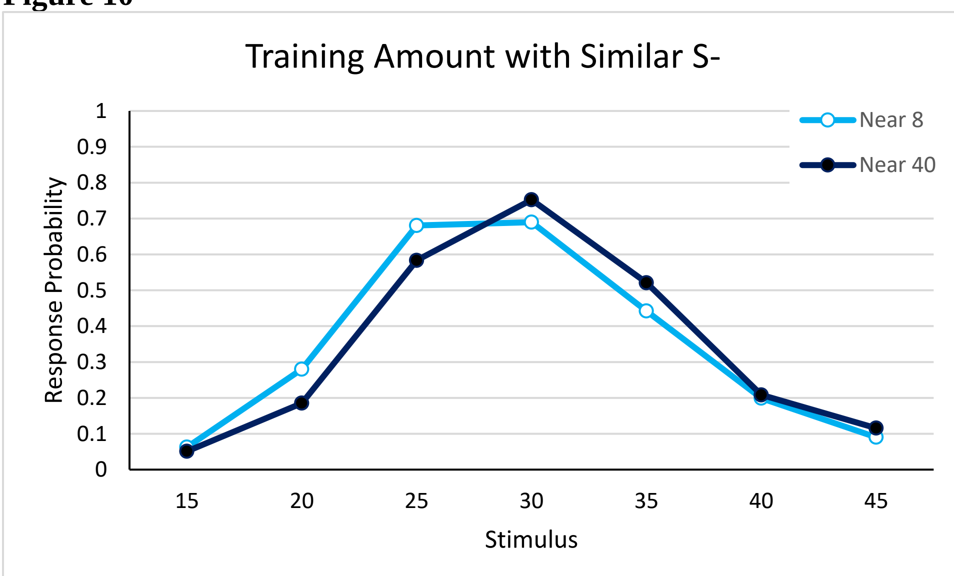


Figure 10



One possible explanation is that perceptions of the stimulus do not correlate linearly with the stimulus score within the range that was used. Specifically, it may be that the difference between S25 and S30 is harder to detect than the difference between S30 and S35. While this would explain the leftward bias in these gradients, it does not match what has been observed in research on perception, which typically shows that uniform differences become harder to detect as the base values increase (Hughes, 2001). This effect is present when one observes the difference in responses between S20 and S25 as compared to S35 and S40. In both the N8 and F8 conditions, responding changes more dramatically at the lower end of the stimulus range than the higher end, leading to steeper slopes and skewed gradients. The reason that both of these gradients exhibit means that are less than the S+, despite their rightward skew, is the high responding to S25. It is therefore likely that their shapes are the result of a combination of non-linear perceptions of difference, and another mechanism that has caused the perceived S+ to shift leftward. The similarity in responding to S25 that occurred in the N8 (M = .68) and F8 (M = .69) groups would suggest that this mechanism is not affected by the value of the S-.

The Near S- condition (Figure 10) also produced gradients that varied depending on the amount of training underwent. Again, the 8-trial condition resulted in a gradient that was shifted left of the S+ and the 40-trial condition, with participants being approximately as likely to respond to S25 (S-) as they were to respond to S30 (S+). However, unlike the Far S- conditions, responding toward S15 and S45 was similar, regardless of the amount of training.

Another explanation for the leftward shift in the 8-trial conditions, which are presented together in Figure 11, is that participants were simply unable to recognize the difference between the S+ and S- in training, either due to inattentiveness toward the stimuli or toward the feedback, leading to conflation of the two. Should this be the case, participants would perceive that they were being presented with the same stimulus in each trial, and that the feedback they were receiving was purely random. This would explain the leftward shift, and would reasonably fit with the N8 group’s training accuracy (which was not significantly greater than chance at the end of training), but not with the F8 group’s training accuracy (which did exceed chance).

Figure 11

As can be seen in Figure 12, the F40 and N40 gradients differed in a number of ways, particularly the rightward bias and overall higher responding that was observed in the F40 condition. Considering the relative similarity of the F8 and N8 gradients, this would suggest that the value of the S- becomes more meaningful when more training is undergone.

Figure 12

Based on these comparisons, it would appear that training amounts and S- similarity interact in several ways. First, under conditions of minimal training, (i.e., 8 training trials), gradients were shifted slightly toward the S-, with the similarity between the S- and S+ having little effect. An important note is that in both cases, the S- was less than the S+. As such, it may be that, rather than showing a shift toward the S-, these gradients demonstrate a shift toward lower values. As previously discussed, this second explanation is difficult to reconcile with what has generally been observed in research on just noticeable differences (Hughes, 2001). However, an area shift in the direction of the S- also does not integrate well with past observations and explanations regarding peak shift. The explanation provided previously (i.e., that participants conflated the S+ and S-) is supported by the effect being somewhat more present in the N8 condition (in which the S+ and S- are more similar, and thus more likely to be mistaken for the same stimulus) than the F8 condition.

In contrast to these, both the N40 and F40 conditions resulted in gradients that showed a rightward bias. As will later be discussed, this difference was largely, but not wholly, due to the increased probability that these conditions would lead participants to develop monotonic response patterns, resulting in sigmoidal gradients. Regardless, this effect was more present in the F40 condition, which also showed greater response probabilities for all stimuli. These observations are in opposition to what would be expected under adaptation-level theory, which predicts that greater training should anchor the adaptation level to its location developed during training, and thus minimize area shift. By comparing the F8 and F40 gradients, it is clear that greater training led to a greater shift in the direction that AL-theory would predict. However, these observations also do not fit well with associative models, which would predict greater shift when the S- is similar to the S+ than when it is dissimilar. Comparing the F40 and N40 conditions reveals that rightward shift was far greater when the S- was dissimilar than when it was similar.

Based on these findings, it would seem that several effects are present, which have not previously been well studied, if observed at all. It is possible that a combination of S+/- conflation, resulting from complex stimuli in conditions of minimal training, and the development of monotonic response patterns, occurring when the amount of training and S- dissimilarity are sufficient for participants to learn the discrimination, would explain these results.

Testing Iteration

As with previous analyses, visual comparisons were also utilized. These involved gradients constructed from responses made at different points in the generalization test, in order to determine how responding changed over the test. Of particular interest were the gradients formed during the first and last iterations, which can be viewed in Figure 13.

Generalization gradients under the F8 condition showed some progressive rightward shift in the peak. Response probabilities peaked at S25 in Iteration 1, and moved to S30 by Iteration 18. There was also some slight leftward area shift over the testing phase.

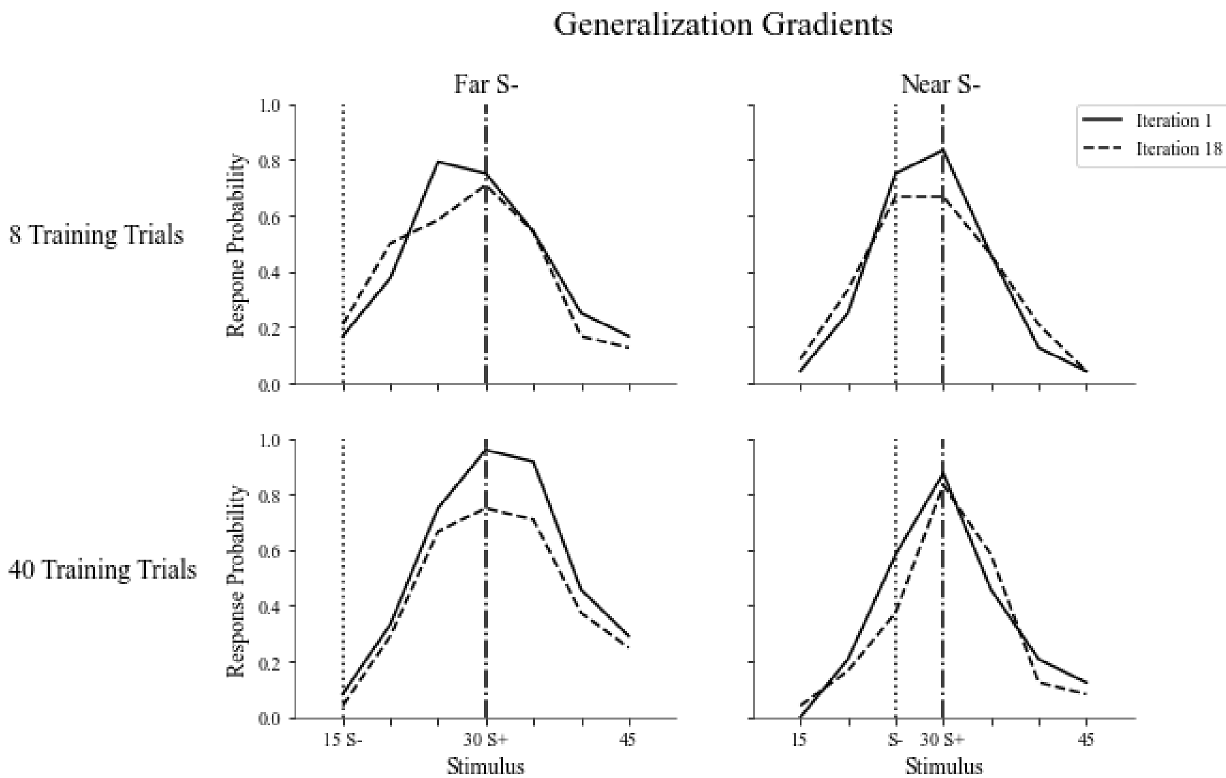
Response probabilities generally decreased for S25 through S40, however this change was not uniform, being the greatest for S25 (Iteration 1: 79.2%, Iteration 18: 58.3%).

Under the F40 condition, response probabilities for each stimulus decreased over the testing phase. This was most strongly observed in S30 and S35, while response probabilities for more extreme stimuli (S15 and S45) changed very little if at all.

N8 was the only condition where a degree of progressive expansion was noticeable in the gradient, though this change was slight. As with conditions F8 and F40, the most salient change was the depression of response probabilities near the S+ (specifically S25 and S30).

N40 was the only condition that showed a degree of progressive rightward area shift, though not peak shift.

Figure 13



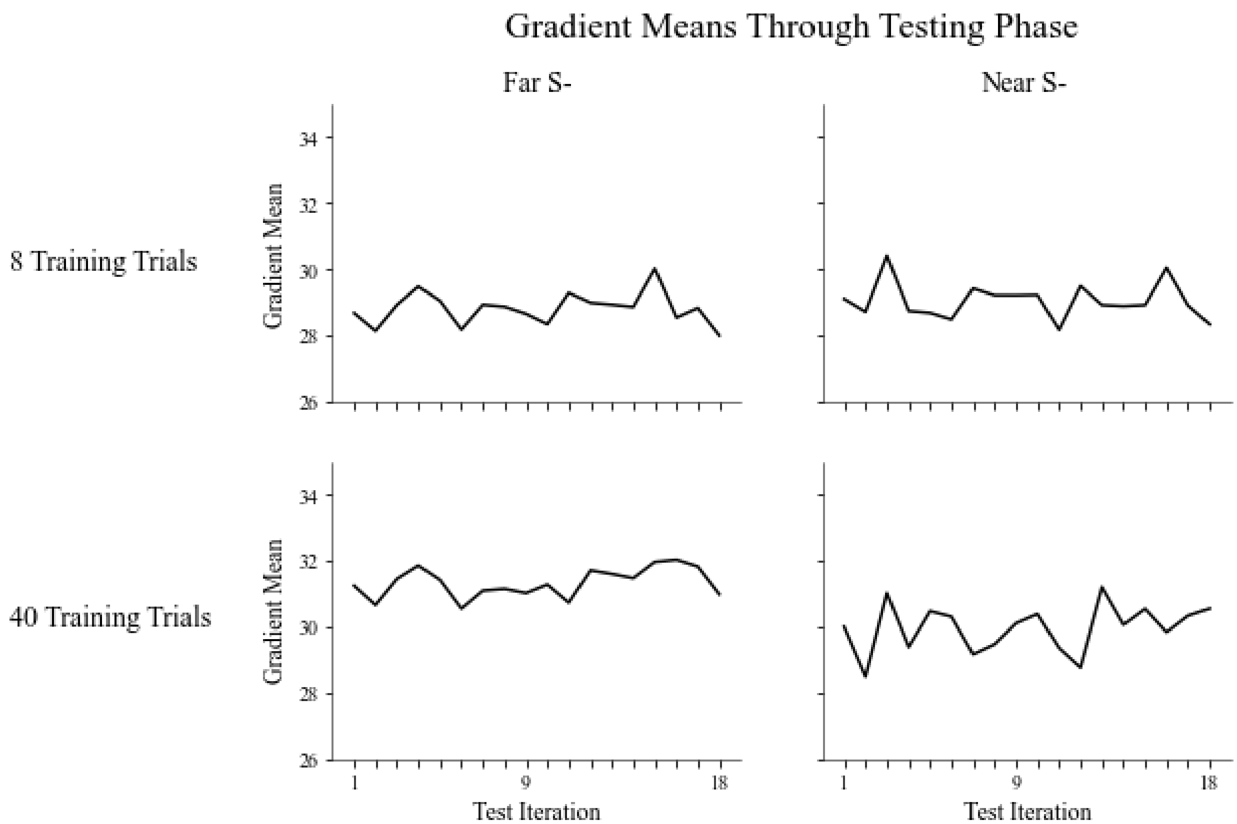
Counter to proposed hypotheses, neither translational shift nor gradient expansion were consistently observed. With the exception the F8 condition, which showed a modal shift from S25 to S30, all of the gradients peaked at the S+ in both Iteration 1 and Iteration 18. This stability is further shown in Figure 14, which shows some difference in gradient means between conditions, but very little change over the course of the testing phase. This type of resilience against peak shift is typically associated with anchoring as a result of prolonged training, but neither 8 training trials nor 40 would be expected to be sufficient for the production of this effect.

One explanation would be that the difference between adjacent stimuli was great enough that the dimension was not adequately sensitive to detect the gradient shift. However, given the difficulty that participants experienced during discrimination training with the S+ (S30) and the relatively similar S- (S25), as well as the relatively wide gradients, it is unlikely that the difference between adjacent stimuli was too great to measure differences.

A more plausible explanation is that the range of the stimulus dimension was not great enough to induce significant shift in the adaptation level. Generally, given an S- with a value of 15 and an S+ of 30 (assuming a symmetrical stimulus range), AL theory would predict a peak shift to S37.5. However, this is typically shown in experiments where the stimulus dimension is wide enough to fully display the generalization gradient. In this case, responding approached but did not reach 0.0 for the most extreme stimuli used. As a result, while the stimulus dimension was wide enough to induce a peak shift that would be measurable under AL theory, the relative difficulty of the discrimination task may have prevented this.

Another possibility is that the confined nature of the stimulus may have limited AL-like effects. Because the bars used within the stimulus were inherently limited in terms of how long they could be, with the borders of the graph being clearly identified, this may have prevented any contextual shift from occurring between the training and testing phase. In effect, because participants could see the absolute minimum and maximum scores that could be represented from the first stimulus presentation, the adaptation level may have been created independently of the stimulus values used in training. This adaptation level, existing as a function of the nature of the stimulus rather than the specific stimuli observed, would then be stable and unaffected by changes in the presented stimuli. Should this be the case, one would expect there to be little difference in the gradients produced in Experiment 1, and those of Experiment 2. As will later be discussed, this was not the case.

Figure 14



Additional Tests

Additional tests were performed based on trends observed in the data. These tests were not planned prior to data collection, and should be treated as exploratory.

Accuracy

S+/- Correlations

As previously discussed, while both S+ and S- accuracies trended upward, improvements in one were often associated with decreases in the other. This is visible in the graphs on the right side of Figure 5, particularly the Near S- condition, where local maxima for the S+ was usually within one stimulus presentation of local minima for the S-, and vice versa (e.g., Stimulus Presentation 4, 8, and 13). To explore the relationship between S+ accuracies and S- accuracies, correlational analyses were run using the 40-trial conditions (due to the lack of data points within the 8-trial conditions), while controlling for the stimulus presentation number. Here, the stimulus presentation number refers to how many instances of the S+ or S- the participant had been exposed to. For example, it would be expected that accuracies toward the 20^{th} S+ would be greater than those toward the 2^{nd} S+, simply due to the development of greater perceptual accuracy. By controlling for the number of trials, this analysis would remove the correlational contribution provided by the learning of participants. Should changing response-conservativeness play a role in the relationship between S+ and S- accuracy, one would expect this analysis to yield a negative correlation. For example, should conservativeness decrease due to recent experiences, one would expect subsequent S+ accuracies to increase and S- accuracies to decrease. It was found that the correlation between S+ and S- accuracy was near zero in the Far S- condition, $r(17) = .045$, $p = .855$, and slightly negative in the Near S- condition, $r(17) = -.279$, $p = .247$. While neither of these correlations were significant, this result should be treated as highly tentative, as this analysis was exploratory in nature, and had a number of limitations. Regardless, it would appear that any apparent relationship between S+ and S- accuracy, separate from general perceptual accuracy, is illusory.

Accuracy and the Generalization Gradients

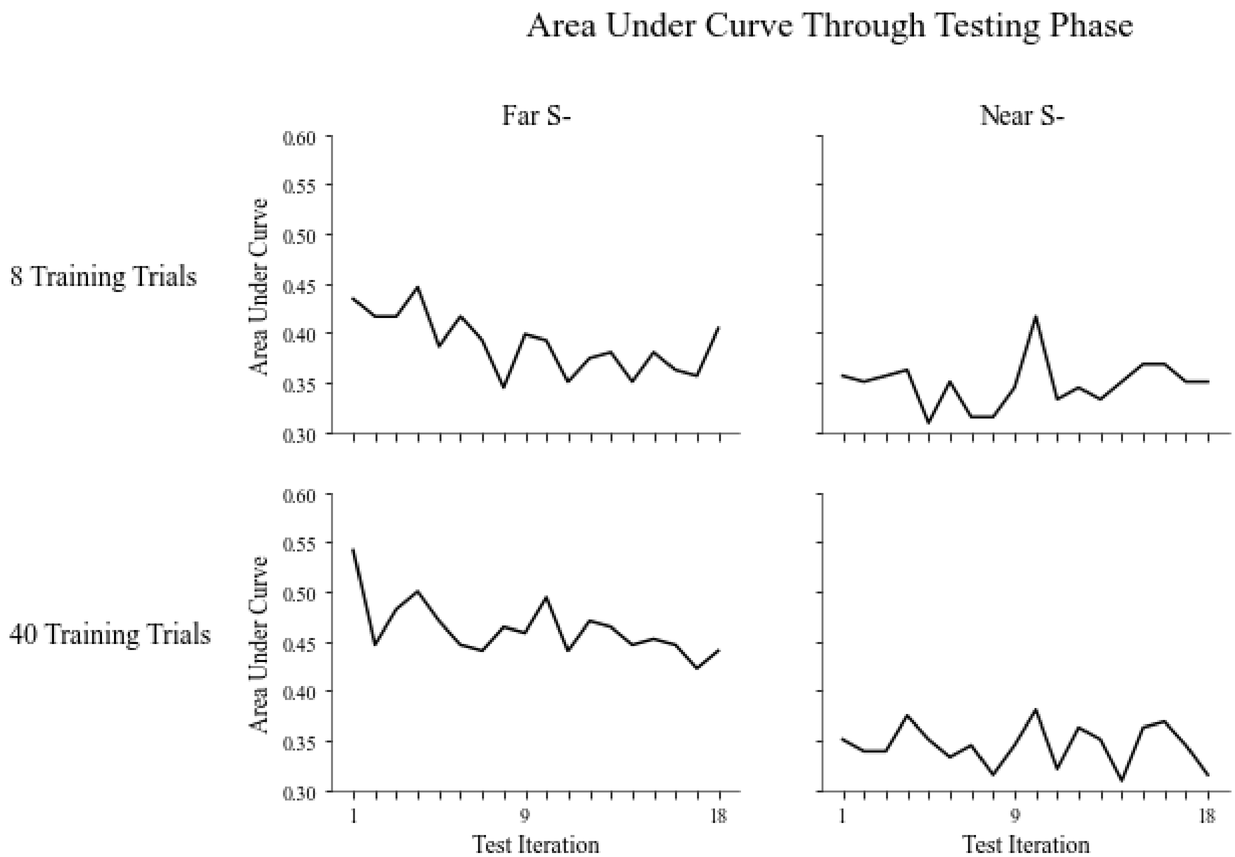
To examine the relationship between discrimination accuracy and the generalization gradients, correlational analyses were performed. Controlling for experimental condition, a significant correlation was found between overall training accuracy and gradient mean ($r = .202$, *p* = .050). This correlation increased to .310 (*p* = .002) when accuracy was limited to the last 4 trials. The correlation between width and accuracy was significant when total accuracy was used (r = -.261, p = .011), but not when only the last 4 trials were used (r = -.160, p = .124). These findings are supportive of the previous explanation that the leftward shift observed in the 8-trial conditions was due to a conflation of the S+ and S-, as one would expect this leftward shift to be more present in participants who were less effective in discriminating between these stimuli.

The absence of a correlation between final accuracies and gradient widths is somewhat more surprising, as it would be reasonable to expect those who are effective in discrimination to exhibit narrower gradients. Further, it appears that this relationship is stronger when all of the training data is used, rather than only using final accuracies. It may be that those who developed relational decision rules over the course of the training phase, resulting in sigmoidal response distributions, were more accurate in discrimination training than those who did not. As a result, the decisional rule developed by the participant would moderate the relationship between accuracy and gradient width. Further investigation is necessary to make this determination.

Area

Due to the apparent trend of progressive gradient depression, gradient areas were analyzed over the course of the testing phase. These are described as the average number of responses within a testing iteration, divided by 7 (the maximum number of responses that could be made), so that areas are represented as proportions between 0 and 1. These can be viewed in Figure 15.

Figure 15



Overall, the number of responses within each testing iteration differed depending on the experimental condition, and decreased as the testing phase progressed. This can be seen in Figures 13 and 15. The condition that resulted in the greatest number of responses was F40, which involved extended training with a dissimilar S-. Over the testing phase, the proportion of positive responses decreased from .542 to .440. Similarly, the F8 conditions resulted in an initial area of .435, which decreased to .405. Training with a similar S- resulted in similar trends, with areas under the N8 condition decreasing from .357 to .351, while the N40 condition showed a decrease from .351 to .315.

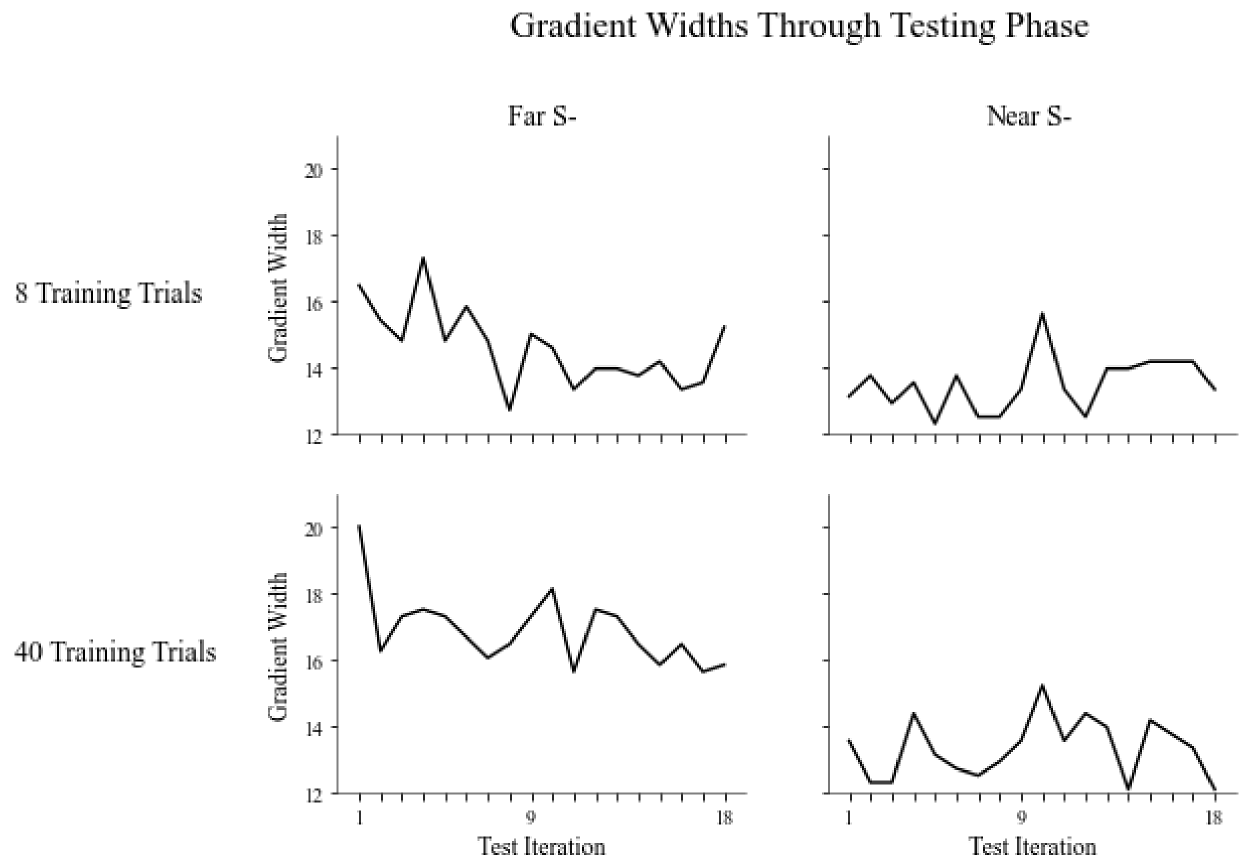
These results are in line with what might be expected given that the generalization test was carried out under conditions of extinction, in which reinforcement was no longer provided.

However, the magnitude of this effect is somewhat surprising. In animal studies, extinction is a logical result of the cessation of reinforcement; as animals are exposed to the S+ without reinforcement of responding, it is reasonable that inhibition would develop and that the animal would eventually stop responding. However, in this case the participants were never being provided with a physiologically salient appetitive stimulus. Rather, the reinforcement was derived from affirming feedback. Further, participants were aware that this feedback would cease once they entered the testing phase. Last, the effort involved in making an affirmative response was no greater than the effort required to make a negative response. Taking these factors into consideration, it seems unlikely that the observed response-depression is a result of extinction per se. Indeed, an examination of the gradient widths suggests that the changing area is likely not due to overall depression of the gradients, but is rather due to the narrowing of gradients that was not visible in the aggregate.

Width

The average widths of the generalization gradients, shown in Figure 16, closely followed trends observed in gradient areas, as widths differed between conditions and across the testing phase. The F40 condition showed the greatest average gradient width, which decreased from 20.00 to 15.83. The F8 condition resulted in an initial average width of 16.46, which decreased to 15.21. The N40 condition showed an initial average width of 15.54, which decreased to 12.08, while the N8 condition resulted in an average width that increased from 13.13 to 13.33.

Figure 16

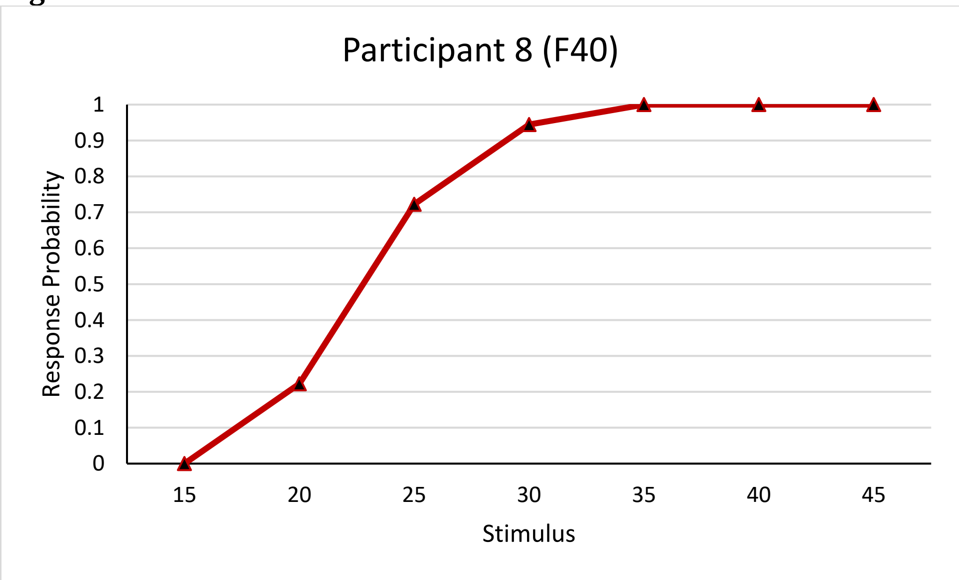


Monotonicity

By examining gradients produced by individual participants, it becomes apparent that the rightward bias within the F40 condition is partially due to some participants developing distinct response patterns. Specifically, participants in the F40 condition were far more likely to develop monotonic response patterns, in which response probabilities increase as the stimulus becomes more extreme, rather than forming peaked distributions. This monotonic responding resulted in sigmoidal gradient distributions, in which probabilities initially increased at an increasing rate, and then increased at a decreasing rate. An example of this can be seen in Figure 17, which comes from a member of the F40 group. Because the majority of participants within the F40 group still showed peaked gradients centered over the S+, the aggregate gradient retains this peaked shape, with the caveat that the response probabilities do not decrease symmetrically, leading to a rightward bias. None of the other groups showed such a noticeable bias. They were, on average, approximately just as likely to make an affirmative response to S15 (M = .08) as S45 (M = .10).

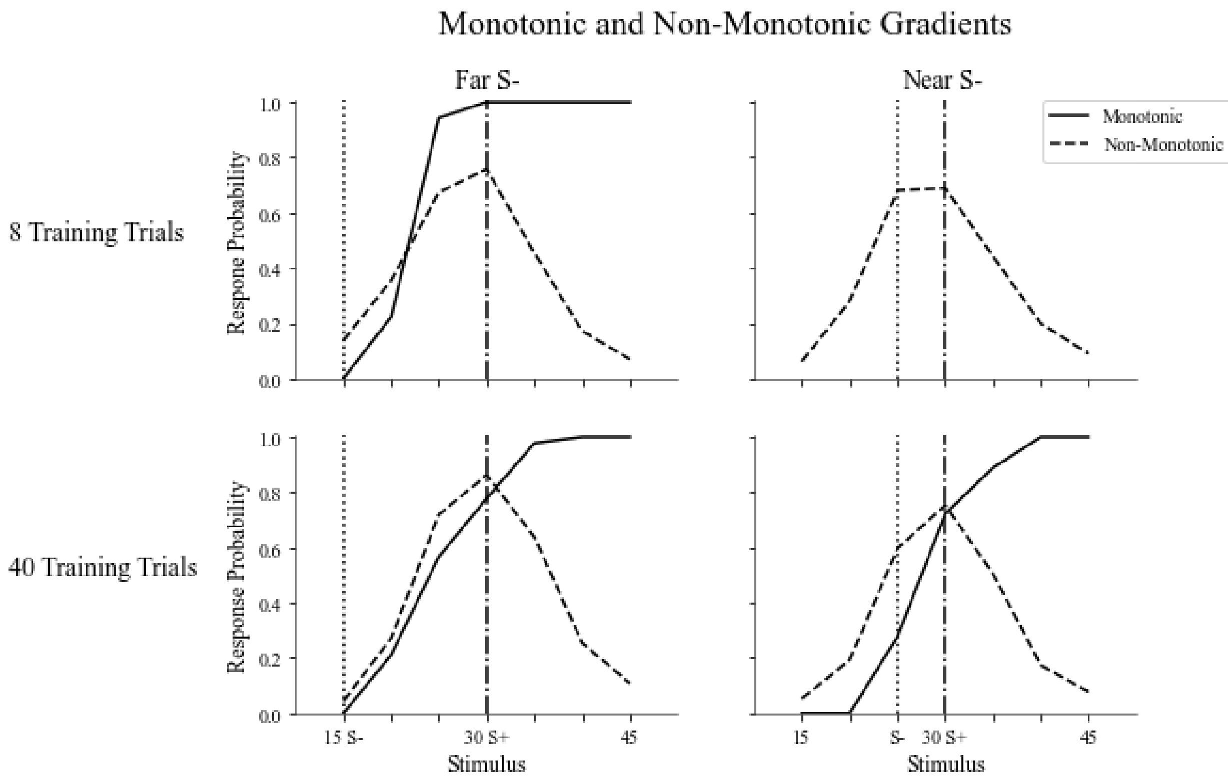
To investigate differences in the prevalence of monotonic responding, participants were classified as either being monotonic or non-monotonic. Monotonic responding was defined as a gradient in which the probability of an affirmative response increased, or did not change, as the stimulus score increased (i.e., the probability of an affirmative response to S45 was greater than, or equal to, the probability for S40, etc.). Overall, 7 participants met these criteria for monotonic responding: 5 from F40, 1 from F8, 1 from N40, 0 from N8. The gradients that result from these conditions can be found in Figure 18.

Figure 17



Note. This figure shows the overall generalization gradient of a participant who underwent 40 training trials using a dissimilar S-. This gradient is provided as an example of monotonic responding.

Figure 18



Note. These gradients were constructed by separating monotonic and non-monotonic responders. Because no participants within the N8 condition showed a monotonic response pattern, this gradient is unchanged.

The data would suggest that monotonic responding was more likely to develop when the S- was dissimilar and when training was more extensive. This is difficult to explain, as relatively little work has been conducted relating to the factors that control whether monotonic or peaked response patterns emerge. One possibility is that participants are more likely to develop definite response rules (e.g., “The greater the stimulus, the more I should respond.”), rather than relative rules (e.g., “The more similar the stimulus is to my target, the more I should respond.”) when the S- is more easily discerned from the S+. This would explain why monotonic responding was more common in the Far S- condition, and why monotonic responding was more common in the F40 group than the F8 group. If monotonic responding develops as a function of discriminability, then it is logical that this response pattern would be more common when the S+ and S- are more dissimilar, and that this would increase as participants became more effective at the discrimination task.

Another possibility is that the variability of the stimulus was the cause. It may be that participants perceived the S+ and S- scores as ranges rather than discrete points, or even that the S+ was perceived as a range while the S- was perceived as a point. If this were the case, these participants could come to the understanding that they should make an affirmative response to any stimulus that is greater than the S-, or at least as great as the S+. Increased training would likely reinforce this understanding, as this would provide more potential instances in which a presentation of an S+ might be perceived to be greater than other instances of the S+. In effect, participants would always be reinforced for making an affirmative response to a stimulus that they perceived to be greater than the S+, leading to the understanding that they should provide this response to all stimuli that are greater than the S+.

These explanations are speculative, as it cann7ot be stated with any certainty that the observed differences between the groups are not due to sampling error or some other confound.

Experiment 2

Method

Experiment 2 followed the general method previously discussed, with two specifications. First, the Far S- was a stimulus with a score of 0, while the Near S- was a stimulus with a score of 20. Second, the generalization test exclusively involved presentations of stimuli with the following scores: 0, 10, 20, 30, 40, 50, 60.

Results and Discussion

Overview

Analyses largely followed what was conducted in Experiment 1. Particular attention was paid to determining whether meaningful differences between the groups emerged during training or testing, and what progressive changes occurred over the course of the generalization test.

Training Accuracy

Accuracies were calculated using the same method described in Experiment 1. On average, accuracies over the first four training trials (M = .471, SD = .235) were not significantly different from chance, $t(95) = -1.196$, $p = .235$. However, this was not true of each condition. The F8 condition showed an initial accuracy (M = .583, SD = .141) that was significantly greater than chance, $t(23) = 18.025$, $p < .001$. Initial accuracies under the F40 condition (M = .531, SD = .112) were also greater than chance, though not significantly, $t(23) = 1.366$, $p = .185$. Initial accuracies under the N8 condition (M = .354, SD = .285) were significantly less than chance, $t(23) = -2.509$, $p = .020$. Accuracies under the N40 condition (M = .417, SD = .282) were initially below chance, but this difference was not significant, $t(23) = -1.446$, $p = .162$. These data follow the same trends observed in Experiment 1, with the Far S- conditions showing initial accuracies above chance, and the Near S- conditions resulting in initial accuracies below chance. However, while initial accuracies in the Far S- conditions were greater than chance in Experiment 1, they were not significantly greater. The difference is almost certainly due to the stimuli used, as the E1 Far S- was S15, while the E2 Far S- was S0.

Final accuracies were, on average (M = .846, SD = .250) greater than chance; to determine whether this was true of all groups, single sample t-tests were conducted. Final accuracies under the F8 condition (M = .927, SD = .116), were significantly greater than chance, *t*(23) = 18.025, *p* < .001. Those of the F40 condition (*M* = .979, *SD* = .102) were also greater than chance, *t*(23) = 23.000, *p* < .001. Final accuracies under the N8 condition *(M* = .604, *SD* = .312) were greater than chance, though this difference was not significant, *t*(23) = 1.635, *p* = .116. Those of the N40 condition (*M* = .875, *SD* = .221) were significantly greater than chance, *t*(23) = 8.307, *p* < .001.

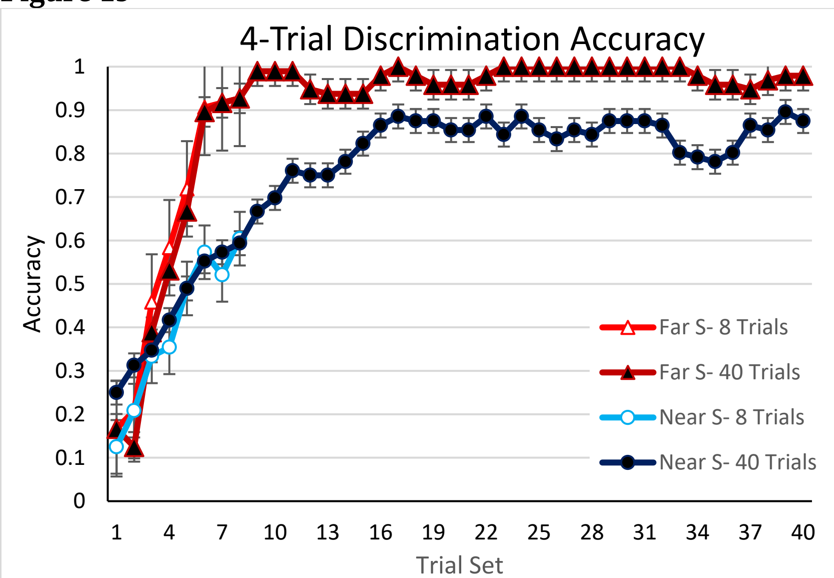
As with Experiment 1, although the final accuracies of the N8 condition were not significantly greater than chance, a paired samples t-test using initial (*M* = .354, *SD* = .285) and final *(M* = .604, *SD* = .312) accuracies revealed that significant improvement did occur, *t*(23) = 4.153, *p* < .001.

Over the course of the training phase, these accuracies (calculated using the 4 most recent trials) formed learning curves (visible in Figure 19) that were similar to those of Experiment 1.

To determine whether apparent differences were statistically significant, a two-way ANOVA was conducted using the final accuracies, which were collected over the last 4 training trials. Assumptions of homogeneity of variance were not met, however observed differences were large enough that this is unlikely to meaningfully affect analyses. Regardless, this should be considered as a potential source of error.

It was found that accuracies under the Far S- (*M* = .953) were significantly greater than those of the Near S- (*M* = .740), *F*(1, 92) = 25.724, *p* < .001, and accuracies under the 40-training trial condition (M = .927) were significantly greater than those of the 8-trial condition (*M* = .766), *F*(1, 92) = 14.706, *p* < .001). Last, a significant interaction was observed between S-similarity and training amount, *F*(1, 92) = 6.749, *p* = .011. This interaction is due to differences in the amount of improvement that occurred as a result of greater training. While both the Far and Near S- conditions were more accurate after experiencing 40 training trials than 8, the difference between the accuracies of the F8 condition (*M* = .927) and the F40 condition (*M* = .979) was less than the difference between the N8 condition (*M* = .604) and the N40 condition (*M* = .875). This interaction can be attributed to a ceiling effect; because discrimination was so easy in the Far S- condition, participants approached the maximum accuracy (1.0) within the first 8 trials. The Near S- condition, on the other hand, was more difficult, and participants thus had more room to improve between the 8th training trial and the 40th.

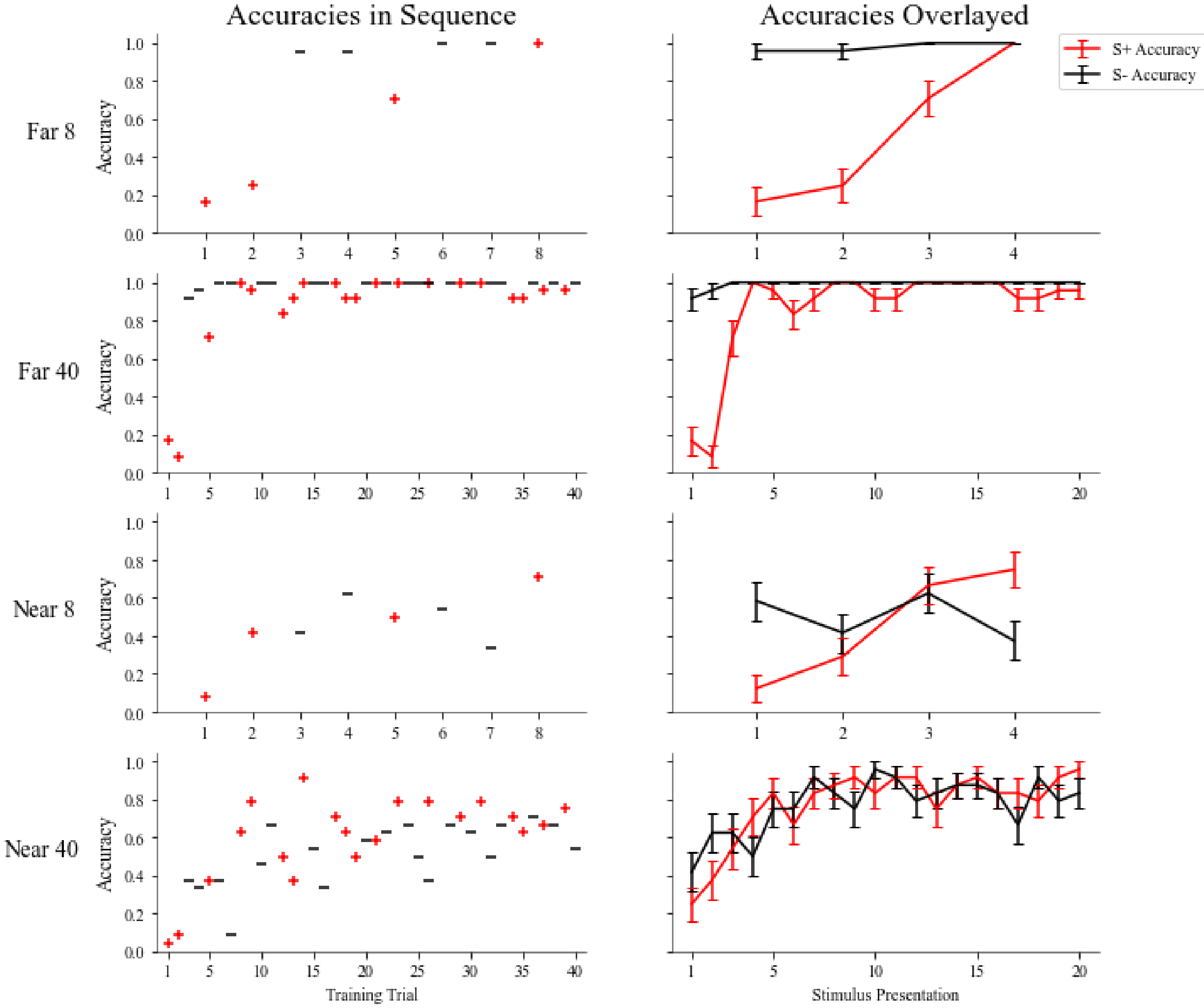
Figure 19



Note. Accuracies are presented as running averages of the four most recent training trials. For example, Trial Set 2 represents accuracies in response to trials 1 and 2, while Trial Set 40 represents average accuracies in response to trials 37 – 40.

To examine differences between the groups, general accuracies were broken down into S+ and S- accuracies. These can be viewed in Figure 20.

Figure 20
S+/- Accuracies in Sequence and Overlayed



Note. The left set of graphs represents the training sequence of each condition, in which the stimulus presentation order was the same for all participants. The right set of graphs shows the same data, but with accuracies toward successive presentations of the S+ and S- overlayed, so that trends may be more easily compared.

One noteworthy difference is in the S+ accuracies of the Far S- and Near S- conditions. The N40 group did not achieve an accuracy greater than .9 until the 9th presentation of the S+, and failed to meet this same accuracy in 6 of the following S+ trials. In contrast, the F40 condition achieved an accuracy of 1.0 with the 4th presentation of the S+, and this group only showed an accuracy below .9 in response to 1 of the following S+ presentations. While these groups differed in terms of the S- stimuli that they were shown, the S+ stimuli were identical between the groups. This lends support to the explanation, mentioned in the Results and Discussion section of E1, that participants adjust conservativeness depending on recent experience. These data suggest that participants quickly become more liberal in terms of what they consider to match the S+, when the S- is extremely different. Should this be the case, generalization gradients would be wider under the Far S- conditions than the Near S- conditions, particularly when more training is undergone, as this would provide more opportunities to increase response-liberalness. As will later be discussed, data presented in Figures 21 and 22 support this interpretation.

Another interpretation is that, when the S- is very dissimilar to the S+, participants begin to avoid the S-, rather than seeking out the S+. Should this be the case, one would expect participants to make an affirmative response to any stimulus that is noticeably different from the S-, regardless of whether that stimulus is similar to the S+. During the generalization test, this type of avoidance would result in the monotonic response pattern that was previously discussed. The data presented in Figure 33, which will later be discussed, would support this conclusion as well.

In short, in the context of this experiment, it is likely that the S- influenced responding to the S+ through a combination of changing conservativeness and a switch from S+ based responding to S- based responding.

Generalization Test

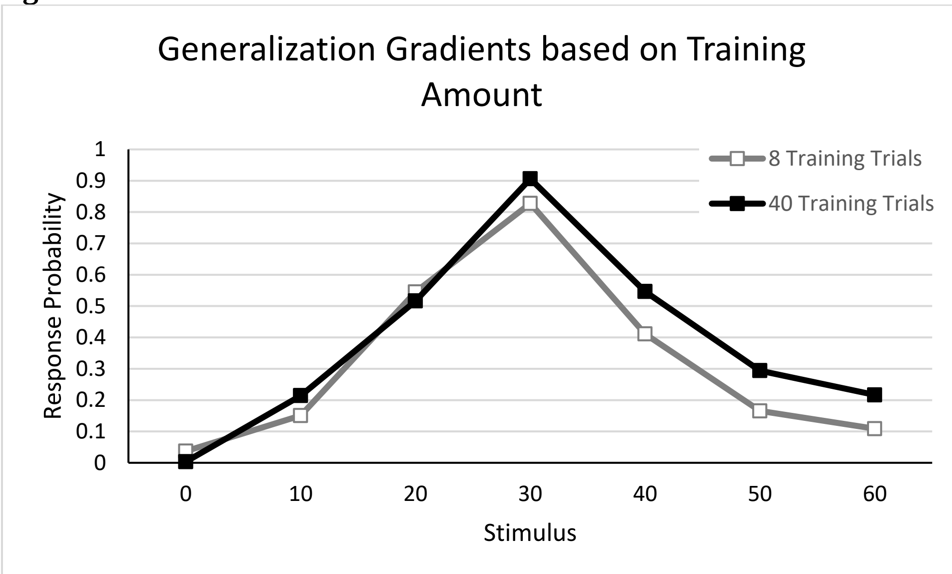
Data were analyzed in the same manner as described in Experiment 1, using a repeated measures MANOVA, with gradient widths and gradient means serving as dependent variables. Through preliminary analyses, it was found that assumptions of normality were not met, however MANOVA are relatively robust against violations of this assumption. One participant was identified as a significant outlier (Mahalanobis distance = 17.24, critical cut-off = 13.82), and was excluded from the analysis. The correlation between the dependent variables (r = .496) was within acceptable range (.2 - .9).

Amount of Training

MANOVA results showed that the effect of S- similarity approached, but did not reach, significance, F (2, 90) = 2.945, p = .058; Wilk's Λ = 0.939, partial η^2 = .061. As such, this effect was not subjected to follow-up statistical tests. However, as with Experiment 1, a number of differences were visually apparent within the generalization gradients.

When data was combined with respect to S- location, gradients produced after 40 training trials showed a rightward area shift that was not present in the 8-trial condition. The peaks of both the 8-trial and 40-trial conditions were located over S30 (S+), and the gradients were nearly identical to the left of this stimulus. However, the 40-trial condition showed a slight rightward bias that the 8-trial condition did not. As is visible in Figure 21, the 8-trial condition resulted in a gradient that was roughly symmetrical. In this condition, some differences were visible between response probabilities toward corresponding stimuli, such as S20 (M = .545) and S40 (M = .411), but participants were generally as likely to respond to stimuli below the S+ as they were to respond to stimuli above the S+. This resulted in a gradient mean (M = 29.614) that was very near the center of the stimulus range. However, the 40-trial condition is less symmetrical, with response rates declining more slowly on the right side of the gradient than the left. As a result, the 40-trial gradient mean (M = 31.949) is slightly greater than the center of the stimulus range. No other noticeable differences between the gradients existed.

Figure 21



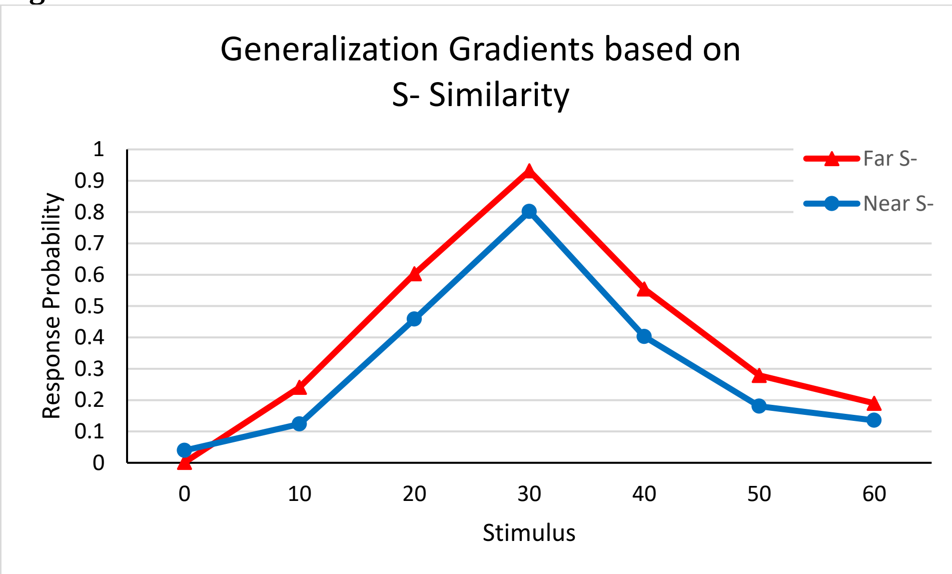
Note. Generalization gradients were constructed by merging the conditions with respect to S-similarity, so that only the effect of training amount is shown.

As was observed in Experiment 1, this difference in the gradients is primarily due to the higher probability of participants developing a sigmoidal gradient, in which participants are more likely to respond to higher-value stimuli, regardless of whether those stimuli are greater than the S+.

S- Similarity

The location of the S- during training did not have a significant effect on either gradient means or widths, F (2, 90) = 1.657, p = .197; Wilk's Λ = 0.964, partial η^2 = .0036. Again, replicating the results of Experiment 1, the modes and means of the gradients were approximately the same. However, as shown in Figure 22, responding was generally lower following training with the Near S- than the Far S-. This difference was somewhat less present at the extreme ends of the stimulus range, as participants were not likely to respond to S0 or S60 regardless of previous S- similarity. Participants who had been trained using the Far S- were consistently more likely to make an affirmative response to all other stimuli.

Figure 22



Note. Generalization gradients were constructed by merging the conditions with respect to training amount, so that only the effect of S- similarity is shown.

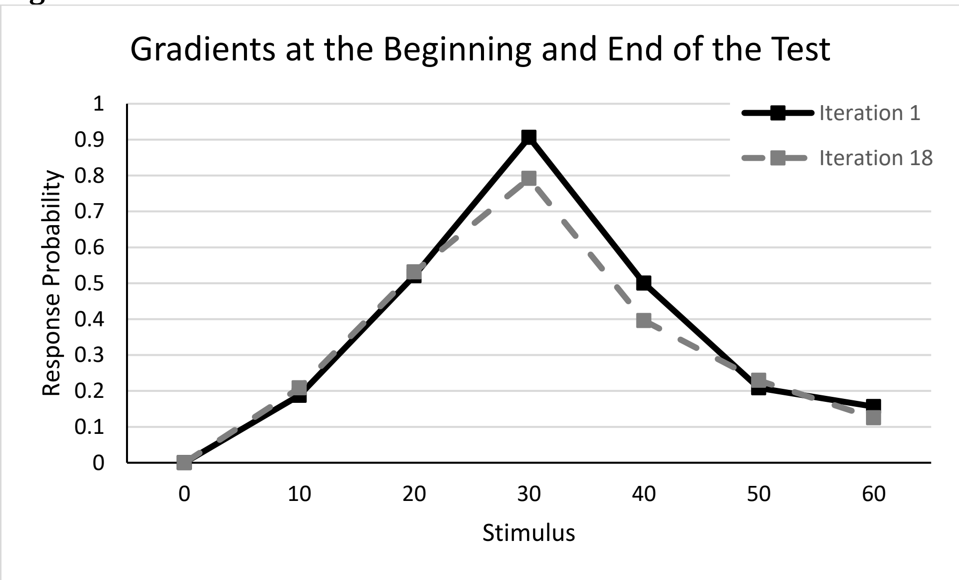
Testing Iteration

The main effect of testing iteration was not significant, F (34, 58) = 1.492, p = .089; Wilk's Λ = 0.533, partial η^2 = .467. While there was no progressive shift in the gradient mean over the course of the test, and changes in the gradient widths were not significant, Figure 23 shows that there was a slight decrease in responding between the first and last trials. This difference only occurred for S30 and S40.

Simple Effects and Interactions

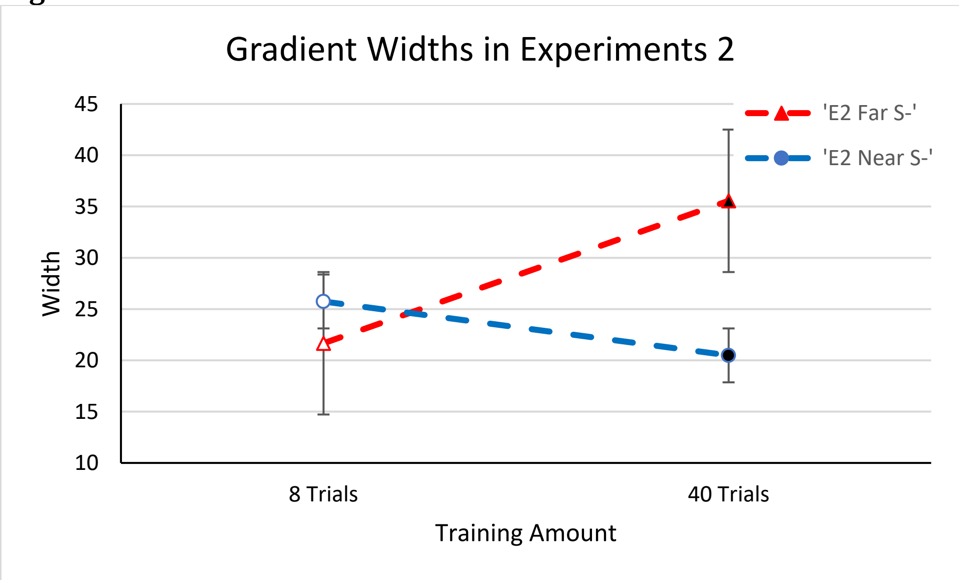
MANOVA results showed a significant interaction between S- similarity and the amount of training, F (2, 90) = 6.570, p = .002; Wilk's Λ = 0.873, partial η^2 = .127. As such, homogeneity of variance was tested for both dependent variables using Levene's F test. It was found that these assumptions were satisfied for all gradient means, however gradient widths showed significant violation of the assumption within Iterations 3, 5, 9, 15, and 18. Because this test is relatively robust against violations of this assumption, follow-up ANOVAs were conducted. These revealed a significant effect on the gradient width, F (1, 91) = 10.142, p = .002, partial η^2 = .100. Under the Far S- condition, the 8-trial condition (M = 21.667) resulted in narrower gradients than the 40-trial condition (M = 35.556). Conversely, under the Near S- condition, the 8-trial condition (M = 25.741) resulted in wider gradients than the 40-trial condition (M = 20.942). These differences are visible in Figure 24.

Figure 23



Note. Generalization gradients were constructed by merging all conditions to display overall differences in responding between the first and final iterations.

Figure 24



As with Experiment 1, qualitative analyses were performed between the generalization gradients that resulted from the different training conditions.

Figure 25 shows that, when the S- is dissimilar to the S+, increased training leads to an overall increase in responding, as well as the same rightward bias that has previously been observed. In this case, the S- had a score of 0 (no bars were visually present); due to the restricted nature of the stimulus dimension, this means that the S- was as dissimilar as was possible with the given S+. This is noteworthy because participants could not reasonably be expected to confuse the S- for the S+, regardless of the amount of training underwent. Indeed, in classifying the S-, both the F8 and F40 groups achieved accuracies of 1.0 by the time the third S- was presented. Considering this, it is unlikely that the differences between the gradients are simply due to different discrimination proficiencies developed after different amounts of training. Thus, these data provide support for the notion that learning does not cease once participants achieve maximum discrimination proficiency, and accuracy alone is not sufficient to predict the characteristics of generalization gradients.

Figure 25

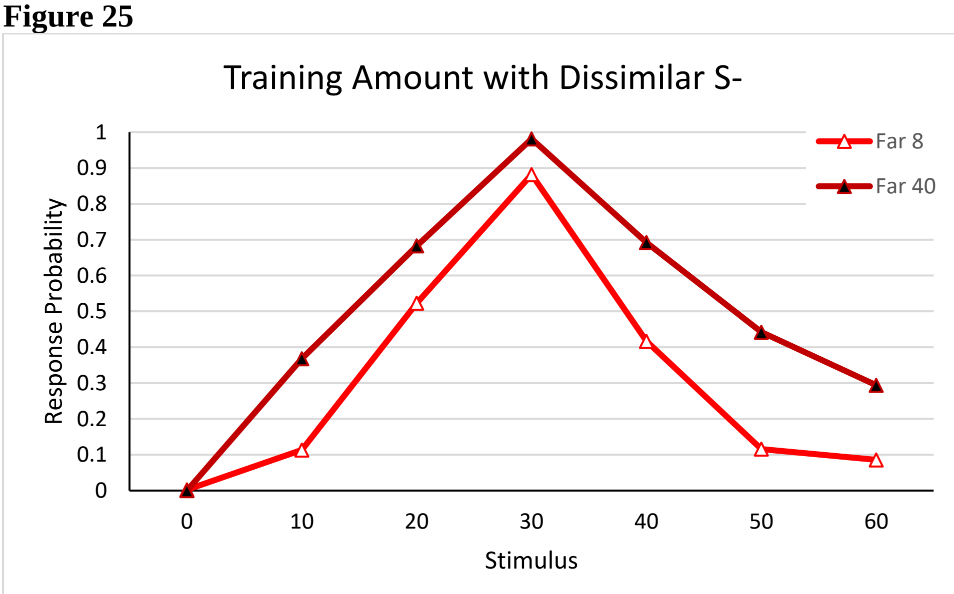
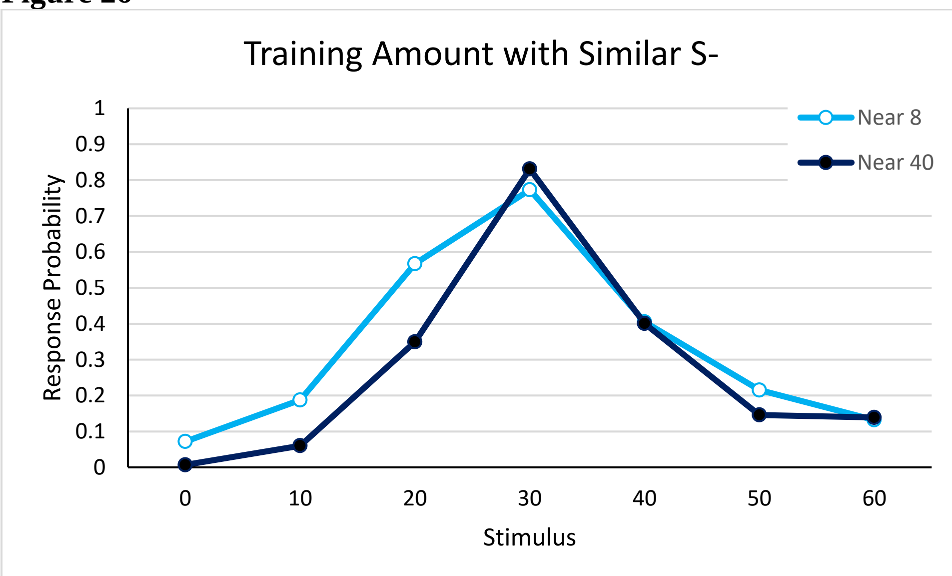


Figure 26 shows that the N8 condition resulted in a gradient with a slight leftward shift, while the N40 condition was roughly symmetrical. While the N8 condition shows the same leftward bias observed in the corresponding group of Experiment 1, supporting the interpretation that conflation of the S+ and S- may have occurred in some participants, the peaks of both gradients are definitively located at S30.

Figure 26

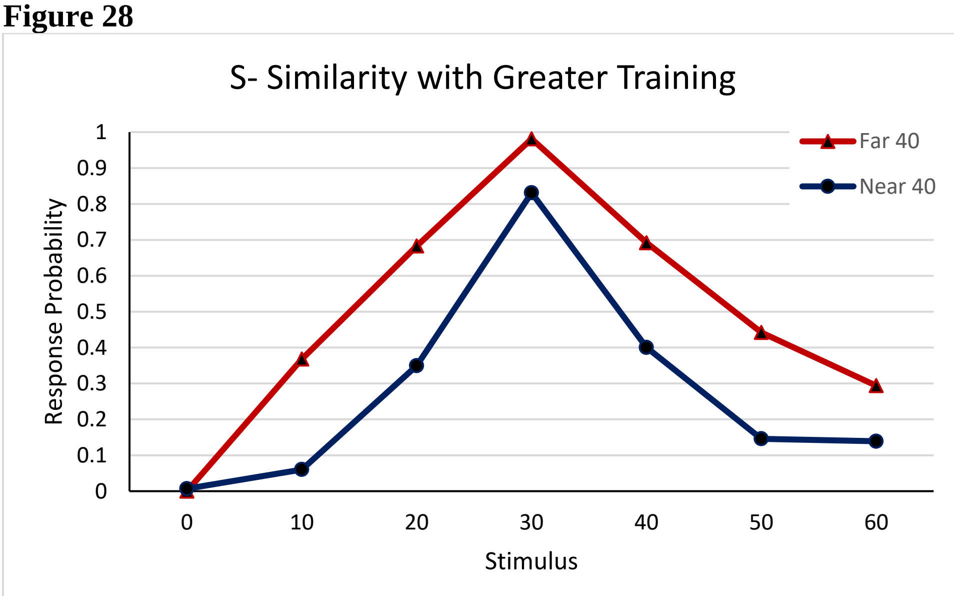


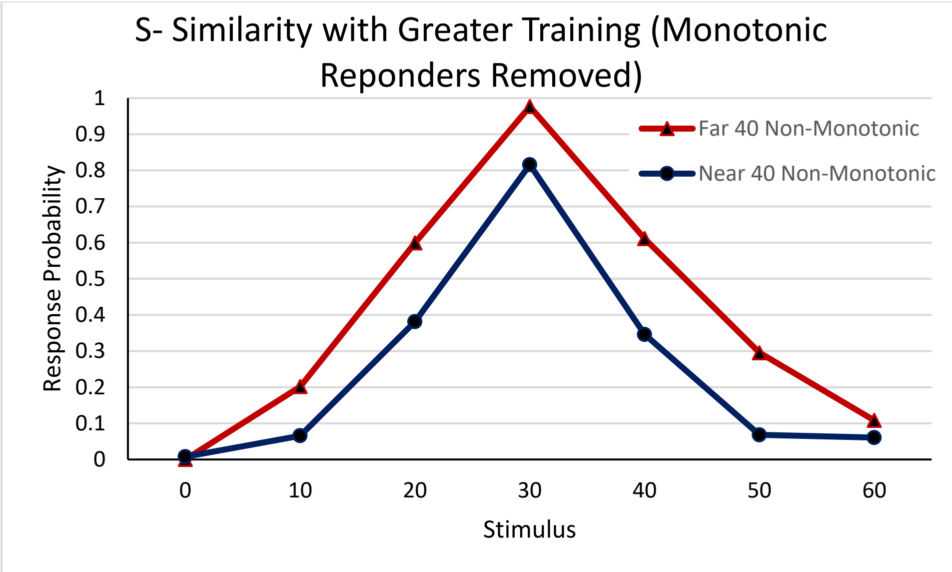
The F8 and N8 gradients, shown in Figure 27, are largely similar. The greatest difference being that the N8 gradient is somewhat lower and wider. Both gradients show a clear peak over S30.

The F40 and N40 gradients, shown in the first graph of Figure 28, represent the clearest contrast of these simple comparisons. While both of these show the rightward bias that has previously been described, there is far more generalization within the F40 condition than the N40 condition. This difference is somewhat surprising considering that the N8 and F8 gradients were highly similar in terms of generalization, with the N8 gradient being slightly wider. As visible in the second graph of Figure 28, this difference persists even when monotonic responders are excluded from the gradients, though to a lesser extent.

Figure 27

Figure 28





Note. *The top graph displays the gradients for all participants, while the bottom graph excludes participants who met criteria to be considered monotonic responders (response probabilities always increased or did not change as stimulus scores increased). This was done to show how the gradients differed in ways that were not attributable to the greater presence of monotonic responders in the F40 condition.*

Testing Iteration

As with previous analyses, visual comparisons were also made between the beginning (Iteration 1) and end (Iteration 18) of the generalization test in order to determine how responding changed. These gradients can be viewed in Figure 29.

The F8 and N8 gradients both showed some progressive depression in responding near the S+. The only change occurring in the F40 gradient was a decrease in responding toward S40, while the N40 gradient remained essentially unchanged. Neither translational shift nor gradient expansion can be said to have taken place to any noticeable extent.

The gradient means, presented in Figure 30, would suggest some progressive leftward shift in the gradient means, though the variation from one iteration to the next, combined with the slightness of this change, makes it difficult to say whether this effect is real and meaningful.

Figure 29

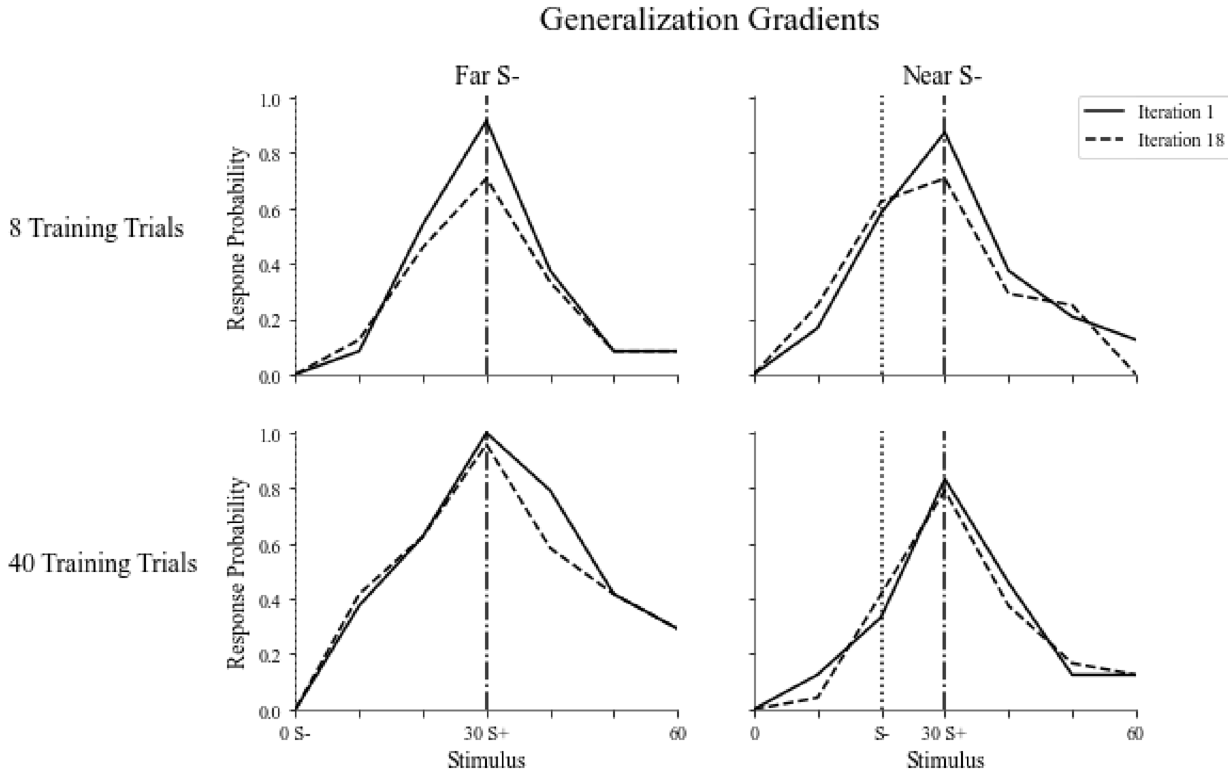
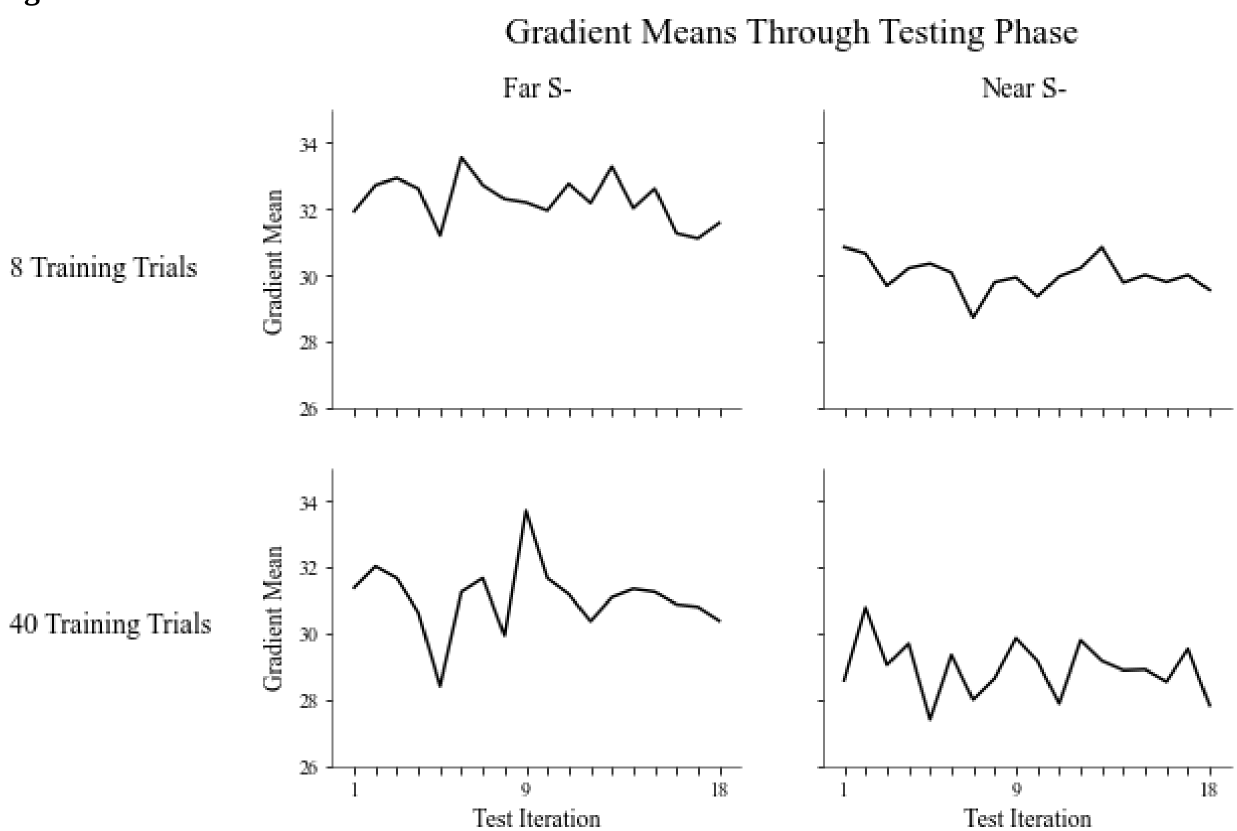


Figure 30



Additional Tests

To facilitate comparisons, the same additional tests that were performed in Experiment 1 were also performed her. Again, these tests were not planned prior to data collection, and should be treated as exploratory.

Accuracy

S+/- Correlations

Correlational analyses were performed using the S+ and S- accuracies for the 40-trial conditions. It was found that the correlations produced under both the F40, $r(17) = .850$, $p < .001$, and N40, $r(17) = .566$, $p = .012$, conditions were significantly positive. These results support the conclusion, described in the Results and Discussion section of Experiment 1, that the relationship between S+ and S- accuracies is not well explained by changing conservativeness alone. If this were the case, these correlations would be expected to be negative, as a change conservativeness would affect subsequent S+ accuracies in an opposite manner to the effect on S- accuracies. In this case, the significant positivity of the correlations would suggest that a shared factor, likely simple perceptual accuracy, is the primarily responsible for the relationship between these variables.

Accuracy and the Generalization Gradients

To examine the relationship between discrimination accuracy and the generalization gradients, correlational analyses were performed. Controlling for experimental condition, it was found that the correlations between overall training accuracy and gradient means were not significant ($r = -.059$, $p = .570$), nor were the correlations when only accuracies over the last 4 trials were used ($r = .052$, $p = .618$). The correlation between width and accuracy was significant when total accuracy was used (r = -.283, p = .006), and when only the last 4 trials were used (r = -.229, p = .026).

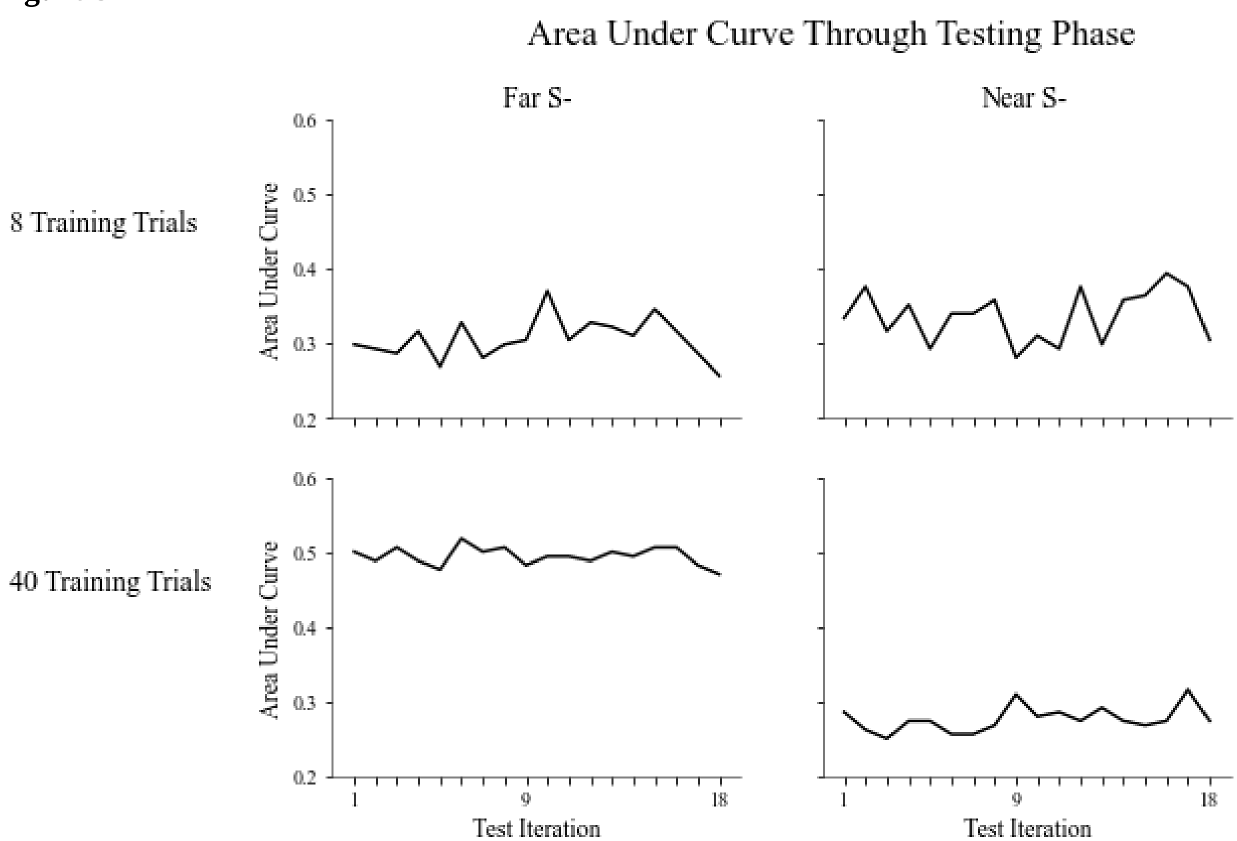
In contrast to Experiment 1, no relationship between accuracy and gradient mean was observed. It is possible that this is due to the relative ease of the discrimination tasks in Experiment 2, which may have made conflation of the S+ and S- less likely. However, it is also possible that the wider stimulus range used in E2 may have played a role, as these gradients would be less sensitive to small changes.

The relationship between accuracy and gradient width is far more similar to what was observed in E1, as a negative relationship was observed between accuracy and gradient width, and this relationship was stronger when all of the training trials were taken into account, as opposed to only the last 4 trials. This is somewhat surprising, as it would be reasonable to expect that final accuracies would be more relevant to performances in the generalization test than accuracies that include earlier trials. One possibility is that this is due to limitations that come with using only 4 training trials to calculate accuracy, as each participant's accuracy could only be 1 of 5 values (i.e., .0, .25, .5, .75, or 1.0). It is also possible that experiences earlier in training are important in determining gradient width. That is, it may be that gradient widths are not only predicted by a participant's perceptual accuracy at the end of training, but also the number of errors that occurred during the development of that accuracy. Because the correlation was more strongly negative when all training trials were used to calculate accuracy, this interpretation would suggest that participants who attained high accuracies quickly tended to produce narrower gradients than those who attained high accuracies slowly.

Area

As shown in Figure 31, little progressive change occurred in gradient areas over the course of the testing phase. Under the 40-trial condition, areas were somewhat more consistent from one trial to the next, but the 8-trial gradients were also relatively stable. Of the four conditions, the F40 areas are noticeably greater than the remaining three, which did not differ meaningfully.

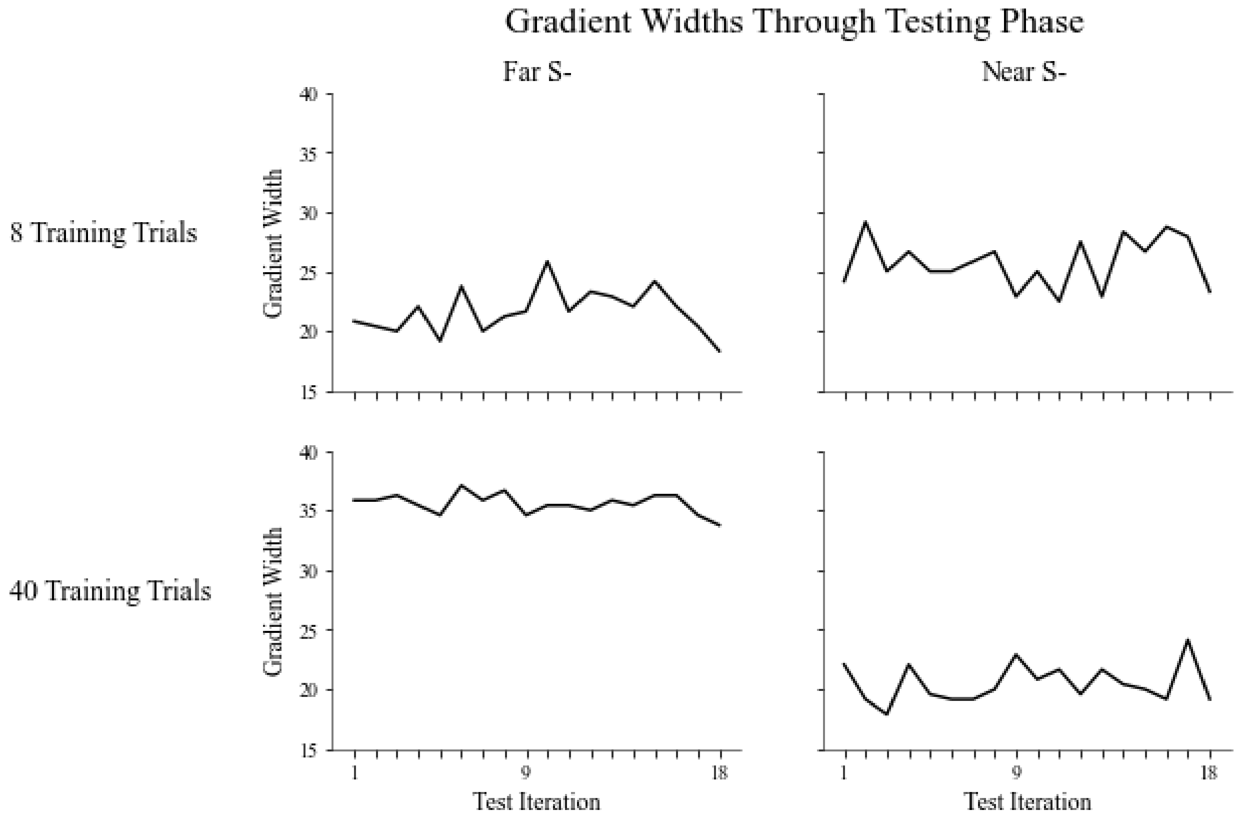
Figure 31



Width

As was observed in Experiment 1, trends in gradient widths (Figure 32) closely matched those of gradient areas. In this case some differences between experimental conditions were observed, but little change occurred over the course of the generalization test. The F40 group, in particular, showed wider gradients than the other groups. The N8 group showed slightly wider gradients than the N8 and N40 groups, which did not differ noticeably.

Figure 32



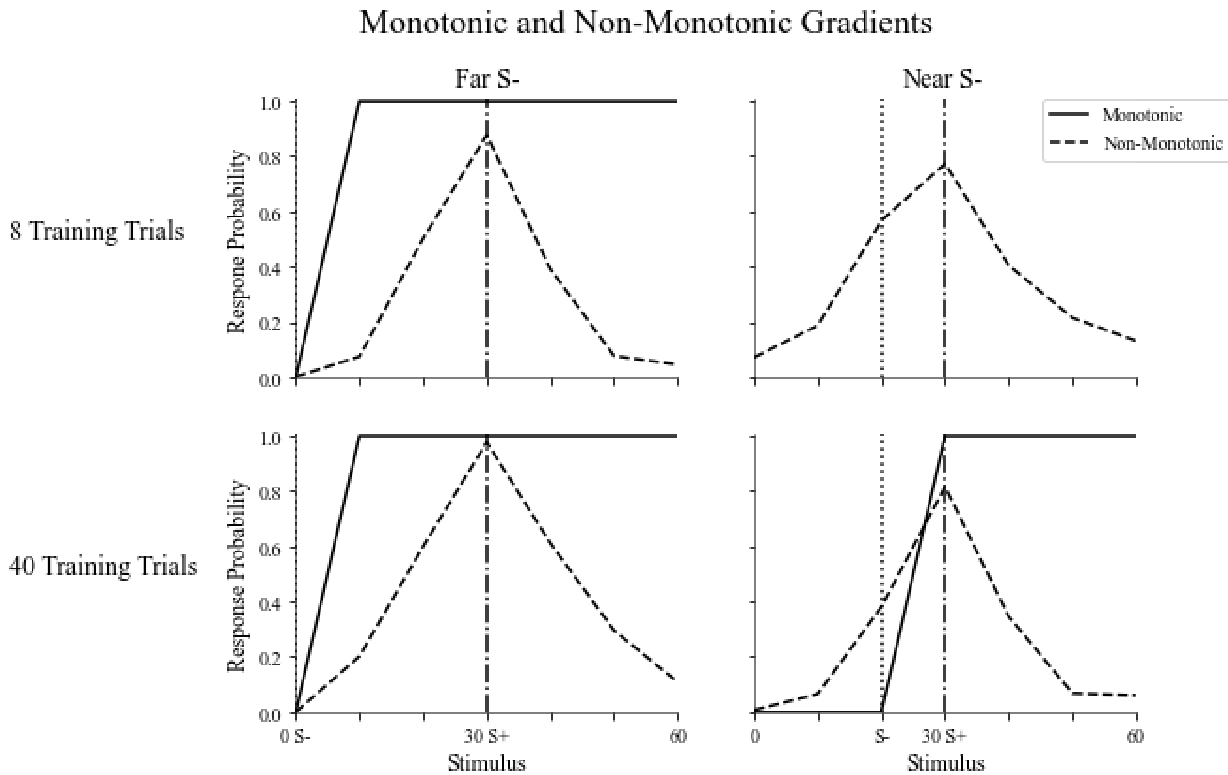
Monotonicity

The number of participants who developed monotonic response patterns was the same in E2 as was observed in E1. Specifically, 7 participants met the criteria for monotonic responding: 5 from F40, 1 from F8, 1 from N40, 0 from N8. The gradients that result from these conditions can be found in Figure 33.

While the prevalence of monotonic responders was the same in E2 as E1, the gradients were noticeably different. Rather than changing gradually, the response probabilities of the monotonic participants in E2 were either 0 or 1.0, without any intermediate responding. Within the Far S- condition, only the S- (S0) resulted in a response probability of 0.0, while all other stimuli received affirmative responses in all presentations. This suggests that these participants were utilizing a categorical response rule based on the S- rather than one based on the S+, as it is unlikely that these participants perceived S10 to be the same as S30 (S+). As such, it would seem that, even among monotonic responders, there are distinct types of responding that may be more or less likely to develop. In this case, all participants who developed monotonic responding also followed an “all or nothing” rule that was not present in the monotonic responders of E1. While this might be partially attributable to the greater sensitivity afforded in E1 by the narrower testing range, it is unlikely that this is the sole reason for the difference. The consistency of all-or-nothing responding among E2’s monotonic participants, in contrast with E1’s total absence of this response-type, would suggest that the S- (particularly S0) is a primary factor. It is possible that this is simply due to the ease with which S0 could be identified, but the more plausible explanation may be that the type of stimulus is also relevant. In cases where single-stimulus reinforcement training has been carried out using stimuli such as a light, in which the light was either illuminated or not, all-or-nothing responding was not seen during the generalization test (e.g., Switalski et al., 1966). In these cases, a light that is off would presumably be as easy to discriminate from the S+ as S0 was in the present experiment. As such, it is likely that some factor, such as stimulus-complexity, is important.

Regardless, as was seen in Experiment 1, these data suggest that participants are more likely to develop monotonic response patterns when the S- is dissimilar, and when 40 training trials are undergone (rather than 8).

Figure 33



General Discussion

Background and Goals

Animals and humans that have been conditioned to provide a particular behavior in the presence of a specific stimulus (S+) will often demonstrate the same behavior in the presence of similar stimuli (Hanson, 1957, 1959); this is known as generalization. Shown graphically, with the stimulus dimension presented along the X-axis and the measure of responding along the Y-axis, responding will peak over the S+, and decrease symmetrically as the stimuli become more dissimilar to the S+. These distributions of responding are known as "generalization gradients".

However, it has often been observed that when an unreinforced stimulus (S-), which exists along the same dimension as the S+, is shown during conditioning, stimulus generalization can be affected. Specifically, humans and animals tend to demonstrate the conditioned behavioral response more often in the presence of novel stimuli, which are less similar to the S- than the S+ is (Hanson, 1959). In effect, the generalization gradient will no longer be centered at the S+, and will instead be displaced toward the side of the S+ that is opposite of the S-. This phenomenon is known as peak shift (Hanson, 1959).

Adaptation-level theory has been proposed to explain peak shift (Helson, 1947; Thomas, 1993). This theory suggests that participants create a reference point (AL) between the S+ and S- during training, which they use to form decisional criteria. This AL represents the average of encountered stimuli, and is therefore subject to change when new stimuli are encountered. As such, the AL shifts during the generalization test to the new average stimulus, which in turn shifts the generalization gradient.

Early on (e.g., Doll & Thomas, 1967; Thomas & Jones, 1962) Thomas found evidence that the adaptation level controls human responding during the generalization test. Further research provided an abundance of evidence that one could greatly alter the generalization gradient simply by altering which stimuli from the dimension appear during the test and how frequently each appears – at least in research using relatively simple stimulus dimensions (Bizo & McMahon, 2007; Hebert & Capehart, 1969; Hebert et al., 1974; Thomas et al., 1985; Thomas et al., 1992). However, Thomas did not produce a comparable body of support for other predictions made by adaptation-level theory or show that the theory works equally well with other, more complex sets of stimuli. When others have attempted to examine other aspects of adaptation-level theory or examine adaptation-level theory with different stimuli than the ones used in early research, the result has been at best mixed support for adaptation-level theory (Derenne, 2006; Gallagahar et al., 2020; Spetch et al., 2004; Verbeek et al., 2006). Further, Ghirlanda & Enquist (2007) have shown that adaptation-level theory may not be needed to explain even those things it does well (i.e., range effects).

As such, two experiments were conducted to test the predictions of AL-theory. In these experiments, S- similarity to the S+ and the amount of training that was undergone by participants were manipulated to observe effects on the generalization gradients. These experiments also used stimuli (i.e., horizontal bar graphs) that were conceptually similar to previously used stimuli (i.e., lines of varying length) but were more complex. Under these circumstances, AL theory would predict: 1) greater gradient shift when the S- was dissimilar, 2) greater gradient shift when less discrimination training was undergone, 3) progressive shift in the generalization gradient over the course of the generalization test, 4) greater gradient shift when a wider range of stimuli was used in the generalization test, 5) progressive gradient expansion, which would be greater when a wider range of stimuli was used during testing.

It was hypothesized that a dissimilar S- would lead to greater shift than a similar S-, and that this difference would be greater when less training was undergone. It was also hypothesized that gradient shift and gradient expansion would occur progressively, and that these effects would be more pronounced when a wider stimulus range was used in testing. As will be discussed, the results of these experiments offered little support for AL-theory, and may indicate important directions for future study of the topic.

General Results and Comparisons between Experiments

The results of these experiments were unusual in a number of ways. Foremost was the general lack of peak/area shift, progressive or otherwise, that was observed. As shown in Figures 13 and 30, the stimulus that resulted in the highest response probabilities was either S4 (S+) or S3, which was more similar to the S- than the S+ was. This was unexpected as adaptation-level theory would not predict gradient shift in the direction of the S- in these circumstances. One explanation is that participants failed to learn the discrimination that was required of them during training, and thus the resultant mechanism(s) responsible for peak shift could not emerge. However, this conclusion is not supported by participants' discrimination accuracies. Both experiments showed that the F40 group had attained average accuracies near 1.0 by the end of training, so that if gradient shifts were simply contingent on learning the discrimination, these groups should have shown that shift. Further, the value of the Far S- differed between these experiments, with the E1 Far S- (S15) being more similar to the S+ (S30) than the E2 Far S- (S0). As such, while both Far S- groups achieved accuracies near 1.0, the number of trials that were required to do this differed between the experiments. This would suggest that, if gradient shift occurred only when the discrimination task was both challenging and well-learned, then this shift would have been expected in E1 but not E2; this was not the case.

Similarly, the N40 conditions of both experiments showed accuracies that had plateaued well before the cessation of training (though at lower maximum accuracies than the F40 groups), and these groups also failed to demonstrate shift. It is thus unlikely that the lack of gradient shift was the result of either a failure to achieve accuracies near 1.0, or failure to achieve accuracies near the participants' maximum proficiency (which differed depending on the difficulty of the task).

It also cannot be stated that gradient shifts failed to emerge due to overtraining, as the 8-training trial groups did not achieve the same high accuracies as their 40-trial counterparts, and this was more true of the N8 groups than the F8 groups.

In short, it is unlikely that either an excessive presence or absence of discrimination proficiency was the cause. However, it may be that, although accuracies plateaued before 40 trials were completed, some learning was still occurring. That is, while accuracies may cease to show notable improvement regardless of further training, it is possible that the mechanism(s) responsible for gradient shift are still developing. Should this be the case, participants within the 40-trial condition would be undertrained, despite having achieved high accuracies. It cannot be stated whether this is the case within the specific context of these experiments, however previous research has shown gradient shift after far fewer training trials (Galizio, 1985).

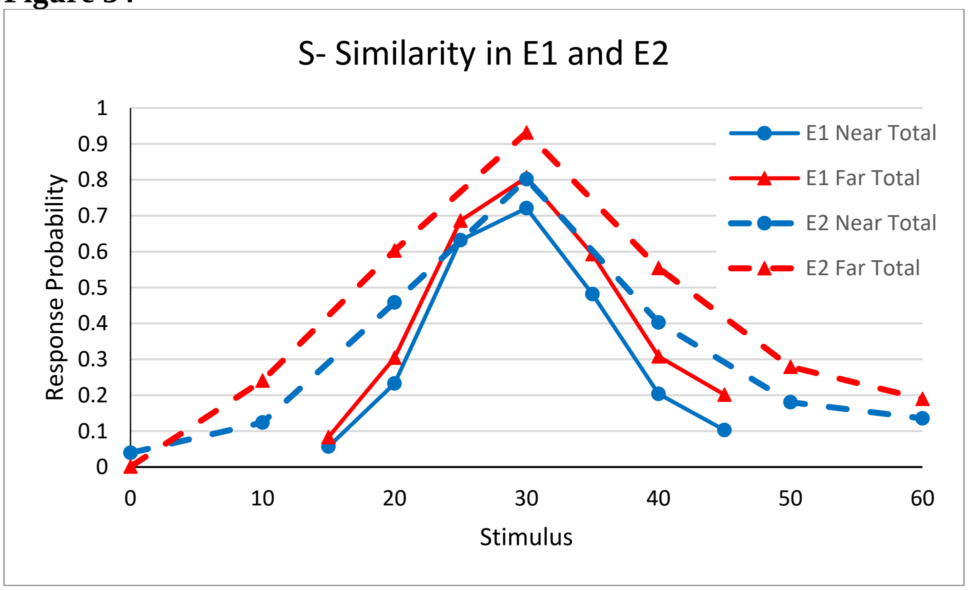
Given these considerations, it is likely that another factor is partially or wholly responsible for the lack of gradient shift; this may be the stimulus that was used. While the current stimulus (a horizontal bar graph) is conceptually similar to lines of varying length, a stimulus that has often been used in the production of gradient shift (Derenne, 2019; O'Donnell et al., 2000; Okouchi, 2003), the bar graphs are far more complex, and could be understood in different ways by participants. For example, these graphs could be perceived in terms of the total length of all 6 bars, or the average length, or the total area covered by the bars, or the percentage of the graph area covered by these bars. Any of these approaches, or combinations there-of, may have been adopted by participants, and it is not known how these differences would affect gradient shift or other characteristics of the gradients. Based solely on AL-theory, there is relatively little reason to expect these different approaches to yield different results, as each of these are simply different ways of describing the same information, albeit in simpler or more complex terms. However, there is some evidence that how participants conceptualize the information they are presented with can moderate gradient shift. Specifically, it has been shown that when participants are perceiving line angles, peak shift disappears once these stimuli are described as hands on a clock (Thomas & Thomas, 1974; Gallaghar et al., 2020). Thus, if participants were able to conceptualize the current stimuli in terms of some other task that they are more familiar with, gradient shift would become unlikely.

While this is possible, this explanation appears somewhat less likely given the difficulty that participants experienced during discrimination training, as well as the width of the gradients that were demonstrated during testing. If participants were able to relate the task to something, to which they had already developed a high degree of perceptual sensitivity, it would be expected that they would attain near-perfect discrimination accuracies quickly, and that generalization gradients would be quite narrow; this was not the case.

It was previously suggested that the bounded nature of the stimulus range may have played a role. In the stimuli used, the graph area was bordered so that it would be visually apparent how long each bar could be. In effect, while these were not presented during training, participants would have been aware of what the maximum and minimum graph “scores” were upon their first observation of the stimulus. During discrimination training, the adaptation level is typically assumed to develop at the midpoint between the S+ and the S-, as it represents the average of the stimuli that have been observed. In this case, providing context as to the range of the stimulus dimension could have allowed participants to develop an adaptation level that was independent of the specific stimuli that were presented. For example, if a participant noted that the bars of the S+ covered approximately half of the graph area, then this indicator could act as the adaptation level that they might use to make their decisions.

As noted in the Results and Discussion of Experiment 1, this explanation would be supported if the gradients were extremely similar between E1 and E2, as the S+ was the same in both cases. However, as shown in Figure 34, these gradients differed noticeably in terms of width and maximum response probabilities, with the E2 gradients being wider and higher than their counterparts in E1. If participants were forming decisional rules based on the context provided by the bordered graph area, then neither the different S- values nor the testing range would be expected to affect the generalization gradients under AL-theory. In essence, an AL that was established without regard to the S- would not differ depending on the S- values, and would not be affected by the range of stimuli presented in testing. Therefore, while it is possible that an AL may have developed purely due to the nature of the stimulus, it cannot be said that the specific stimuli encountered did not play a role in the development of generalization.

Figure 34



Note. This graph displays gradients from Experiment 1 and Experiment 2. Gradients were combined with respect to Training Amount.

Another explanation that was posited for the lack of area shift observed in Experiment 1 was that the stimulus dimension was not wide enough to induce a shift in the adaptation level. The results of Experiment 2 would discredit this explanation, as the range of stimuli used in Experiment 2 (S0 – S60) was, in absolute terms, double that of Experiment 1 (S15 – S45).

However, the rightward shift predicted by AL-theory still did not occur. Further, E2 showed limited evidence of a leftward shift, in the direction opposite of what AL-theory predicts.

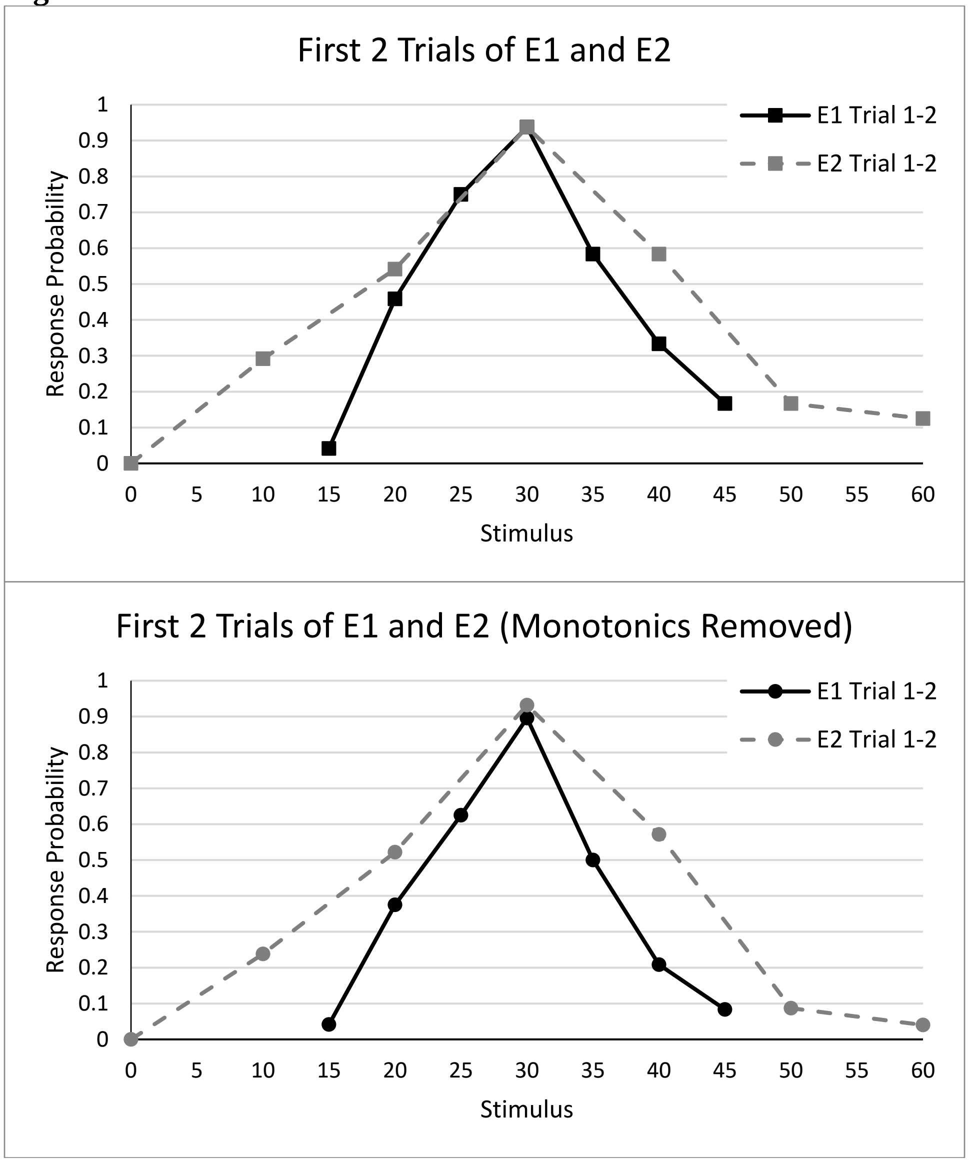
Acknowledging this, it is still true that the stimulus dimension used in E2 was not noticeably wider than the generalization gradients, as responding only approached 0 in the presence of the most extreme stimuli. As such, the previously posited explanation that the relative difficulty of the task prevented area shift cannot be dismissed entirely.

While peak shift was not observed in these experiments, these findings are relevant to the topic of peak shift. Generally, the stimuli used to study peak shift have been simple, consisting of 1 relevant attribute, while multidimensional stimuli are far less common in the literature (Spetch et al., 2004). These stimuli have also been such that the range of the stimulus dimension was only limited by practical considerations (e.g., lights of various wavelengths would be constrained to the visible spectrum), and did not exist within bounds that would be immediately apparent to participants. This is not true of all stimuli in the natural environment, which can often be complex and exist within explicit bounds. Further research is necessary to determine whether peak shift would fail to occur in these cases, or if the present findings are due to a unique aspect of the stimuli employed.

Another surprising result was the overall lack of progressive changes in the gradients over the course of the generalization test, particularly the absence of gradient expansion. As previously mentioned, the gradients of E1 and E2 differed notably in terms of width, but the first and last testing iterations within each experiment did not. The difference between experiments is not attributable to the use of more or less similar S-s, as the Near S- of E2 (S20) led to gradients that were wider than either the Near S- of E1 (S25) and the Far S- of E1 (S15). Similarly, the Far S- of E1 (S15) produced gradients that were narrower than either the Near S- of E2 (S20) and the

Far S- of E2 (S0). Both experiments involved one S- that was greater than an S- of the other experiment, and one S- that was less than an S- of the other experiment; while gradient widths did differ depending on S- similarity, with more dissimilar S-s leading to wider gradients, the greatest effect was apparently by the range of the stimulus dimension that was used during testing. This suggests that gradient widths were primarily determined during testing, rather than training. However, the absence of progressive change over the course of the test suggests that expansion occurred and stabilized within the first iteration (7 stimulus presentations) of the generalization test. As 3 of the stimuli (S20, S30, and S40) were identical between E1 and E2, the observed gradient expansion would be due to, at most, 4 stimuli that differed between the first iteration of E1 and E2. In fact, as displayed in Figure 35, when gradients were constructed using only the first 2 testing trials, the difference between the experiments was still apparent, regardless of whether monotonic responders were included or excluded. However, this comparison should be made somewhat tentatively, as presentation orders were not identical between the experiments. That is, participants within Experiment 1 who viewed S20 or S40 within their first 2 stimulus presentations were shown these in trial 1, while participants in Experiment 2 were shown these in trial 2 (presentation orders are described in detail in Table 1). Regardless, Figure 35 suggests that these effects take place rapidly, with only 1 exposure to a relatively extreme stimulus (in the case of E2, S10 and S50) being sufficient to increase responding to S20 and S40 (as compared to E1, where S20 and S40 were shown first to these participants).

Figure 35



Note. These graphs display gradients constructed using the first 2 stimulus presentations of the generalization test. As such, each participant responded to only 2 of the 7 possible stimuli, which depended on the presentation order that they had been randomly assigned to. A detailed explanation of this can be found in Table 2. The top graph displays the gradients for all participants, while the bottom graph excludes participants who met criteria to be considered monotonic responders (response probabilities always increased or did not change as stimulus scores increased).

This is difficult to reconcile with AL-theory, as the AL is described as the average of encountered stimuli. Particularly in the 40-training trial conditions, the effect of the first 4 stimuli encountered in the generalization test would be expected to be minimal. It would also be expected that any changes that occurred within the first testing iteration would continue over subsequent iterations; however, this was not the case. It is reasonable that more recent experiences would affect generalization more than older experiences, so that the 7 stimuli encountered during the first testing iteration would affect generalization more than, for example, the first 7 stimuli encountered during discrimination training. It is also possible that participants would consider stimuli encountered during the generalization test to be more important in determining generalization than those encountered during training, simply because the former were a part of the test. That is, should participants adjust their response-conservativeness depending on the range of stimuli they expect to encounter, then stimuli presented during the test would inform these expectations far more than the stimuli observed during training.

While both of these explanations may have been partially responsible for the differences between the E1 and E2 gradients, the second would better account for the immediacy of the change. Should increases in generalization be the result of updating expectations, and thus conservativeness, these changes could occur very quickly, possibly even in the time between the observation of a stimulus and the response to that stimulus.

While there was little if any progressive change in the gradient mean, and gradient expansion would appear to have taken place within the first testing iteration, this is not to say that there were no progressive changes under any of the conditions. Most notably was an observed decrease in the area under the curve that occurred in the Far S- condition of Experiment 1 (Figure 15), but not at all in Experiment 2 (Figure 31). As noted in the Results and Discussion section of E1, a degree of extinction was expected to occur; however this change was expected to be relatively slight due to the weakness of the appetitive stimulus used in training (affirming feedback) and the participants' knowledge that this stimulus would cease at the beginning of the generalization test. While no explanation was posited during the discussion of E1, comparing the results of E1 and E2 may shed some light on this trend. That is, it would appear that this decrease in responding was an effect of the range of the stimulus dimension, possibly moderated by S- similarity. In E1, participants were subjected to a narrower stimulus range during testing than those of E2; it was hypothesized that this would lead to an expansion of the gradients in E2, which would not occur in E1. It would appear that the reverse was the case; gradients under E2 did not expand, but gradients under E1 (specifically, those of the Far S- groups) contracted. This change is not readily apparent within the aggregate gradients displayed in Figure 13, which show a decrease in response probabilities near the S+, but little change in the gradients' widths. However, Figure 16 shows that the average width of these gradients did decrease over the course of the test. This discrepancy is likely due to differences between individual participants, which are not visible in the aggregate. Because some participants showed a leftward shift in their gradients, while others showed a rightward shift, the decrease in area under the curve appears to be a result of lower overall responding, rather than a narrowing of the individual gradients.

Again, these trends may be explained by participant expectations. As the generalization test progressed, it would become increasingly apparent that the range of the stimulus dimension used in E1 did not encompass the full range that participants may have inferred. This would also explain why this trend only occurred in the Far S- condition, and not the Near S- condition. The Near S- was always the most similar stimulus to the S+ that was presented in testing (S25 in E1, S20 in E2), while the Far S- was the most dissimilar stimulus presented (S15 in E1, S0 in E2). As such, those who underwent training with the Near S- had been trained not to make an affirmative response to the most similar stimulus prior to the initiation of the generalization test, while those who were trained with the Far S- were encountering these similar stimuli for the first time.

Therefore, it would be reasonable to expect those who had been trained with a dissimilar S- to consider stimuli that were somewhat similar to the S+ to be sufficiently similar. Under this interpretation, it would appear that "sufficiently similar" is a criteria that is partially controlled by expectations as to what would later be encountered.

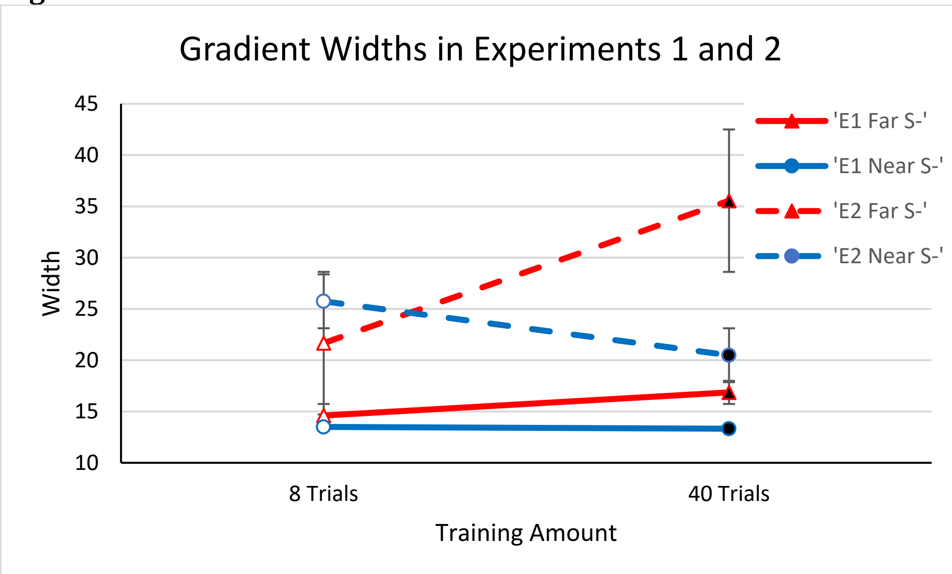
Additionally, this would explain why the apparent gradient contraction was slower to develop (occurring across all 18 testing-iterations) than the expansion observed in Experiment 2. The presentation of extreme stimuli would immediately inform participants' expectations as to the width of the stimulus range, while the absence of these stimuli would require more presentations to ascertain.

Two situations might be expected to produce this gradient contraction: 1) when the properties or nature of the stimulus would lead participants to assume, with some confidence, that the generalization test would employ a wider range of stimuli than would actually be the case; 2) when discrimination training involved an S- that was outside of the range of stimuli that would be used during testing. While AL-theory would predict that the latter condition would lead to gradient contraction, this is not true of the former condition. In the case of the present experiments, because the Far S- was equal to the most dissimilar stimulus presented during testing, AL-theory would not predict gradient contraction. Rather AL-theory would predict gradient expansion in the Near S- condition, and no change in gradient width in the Far S- condition.

Assuming that gradient widths are partially controlled by participants' expectations, both a stimulus that implies a wider test-range and an S- that is outside of the test-range would lead some individuals to assume that the generalization test would consist of a wider range of stimuli than it actually would. As such, these expectations would lead participants to respond more liberally at the beginning of the generalization test, and to become more conservative as their initial expectations were disconfirmed. As previously stated, stimuli for which participants might confidently form expectations as to the range of the stimulus dimension have not been used in other studies on peak shift and, to the author's knowledge, discrimination training has not been carried out with a stimulus that was outside of the testing range. As such, gradient contraction has not been described previously.

Notably, an interaction was observed between training amount and S- similarity in determining gradient widths; this is displayed in Figure 36. While this interaction was only significant in Experiment 2 (p = .002), gradients obtained in Experiment 1 showed similar trends. This interaction can be explained by the previously discussed proposition that response-conservativeness, and not just perceptual sensitivity, is affected by discrimination training. In this case, training with a similar S- would more often punish liberal responding (as a stimulus that appeared to be somewhat similar to the S+ would often be an instance of the S-), while training with a dissimilar S- would more often punish conservative responding. As such, it is logical that prolonged training would offer more opportunities to reinforce or punish conservativeness (depending on the similarity of the S-). While this explanation was not supported by the previously discussed correlations between S+ and S- accuracies, the limitations of that analysis are sufficient to prevent the dismissal of this theory. Further investigation, involving randomized training presentation orders, would be necessary to conclude whether this interaction is the result of response-conservativeness changing through training.

Figure 36



Another way that the results of E1 and E2 differed was in the gradients of monotonic responders, with those of E1 (Figure 18) showing relatively gradual increases in response probabilities, while those of E2 (Figure 33) never responded to stimuli below some point and always responded to stimuli above that point (referred to as "all-or-nothing" responding). While E1 involved testing-stimuli that were nearer to each other, thus allowing for greater ability to detect gradual changes in response probabilities from one stimulus to the next, this difference is likely due to the stimuli that were used as S-s, particularly the Far S- of E2. In this condition, all of the participants who developed a monotonic response pattern did not respond to the S- (S0), but did respond to every other stimulus in every presentation. It may be that both the monotonic responders of E1 and those of E2 were identifying the S- and attempting to respond to all stimuli that were greater, and that the consistency observed in E2 was due to the ease with which the S- could be distinguished from all other stimuli. However, Figure 18 shows that monotonic participants in the Far S- of E1 often failed to respond affirmatively to S30. This was particularly true of the F40 group, which showed final training accuracies (M = .948) that were far greater than their response probability toward S30 (*M* = .778) during the generalization test. This would suggest that some or all of the monotonic responders of E1 were attempting respond to stimuli that were greater than or equal to the S+, while monotonic responders of E2 were attempting to respond to all stimuli greater than the S-. Should this be the case, it would indicate that an S- of 0 consistently engendered an understanding of the task that was distinct from an S- of 15. However, it can not be stated based on these data whether the S- avoidance observed in E2 increases continuously as the S- approaches 0, or if this avoidance is unique to this S-, and would not appear for any S- that is visually discernable from 0.

In Summary, these experiments revealed a number of results that were unusual within the literature. Among these were the absence of peak shift or area shift regardless of S- similarity or training amount, the speed with which gradient widths were established in the different experiments, and the apparent contraction that was observed in Experiment 1. While these unexpected results made it impossible to draw meaningful conclusions regarding whether peak shift is determined by an interaction between the amount of discrimination training and the similarity of the S-, they do support the conclusion that AL-theory is not always sufficient to describe or predict generalization. These experiments also indicate several avenues for further research. Among these, it remains to be determined whether the absence of peak shift was a result of the stimulus-complexity, the bounded nature of the stimulus, or some other factor. Should complexity be the determinant, the question is raised as to whether peak shift would emerge under greater training. It is also unknown whether the gradient contraction observed in Experiment 1 was due to participants' expectations regarding the width of the stimulus dimension, and whether this effect can be reproduced with other bounded stimuli. Last, it remains to be determined whether the all-or-nothing responding observed in Experiment 2 was the result of the stimulus-type, the value of the S-, the amount of training, or a combination of these factors.

Concluding Remarks

Adaptation-level theory provides a framework for understanding several aspects of stimulus generalization in human participants, including why peak shift occurs and how generalization gradients are affected by features of the generalization test. The present research examined several predictions of adaptation-level theory that have received limited support due either to a lack of previous research or conflicting past findings. Specifically, the present research considered how gradients might be affected by the relative similarity of the S+ and S- used during discrimination training, the number of discrimination training trials that participants received, and the range of stimuli included in the generalization test. Adaptation-level theory suggests that the stimuli used during discrimination training affects the initial adaptation level and that the number of training trials affects how quickly the adaptation level will change during the generalization test. Together these variables should control when, and to what extent, peak shift occurs during the generalization test. However, none of the generalization gradients produced by the manipulation of these variables showed the effects that AL theory predicts (i.e., a progressive expansion of the gradient and a shift away from the S-). Instead, the gradients peaked at S+ in all conditions, and the widths of these gradients did not show progressive increase regardless of the number of training trials that participants received or the point in time during the generalization test when the gradient was measured. The one variable that consistently affected stimulus generalization in this research was the range of values used during the test. Specifically, increased amounts of generalization accompanied the use of wider test ranges.

Notably, this effect was as apparent in the initial responses during the generalization test as it was with the terminal and overall gradients.

In total, the results do not provide support for two important predictions of adaptation-level theory: that the gradient should be partly dependent on the initial adaptation level and that the adaptation level should be constantly recalibrated (potentially causing changes in the gradient) with each exposure to a test stimulus. Instead, the results show that a stable pattern of responding can emerge at or near beginning of the generalization test, and that this pattern of responding may be more sensitive to the initial test stimulus values than it is to the difference between S+ and S- during training or the amount of training that participants received. Of course, it is unclear to what extent these results are generally true of stimulus generalization in human participants, as opposed to being a result of special features of the present procedure (e.g., the relatively visually complex stimuli used to measure generalization). Regardless, the results confirm that adaptation-level theory is unable to explain all features of stimulus generalization in human participants, and that further theoretical advances are needed.

References

Baron, A. (1973). Postdiscrimination gradients of human subjects on a tone continuum. *Journal of Experimental Psychology*, *101*(2), 337–342.

Bizo, L. A., & McMahon, C. V. (2007). Temporal generalization and peak shift in humans. *Learning & Behavior*, *35*(2), 123–130.

Boneau, C. A., & Cole, J. L. (1967). Decision theory, the pigeon, and the psychological function. *Psychological Review*, 74, 123-135.

Derenne, A. (2006). Effects of S+ and S- separation on gradient shifts in humans. *Journal of General Psychology*, *133*(2), 163–173.

Derenne, A. (2019). Gradient shift is sensitive to the width of the generalization test range. *The Psychological Record*, 69(3), 401–414.

Derenne, A., Veenstra, C. C., & Ruthig, J. C. (2022). Effects of stimulus discriminability on peak shift: An investigation with COVID-19 risk assessments. *Learning and Motivation*, 101813.

Doll, T. J., & Thomas, D. R. (1967). Effects of discrimination training on stimulus generalization for human subjects. *Journal of Experimental Psychology*, *75*(4), 508.

Gallaghar, N. J., Provost, S., & Bizo, L. A. (2020). A jump to the left and a step to the right: A test of two accounts of peak shift. *The Psychological Record*, *70*(1), 11–20.

Galizio, M. (1985). Human peak shift: Analysis of the effects of three-stimulus discrimination training. *Learning and Motivation*, *16*(4), 478–494.

Galizio, M., & Baron, A. (1979). Human postdiscrimination gradients: The effects of three-stimulus discrimination training. *Animal Learning & Behavior*, *7*(1), 53-56.

Ghirlanda, S., & Enquist, M. (2007). How training and testing histories affect generalization: a test of simple neural networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1479), 449-454.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, *51*(1), 79–88.

Hansen, G., Tomie, A., Thomas, D. R., & Thomas, D. H. (1974). Effects of test stimulus range on stimulus generalization in human subjects. *Journal of Experimental Psychology*, 102, 634–639.

Hanson, H. M. (1957). Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*.

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *58*(5), 321.

Hearst, E. (1962). Concurrent generalization gradients for food-controlled and shock-controlled behavior. *Journal of the Experimental Analysis of Behavior*, *5*(1), 19-31.

Hebert, J. A., Bullock, M., Levitt, L., Woodward, K. G., & McGuirk, F. D. (1974). Context and frequency effects in the generalization of a human voluntary response. *Journal of Experimental Psychology*, *102*(3), 456.

Hebert, J. A., & Capehart, J. E. (1969). Generalization of a voluntary response as a function of presentation frequency of the training stimulus in testing. *Psychonomic Science*, *16*(6), 315-316.

Hebert, J. A., Origlio, D. P., & Mcguirk, F. D. (1972). Training and testing effects in the generalization of a voluntary response. *Psychonomic Science*, *26*(4), 209-210.

Helson, H. (1947). Adaptation-level as frame of reference for prediction of psychophysical data. *The American journal of psychology*, *60*(1), 1-29.

Helson, H., & Avant, L. L. (1967). Stimulus generalization as a function of contextual stimuli. *Journal of Experimental Psychology, 73*(4p1), 565.

Honig, W. K., & Slivka, R. M. (1964). Stimulus generalization of the effects of punishment. *Journal of the Experimental analysis of Behavior*, *7*(1), 21-25.

Hughes, B. M. (2001). Just noticeable differences in 2D and 3D bar charts: A psychophysical analysis of chart readability. *Perceptual and Motor Skills*, *92*(2), 495–503.

Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2018). Peak shift and rules in human generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1955.

Malone, J. C., Armento, M. E., Nemeth, R. M., Billington, E. J., Carpenter, C. N., & Andrews, K. B. (2004). Adaptation level effects in discrimination of flicker frequency. *Behavioural Processes, 65*(1), 95-109.

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II Generalization and discrimination. *Animal Learning & Behavior, 30*(3), 177–200.

Newlin, R. J., Rodgers, J. P., & Thomas, D. R. (1979). Two determinants of the peak shift in human voluntary stimulus generalization. *Perception & Psychophysics*, *25*(6), 478-486.

O'Donnell, J., Crosbie, J., Williams, D. C., & Saunders, K. J. (2000). Stimulus control and generalization of point-loss punishment with humans. *Journal of the Experimental analysis of Behavior, 73*(3), 261-274.

Okouchi, H. (2003). Stimulus generalization of behavioral history. *Journal of the Experimental Analysis of Behavior, 80*(2), 173-186.

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*(3), 221-241.

Switalski, R. W., Lyons, J., & Thomas, D. R. (1966). Effects of interdimensional training on stimulus generalization. *Journal of Experimental Psychology*, *72*(5), 661–666.

Thomas, D. R., & Bistey, G. (1964). Stimulus generalization as a function of the number and range of generalization test stimuli. *Journal of Experimental Psychology*, 68, 599–602.

Thomas, D. R., & Decapito, A. (1966). Role of stimulus labeling in stimulus generalization. *Journal of Experimental Psychology*, *71*(6), 913.

Thomas, D. R., & Hiss, R. H. (1963). A test of the "units hypothesis" employing wave-length generalization in human subjects. *Journal of Experimental Psychology*, 65, 59–62.

Thomas, D. R., & Jones, C. G. (1962). Stimulus generalization as a function of the frame of reference. *Journal of Experimental Psychology*, *64*(1), 77.

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*(6), 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*(2), 130.

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*(2), 210.

Thomas, D. R., & Thomas, D. H. (1974). Stimulus labeling, adaptation level, and the central tendency shift. *Journal of Experimental Psychology*, *103*(5), 896–899.

Thomas, D. R., Windell, B. T., Williams, J. L., & White, K. G. (1985). Stimulus presentation frequency in brightness discrimination and generalization: A test of adaptation-level and signal-detection interpretations. *Perception & Psychophysics*, *37*(3), 243-248.

Tomie, A., & Thomas, D. R. (1974). Adaptation level as a factor in human wavelength generalization. *Journal of Experimental Psychology*, *103*(1), 29.

Verbeek, E. L., Spetch, M. L., Cheng, K., & Clifford, C. W. (2006). Determinants of range effects in face recognition. *Learning & Behavior*, *34*(3), 229-240.

Wisniewski, M. G., Church, B. A., & Mercado, E. (2010). Temporal dynamics of generalization and representational distortion. *Psychonomic Bulletin & Review*, *17*(6), 809-814.

Appendix A

The deviation of each bar from the graph's mean was systematic, with one bar being .5 standard deviations below the graph's mean, one 1.0 standard deviations below the mean, and one 1.5 standard deviations below the mean. The other three bars were dispersed in the same manner above the mean. Each bar's ordinal position within the graph was randomized so that no two graphs were identical.

The standard deviation of each graph was variable depending on the sum score of that graph. This was determined through the following formula:

ObservedSD = MaxSD – MaxSD x [SQRT(SumScore – 30) ^ 2 / 30] (1)

This was done so that the deviation in each graph would be proportionate to that graph's distance from a maximum or minimum score. Bar deviation was at its maximum when the sum score was 30, and approached 0 as the sum score approached either 0 or 60. This is similar to the distribution pattern that emerges within randomly generated scores that meet the same parameters for sum score and maximum/minimum bar-length.

Three maximum standard deviations were used: 1.67, 2.5, and 3.33. The 2.5 variation was chosen as this would produce graphs that were representative of the average graphs that would be produced from a true-random distribution, and 3.33 was chosen as this represented the maximum standard deviation in which the previously described criteria would be possible without any bars exceeding the 0 to 10 range. 1.67 was chosen for the final variation so that the average standard deviation across all of the graphs would equal 2.5.

The reason that three maximum standard deviations were used, rather than using only one, was to prevent participants from relying on simpler criteria during discrimination. If only one standard deviation were used, then participants hypothetically would only need to observe the longest or shortest bar within each graph. By using multiple standard deviations, it becomes necessary for participants to consider the sum of all bars, thus ensuring a task without any simplifying short-cuts.

Table 1

Presented Stimuli During the First Two Trials of Each Testing Iteration

Stimulus Score		Iteration																	
Exp. 1	Exp. 2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
									Group 1										
15	0	x				x			x				x			x			
20	10		x				x			x				x			x		
25	20			x				x			x				x			x	
30	30	x			x				x			x				x			x
35	40		x			x				x			x				x		
40	50			x			x				x			x				x	
45	60				x			x				x			x				x
									Group 2										
15	0				x			x				x			x				x
20	10	x				x			x				x			x			
25	20		x				x			x				x			x		
30	30			x				x			x				x			x	
35	40	x			x				x			x				x			x
40	50		x			x				x			x				x		
45	60			x			x				x			x				x	
									Group 3										
15	0			x			x				x			x				x	
20	10				x			x				x			x				x
25	20	x				x			x				x			x			
30	30		x				x			x				x			x		
35	40			x				x			x				x			x	
40	50	x			x				x			x				x			x
45	60		x			x				x			x				x		
									Group 4										
15	0		x			x				x			x				x		
20	10			x			x				x			x				x	
25	20				x			x				x			x				x
30	30	x				x			x				x			x			
35	40		x				x			x				x			x		
40	50			x				x			x				x			x	
45	60	x			x				x			x				x			x

Note. *This table describes which stimuli are presented to each of the groups during the first two trials of each testing iteration. For example, in Experiment 1 Iteration 1, Group 1 were presented with Stimuli 15 and 30, Group 2 were presented with 20 and 35, and so on.*