Assessment of Safety Generalization and Demonstration as a Function of Various Training Stimulus Parameters

By

Matthew Arthur Taylor

A dissertation submitted to the Graduate Faculty in Psychology: Learning Processes and Behavior Analysis in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2010

This manuscript has been read and accepted for the
Graduate Faculty in Psychology: Learning Processes and Behavior Analysis in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Alicia M. Alvero

Date — Chair of Examining Committee

Dr. Maureen O' Connor

Date — Executive Officer

Dr. Bruce Brown

Dr. Robert Lanson

Dr. Angela Lebbon

Dr. Oliver Wirth

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Assessment of Safety Generalization and Demonstration as a Function of Various Training Stimulus Parameters

by

Matthew Arthur Taylor

Adviser: Professor Alicia M. Alvero

Two experiments explored training of safe leg position behavior in an office setting. Both experiments measured (a) generalization of safety-related stimuli along a generalization gradient and (b) demonstration of the limits of "safe" behavior. Experiment 1 used a 2 x 2 between-subject factorial design plus a non-differential stimulus training group to assess the effects of (a) safe [S+] versus both safe and at-risk stimulus classes [S+ and S-] and (b) the similarity between the safe [S+] and at-risk [S-] stimuli along a continuum. The results show that training with stimuli from both classes, and greater similarity between S+ and S- stimuli, will increase correct classification of at-risk stimuli (i.e., decrease responding to S-) and decrease over-extension of the limits of safe behavior. Experiment 2 used a 2 x 2 between-subject factorial design to determine the effects of (a) the relative frequency (i.e., ratio) of safe to at-risk training trials and (b) the absolute frequency of safe and at-risk training trials. The results show that training should use a lesser S+ relative frequency to increase correct classification of at-risk stimuli and decrease over-extension of the limits of safe behavior. The absolute frequency of trials did not show an effect on either of the generalization or demonstration dependent variables.

Acknowledgements

I offer many people my sincerest gratitude for their many hours of help, support, and guidance. First, I would like thank my advisor, Dr. Alicia Alvero, for her unwavering mentorship dating back as far as our days in Kalamazoo, MI. Next, I am grateful for the opportunity that I've been given to learn from the many bright and giving faculty in the Learning Processes and Behavior Analysis Program, especially Bruce Brown, Robert Lanson, and Nancy Hemmes. In addition to the faculty, many current and former graduate students have contributed a great deal to my completion of this degree: John Ward-Horner, Adrienne Robek, Joseph Jacobs, Kathie Mangiapanello, Danielle Tittelbach Geissler, Kristen Rost, Amanda Mentzer, Liliane de Aguiar-Rocha, Melody Berkovits, Erich Grommet, Michelle Garruto, Patricia Moss, Michael Marroquin, Paul Argott, Jen Juska, Kristen Struss, Laura Seiverling, Jason Rockwell, Mari Watanabe-Rose, Roxana Nedelcu, and Kim Shamoun. I also want to thank the many undergraduate students who tirelessly helped me with this dissertation and several other projects: Andreas Skourides, Marc Olvina, Shoshana Nusbacher, Abu Mohammed, Emily Linder, Patrick Miao, Angela Cruells, Julieanne Guadalupe, Kristina Rodriguez, Velika Lawrence, Melissa Cuartas, Arielle Kaish, Stephanie Sanchez, John Sciarappo, Monalisa Ribeiro, Eva Rappaport, Katherine Urena, and Grace Nagaur. I would also like to express thanks to Siggi Sigurdsson for his advice and assistance. Additionally, I appreciate the time and helpful comments from Oliver Wirth and Angie Lebbon, my dissertation outside committee members. Finally, I dedicate this to my loving and always supporting family: Rachel, Mary, William, Roderick, and Kelly, who will see and hear much more from me since I began this six year sabbatical from normal life.

Table of Contents

Lists of Tables

Lists of Figures

Assessment of Combined Safety Generalization and Demonstration as a Function of Various Training Stimulus Parameters

General Introduction

Health Risks, Costs, and Future Direction

One-third of the annual U.S. work related injuries and illnesses are musculoskeletal disorders (MSDs) (U.S. Department of Labor, 1999). According to the National Research Council and the Institute of Medicine (2001), work-related MSDs and resulting costs are currently and will continue to be a substantial financial concern in the U.S. The direct and indirect costs associated with MSDs are estimated to be as much as $45-$54 billion per year (U.S. Department of Labor, 1999). Research has established that MSDs are preventable through reduced exposure to ergonomic risk factors (U.S. Department of Labor, 1999).

The National Occupational Research Agenda for musculoskeletal disorders has expressed a need for controlled research (i.e., laboratory settings that minimize confounding variables) that assesses intervention effectiveness (National Institute for Occupational Safety and Health [NIOSH], 2001). More specifically, there is a need for training and educational research that demonstrates a change in behavior that lessens exposure to health risks, rather than only relying on measures of health (NIOSH, 2001). Unfortunately, research is not always able to demonstrate experimentally a change in health due to ethical reasons or a lack of quality and controlled measures.

Types of Safety Training

Burke et al. (2006) categorized safety training into three levels of engagement (i.e., activity) that the training procedures require from the trainees: (a) least, (b) moderate, and (c) most. Least engaging trainings are a passive format, where a person learns from information typically given through lectures, videos, or handouts. Moderately engaging trainings are programmed instruction or discrimination training (DT) formats, where a person is given feedback on their answers. Moderately engaging formats are often conveyed through a workbook or are computer-based. The most engaging training is a behavior skills training (BST) format (i.e., behavior modeling), where a person learns while engaging in the target behavior. The most engaging formats are conducted by a human trainer in the target or a simulation setting.

It is important to note that both the DT and BST formats use the following components: instructions, modeling, practice, and corrective feedback of practice. Despite these common components, there are two primary differences between the two training formats. The first difference is how the practice component is conducted. The DT practice component requires that the participant discriminate between correct and incorrect stimuli (e.g., classify pictures of safe [S+] and at-risk [S-] behavior). In contrast, during the BST practice component a participant engages in the target behavior (e.g., demonstrates the safe and/or at-risk behavior). A second difference is that DT is often conducted entirely with a computer and requires little or no human trainer involvement. The BST format often requires a human trainer to provide feedback.

Burke et al. (2006) conducted a meta-analysis by coding 95 field training studies (number of participants = 20991) across the three categories of engagement for the following variables: (a) safety knowledge (i.e., test of knowledge); (b) safety behavior (e.g., self rating or data collected through behavioral observation); and (c) health and safe outcome (i.e., injuries). Across the three variables, the researchers reported “that moderately and highly engaging training methods are, on average, more time consuming and probably more expensive in the short term but that they are potentially less costly and more effective in the long term while better ensuring worker and public safety" (Burke et al., 2006, p. 321). Although both the moderate and most engaging training formats were found to be effective (Burke et al., 2006), it is much more cost effective for some companies to provide computer-based training (i.e., training with moderate engagement), rather than using a human trainer (i.e., training with most engagement) (Anger et al., 2006). The next section will describe DT in detail, which is a procedure often used within computer-based training.

Discrimination and Non-Differential Stimulus Training: Procedure and Effects

In non-differential stimulus (ND) training a behavior occurs only in the presence of one stimulus and is reinforced in the presence of that one stimulus. In other words, during ND training the behavior is never extinguished in the presence of any stimuli. Non-differential training produces a flattening effect on a generalization gradient when compared to DT. A gradient is a graphic representation of responding that is often derived from generalization assessment. The assessment is comprised of stimulus presentations that have a common dimension. The gradient shows the extent of responding along each value of the stimulus dimension that was presented during the assessment (Catania, 2007). The gradient allows one to easily inspect the similarities or differences in responding from one value to the next along the stimulus dimension. Responding during ND training is more similar across the stimuli when compared to DT. For an example of a flat generalization gradient, see Figure 1 and compare Group 1 (DT) to Group 2 (ND Training).

A standard DT procedure is operationally defined as requiring "one behavior and two antecedent stimulus conditions. Responses are reinforced in the presence of one stimulus condition [S+], but not in the presence of the other stimulus [S-]" (Cooper, Heron, & Heward, 2007, p. 705). The differences between S+ and S- stimuli can be (a) intra-dimensional, (b) inter-dimensional, and (c) extra-dimensional. Intra-dimensional stimulus differences are within a dimension (Catania, 2007). Inter-dimensional stimulus differences are between dimensions (Tomie, Davitt, & Thomas, 1975). Extra-dimensional stimulus differences are a compound that varies inter-dimensionally, yet have an identical value along another dimension (Tomie et al., 1975). A standard generalization assessment that follows discrimination training would present stimuli that vary along one dimension. After inter-dimensional stimulus training, the one dimension used in the generalization assessment would be the S+ dimension. After extra-dimensional stimulus training, the one dimension used in the generalization test would be the dimension that was held constant during training.

Discrimination training can have four categories of effect on a generalization gradient: (a) narrowing of the gradient [e.g., Jenkins & Harrison, 1960], (b) peak shift along the gradient [e.g., Doll & Thomas, 1967; Thomas et al., 1973], (c) area shift along the gradient [e.g., Derenne, 2006; Hanson, 1959], and (d) an increased height of the gradient's peak [e.g., Hearst & Koresko, 1968; Jenkins & Harrison, 1960]. The narrowing of a gradient is a lessening of the distance between the right and left side of the gradient along the stimulus dimension, when compared to a ND training group. For an example of a narrowed gradient, see Figure 2 and compare Groups 1 (ND training) and 2 (narrowed gradient). A peak shift along a gradient occurs when the point of maximum responding has moved, typically, in the opposite direction from the S- value. Figure 2 shows an example of a peak shift by comparing Group 1 (ND training) to Group 3 (peak shift). An area shift along a gradient occurs when maximum responding is near the S+ value, but a greater proportion of responding now occurs more on one side, typically opposite the S- value, than the other. Figure 2 shows an example of an area shift by comparing Group 1 (ND training) to Group 4 (area shift). An increased height of a peak occurs when responding is greater above S+ value than when compared to a ND training group. The height of the peak can be plotted according to (a) absolute responding or (b) relative responding. Absolute responding is plotted as the rate or total number of responses per stimulus value. Relative responding is plotted as a percent of the total number of responses across the assessment stimuli. For an example of a greater peak height, compare Group 1 (ND training) and Group 5 (higher peak) in Figure 2.

Researchers can specifically assess the extent of the four effects of DT; however, three of the effects are generally summarized as a sharpening effect on a generalization gradient: (a) increased peak height, (b) narrowing of the gradient, and (c) area shift away from the S- stimulus direction. Discrimination training increases the sharpness of the generalization gradient around the S+ and decreases behavioral control by non S+ stimuli (Schwartz, Wasserman, & Robbins, 2002). Discrimination training is important because of the potential influence on the safety behaviors that underlie the behavioral safety approach.

Implications of Training from a Behavioral Safety Approach Perspective

Types of behavior that are trained and used in a behavioral safety approach. Safety discrimination training can influence discrimination and demonstration types of safety-related behavior. Safety-related discrimination behavior is defined as requiring a person to observe stimuli or situations and classify them as being safe or at-risk (i.e., unsafe). Through training, a person learns to perform specific safety skills in the presence of specific stimuli, which is establishing of the S+ and S- stimuli. The S+ stimuli are correlated with the availability of response contingent reinforcement, and thereby set the occasion for responding (Catania, 2007). The S- stimuli do not set the occasion for responding because responses are never reinforced in the presence those stimuli (Catania, 2007). Overall, behavior is often described in terms of stimulus control, which refers to the discriminative control of behavior (Catania, 2007, p. 412).

Safety-related demonstration behavior is defined as requiring a person to actively engage in a task. Training can inform a worker of the behavior required to attain response contingent consequences. A response class is a group of functionally equivalent behaviors that has been formed by their subsequent consequences (Catania, 2007). The response class can be defined by topography, magnitude, rate, latency, etc. Any behavior that does not result in the response contingent reinforcement of "safe behavior" is not a part of the response class. Overall, behavior is often described in terms of the extent of differentiation, which refers to the proportion of responding within a response class compared to the proportion of responding outside the response class limits (Catania, 2007).

The discrimination and demonstration types of behavior are vital to the behavioral safety approach. The behavioral safety approach is used to influence a person to improve and/or maintain their level of safety in the work environment. According to McSween (2003), the components used within the behavioral safety approach are an "observation and feedback process, formal review of observation data, improvement goals, and reinforcement for improvement and goal attainment" (p. 17).

Types of persons involved in a behavioral safety approach. The behavioral safety approach involves two primary types of persons: (a) the person who is targeted for an improvement and/or maintenance of the level of safety in the work environment, and (b) the person who is collecting data for the purpose of feedback, determining if goals are met, and determining when to implement contingent consequences. The following paragraphs describe how discrimination and demonstration behavior are vital to the persons involved in the behavioral safety approach.

The persons who are targeted by the behavioral safety approach use demonstration behavior (a) in a way that increases safety while using equipment, (b) to correct and/or avoid at-risk situations, and (c) to correctly behave within the limits of safe behavior. Additionally, the persons who are a target of the behavioral safety approach would use discrimination behavior to identify stimuli/situations (e.g., classify dangerous work conditions that could lead to safety and health issues) that set the occasion for a demonstration type of safety-related behavior.

Overall, the persons who are targeted within the behavioral safety approach learn through training which stimulus and response classes are correlated with reinforcement. The greater proportion to responding occasioned by the stimulus classes and occurring within the response classes that are correlated with reinforcement increases the likelihood of the safety-related behavior being reinforced. Therefore, safety training facilitates the behavioral safety approach from the perspective of the persons who are targeted for an improvement and/or maintenance of the level of safety in the work environment.

The persons who are collecting the data used within behavioral safety approach traditionally would not use demonstration behavior unless it was used to help give a person feedback. The data collector might show the correct way to engage in a behavior if the target person incorrectly engaged in the demonstration type of behavior. The discrimination type of behavior is commonly used to observe stimuli or situations and classify them as being safe or at-risk (i.e., unsafe). The extent of correct classification of safety-related situations during the data collection process affects the integrity of (a) feedback, (b) determining if goals are met, and (c) implementation of contingent consequences. Overall, the persons who collect data used within the behavioral safety approach are the persons who directly or indirectly shape the target persons' safety-related behavior.

The safety training for the data collectors can increase their extent of correct identification of the stimuli and responses within the stimulus and response classes, respectively, which would increase the likelihood that the persons targeted by the behavioral safety approach would be properly affected by the components within the behavioral safety approach. Essentially, the concern is that the data collectors could, either directly or indirectly, reinforce at-risk behavior or withhold reinforcement of safe behavior for the persons who are targeted within the behavioral safety approach, due to incorrect classification during data collection and/or incorrect demonstration behavior during feedback. Therefore, the safety training for data collectors can have implications for the extent of effectiveness of the behavioral safety approach for the target persons.

Training Stimulus Parameters to be Considered during Safety Discrimination Training

Behavioral safety researchers have not yet thoroughly established the most effective training stimulus parameters. There is a need to establish more efficient and effective training stimulus parameters within a DT format because they could potentially have a dramatic effect on the outcome of a behavioral safety approach. The following sections will describe some basic research demonstrating the effects of stimulus parameters and propose ideas to be explored within future safety research. The basic research on stimulus control is relevant to the safety research because basic research is often where principles were developed and later evaluated in an applied context. The stimulus parameters that are described in the following sections have yet to be thoroughly evaluated in a safety related experiment. Therefore, it is appropriate to use basic research as the springboard for providing support for evaluating the same training stimulus parameters in the context of safety. The effects of the stimulus parameters described in the upcoming subsections will be, for the most part, described according to four categories of effect on a generalization gradient (described in the aforementioned discrimination and non-differential stimulus training section).

Positive and negative stimuli within training. Both basic and applied research have empirically demonstrated the importance of using both S+ and S- stimuli (i.e., DT or training with differential stimuli), rather than one class of stimuli (i.e., ND or non-differential stimuli training).

Jenkins and Harrison (1960) conducted two experiments that evaluate the effects of DT and ND training. Experiment 1 compared stimulus generalization by pigeons assigned to one of two groups: (a) uninterrupted presentation of S+, or (b) random order presentations of inter-dimensional S+ and S-. The S+ was a tone (i.e., 1000 cycles per second) and the S- was no tone. The generalization assessment presented a tone that varied in frequency (cycles per second). The researchers found that pigeons trained with both S+ and S- (i.e., DT) exhibited a sharper generalization gradient through an increased peak height and narrowing of the gradient, i.e., more stimulus control, than those trained with only an S+ (ND).

The narrowing of the gradient has a nearly symmetrical form and showing notable stimulus control when compared to ND, despite being trained inter-dimensionally. An argument could be made that the frequencies above and below the S+ value in the generalization assessment were untrained, since the S- value was "no tone." The effect on stimulus control in the untrained frequencies is especially interesting because of the possible implications that that this effect could have with humans on a socially valid dimension. This could suggest that perhaps safety responding along an untrained area of a stimulus dimension could be sharpened as a result of safety training in another area, therefore creating an emergent effect in the untrained area. In summary, the findings in Experiment 1 of Jenkins and Harrison show greater stimulus control through DT, rather than ND, and in an untrained area of a dimension, which could have implications for human safety training.

Experiment 2 used different S+ and S- stimulus values from Experiment 1, thereby ruling out an argument that the tones near the end of the continuum were less detectable than the tones presented near the center of the continuum. The second experiment assessed stimulus generalization in two groups of pigeons that received either (a) random order presentation of two different S+ stimuli, or (b) random order of two different S+ stimuli and one S-. The S+ were two fixed tones (i.e., 450 and 250 cycles per second) and the S- was no tone. The effects on the generalization gradient were similar for Experiments 1 and 2. Pigeons exposed to DT showed more stimulus control, a sharper generalization gradient through an increased peak height, than those trained with non-differential stimulus (ND) training. In summary, the aforementioned study along with other well designed animal research (e.g., Hanson, 1959; Lyons, & Thomas, 1967; Switalski, Lyons & Thomas, 1966; Tomie et al., 1975) suggests that training with stimuli correlated with reinforcement/nonreinforcement, DT, will increase stimulus control.

Another effect reported by Jenkins and Harrison, in Experiment 2, was an effect of repeated testing under extinction. The authors reported an increased amount of stimulus control, of mild inconsistency, within the greater numbered blocks within a test and the greater numbered tests. Despite the reported findings by Jenkins and Harrison, the authors note a fair amount of inconsistent findings across other cited research. Jenkins and Harrison cited several other studies that showed the following range of effects on stimulus control with an increased amount of testing under extinction: no effect, decreased, and increased. Based on the basic literature, increased testing under extinction could have varying effects on the extent of safety related stimulus control exhibited by humans. Overall, despite the established findings within the basic literature, little research has replicated the DT vs. ND training phenomenon within a typical functioning human population.

Taylor, Olvina, & Alvero (2008) used human participants and expanded upon previous DT research by assessing the effects on both stimulus generalization and response generalization assessments in the context of occupational safety. Their study evaluated (a) the extent of correct classification of a person’s video recorded behavior into a safe (S+) or at-risk (S-) class, and (b) the extent of correct demonstration of the trained behavior by showing the topographical limits of the safe response class. The researchers used four groups of training: (a) none; (b) safe [S+] only; (c) at-risk [S-] only; and (d) safe [S+] and at-risk [S-]. The researchers found that Group 4, DT group, showed significantly greater stimulus control and differentiation of the safe response class limits than the ND training (i.e., safe only and at risk only) and no training groups. Therefore, the differential effects between DT group and ND training groups are consistent with basic research findings.

Taylor et al. (2008) recommended that safety researchers investigate the effects of other S+ and S- parameters (e.g., similarity, relative frequency, and absolute frequency) through the use of both stimulus generalization and response generalization assessments. The following section will describe DT research that considers the similarity of S+ to S- as a stimulus parameter and the implications of the stimulus parameter for the safety discipline.

Similarity of S+ to S- within training. Basic research with animals and humans has empirically demonstrated that the S+ to S- similarity (i.e. discrimination difficulty) during DT affects stimulus control. Hanson (1959) assessed the effects of the similarity of S+ and S- training stimuli on stimulus generalization. Hanson (1959) used pigeons assigned to one of five groups. One group had S+ only training and four other groups were trained with both S+ and S- wavelength. The S+ location was held constant along the wavelength continuum, but the S- location varied in degree of similarity across the four experimental groups. Hanson (1959) found that DT with an S- with greater similarity to an S+ influenced a larger area shift away from the S- stimulus. In other words, the greater extent of S+ and S- similarity increased the percentage of responding in the direction opposite of the S-. The study by Hanson (1959) plus additional research (e.g., Perkins, Hershberger, & Weyant, 1959; Pierrel & Sherman, 1962; Thomas, 1962; Tomie et al., 1975; Weinberg, 1973) clearly demonstrated an effect of training stimulus similarity; however, each used animals as their subjects rather than humans.

Derenne (2006) conducted a study with humans to assess the effect of intra-dimensional DT stimulus similarity on a generalization assessment. The participants were assigned to three groups of training: (a) no S-, (b) S- with lesser similarity, and (c) S- with greater similarity. During the training and generalization assessment the participants classified line lengths (i.e., the lengths of the lines were determined by the number of ASCII characters) presented on a computer screen. The training and generalization assessment used a delayed matching-to-sample procedure, where a press and no press of a space bar indicated a “match” and “no match” classification, respectively. S+ values were held constant across participants. The study reported a statistically significant area shift away from the S- value for only the greater similarity S- group. Baron (1973) used humans and also demonstrated a larger area shift with training tone stimuli of greater similarity. Visual analysis suggests that S- with greater similarity to the S+ produced a steeper gradient on the S- side of the S+ for both Derenne (2006) and Baron (1973).

Despite all of the research that suggests greater S+ and S- similarity will increase gradient steepness and/or area shift, there is also some human participant research that opposes this outcome (e.g., Doll & Thomas, 1967; Thomas, Svinicki, & Vogt, 1973; Thomas, Mood, Morrison, & Wiertelak, 1991). Doll and Thomas (1967) ran a DT similarity experiment and assessed the effects on a generalization assessment. Adult students were assigned to one of two DT similarity groups: (a) lesser similarity S-, or (b) greater similarity S-. The DT and generalization assessment employed S+ and S- stimuli along a visual wavelength continuum. The results show a mode (peak) of responding by the lesser similarity group, when compared to the greater similarity group, to a wavelength further in the opposite direction of the S- value. Therefore, the researchers found that the participants in the lesser similarity group had a greater peak shift when compared to the greater similarity group.

An explanation for the above findings by Doll and Thomas (1967) would be the participants' possible use of labels (e.g., blue, green, yellow, etc.) to classify the wavelengths rather than an absolute stimulus location along the continuum (Bornstein, 1974; Derenne, 2006). Thomas et al. (1973) remedied the labeling concern within the Doll & Thomas (1967) study by using S+ and S- stimuli along a light intensity continuum, yet had unreliable effects according to Derenne (2006). The researchers showed a larger gradient shift with increased S+ and S- separation when the S+ was more intense than the S-. Unfortunately, the same stimulus manipulation showed that similarity did not produce different results between the two similarity groups when the S- was more intense than the S+. Therefore, stimulus similarity does not have a clear effect when using an intensity continuum. Thomas et al. (1991) improved upon Doll and Thomas (1967) and Thomas et al. (1973) by using stimuli varying in angular orientation.

Thomas et al. (1991) used a generalization assessment with human participants to determine the effects of S+ and S- similarity during DT. The participants were split into two groups during DT: (a) lesser similarity, or (b) greater similarity. During DT, the S+

and two S- stimuli were fixed line angles on a black background. The lesser and greater similarity S- groups had a 15 and 10 degree angle difference from the S+, respectively. The generalization assessment was a random order presentation of 9 different line angles, each with a five degrees difference. The researchers reported that the lesser S+ and S- similarity influenced a significantly greater peak shift away from the S- value, when compared to greater similarity training.

In summary, the foregoing similarity research has shown support for a greater peak or area shift with both lesser and greater similarity manipulations, but the majority of the research supports a greater effect with a greater similarity of the S+ and S-stimuli. Therefore, according to most of the foregoing cited research, occupational safety training would likely benefit from the application of training stimuli that have greater similarity in order to increase stimulus control and differentiation. The similarity of the S+ (safe) and S- (at-risk) training stimuli could interact with antecedents and consequences, and influence behavior, which underlie the behavioral safety approach. For example, if the at-risk training stimuli have less similarity to the safe stimuli, then the participants might demonstrate less stimulus control and differentiation of safe stimuli and safe behavior, respectively (i.e., will incorrectly report safe stimuli as at risk and more likely to respond outside response class limits). The next section highlights DT research about another stimulus parameter, the relative frequency of S+ and S- stimulus presentations, and suggests potential implications for occupational safety.

The relative frequency of S+ and S- stimulus presentations within training. Basic research has empirically demonstrated the effects of the relative frequency of S+ to S- stimulus presentations (i.e., ratio of S+ to S- presentations) during DT. Research shows that training with a greater relative frequency of S- stimuli creates a sharper gradient. In other words, DT with more S- trials, relative to S+, would increase stimulus control and may increase safety.

Thomas and Vogt (1983) used a stimulus generalization assessment to evaluate the effects of the relative frequency of S+ to S- stimuli during operant DT. College students were assigned to a DT group with one of the following relative frequencies of S+ to S- stimuli: (a) 3:1 (n = 18:6), or (b) 1:3 (n = 6:18). During DT, each group was exposed to the same S+ and S- stimulus values, which differed intra-dimensionally along a light intensity continuum. During the DT and generalization assessment the light stimuli were projected onto a screen in a random order and the participants were instructed to classify the stimuli as “same” or “different” to the S+ through a verbal response. The generalization assessment used seven different values of light intensity. The results showed that the 1:3 group, fewer S+ trials relative to S-, had significantly greater S- classifications of the novel stimuli between the S+ and S- values, when compared to the 3:1 group. Furthermore, the researchers reported that the 1:3 group had a significantly greater area shift of the gradient, when compared to the 3:1 group. The findings were consistent no matter whether the S+ value was of greater or lower intensity than that of S-. In summary, Thomas and Vogt (1983) showed that a presentation of fewer S+ than S- stimuli during training increased stimulus control, which could have implications for occupational safety training.

Thomas, Windell, Williams, and White (1985) also used college students and assessed the function of the relative frequency of S+ to S- during DT on a stimulus generalization assessment. The students were assigned to one of three groups for the following relative frequency of S+ to S- training stimuli: (a) 3:1 (*n* = 18:6), (b) 1:1 (*n* = 12:12), or (c) 1:3 (*n* = 6:18). During DT, each group was exposed to the same S+ and S- stimulus values, which differed intra-dimensionally along a light intensity continuum. The light intensity stimuli were projected onto a screen in random order during the training and assessment. The generalization assessment used nine different values of light intensity. The findings show that the 1: 3 group, with more S- trials, had significantly greater stimulus control through greater S- classifications of the novel stimuli between the S+ and S- values, when compared to the 3:1 and 1:1 groups. Interestingly, the 1:1 group also showed significantly greater stimulus control, through greater S- classifications of the novel stimuli between the S+ and S- values, than the 3:1 group. Thus, number of S- classifications increased systematically with proportion of S- trials during training.

Both of the foregoing DT studies (Thomas & Vogt, 1983; Thomas et al., 1985) support the use of a lesser frequency of S+ presentations, relative to S- presentations, in order to increase stimulus control. The training procedures used in occupational safety should consider the findings from the relative frequency of S+ to S- training research, which suggest that occupational safety DT should use a lesser frequency of safe stimuli relative to at-risk stimuli to increase worker safety. Future research should also consider evaluating the most effective and efficient relative frequency of S+ to S- trials.

The absolute frequency of S+ and S- stimulus presentations within training. Basic research has systematically evaluated the effects of the absolute frequency of S+ and S- stimuli during DT. The research suggests that DT should include a greater absolute frequency of stimulus trials, rather than a lesser frequency.

Hearst and Koresko (1968) conducted two experiments. Experiment one assessed the effects of the absolute frequency of S+ and S- stimuli on a stimulus generalization assessment. The pigeons were randomly assigned to one of four discrimination training groups: (a) 2, (b) 4, (c) 7, or (d) 14 days. The S+ was a vertical line projected on a key and the S- was a dark key. The generalization assessment presented the S+ and six other line angles on the key in random order. The researchers reported that the peak height for the absolute and relative generalization gradients increased (i.e., an increase of stimulus control) as a function of an increased amount of training. A second experiment by Hearst and Koresko (1968) used a different strain of pigeon, but otherwise was a replication of the first experiment. The findings from the second experiment fully support the first experiment.

Thomas (1993) measured the effects of the absolute frequency of DT trials on a stimulus generalization assessment. The researchers assigned human observers to two groups: (a) 12, or (b) 36 trials. The S+ and S- were different intra-dimensional values along a light intensity continuum, and the stimulus values were fixed across all groups throughout DT. The stimulus generalization assessment presented 11 values of light brightness. Thomas (1993) did not conduct quantitative analyses of the data. Visual analysis suggests that the participants in Thomas (1993) showed an increased peak height and less of an area shift away from the S- with an increased frequency of trials. Regardless of the frequency of training trials, both groups exhibited the same amount of peak shift. Overall, the researchers showed that an increase of the absolute frequency of DT trials produced an increase of stimulus control.

The results from the absolute frequency of S+ and S- training trials research (i.e., Hearst & Koresko, 1968; Thomas, 1993) could be generalized to the training procedures used in occupational safety. Within the context of occupational safety, the use of a greater absolute frequency of DT stimuli may increase worker safety.

Purpose

Discrimination training has been established as an effective procedure for improving safety behavior, yet there has been a lack of research that bridges the aforementioned mostly basic training stimulus parameters research and safety research. Therefore, there is a need for research on the aforementioned training stimulus parameters in the context of generalization and demonstration of occupational safety related skills. The context of the current study is training of leg safety behavior while a person is sitting in an office chair. The goal of the present study was to assess generalization of safety related stimuli along a leg angle continuum [safe class, and trained and untrained at-risk classes] and demonstration of the leg safety limits [trained and untrained at-risk directions], under extinction with repeated testing, as a function of following training manipulations: (a) the use of a safe class [S+] or both safe and at-risk stimulus classes [S+ and S-], (b) the similarity of the at-risk [S-] stimuli to the safe [S+] stimuli along a continuum, (c) the relative frequency of safe [S+] to at-risk [S-] training trials, and (d) the absolute frequency of safe [S+] and at-risk [S-] training trials.

Based on the aforementioned literature, it is hypothesized that the following training stimulus parameters will be optimal in the context of safety: (a) the use of both safe and at-risk stimulus classes [S+ and S-], (b) greater similarity of the at-risk [S-] stimuli to the safe [S+] stimuli along a continuum, (c) a lesser relative frequency of safe

[S+] to at-risk [S-] training trials, (d) a greater absolute frequency of safe [S+] and at-risk [S-] training trials. It is also hypothesized that the assessments will show: (a) either an increased effect, no effect, or decreased effect with increased testing under extinction [previous research supports all hypotheses, without a consistent outcome], and (b) increased effect in both trained and untrained classes/directions following DT, when compared to ND training.

EXPERIMENT 1

The goal of Experiment 1 was to assess generalization of safety related stimuli along a leg angle continuum [safe class, and trained and untrained at-risk classes] and demonstration of the leg safety limits [trained and untrained at-risk directions], under extinction with repeated testing, as a function of following training manipulations: (a) the use of a safe class [S+] or both safe and at-risk stimulus classes [S+ and S-], and (b) the similarity of the at-risk [S-] stimuli to the safe [S+] stimuli along a continuum.

Based on the aforementioned literature, it is hypothesized that the following training stimulus parameters will be optimal in the context of safety: (a) the use of both safe and at-risk stimulus classes [S+ and S-], and (b) greater similarity of the at-risk [S-] stimuli to the safe [S+] stimuli along a continuum. It is also hypothesized that the assessments will show: (a) either an increased effect, no effect, or decreased effect with increased testing under extinction [previous research supports all hypotheses, without a consistent outcome], and (b) increased effect in both trained and untrained classes/directions following DT, when compared to ND training.

Method

Participants, Recruitment, and Duration

A total of one hundred twenty undergraduate participants were assigned to five groups (n=24 per group). Participants were recruited through announcements in classrooms or fliers on campus. The participants were compensated for their time by (a) fulfilling an academic requirement, (b) earning course extra credit, or (c) being paid $10. The assignment of participants by type of compensation was balanced across groups. Each session ranged between 40 and 60 min, depending on the group. Each participant was run for one session.

Settings, Apparatus, and Materials

The participants were run in a Queens College laboratory. The experimental room was 6 ft by 10 ft and furnished with a computer on a desk, a chair, and a camera (Sony DSC-V1) station to the right of the computer desk. The computer was a Dell Optiplex GX270 equipped with Windows XP version 2002, a 17 in. monitor, and Microsoft PowerPoint 2002. The Microsoft PowerPoint 2002 software was used to present participants with training and a generalization assessment. Additionally, participants were provided with data collection sheets for the practice training component and generalization assessment. The experimenters used scripts to provide verbatim instructions throughout the experiment (e.g., Appendix A and Appendix B). Details about the materials mentioned above are provided in the sections below.

Research Assistant Training

The primary experimenter trained the research assistants to help run sessions and measure the demonstration variable using a digital protractor (see the Dependent Variable section for further detail). Throughout the study the research assistants were kept blind to the various factors and hypotheses, thereby reducing the probability of bias during data collection.

Design and Group Assignment

The study used a 2 x 2 between-groups factorial design (Leg position x Similarity) to assess four experimental groups: (a) at-risk stimuli in a leg forward position and greater similarity to the safe stimuli; (b) at-risk stimuli in a leg forward position and lesser similarity to the safe stimuli; (c) at-risk stimuli in a leg backward position and greater similarity to the safe stimuli; and (d) at-risk stimuli in a leg backward position and lesser similarity to the safe stimuli. Additionally, the study used a non-differential stimulus (ND) training group of participants that were trained using only safe stimuli. All groups were exposed to the same safe stimuli. The participants were assigned to one of the five groups using randomized blocks of five based on their chronological order of participation.

Dependent Variables: Definitions, Descriptions, and Measurements

Generalization. The generalization assessment presented stimuli (i.e., pictures), one per trial, showing a person sitting on a chair in front of a computer desk. The stimuli consisted of pictures of the person's right lower leg varying in position over an angular range of 27-165 degrees (90 degrees was a vertical lower leg position). During the assessment, the participants were instructed to classify each stimulus as showing "safe" or "at-risk" leg behavior. The participants were not given feedback during the generalization assessment.

The dependent measure was the number of stimuli correctly classified as safe or at risk. The safe stimulus class was leg angles from 80-100 degrees and the at-risk stimulus classes were leg angles from 27-79 and 101-165 degrees (safe and at-risk angular ranges were derived from information presented in OSHA, 2009). Stimuli that represented each safe angular value were presented in one degree increments during the assessment. The at-risk stimuli were presented in odd numbered two degree increments during the assessment. Because fatigue was a concern, not all at-risk angular values were used during the assessment. Appendix C illustrates the safe and at-risk classes of stimuli. The effect of *Leg Angle* within each stimulus class was analyzed in the results section.

The 81 assessment stimuli, 21 safe and 60 at risk, each presented three times totaling 243 trials during the assessment. The 243 assessment trials were presented in non-repeated randomized blocks of 81. The first, second, and third blocks of 81 stimuli were presented to all participants in the same order (i.e., Assessment Order) of A, B, and C, respectively. The effect of the *Assessment Order* is analyzed in the results section.

The operational definition presented to participants for safe lower leg behavior stated that "lower legs are positioned anywhere between somewhat forward and somewhat backward (ignore the feet)." The operational definition of at-risk lower leg behavior presented to participants stated that "lower legs are positioned substantially forward or substantially backward (ignore the feet)."

Demonstration. Participants were instructed to demonstrate the limits of safe lower leg behavior. More specifically, the participants were instructed to show the experimenter the farthest that he or she could move their right leg forward and backward, while ensuring that it remains "safe." The participants were not given any feedback throughout the demonstration assessment. The left leg was not used because the location of the furniture within the setting obstructed the camera view. Pictures of the participant's forward and backward lower leg positions were recorded by a camera and later measured using a digital protractor. The dependent measure was the angle in degrees of the right lower leg in the forward and backward position. More specifically, the measure was calculated as the number of degrees performed above or below the limits of safe leg behavior, which is referred to as over or under-extension.

The order of demonstration of the lower leg response was either sequence I (i.e., leg forward and then backward) or II (i.e., leg backward and then forward). Participants within each group were assigned in randomized blocks of two to sequence I or II. The experimenter read a script that instructed the participants to engage in the appropriate sequence (e.g., see Appendix B). Prior to each forward or backward demonstration, the participant started from a 90 degree leg position (a vertical and neutral position). All participants engaged in their assigned demonstration sequence (i.e., Assessment Order) three times (i.e., 3 forward and 3 backward positions) to assess consistency of the behavior.

Reliability of Dependent Variables

Two independent observers measured 30% (i.e., measured/total n = 36/120) of the generalization data collection sheets in Experiment 1. The 36 of 120 generalization data collection sheets (each participant had one generalization data collection sheet) were randomly selected for reliability measures. Each data collection sheet had 243 trials, therefore there were 8748 total trials assessed for reliability of data collection. An agreement between observers was defined as both persons identifying a trial as correct or incorrect. The percentage of interobserver agreement was calculated by dividing the agreements by the total agreements plus disagreements, and then multiplying by 100. The scoring of trials on the data collection sheets was 99.98% in agreement (n of trials in agreement/total = 8746/8748). All generalization and demonstration data were calculated using point by point reliability for safe and at risk data.

Two independent observers measured 100% of the pictures within the demonstration variable for each participant. The pictures taken of all 120 participants were measured for reliability and each participant had 6 pictures, therefore 720 total pictures were assessed for reliability of data collection. An agreement of leg angle measure between observers was defined as a difference of two degrees or less. The observers disagreed upon 37 pictures and therefore agreed upon measures for 94.86% of the pictures (n of pictures in agreement/total = 683/720). The observers re-measured the 37 disagreed upon pictures, which did not include retraining or any novel instructions, due to an initial disagreement. Following the re-measure all observations were 100% in agreement (n of pictures in agreement/total = 720/720).

Independent Variables: Factors and Levels

Through DT, the study manipulated two factors, *similarity* and *leg position*, between groups. The two levels of similarity, *greater* and *lesser,* were the angular separation between the person's safe and at-risk leg position. The at-risk stimuli were trained in the *forward* or *backward* levels of right leg position. The trained and untrained at-risk stimulus classes were determined by whether the at-risk stimuli were presented in the forward or backward levels. The trained at-risk class was the leg position level, either forward or backward, where the at-risk stimuli were presented. The untrained at-risk class was the leg position level, either forward or backward, where the untrained at-risk stimuli were *not* presented. For example, if the trained at-risk stimuli were presented in the forward position, then the 101-165 degrees were considered the trained at-risk class and the stimuli in the backward position, 27-79 degrees, were considered the untrained at-risk class.

The safe stimuli used during training were held constant, across all groups, at 80 and 100 degrees. The at-risk stimuli were also represented by two stimuli with a range of 4 degrees. The greater and lesser similarity between the safe and at-risk training stimuli were defined as a 4 and 34 degree separation along the angular dimension, respectively. The participants were assigned to the following groups (Appendix D shows a picture that conveys the levels within each factor):

1. At-risk stimuli showed the leg in the forward position and had a greater similarity, 4 degree separation, to the safe leg stimuli. The at-risk angular values were 105 and 109 (range = 4 degrees).
2. At-risk stimuli showed the leg in the forward position and had a lesser similarity, 34 degree separation, to the safe leg stimuli. The at-risk angular values were 135 and 139 (range = 4 degrees).
3. At-risk stimuli showed the leg in the backward position and had a greater similarity, 4 degree separation, to the safe leg stimuli. The at-risk angular values were 71 and 75 (range = 4 degrees).
4. At-risk stimuli showed the leg in the backward position and had a lesser similarity, 34 degree separation, to the safe leg stimuli. The at-risk angular values were 41 and 45 (range = 4 degrees).

5. Non-differential stimulus (ND) training: only safe leg stimuli were presented during training.

There were four components (i.e., instructions, modeling, practice, and feedback) within the DT. During both the model and practice components the participants were exposed to 32 safe stimuli and 8 at-risk stimuli (40 trials per component). The purpose of the unequal relative frequency was to allow for a comparison to Experiment 2. The safe and at-risk classes of stimuli were each represented by two training values for all groups.

Integrity of the Independent Variables

During Experiment 1, 25% (i.e., n of measured/total = 30/120) of the sessions had two researchers present to collect integrity measures of correct implementation of the procedures. An agreement between observers was defined as both persons identifying a procedure as being correctly implemented. The researchers monitored the correct implementation of scripts for participant instructions (e.g., see Appendix A and Appendix B), PowerPoint training for the respective condition, and participant data collection sheet for the respective condition. The researchers agreed on correct implementation of the four procedures for 100% of the sampled sessions.

Furthermore, two independent observers verified that the material within all PowerPoint presentations was 100% correct to ensure the correct stimulus order, formatting, and duration. The PowerPoints were “read only” files to prevent accidental changes throughout running of the experiment.

Finally, the leg angles in picture stimuli used throughout training and generalization assessment were measured by two independent observers with 100% agreement. The picture stimuli were all agreed to be within the 0.5 degrees criterion of the target leg angle for any one picture stimulus. Any initial disagreements between the two observers were re-measured and if an agreement was not reached, then the disagreed upon stimuli were modified and then measured for integrity.

Procedure

Prior to starting the automated PowerPoint, the experimenter read the instructions verbatim from Appendix A to each participant. All groups were exposed to the following training components and generalization procedure through a PowerPoint presentation. The dependent variables were measured during testing procedures following the training components. The order of testing was the generalization procedure followed by the demonstration procedure. Within the training components, the ND training group was exposed to only safe information and stimuli.

1. Instructions: participants were provided with the written operational definitions of safe and at-risk lower leg behavior prior to and throughout the following training components and generalization procedure.
2. Modeling: participants were shown picture stimuli of the person's leg behavior that were identified as both safe and at risk. The person's body was shown in full view with a similar position throughout all pictures, except for the angle manipulation of the lower leg behavior. The leg angle manipulation varied according to the group assignment. Each stimulus trial was for a duration of 5 s and was presented consecutively without an inter-stimulus interval (ISI).

   During the modeling component, training stimuli from each of the safe and at-risk stimulus classes were presented separately from each other. The presentation order of the stimulus classes within the modeling component were counterbalanced within each group, in randomized blocks of two. The two stimuli within each class were assigned to an order in randomized blocks of two.

3. Practice: the participants viewed the same picture stimuli shown during modeling, yet in a different sequence, and were instructed to classify them as safe or at risk. Prior to classifying the stimuli, the participants were provided with trial data collection training. The participants were instructed to use a data collection sheet to classify the stimuli as safe or at risk by circling an "S" or "AR", respectively. Each stimulus trial duration was 5 s, and trials were presented consecutively without an inter-stimulus interval (ISI).

   A symbolic-matching to sample procedure was used during this practice component (e.g., S or AR on the data collection sheet was matched with a specific topography of the leg response). The extra-stimulus prompt (word "safe" or "at risk" next to the picture) used during the modeling component was no longer presented during the practice component. During the practice component, the safe and at-risk classes were presented together. The 32 safe and 8 at-risk stimuli were assigned in randomized blocks, four safe stimuli and one at-risk stimulus per block.

4. Feedback: after classifying all stimuli in the practice component, the participants' accuracy of practice responses was conveyed to them by re-showing each picture (same order) and along with the correct classification (i.e., safe or at risk). The participants were instructed to compare their responses on the data collection sheets to the answers presented in the PowerPoint slideshow. They were instructed to highlight all correct answers on the data collection sheet—this acted

as a measure of their understanding of the feedback. Each feedback slide duration was 8 s, and was presented consecutively without an ISI.

5. Generalization assessment: participants' generalization was assessed following the aforementioned training components. The assessment procedure was identical to the practice component.
6. Demonstration assessment: the participants demonstrated the limits of safe lower leg behavior. Appendix B was used as verbatim instructions during this testing procedure.

Human Subjects Review

The Queens College Human Subjects Institutional Review Board (IRB) approved this study. Prior to each session, all participants gave consent to participate in this study by signing a consent form approved by the IRB.

Results

Generalization: Similarity, Leg Position, and Assessment Order Factors

Statistical analyses were used to determine the extent of the effect that similarity, leg position, and assessment order factors had on each of the three classes of stimuli: trained at risk, untrained at risk, and safe. The dependent variable used in the statistical analyses was the number of trials correctly classified within each class of stimuli, collapsed across the leg angle factor. The at-risk stimuli ranged between 27-79 and 101-165 degrees during the generalization assessment; however, only data stimulus angles 27-79 and 101-153 degrees (i.e., 27 data points or 53 degrees per class) were used during all statistical analyses. The data sets from the at-risk classes were made to be of equal size during the analyses so that the classes could be collapsed across leg positions.

Performance within each class of stimuli was assessed by using a 2 x 2 x 3 (Similarity x Leg Position x Assessment Order) mixed factorial ANOVA of the number of trials correctly classified. The similarity and leg position were between groups variables. Assessment order (p. 24 has a description) was a within-subject variable. For the Experiment 1 generalization assessment, see Table 1 for the group mean number of trials correct and Table 2 for a summary of the ANOVAs.

Figure 3 shows the percentage of trials classified as safe across the leg angles in degrees as a function of training with at-risks stimuli in the leg *forward* position for each of the similarity groups. Figure 4 shows the percentage of trials classified as safe across the leg angles in degrees as a function of training with at-risks stimuli in the leg *backward* position for each of the similarity groups. The following subsections ignore the leg angle factor, therefore Figures 5-7 better represent the statistical analyses within this section (the analyses that use the leg angle factor are provided in the *Generalization: Discrimination Versus Non-differential Stimulus Training, Type of Training x Leg Angle* section, p. 35). The aforementioned figures and tables will be referred to in the appropriate result sections below.

Trained at-risk class. Figure 5 shows the percentage of trained at-risk stimuli classified correctly according to the levels within the similarity and leg position factors. There was a main effect of similarity on the correct classification of the trained at-risk stimuli (Table 1 shows the group mean number of trials correct and Table 2 shows a summary of the ANOVA). Participants correctly classified the trained at-risk stimuli significantly more frequently when trained with a greater similarity than a lesser similarity.

There was also a main effect of leg position for the correct classification of the trained at-risk stimuli. Participants' level of correct classification of trained at-risk stimuli was significantly greater when trained in the leg forward position than in the leg backward position. There was a significant interaction effect of similarity with the leg position for correct classification of the trained at-risk stimuli. The effect on correct classification of the trained at-risk stimuli was greater in the forward leg position level over the backward leg position level, for the lesser similarity level rather than the greater similarity level.

There was a main effect of assessment order for the correct classification of the trained at-risk stimuli. The LSD post-hoc test showed that the participants classified the trained at-risk stimuli significantly more correctly during the first set than the second and third sets.

Untrained at-risk class. Figure 6 illustrates the percentage of untrained at-risk stimuli classified correctly according to the levels within the similarity and leg position factors. There was no effect of similarity for the correct classification of the untrained at-risk stimuli (Table 1 shows the group mean number of trials correct and Table 2 shows a summary of the ANOVA). There was a main effect of leg position for the correct classification of the untrained at-risk stimuli. Participants' classification of untrained at-risk stimuli was significantly greater when the untrained at risk class was the leg extended in the forward position than the backward position.

There was a main effect of assessment order for the correct classification of the untrained at-risk stimuli. According to the LSD post-hoc test, the participants classified the untrained at-risk stimuli correctly at a significantly higher level during the first set than the second and third sets. There was a significant interaction effect of leg position with the assessment order for correct classification of the untrained at-risk stimuli. The effect on correct classification of the untrained at-risk stimuli was greater in the first assessment set over the second and third sets, for the forward position rather than the backward position.

Safe class. The percentages of safe stimuli classified correctly according to the levels within the similarity and leg position factors are shown in Figure 7. There was no significant effect of similarity or leg position on the level of correct classification of the safe stimuli (Table 1 shows the group mean number of trials correct and Table 2 shows a summary of the ANOVA). There was no significant effect of assessment order for the correct classification of the safe stimuli.

Generalization: Discrimination Versus Non-differential Stimulus Training, Type of Training and Leg Angle

The DT groups and ND training group were compared to each other by using a 3 x 27 or 3 x 21 (Type of Training x Leg Angle) mixed factorial ANOVA for each class of stimuli (i.e., trained at risk, untrained at risk, and safe). The dependent variable used in the statistical analyses was the percentage of trials correctly classified for each leg angle within each class of stimuli. The *type of training* variable (i.e., greater similarity, lesser similarity, and ND training) was manipulated between groups. Leg angle was a within-subject variable that had 27 levels in the trained and untrained at-risk classes, and 21 levels in the safe class. The first set of three ANOVAs was a comparison of the similarity groups who, were trained with at-risk stimuli in the leg forward position (i.e., Group 1 and Group 2), and the ND training group. The second set of three ANOVAs was a comparison of the similarity groups, who were trained with at-risk stimuli in the leg backward position (i.e., Group 3 and Group 4), and the ND training group.

When a type of training x leg angle ANOVA revealed an interaction, then the levels of leg angle were split into two ranges in order to parse the interaction. One way ANOVAs were used to show the significant differences between the type of training levels within limited ranges of the leg angle factor. The forward leg position at-risk class was split into degree ranges (a) 101-125, and (b) 127-153. The backward leg position at-risk class was split into degree ranges (a) 27-53, and (b) 55-79. For those who were trained with at-risk stimuli in the forward leg position, the safe class was split into degree ranges (a) 80-90, and (b) 91-100. The safe class was split into degree ranges (a) 80-89, and (b) 90-100, for those who were trained with at-risk stimuli in the backward leg position. The levels within the type of training factor were compared using Tukey's HSD as the post-hoc test. A post-hoc test was not used to compare the levels within the leg angle factor due to the lack of useful information for the occupational safety field. Determining which leg angle levels (i.e., degrees) significantly differ from each other is not the emphasis of the current study.

During the following analyses the leg position factor was not collapsed because the ND training group would not represent independent samples while the DT groups would represent independent samples. Therefore, owing to the independent samples *F*-test assumption of independent measures, the two forward leg position DT groups were compared to the ND training group separately from the two backward leg position DT groups. Additionally, the assessment order factor was collapsed within the following analyses.

Figure 3 shows the percentage of trials classified as safe across the leg angles in degrees as a function of training with at-risks stimuli in the leg *forward* position for each of the similarity groups. Table 3 shows the groups' mean percentage of trials correct and Table 4 provides a summary of the ANOVAs in the *leg forward position* level for each stimulus class. Figure 4 shows the percentage of trials classified as safe across the leg angles in degrees as a function of training with at-risks stimuli in the leg *backward* position for each of the similarity groups. Table 5 presents the group mean percentage of trials correct and Table 6 displays a summary of the ANOVAs in the *leg backward position* level for each stimulus class. The aforementioned figures and tables will be referred to in the appropriate result sections below.

Group 1 and group 2 forward trained at-risk classes versus non-differential stimulus training. There was a main effect of type of training for the correct classification of the trained at-risk stimuli (Figure 5 conveys the type of training main effect, Table 3 shows the group mean percentage of trials correct, and Table 4 shows a summary of the ANOVA). According to the post-hoc test, the participants who were in either the greater or lesser similarity groups, and trained with forward position at-risk stimuli, correctly classified the stimuli significantly more often than did the ND training group. Additionally, the post-hoc test showed that the group trained in the greater similarity stimulus level correctly classified the trained at-risk stimuli in the leg forward position significantly more often than did the group trained in the lesser similarity stimulus level. Finally, there was a main effect of leg angle and type of training x leg angle interaction (Figure 3 shows the type of training x leg angle interaction).

Group 1 and group 2 versus non-differential stimulus training for backward untrained at-risk classes. The statistical analysis did not yield a significant main effect for the type of training factor within the untrained at-risk class of stimuli (Figure 6 conveys the lack of type of training main effect, Table 3 shows the group mean percentage of trials correct, and Table 4 shows a summary of the ANOVA); however, there was a main effect of leg angle and an interaction of type of training x leg angle (Figure 3 shows the type of training x leg angle interaction). The levels within the type of training factor were compared across the 27-53 degree range and the post-hoc test showed that participants trained in the greater similarity level correctly classified the untrained at-risk stimuli significantly more often than the ND training group. None of the type of training levels classified the untrained at-risk stimuli significantly different from each other within the 55-79 degree range. This pattern of results indicates a steeper generalization gradient for the greater similarity group than for the ND group.

Group 1 and group 2 safe classes versus non-differential stimulus training. There was no main effect of type of training or leg angle, or an interaction of type of training x leg angle for the correct classification of the safe stimuli (Figure 7 conveys the lack of type of training main effect, Figure 3 shows the lack of type of training x leg angle interaction, Table 3 shows the group mean percentage of trials correct, and Table 4 shows a summary of the ANOVA).

Group 3 and group 4 backward trained at-risk classes versus non-differential stimulus training. There was a main effect of type of training for the correct classification of the trained at-risk stimuli (Figure 5 conveys the type of training main effect, Table 5 shows the group mean percentage of trials correct, and Table 6 shows a summary of the ANOVA). According to the post-hoc test, the participants trained with backward position at-risk stimuli at the greater similarity level correctly classified the trained at-risk stimuli significantly more often than did the lesser similarity and ND training groups. Additionally, there was a main effect of leg angle and a type of training x leg angle interaction (Figure 4 shows the type of training x leg angle interaction). The post-hoc test revealed that the lesser similarity group correctly classified the trained at-risk stimuli significantly more often than the ND training group in the 27-53 degree range; however, there was no significant difference in the 55-79 degree range. This pattern of results indicates a steeper generalization gradient for the lesser similarity group than for the ND group.

Group 3 and group 4 versus non-differential stimulus training for forward untrained at-risk classes. The statistical analysis did not yield a significant main effect for the type of training factor within the untrained at-risk class of stimuli; however, there was a main effect of leg angle and an interaction of type of training x leg angle (Figure 6 conveys the lack of type of training main effect, Figure 4 shows the type of training x leg angle interaction, Table 5 shows the group mean percentage of trials correct, and Table 6 shows a summary of the ANOVA). The post-hoc test showed that the participants trained in the greater similarity group correctly classified the untrained at-risk stimuli significantly more often than the ND training group, in the 101-125 degree range. There was no significant differences between the type of training levels in the forward position untrained at-risk stimuli in the 127-153 degree range. This pattern of results indicates a steeper generalization gradient for the greater similarity group than for the ND group.

Group 3 and group 4 safe classes versus non-differential stimulus training. There was no main effect of type of training or leg angle for the correct classification of the safe stimuli; however, there was a type of training x leg angle interaction (Figure 7 conveys the measures for type of training main effect, Figure 4 shows the type of training x leg angle interaction, Table 5 shows the group mean percentage of trials correct, and Table 6 shows a summary of the ANOVA). According to the post-hoc test, the lesser similarity group correctly classified the safe stimuli significantly more often than the greater similarity group, in the 90-100 degree range. The participants in the type of training levels did not classify the safe stimuli significantly different in the 80-89 degree range. This pattern of results indicates a steeper generalization gradient for the greater similarity group than for the lesser similarity group.

Demonstration: Similarity, Leg Position, and Assessment Order Factors

Statistical analyses were used to determine the nature of the effect that similarity, leg position, and assessment order factors had on demonstration of each of the two directions of leg extension (i.e., trained at-risk direction and untrained at-risk direction). The number of degrees the leg was extended from the limits of safe leg behavior (i.e., 80 and 100 degrees) was the dependent variable used in the statistical analyses. Each direction of leg extension was assessed by using a 2 x 2 x 3 (Similarity x Leg Position x Assessment Order) mixed factorial ANOVA. The similarity and leg position variables were measured between groups. Assessment order (p. 25 has a description) was a within-subject variable. For the Experiment 1 demonstration assessment, Table 7 shows the groups' mean degree of leg extension and Table 8 presents a summary of the ANOVAs. The aforementioned tables will be referred to in the appropriate result sections below.

Prior to all statistical analyses for the demonstration variable, outliers were identified and removed from the data sets because they could have a substantial effect on the type I and type II error rate during statistical analyses (Wilcox, 2001). Wilcox (2001) recommended that outliers be identified by using the boxplot method. The method determined the interquartile range (i.e., range between 1st and 3rd quartile limits) and then multiplied it by 1.5. The product was subtracted from the 1st quartile limit and added to the 3rd quartile limit. Any data points within a data set that were beyond those two values were considered as outliers and then removed from the data set prior to statistical analyses.

The method used to identify outliers was applied within each of the four experimental groups and across all levels of assessment order. For example, the method was applied to group 1 (i.e., greater similarity/leg forward group) and across all three assessment order levels for the trained at-risk direction (i.e., forward leg direction). Therefore, the outlier method was applied to eight different data sets (i.e., 4 experimental groups x 2 directions of leg extension). Overall, the outlier method surveyed 72 observations per application (i.e., 24 participants x 3 leg extensions). Across the eight applications of the outlier method, 3.47% (i.e., 20/576) of data points were identified as outliers and subsequently removed.

Trained at-risk direction. Figure 8 displays the mean number of degrees the leg was extended in the trained at-risk direction, from to the safe limits of leg behavior, as a function of the levels in the similarity and leg position factors. The statistical analysis showed a main effect of similarity for the leg extension behavior (Table 7 shows the group mean degree of leg extension and Table 8 shows a summary of the ANOVA). Participants trained in the lesser similarity level over-extended their leg significantly further beyond the safe limits, in the trained at-risk direction, than those trained in the greater similarity level. There was no main effect of leg position or assessment order; however, there was a significant interaction of leg position x assessment order for leg extension behavior in the trained at-risk direction. Over-extension of leg behavior was greater in the first extension assessment order over the second and third extensions, for the backward position.

Untrained at-risk direction. Figure 9 presents mean number of degrees the leg was extended in the untrained at-risk direction, from the safe limits of behavior, according to the levels in the similarity and leg position factors. There was no main effect of similarity or leg position for the leg extension behavior in the untrained at-risk direction (Table 7 shows the group mean degree of leg extension and Table 8 shows a summary of the ANOVA). Nevertheless, there was a significant main effect of assessment order for leg extension behavior in the untrained at-risk direction. The LSD post-hoc test showed that the participants extended their leg significantly further beyond the limits of safe behavior, in the untrained at-risk direction, for the first extension assessment when compared to the second extension assessment.

Demonstration: Discrimination Versus Non-differential Stimulus Training, Type of Training and Assessment Order

Statistical analyses were used to evaluate performance differences between four experimental groups and the ND training group along each of the two leg extension directions (i.e., trained at-risk direction and untrained at-risk direction). The four experimental groups were individually compared to the ND training group, rather than collapsing the leg position factor.

The dependent variable used in the statistical analyses was the number of degrees the leg was extended from the limits of safe leg behavior (i.e., 80 and 100 degrees). Each direction of leg extension was assessed by using a 3 x 3 (Type of Training x Assessment Order) mixed factorial ANOVA. The type of training variable (i.e., greater similarity, lesser similarity, and ND training) was measured between groups. Assessment order was a within-subject variable.

The first set of two ANOVAs was a comparison of the similarity groups who were trained with at-risk stimuli in the leg forward position (i.e., Group 1 and Group 2) to the ND training group. Table 9 shows the groups' mean degree of leg extension and Table 10 presents a summary of the first set of two mixed factorial ANOVAs. The second set of two ANOVAs was a comparison of the similarity groups who were trained with at-risk stimuli in the leg backward position (i.e., Group 3 and Group 4) to the ND training group. Table 11 displays the groups' mean degree of leg extension and Table 12 provides a summary of the second set of two mixed factorial ANOVAs. The aforementioned tables will be referred to in the appropriate result sections below.

The statistical analyses described below are intended for a comparison of the similarity groups to the ND training group, not a comparison of the similarity levels to each other. The similarity groups were compared in the previous section (the *Demonstration: Similarity, Leg Position, and Assessment Order Factors* section on p. 40). Tukey's HSD was used as the post-hoc test to compare the similarity groups to the ND training group. Finally, the statistical analyses were not intended to evaluate assessment order because that factor, except for the ND training level, was already discussed within the preceding section (the *Demonstration: Similarity, Leg Position, and Assessment Order Factors* section on p. 40).

Prior to all statistical analyses for the demonstration variable, outliers were identified and removed from the data sets. For further information about the method used to identify outliers, see the *Demonstration: Similarity, Leg Position, and Assessment Order Factors* section on p. 40. The outlier method was applied 10 different data sets (i.e., 4 experimental groups + 1 ND training group x 2 directions of leg extension). Overall, the outlier method surveyed 72 observations per application (i.e., 24 participants x 3 leg extensions). Across the 10 applications of the outlier method, 3.33% (i.e., 24/720) of data points were identified as outliers and subsequently removed.

Group 1 and group 2 forward trained at-risk direction versus non-differential stimulus training. Figure 8 shows the mean number of degrees the leg was extended in the trained at-risk direction, from the safe limits of leg behavior, and as a function of the levels within the type of training factor. The statistical analysis showed a main effect of type of training for the leg extension behavior (Table 9 shows the groups' mean degree of leg extension and Table 10 shows a summary of the ANOVA). The post-hoc test revealed that the participants trained in the ND training group over-extended their leg significantly further beyond the safe limit, in the trained at-risk direction, than did the greater similarity group. The lesser similarity group when compared to the ND training group did not perform significantly different from each other for leg extension behavior in the trained at-risk direction.

Group 1 and group 2 backward untrained at-risk direction versus non-differential stimulus training. Figure 9 presents the mean number of degrees the leg was extended in the untrained at-risk direction, from the safe limits of leg behavior, as a function of the levels within the type of training factor. The statistical analysis did not detect a significant main effect for the type of training factor for the untrained at-risk leg extension direction (Table 9 shows the groups' mean degree of leg extension and Table 10 shows a summary of the ANOVA). In other words, the participants trained in either the greater similarity or lesser similarity groups did not extend their leg significantly different from the ND training group in the backward untrained at-risk direction.

Group 3 and group 4 backward trained at-risk direction versus non-differential stimulus training. Figure 8 shows the mean number of degrees the leg was extended in the trained at-risk direction, from the safe limits of leg behavior, and as a function of the levels within the type of training factor. The greater and lesser similarity groups, when compared to the ND training group, did not perform significantly different from each other for leg extension behavior in the backward trained at-risk direction (Table 11 shows the groups' mean degree of leg extension and Table 12 shows a summary of the ANOVA).

Group 3 and group 4 forward untrained at-risk direction versus non-differential stimulus training. Figure 9 presents the mean number of degrees the leg was extended in the untrained at-risk direction from the safe limits of leg behavior, as a function of the levels within the type of training factor. The statistical analysis detected a significant main effect for the type of training factor for the untrained at-risk leg extension (Table 11 shows the groups' mean degree of leg extension and Table 12 shows a summary of the ANOVA). According to the post-hoc test, the participants trained in the ND training group extended their leg significantly further beyond the limit of safe behavior in the forward untrained at-risk direction when compared to the greater similarity group. The lesser similarity group when compared to the ND training group did not perform significantly different from each other for leg extension behavior in the forward untrained at-risk direction.

Discussion

The discussion section for Experiment 1 (a) describes an interpretation of the effects of the factors on the generalization and demonstration data, (b) lists strengths/limitations, and (c) provides a summary of the findings.

Generalization

The greater stimulus similarity training had a significant effect, when compared to lesser stimulus similarity training, on the correct classification of the trained at-risk stimuli. The greater and lesser similarity training groups, in the forward or backward positions, correctly classified a greater percentage of trained at-risk stimuli than the ND training group. The greater similarity of S+ and S- training showed an effect on correct classification when compared to lesser similarity, which supports previous research (Baron, 1973; Derenne, 2006; Hanson, 1959). Additionally, the greater effect of DT (i.e., both S+ and S- training) when compared to ND training (i.e., S+ only), showed findings aligned with previous basic and applied research (Hanson, 1959; Jenkins & Harrison, 1960; Lyons, & Thomas, 1967; Switalski, Lyons & Thomas, 1966; Taylor, Olvina, & Alvero, 2008; Tomie et al., 1975).

The greater similarity stimulus training group, in the forward or backward positions, correctly classified significantly more untrained at-risk stimuli than the ND training group. The lesser similarity stimulus training group and ND training groups did not perform significantly different from each other across the untrained at-risk classes. The finding in the current study that suggest that DT with greater stimulus similarity training yielded a sharper generalization gradient than did ND training in the untrained S-stimulus class, in the opposite direction of the trained S-, has been shown in the animal literature (e.g., Jenkins & Harrison, 1960), but not the occupational safety literature. It could be argued that the effect of the trained S- class generalized to the untrained S- class.

Only the greater similarity group trained with at-risk stimuli in the backward position classified the safe stimuli significantly different from the ND training group. The ND training groups correctly classified the safe stimuli significantly more often than the greater similarity stimulus training group, in the 90-100 degree range. Counter intuitively, the difference between the greater similarity and ND training groups was in the 90-100 safe degree range, which is the side of the safe class opposite from the trained at-risk stimuli. The difference between the greater similarity and ND training groups in the 90-100 degree range could be explained by differing discriminability along the degree continuum.

The main effect of leg position within the 2 x 2 (Similarity x Leg Position x Assessment Order) ANOVA showed that either (a) the stimuli were not equally discriminable across the forward and backward positions, or (b) the differential range of values in the forward (i.e., 101-165 degrees) and backward (i.e., 27-79 degrees) positions could have affected the distribution of safe and at risk classification responses (Thomas et al., 1991).

The main effect of assessment order in the trained and untrained at-risk classes shows that the stimuli were classified significantly more correct during the earlier assessment sets than the later assessment sets. The assessment order finding could be explained through repeated testing under extinction (Jenkins & Harrison, 1960).

Demonstration

The participants trained with the lesser similarity level extended their leg significantly further beyond the limits of safe behavior, when compared to the greater similarity level, in the trained at-risk direction. The greater similarity groups also performed significantly more accurately than the ND training group in the forward trained at-risk direction, and exhibited a difference that approached significance in the backward trained at-risk direction. The lesser similarity groups and the ND training group did not perform differently from each other in either the forward or backward trained at-risk directions. Through a literature review, no research was identified that had shown different effects between greater and lesser similarity of S+ and S- training on differentiation. The findings in the current study support previous research (Taylor, Olvina, & Alvero, 2008) by showing different effects between DT and ND on differentiation in a trained response class.

The greater similarity group performed significantly closer to the limit of safe behavior than did the ND training group in the forward untrained at-risk direction, but not in the backward direction. The lesser similarity groups and the ND training group did not differ from each other in either the forward or backward untrained at-risk directions. No difference was observed between the greater and lesser similarity levels within the untrained at-risk direction. The findings show that the difference between the more and lesser similarity levels within the trained at-risk direction did not generalize to the untrained at-risk direction; however, the difference between the greater similarity and ND training in the trained at-risk direction did generalize to the forward untrained at-risk direction. Through a literature review, no research was identified that showed different effects between DT and ND training on differentiation in an untrained response class.

On the one hand, the findings across the generalization and demonstration variables were similarly aligned within the similarity factor and DT vs. ND training, thereby suggesting that the independent variables had a similar effect on both the generalization and demonstration variables. On the other hand, there was no main effect of leg position within the demonstration variable and therefore is inconsistent with the effect of leg position found within the generalization variable. A possible explanation for the different effects of the leg factor is that the learning processes that underlie each of the dependent variables were not equally affected by the training.

The main effect of assessment order in the trained and untrained at-risk directions shows that leg extensions were significantly more correct during the later assessment sets than the earlier assessment set. The assessment order finding could be explained through the lesser at-risk (i.e., more correct) leg extensions requiring a reduced amount of response effort than the greater at-risk (i.e., less correct) leg extensions.

Strengths and Limitations of the Study

Strengths. The current study showed the effect of similarity within the generalization and demonstration assessments along a dimension that is socially important in the occupational safety field. The effect also held up across leg positions (i.e., various places along the continuum) and two different dependent variables that are both important within occupational safety. Finally, the greater similarity condition was more effective than the lesser similarity condition and ND training in both the generalization and demonstration assessments. Despite the aforementioned strengths, there are also limitations to the current study.

Limitations. It should be noted that a likely reason for the data from the greater similarity group only approaching significance when compared to the data from the ND training group in the backward trained at-risk direction could have been because of the smaller number of participants in the training group x leg angle analyses relative to the similarity x leg position x assessment order analyses. The similarity x leg position x assessment order analyses had 48 participants per level, but the individual comparison of each of the four similarity groups to the ND training group (i.e., training group x leg angle analyses) reduced the analyses to using only 24 participants per group. Lower power owing to the fewer participants likely contributed to the greater similarity group only approaching significance when compared to the ND training group in the backward trained at-risk direction. The splitting up of the similarity groups for a comparison to the ND training group was necessary because of an assumption of independence within the independent samples ANOVA. The independence assumption would not have been an issue if there would have been twice as many participants run in the ND training group and then using half of the participants' data in the forward position and the other half of participants' data in the backward condition, which would have equated the number of independent samples across the similarity levels. The advantage of comparing each of the four similarity groups to the ND training group was the increased replication of the phenomenon, which held up for the most part across both dependent variables and all greater similarity groups.

Experiment 1 Summary

There were two main purposes of Experiment 1: (a) comparison of DT and ND training, and (b) within DT, an examination of the role of stimulus similarity. Experiment 1 expanded upon basic research and similar bridge research by Taylor et al. (2008) by showing an increase of stimulus control and differentiation through the use of DT (i.e., S+ and S-), in comparison with ND training, in both trained and untrained stimulus/response classes. Experiment 1 expanded upon basic DT research that showed an effect of training stimulus similarity. Most of the previous research demonstrated that increasing the similarity of S+ and S- increased stimulus control through such outcomes as an increase of correct stimulus classification and/or a sharpening of the generalization gradient. The results from Experiment 1 further support such findings and bridge basic research with applied research by demonstrating the effect of stimulus similarity in the context of occupational safety. The current study also expanded upon basic stimulus similarity research by measuring the effect on differentiation. The results showed that greater similarity of S+ and S- stimuli increased the participants' differentiation (i.e., correct approximation) of the limits of safe behavior, when compared to lesser stimulus similarity training. Finally, the generalization and demonstration assessments showed an order effect that weakened stimulus control and improved differentiation.

EXPERIMENT 2

There has been a lack of research that bridges the aforementioned mostly basic training stimulus parameters research and safety research. The goal of Experiment 2 was to assess generalization of safety related stimuli along a leg angle continuum [safe class, and trained and untrained at-risk classes] and demonstration of the leg safety limits [trained and untrained at-risk directions], under extinction with repeated testing, as a function of following training manipulations: (a) the relative frequency of safe [S+] to at-risk [S-] training trials, (b) the absolute frequency of safe [S+] and at-risk [S-] training trials, and (c) the use of a safe class [S+] or both safe and at-risk stimulus classes [S+ and S-].

Based on the aforementioned literature, it is hypothesized that the following training stimulus parameters will be optimal in the context of safety: (a) a lesser relative frequency of safe [S+] to at-risk [S-] training trials, (b) a greater absolute frequency of safe [S+] and at-risk [S-] training trials, and (c) the use of both safe and at-risk stimulus classes [S+ and S-]. It is also hypothesized that the assessments will show: (a) either an increased effect, no effect, or decreased effect with increased testing under extinction [previous research supports all hypotheses, without a consistent outcome], and (b) increased effect in both trained and untrained classes/directions following DT, when compared to ND training.

Method

All procedures used in Experiment 2 were identical to those in Experiment 1, unless noted below. The findings from Experiment 1 were mostly consistent across leg positions, therefore Experiment 2 used only the leg forward position for the trained at-risk stimuli.

Participants

Ninety-six undergraduate students participated (24 per group).

Design and Group Assignment

The study used a 2 x 2 between–groups factorial design (Relative Frequency x Absolute Frequency) to assess the four experimental groups described in the below independent variables section. The participants were assigned to one of the four groups using randomized blocks of four based on their chronological order of participation.

Reliability of Dependent Variables

Two independent observers measured 30% (i.e., measured/total n = 29/96) of the generalization data collection sheets in Experiment 2. An agreement between observers was defined as both persons identifying a trial as correct or incorrect. The scoring of trials on the data collection sheets was 99.97% in agreement (n of trials in agreement/total = 7045/7047). The 29 of 96 generalization data collection sheets (each participant had one generalization data collection sheet) were randomly selected for reliability measures. Each data collection sheet had 243 trials, therefore there were 7047 total trials assessed for reliability of data collection. The percentage of interobserver agreement was calculated by dividing the agreements by the total agreements plus disagreements, and then multiplying by 100. All generalization and demonstration data were calculated using point by point reliability for occurrence and non-occurrence (i.e., safe and at risk data).

Two independent observers measured 100% of the pictures within the demonstration variable for each participant. The pictures taken of all 96 participants were measured for reliability and each participant had 6 pictures, therefore there were 576 total pictures assessed for reliability of data collection. An agreement of leg angle measure between observers was defined as a difference of two degrees or less. The observers disagreed upon 22 pictures and therefore agreed upon measures for 96.18% of the pictures (n of pictures in agreement/total = 554/576). The observers re-measured the 22 disagreed upon pictures, which did not include retraining or any novel instructions, due to an initial disagreement. Following the re-measure all observations were 100% in agreement (n of pictures in agreement/total = 576/576).

Independent Variables: Factors and Levels

The study manipulated two factors, *Relative Frequency* and *Absolute Frequency*, of training trials during DT. The two levels of relative frequency were a *4:1* and *1:4,* safe (S+) to at-risk (S-) training stimulus trials. The two levels for absolute frequency of training trials were *10* and *40,* which included both safe and at-risk stimuli.

The values of safe stimuli used during training were held constant across all groups at 80 and 100 degrees. The at-risk stimuli showed the leg in a forward position (i.e., 105 and 109 degrees) with a greater similarity (i.e., 4 degree separation) to the safe stimuli. The participants were run in the following groups:

1. Safe to at-risk relative frequency was 1:4 with 40 absolute frequency of training trials, 8 safe and 32 at risk

2. Safe to at-risk relative frequency was 1:4 with 10 absolute frequency of training trials, 2 safe and 8 at risk
3. Safe to at-risk relative frequency was 4:1 with 40 absolute frequency of training trials, 32 safe and 8 at risk
4. Safe to at-risk relative frequency was 4:1 with 10 absolute frequency of training trials, 8 safe and 2 at risk

Integrity of the Independent Variables

During Experiment 2, 48% (i.e., n of measured/total = 46/96) of the sessions had two researchers present to collect integrity measures of correct implementation of procedures. An agreement between observers was defined as both persons identifying a procedure as being correctly implemented. The researchers monitored the correct implementation of scripts for participant instructions (e.g., see Appendix A and Appendix B), PowerPoint training for the respective condition, and participant data collection sheet for the respective condition. The researchers agreed on correct implementation of the four procedures for 100% of the sampled sessions.

Furthermore, two independent observers verified that the material within all PowerPoints presentations was 100% correct to ensure the correct stimulus order, formatting, and duration. The PowerPoints were "read only" files to prevent accidental changes throughout running of the experiment. Finally, the picture stimuli used throughout training and generalization assessment were identical to Experiment 1 and therefore had 100% integrity (see Experiment 1 for further information about picture stimulus integrity).

Results

Generalization: Relative Frequency, Absolute Frequency, and Assessment Order Factors

Statistical analyses were used to determine the extent of the effect that the relative frequency of trials, absolute frequency of trials, and assessment order factors had on performance with each of the three classes of stimuli: trained at risk, untrained at risk, and safe. The dependent variable used in the statistical analyses was the number of trials correctly classified within each class of stimuli, collapsed across the leg angle factor. The at-risk stimuli ranged between 27-79 and 101-165 degrees during the generalization assessment, however only data from stimulus angles 27-79 and 101-153 degrees (i.e., 27 data points or 53 degrees per class) were used in all statistical analyses. The data sets from the at-risk classes were made to be of equal size for the analyses so that the classes could be collapsed across leg positions.

Performance within each class of stimuli was assessed by using a 2 x 2 x 3 (Relative Frequency x Absolute Frequency x Assessment Order) mixed factorial ANOVA of the number of trials correctly classified. The relative frequency and absolute frequency were between group independent variables. Assessment order was a within-subject variable. Table 13 shows the groups' mean number of trials correct and Table 14 presents a summary of the ANOVAs from the Experiment 2 generalization assessment.

Figure 10 shows the percent of trials classified as safe across the leg angles in degrees as a function of training with an absolute frequency of *10* training stimuli for each of the relative frequency groups. Figure 11 shows the percent of trials classified as safe across the leg angles in degrees as a function of training with an absolute frequency of *40* training stimuli for each of the relative frequency groups. The following subsections ignore the leg angle factor, therefore the Figures 12-14 better represent the statistical analyses within this section (for analyses the that used the leg angle factor, see the *Generalization: Discrimination Versus Non-differential Stimulus Training, Type of Training x Leg Angle* section, p. 58). The aforementioned figures and tables will be referred to in the appropriate result sections below.

Trained at-risk class. Figure 12 displays the percentage of trained at-risk stimuli classified correctly as a function of the relative frequency and absolute frequency factors. There was a main effect of relative frequency on the correct classification of the trained at-risk stimuli (Table 13 shows the groups' mean number of trials correct and Table 14 presents a summary of the ANOVA). Participants correctly classified the trained at-risk stimuli significantly more frequently when trained with a 1:4 relative frequency (S:AR) than a 4:1 relative frequency. There was no main effect of absolute frequency for the correct classification of the trained at-risk stimuli.

There was a main effect of assessment order for the correct classification of the trained at-risk stimuli. The LSD post-hoc test revealed that the participants classified the trained at-risk stimuli significantly more correctly during the first assessment set than during the second and third assessment sets.

Untrained at-risk class. The percentages of untrained at-risk stimuli classified correctly as a function of the relative frequency and absolute frequency factors are shown in Figure 13. The relative frequency and absolute frequency factors did not have a significant effect on the correct classification of the untrained at-risk stimuli (Table 13 shows the groups' mean number of trials correct and Table 14 presents a summary of the ANOVA).

The assessment order did have a main effect on the correct classification of the untrained at-risk stimuli. The LSD post-hoc tests showed that the participants correctly classified the untrained at-risk stimuli significantly more often during the first assessment set than during the second and third assessment sets.

Safe class. Figure 14 illustrates the percentage of safe stimuli classified correctly as a function of the relative frequency and absolute frequency factors. The relative frequency factor had a main effect on the level of correct classification of the safe stimuli (Table 13 shows the groups' mean number of trials correct and Table 14 presents a summary of the ANOVA). Participants correctly classified the safe stimuli significantly more often when trained with a 4:1 (S:AR) relative frequency than a 1:4 relative frequency. The absolute frequency and assessment order factors did not have a significant effect on the correct classification of the safe stimuli.

Generalization: Discrimination Versus Non-differential Stimulus Training, Type of Training and Leg Angle

Comparison of Experiment 1 and Experiment 2 Groups. The ND training group used for comparison was from Experiment 1. The ND training group was appropriate to use as a part of Experiment 2 because the participant samples from each of the two experiments were similar enough to allow comparison. The following paragraph will describe the statistical similarity between the samples used in Experiment 1 and Experiment 2.

Experiment 1, Group 1 and Experiment 2, Group 3 were exposed to identical experimental conditions. The two groups were compared along each of the three stimulus classes (i.e., trained at risk, untrained at risk, and safe) by using a 2 x 3 (Group x Assessment Order) mixed factorial ANOVA for each class of stimuli. There was no significant main effect for the *group* factor (i.e., Experiment 1, group 1 and Experiment 2, group 3) for any of the three stimulus classes. Therefore, it can be argued that the two samples are similar enough to allow the ND training group to be compared with the experimental groups in Experiment 2. Table 15 and Table 16 show the groups' mean number of trials correct and a summary of the ANOVAs, respectively.

Description of variables and procedures used to compare the relative frequency and non-differential stimulus training groups. Since the absolute frequency factor did not have a main effect in the prior generalization analyses (for more information see the *Generalization: Relative Frequency, Absolute Frequency, and Assessment Order Factors* section on p. 56), only two of the DT groups from Experiment 2 and the ND training group were compared. The two DT groups were the 1:4 (safe:at risk) and 4:1 relative frequency groups that were trained in the 40 absolute frequency level or groups 1 and 3, respectively.

The dependent variable used in the statistical analyses was the percentage of trials correctly classified for each degree within each class of stimuli. The DT groups and the ND training groups were compared in a 3 x 27 or 3 x 21 (Type of Training x Leg Angle) mixed factorial ANOVA for each class of stimuli (i.e., trained at risk, untrained at risk, and safe). The type of training variable (i.e., 1:4 relative frequency, 4:1 relative frequency, and ND training) was a between groups factor. Leg angle was a within-subject variable that had 27 levels in the trained and untrained at-risk classes, and 21 levels in the safe class. The assessment order factor was collapsed within the following analyses.

Statistical analyses described below included the leg angle factor, rather than only using the type of training factor, due to the greater likelihood of showing a difference between the levels in the type of training factor. When a type of training x leg angle ANOVA revealed an interaction, then the levels of leg angle were split into two ranges in order to parse the interaction.

One way ANOVAs were used to assess the differences in between the type of training levels within limited ranges of the leg angle factor. The forward leg position trained at-risk class was split into degree ranges (a) 101-125, and (b) 127-153. The backward leg position untrained at-risk class was split into degree ranges (a) 27-53, and (b) 55-79. The safe class was split into degree ranges (a) 80-90, and (b) 91-100. The levels within the type of training factor were compared using Tukey's HSD as the post-hoc test. A post-hoc test was not used to compare the levels within the leg angle factor due to the lack of useful information for the occupational safety field. Determining which leg angle levels (i.e., degrees) significantly differ from each other is not the emphasis of the current study. The relative frequency groups were compared to the ND training group (i.e., type of training factor) by using Tukey's HSD as the post-hoc test.

Figure 11 shows the percent of trials classified as safe across the leg angles in degrees as a function of training with an absolute frequency of *40* training stimuli for each of the relative frequency groups. Table 17 shows the group mean percentage of trials correct and Table 18 provides a summary of the ANOVAs. The aforementioned figures and tables will be referred to in the appropriate result sections below.

Group 1 and group 3 trained at-risk classes versus non-differential stimulus training. There was a main effect of type of training for the correct classification of the trained at-risk stimuli (Figure 12 conveys the type of training main effect, Table 17 shows the group mean percentage of trials correct, and Table 18 shows a summary of the ANOVA). The post-hoc test showed that the participants who were in either the 1:4 or 4:1 relative frequency groups classified the trained at-risk stimuli significantly more correctly than the ND training group. Additionally, there was a main effect of leg angle and a type of training x leg angle interaction (Figure 11 shows the type of training x leg angle interaction). The 1:4, 4:1, and ND training groups were compared across the 101-125 degree rage and the post-hoc test showed that the 1:4 group correctly classified the trained at-risk stimuli significantly more often than the 4:1 and ND training groups, and the 4:1 group performed significantly better than the ND group. The 1:4 and 4:1 relative frequency groups did not classify the trained at-risk stimuli significantly different within the 127-153 degree range, however both groups classified the trained at-risk stimuli significantly more correctly than the ND training group. This pattern of results indicates a steeper generalization gradient for the 1:4 group than for the 4:1 group.

Group 1 and group 3 versus non-differential stimulus training for untrained at-risk classes. The statistical analysis showed a significant main effect for the type of training factor within the untrained at-risk class of stimuli (Figure 13 conveys the type of training main effect, Table 17 shows the group mean percentage of trials correct, and Table 18 shows a summary of the ANOVA). According to the post-hoc test, the participants in the 1:4 relative frequency group correctly classified the untrained at-risk stimuli significantly more often than the ND training group, but not the 4:1 group.

Additionally, there a main effect of leg angle and a type of training x leg angle interaction (Figure 11 shows the type of training x leg angle interaction). The 1:4, 4:1, and ND training groups were compared across the 27-53 degree range and the post-hoc test revealed that the 1:4 and 4:1 groups did not perform significantly differently from each other; however both correctly classified the untrained at-risk stimuli significantly more often than the ND training group. The 1:4, 4:1, and ND training groups did not classify the untrained at-risk stimuli significantly different within the 55-79 degree range. This pattern of results indicates a steeper generalization gradient for the 1:4 and 4:1 groups than for the ND group.

Group 1 and group 3 safe classes versus non-differential stimulus training. There was no main effect of type of training for the correct classification of the safe stimuli (Figure 14 conveys the lack of type of training main effect, Table 17 shows the group mean percentage of trials correct, and Table 18 shows a summary of the ANOVA); however, there was a main effect of leg angle and a type of training x leg angle interaction (Figure 11 shows the type of training x leg angle interaction). The post-hoc test showed that the ND training group correctly classified the safe stimuli significantly more often than the 1:4 relative frequency group in the 91-100 degree range. The 1:4 and ND training groups did not classify the safe stimuli significantly different within the 80-90 degree range. This pattern of results indicates a steeper generalization gradient for the 1:4 group than for the ND group.

Demonstration: Relative Frequency, Absolute Frequency, and Assessment Order Factors

Statistical analyses were used to determine the nature of the effect that the relative frequency, absolute frequency, and assessment order factors had on each of the two directions of leg extension (i.e., trained at-risk direction and untrained at-risk direction). The dependent variable was the number of degrees the leg was extended from the limits of safe leg behavior (i.e., 80 and 100 degrees). Each direction of leg extension (i.e., trained at-risk and untrained at-risk) was assessed with a 2 x 2 x 3 (Relative Frequency x Absolute Frequency x Assessment Order) mixed factorial ANOVA. The relative and absolute frequency variables were measured between groups. Assessment order was a within-subject variable. For the Experiment 2 demonstration assessment, Table 19 presents the groups mean degree of leg extension and Table 20 shows a summary of the ANOVAs. The aforementioned tables will be referred to in the appropriate result sections below.

Outliers were removed from the demonstration variable data sets in Experiment 2. The boxplot method used for identifying outliers was described in the results section of Experiment 1 (see p. 40). Across the eight applications of the outlier method, 1.22% (i.e., 7/576) of data points were identified as outliers and subsequently removed.

Trained at-risk direction. Figure 15 illustrates the mean number of degrees the leg was extended in the trained at-risk direction, from the safe limits of behavior, as a function of the levels within the relative and absolute frequency factors. A main effect of relative frequency was obtained for the leg extension behavior (Table 19 presents the groups mean degree of leg extension and Table 20 shows a summary of the ANOVA).

Participants trained with 4:1 relative frequency (S:AR) over-extended their leg significantly further beyond the safe limit, in the trained at-risk direction, than those trained with the 1:4 relative frequency level. There was no main effect of absolute frequency factor for the leg extension behavior in the trained at-risk direction. There was a main effect of the assessment order factor for the leg extension behavior. According ot the post-hoc test, the participants extended their leg significantly further beyond the limit of safe behavior, in the trained at-risk direction, during the first extension assessment than the second and third extension assessments.

Untrained at-risk direction. Figure 16 diagrams the mean number of degrees the leg was extended in the untrained at-risk direction, from the safe limits of behavior, as a function of the levels within the relative and absolute frequency factors. The statistical analysis did not yield a significant effect of the relative frequency, absolute frequency, or assessment order factors on leg extension behavior in the untrained at-risk direction (Table 19 presents the groups mean degree of leg extension and Table 20 shows a summary of the ANOVA).

Demonstration: Discrimination Versus Non-differential Stimulus Training, Type of Training and Assessment Order

Comparison of Experiment 1 and Experiment 2 Groups. The ND training group from Experiment 1 was used in Experiment 2. The ND training group was appropriate to use as a part of Experiment 2 because the participant samples from each of the two experiments were similar. The following paragraph will describe the statistical similarity between the samples used in Experiment 1 and Experiment 2.

Experiment 1, Group 1 and Experiment 2, Group 3 were exposed to identical experimental conditions. The two groups were compared along each of the two directions of leg extension (i.e., trained at-risk direction and untrained at-risk direction). by using a 2 x 3 (Group x Assessment Order) mixed factorial ANOVA. A significant main effect for the group factor (i.e., Experiment 1, Group 1 and Experiment 2, Group 3) was not obtained for either of the leg extension directions. Therefore, it can be argued that the two samples are similar enough to allow the ND training group to be compared with the experimental groups in Experiment 2. Table 21 and Table 22 show the groups mean degree of leg extension and shows a summary of the ANOVAs, respectively.

Description of variables and procedures used to compare the relative frequency and non-differential stimulus training groups. Statistical analyses were used to determine the nature of the effect that the two experimental groups had by comparison to the ND training group along each of the two leg extension directions (i.e., trained at-risk direction and untrained at-risk direction). The two experimental groups compared to the ND training group were the 1:4 relative frequency (safe:at risk) and 4:1 relative frequency groups that were trained with 40 stimuli or Group 1 and Group 3, respectively.

The dependent variable used in the statistical analyses was the number of degrees the leg was extended from the limit of safe leg behavior (i.e., 80 and 100 degrees). Each direction of leg extension was assessed by using a 3 x 3 (Type of Training x Assessment Order) mixed factorial ANOVA. The type of training variable (i.e., 1:4 relative frequency, 4:1 relative frequency, and ND training) was measured between groups. Assessment order was a within-subject variable.

The statistical analyses described below are intended for a comparison of the relative frequency groups to the ND training group, not a comparison of the relative frequency groups to each other. The relative frequency groups were compared to each other in the aforementioned section (see the *Demonstration: Relative Frequency, Absolute Frequency, and Assessment Order Factors* section on p. 63). Tukey's HSD was used as the post-hoc test to compare the relative frequency groups to the ND training group. Finally, the statistical analyses were not intended to evaluate assessment order because that factor, except for the ND training level, was already a focus within the preceding section (the *Demonstration: Relative Frequency, Absolute Frequency, and Assessment Order Factors* section). Table 23 presents the groups' mean degree of leg extension and Table 24 shows a summary of the ANOVAs. The aforementioned tables will be referred to in the appropriate result sections below.

Outliers were removed from the demonstration variable data sets in Experiment 2. The boxplot method used for identifying outliers was described in the results section of Experiment 1 (see p. 40). Across the eight applications of the outlier method, 1.62% (i.e., 7/432) of data points were identified as outliers and subsequently removed.

Group 1 and group 3 trained at-risk direction versus non-differential stimulus training. Figure 15 illustrates the mean number of degrees the leg was extended in the trained at-risk direction, from the safe limits of leg behavior, as a function of the levels within the type of training factor. The statistical analysis showed a main effect of type of training for the leg extension behavior (Table 23presents the groups mean degree of leg extension and Table 24 shows a summary of the ANOVA). The post-hoc test showed that the participants trained in the ND training group over-extended their leg significantly further beyond the safe limit, in the trained at-risk direction, when compared to either the 1:4 relative frequency or 4:1 relative frequency groups.

Group 1 and group 3 untrained at-risk direction versus non-differential stimulus training. Figure 16 presents the mean number of degrees the leg was extended in the untrained at-risk direction, from the safe limits of leg behavior, as a function of the levels within the type of training factor. A significant main effect was obtained for the type of training factor within the untrained at-risk leg extension direction (Table 23presents the groups mean degree of leg extension and Table 24 shows a summary of the ANOVA). According to the post-hoc test, the participants trained in the ND training group over-extended their leg significantly further beyond the safe limit, in the untrained at-risk direction, when compared to the 1:4 relative frequency group. The 4:1 relative frequency group did not extend their leg significantly different from the ND training group in the untrained at-risk direction.

Discussion

The discussion section for Experiment 2 (a) describes an interpretation of the effects of the factors on the generalization and demonstration data, (b) lists strengths/limitations, and (c) provides a summary of the findings.

Generalization

The levels with a 1:4 relative frequency of training stimuli exhibited significantly greater percent of correct classification of trained at-risk stimuli when compared to the 4:1 relative frequency or ND training levels. The 4:1 relative frequency group showed significantly greater correct classification of the trained at-risk stimuli, when compared to the ND training group. The effect of relative frequency, when comparing 4:1 and 1:4 groups, was consistent with the basic literature (Thomas & Vogt, 1983; Thomas, Windell, Williams, & White, 1985) that showed a lesser S+ relative frequency increased correct classification of the novel stimuli between the S+ and S- values. It could be argued that the relative frequency created a response bias during the generalization assessment, based on the differing proportion of S+ and S- during training (Thomas et al., 1985), rather than increasing stimulus control. Finally, the finding that the DT groups (i.e., 4:1 and 1:4 relative frequency) performed better than the ND training group was consistent with previous basic and applied literature (Hanson, 1959; Jenkins & Harrison, 1960; Lyons, & Thomas, 1967; Switalski, Lyons & Thomas, 1966; Taylor, Olvina, & Alvero, 2008; Tomie et al., 1975).

The results comparing the 4:1 and 1:4 relative frequency group suggest a sharper gradient and show greater correct classifications of the untrained at-risk stimulus class, respectively, than the ND training group. The 4:1 and 1:4 relative frequency groups did not perform significantly differently in the untrained at-risk class. The finding that the DT groups had a greater correct classification of the untrained S- class is supported by the animal literature (e.g., Jenkins & Harrison, 1960) and extends the occupational safety literature. It could be argued that the effect of the trained S- class generalized to the untrained S- class.

The 4:1 relative frequency and ND training groups correctly classified significantly more safe stimuli than the 1:4 relative frequency group. The correct classification difference in the safe class between the 4:1 relative frequency and ND training groups vs. the 1:4 relative frequency group could be due to a response bias caused by the differing proportions of safe to at-risk stimuli during training (Thomas et al., 1985).

Across all classes of stimuli, the analyses did not show an effect of absolute frequency. The findings in the current study are not consistent with previous research (Hearst & Koresko, 1968; Thomas, 1993) which found that an increase of absolute frequency of training increased stimulus control. It is possible that the lack of absolute frequency effect could be due to a ceiling effect or the number of training trials used in each of the absolute frequency levels were not sufficiently different.

Finally, there was an effect of assessment order in the trained and untrained at-risk classes. The groups showed significantly greater correct classification during the first assessment set than the second and third sets. Repeated testing under extinction could explain the assessment order phenomenon (Jenkins & Harrison, 1960).

Demonstration

The participants trained with the 4:1 relative frequency or ND training extended their leg significantly further beyond the limit of safe behavior than the 1:4 relative frequency group, in the trained at-risk direction. Furthermore, the ND training group also extended the leg significantly further beyond the same limit in the trained at-risk direction, than the 4:1 relative frequency group. Through a literature review, no research was identified that had shown different effects between greater or lesser relative frequency of S+ to S- on differentiation. The findings in the current study support previous research (Taylor, Olvina, & Alvero, 2008) by showing a difference between DT and ND conditions on differentiation in a trained response class. The effect of relative frequency in the demonstration assessment is consistent with the findings in the generalization assessment. Therefore, the argument of either response bias or greater stimulus control/differentiation could be made because the findings are consistent across the generalization and demonstration assessments. The bottom line is that the effect is similar across the two assessments which is arguably the most important. Had the results shown that the there was an effect of relative frequency in the generalization assessment, but not the demonstration assessment, then it is more plausible that the effect was due to a response bias. Nevertheless, the effect of relative frequency transfers response modalities, from paper and pencil (generalization assessment) to the leg extensions (demonstration assessment).

The ND training group extended the leg significantly further beyond the safe limit in the untrained at-risk direction, than the 1:4 relative frequency group. Therefore, it appears that the effect of training with a 1:4 relative frequency generalizes from trained at-risk direction to the untrained at-risk direction. The 4:1 relative frequency group did not show significantly different leg extension behavior when compared to the 1:4 relative frequency or ND training groups, in the untrained at-risk direction. Through a literature review, no research was identified that showed different effects between DT and ND training on differentiation in an untrained response class.

Across the trained and untrained leg extension direction, the analyses did not show an effect of absolute frequency, which does not support previous research (Hearst & Koresko, 1968; Thomas, 1993) that found an increase of absolute frequency of training increased stimulus control. The findings in the demonstration assessment are consistent with the findings reported in generalization assessment section. It is possible that the lack of absolute frequency effect could be due to a ceiling effect or the number of training trials used in each of the absolute frequency levels were not sufficiently different.

Finally, there was an effect of assessment order only in the trained at-risk direction. The leg extensions were significantly more correct during the later assessment sets than the earlier assessment set. The assessment order finding could be explained through the less at-risk (i.e., more correct) leg extensions requiring a reduced amount of response effort than the more at-risk (i.e., less correct) leg extensions.

Strengths and Limitations of the Study

Strengths. Experiment 2 used a behavioral dimension that is important within the occupational safety field, leg position, and demonstrated the effects of the relative frequency of training stimuli. The lesser S+ relative frequency showed consistently greater stimulus control and differentiation or response bias across the generalization and demonstration variables, when compared to the greater S+ relative frequency and ND training groups.

Limitations. The current study did not show an effect of the absolute frequency of training trials, which could be the result of a ceiling effect or the number of training trials used in each of the absolute frequency levels were not sufficiently different. Basic DT research (e.g., Hearst & Koresko, 1968; Thomas, 1993) found that an increase of absolute frequency of training increased stimulus control. Another limitation was that the participants in the ND training group were not randomly assigned from the same population because they were from Experiment 1. Finally, the current analyses did not assess correct stimulus classification and bias independent of each other. Further analyses should be considered (e.g., *d'* and logarithm of *β*).

Experiment 2 Summary

There were three main purposes of Experiment 2: (a) comparison of 4:1 and 1:4 relative training stimulus frequencies, (b) an evaluation of the effects of 10 and 40 absolute training stimulus frequencies, and (c) comparison of DT and ND training. Experiment 2 expanded upon basic discrimination training research that explored the effects of the relative frequency and absolute frequency of S+ and S- trials used during DT. Basic research has shown that a lesser S+ relative frequency, when compared to a greater S+ relative frequency, increased correct stimulus classification and/or generalization gradient sharpness. The results from Experiment 2 support basic research by showing a similar effect through increased gradient sharpness and correct classification of the trained at-risk stimuli. The current study was conducted in the context of occupational safety, and therefore bridged basic research with applied research by demonstrating the effect of the relative training stimulus frequencies. The basic research was also expanded upon in the current study by measuring the effect of relative training stimulus frequencies on differentiation. The results showed that the lesser relative frequency of S+ during training, when compared to a greater relative frequency, increased the participants' differentiation (i.e., correct approximation) of the limits of safe behavior. Interestingly, 4:1 and/or 1:4 DT groups were shown to perform more correctly in the untrained at-risk class/direction, when compared to ND training group. The effect of DT in the untrained S- class is supported by basic literature; however, no literature was identified that showed an effect of differentiation in an untrained direction of response class. Additionally, Experiment 2 showed that the absolute frequency of training stimuli did not have an effect on safety generalization and differentiation, which is consistent with the reported effect within the basic research that an increase of absolute frequency of training increases stimulus control. Finally, the generalization and demonstration assessments showed an order effect that weakened stimulus control and improved differentiation.

EXPERIMENT 1 AND EXPERIMENT 2

General Discussion

Overall Summary

The purpose of the present study was to assess generalization of safety related stimuli along a leg angle continuum and demonstration of the leg safety limits. The dependent variables were assessed as a function of following training manipulations: (a) the use of a safe class [S+] or both safe and at-risk stimulus classes [S+ and S-], (b) the similarity of the at-risk [S-] stimuli to the safe [S+] stimuli along a continuum; (c) the relative frequency of safe [S+] to at-risk [S-] training trials; and (d) the absolute frequency of safe [S+] and at-risk [S-] training trials.

The results showed that the participants who were exposed to DT [S+ and S-], rather than ND training [S+], in Experiment 1 and 2 correctly classified the stimuli in trained at-risk class significantly more often. The findings also showed that the ND training group extended their leg significantly further beyond the limits of safe behavior in the trained at-risk direction than the DT groups with greater similarity and 4:1 or 1:4 relative frequencies. Additionally, the DT with greater similarity and either 4:1 or 1:4 relative frequencies showed a greater effect on correct classification of the untrained at-risk stimuli than the ND training group. The ND training group extended their leg significantly further beyond the limits of safe behavior than DT groups with (a) greater similarity and a 1:4 relative frequency and (b) greater similarity and a 4:1 relative frequency in the forward untrained at-risk direction, but not the backward untrained at-risk direction.

The findings revealed that groups given DT with greater stimulus similarity (i.e., 4 degree separation) between safe [S+] and at-risk [S-] stimuli, rather than lesser similarity (i.e., 34 degree separation), significantly increased correct classification of the trained at-risk stimuli. The participants trained with the lesser similarity level extended their leg significantly further beyond the limits of safe behavior, when compared to the greater similarity level, in the trained at-risk direction. The level trained with greater similarity correctly classified the untrained at-risk stimuli significantly more often than the lesser similarity level. Despite the effect of training stimulus similarity on the untrained at-risk generalization data, there was no difference between the greater and lesser similarity levels on the untrained at-risk leg extension direction.

The current study found that the level of 1:4 [S+:S-] relative frequency of training trials correctly classified the trained at-risk stimuli significantly more often than the 4:1 relative frequency level. Also, the participants trained in the 4:1 relative frequency level extended their leg significantly further beyond the limits of safe behavior in the trained at-risk direction, than the 1:4 relative frequency level. Participants in the 4:1 and 1:4 relative frequency levels did not perform different from each other within the untrained at-risk stimulus class in the generalization assessment and untrained at-risk response direction in the demonstration assessment. The effect of relative frequency could be reconceptualized as a proportion safe. The 1:4, 4:1, and ND training groups have a proportion safe of .20, .80, and 1.00, respectively. The findings show that a lower proportion safe increases correct classification of the at-risk stimuli and increases differentiation.

The participants trained with 40 and 10 absolute frequency levels of training trials did not perform differently in either the generalization or demonstration assessments. The findings from comparing the absolute frequency levels are not consistent with the basic literature (e.g., Hearst & Koresko, 1968; Thomas, 1993), which show increased stimulus control with an increased absolute frequency of training trials. An argument could be made that the lack of effect could be attributed to a ceiling effect or the number of training trials used in each of the absolute frequency levels were not sufficiently different.

Potential Impact of the Current Studies on the Industry

Experiment 1 and Experiment 2 indicated that safety training should use (a) both safe [S+] and at-risk [S-] stimuli, (b) greater similarity between the safe [S+] and at-risk [S-] stimuli, and (c) lesser relative frequency of safe [S+] to at-risk [S-] stimuli. Most trainings that are developed could feasibly create pictures and videos of a model showing examples of both safe and at-risk behavior, because the at-risk behavior that a model would demonstrate would not pose a high level of danger. Nevertheless, there will likely be some behavior that is too dangerous to have a model demonstrate for the sake of creating example pictures and videos of at-risk behavior, because the level of danger is too high. In the case where the at-risk behavior could not be demonstrated by a model for development of pictures and video, then perhaps the model could demonstrate “safe” behavior that is approaching the limit of “at risk.” Overall, nearly any training could use safe and at-risk/nearly at-risk examples to enhance the effectiveness of safety DT.

The similarity between the safe and at-risk examples is a parameter that could be used in most trainings. The trainings that cover anything that uses a small or large continuum could apply the use of greater similarity between safe and at-risk examples.

The following is a short list of examples where the similarity parameter could be applied to training safe behavior and conditions: posture sitting at a desk or in a vehicle; posture while lifting, or digging with a shovel; posture or appropriate weight while manually carrying items; walking distance from an edge without a railing; distance from persons and other equipment while heavy equipment is being operated; identifying safe and at-risk conditions like size of crack in the ceiling of a mine, the extent of moisture on a surface, the extent of gas in an enclosed area, the extent of dust particles in the air, the pitch of a roof without a harness, the appropriate angle of a ladder, etc. Aside from the many behaviors and work conditions that the similarity parameter could be applied, there are categorical behaviors and conditions that it could not be applied: wearing seatbelt while vehicle is moving; wearing personal protection equipment while on a job site; putting a gun on “lock” while in a holster; lowering the lift on a forklift before moving; closing/locking fences and gates to mechanically moving, hot, or live electricity areas; no alcohol or guns on job site; etc.

The relative frequency parameter could be applied to nearly any training, so long as time is not a major constraint. The DT portion within the current experiment was approximately 10 minutes or less. The major portion of the 40-60 minute sessions was comprised of the generalization and demonstration assessments, which are not necessary within job training. Therefore, the lesser relative frequency of safe examples could be applied to nearly any training.

Finally, it should be noted that a benefit of targeting behavior or conditions within safety training is that the effects will generalize to untrained areas. The generalization effect to untrained areas could mean that training of correct back posture at a desk would also generalize to different settings (e.g., back posture while sitting in a vehicle) and other posture behaviors (e.g., leg, arm, or shoulder). The generalization effect from training could improve work safety beyond the areas that are specifically targeted within the training.

The aforementioned stimulus parameters could have a notable impact on the effectiveness of safety training. Greater effectiveness of safety training is important due to the potential that the training could influence behavior that is vital to the effectiveness of the behavioral safety approach. The behavioral safety approach has been shown to be very effective in improving worker safety, yet the effectiveness can be notably limited by inadequate training. Therefore, the parameters within training should be manipulated to facilitate the behavioral safety approach.

Recommendations for Future Research

The current study could be expanded upon in several ways. First, it would be important to replicate the findings with a more complex behavior and in an environment that more closely resembles a natural work setting. For example, an experiment could vary the number of components that are a part of chained responding or a smaller range of safe topography, 5 degrees vs. 20 degrees, at a construction or manufacturing site. Second, a parametric manipulation of the similarity or relative frequency values could provide important information for improving safety training. A parametric analysis could use relative frequencies of safe to at-risk training examples of 19:1, 1:19, 9:1, 1:9, 4:1, and 1:4 with absolute frequencies of 160, 80, 40, and 20. Third, a manipulation of the amount of risk associated an at-risk behavior, varied through information, could be an additional parameter to be addressed in training. The amount of risk could be manipulated through informing a participant of the limit of the number of safety violations that are allowable before being permanently or temporarily removed from the money earning component of the experiment. The number limit of violations would be the independent variable while percentage safe/rate of widgets completed could be the dependent variable. Fourth, the accuracy of observational data used for feedback is affected by training, but it is unknown how the feedback accuracy affects worker safety. The accuracy of feedback could be manipulated through over-estimating and under-estimating the participant's actual safety performance. Fifth, future research could assess the possible effects of a safety related stimulus and/or response class size on generalization. A researcher could manipulate the range of stimuli within a stimulus class and topography within a response class to determine the effects on generalization and induction, respectively. Sixth, it could be important to explore the independent variables that moderate the extent of generalization of skills from a trained at-risk dependent variable to an untrained at-risk dependent variable. The untrained measure would be a common measure in many experiments and therefore not require any special procedures. Seventh, the amount of time and money that would be required to train individuals from a DT format versus a behavior skills training (BST) format could be evaluated. For example, a survey could be sent to various consulting firms asking for a quote of their services, which could be DT and/or BST. Finally, future research could compare groups where one is exposed one time to DT and then repeatedly tested under extinction while another group is repeatedly exposed to DT and tested under extinction in tandem. For example, two groups of participants would be given a generalization assessment during each of the five consecutive sessions, although one group would be given training only prior to the first assessment, while the second group would be given training prior to each assessment.

TABLE 1. Number of participants per group, and group mean and SD of number of correctly classified trials for the Assessment Order x Leg Position x Similarity Factorial ANOVAs for Experiment 1, Generalization data.

			Trained At-risk Class			Untrained At-risk Class			_Safe Class_		
AO	Leg	Similarity	M	SD	n	M	SD	n	M	SD	n
1	back	greater	21.96	3.62	24	20.75	7.49	24	20.00	1.77	24
		lesser	14.25	2.58	24	18.75	8.84	24	20.92	0.28	24
		Total	18.10	4.98	48	19.75	8.17	48	20.46	1.34	48
	forw	greater	24.38	2.52	24	16.13	4.32	24	20.38	1.58	24
		lesser	20.87	3.79	24	14.96	5.14	24	20.71	0.86	24
		Total	22.63	3.64	48	15.54	4.73	48	20.54	1.27	48
	Total	greater	23.17	3.32	48	18.44	6.48	48	20.19	1.67	48
		lesser	17.56	4.64	48	16.85	7.40	48	20.81	0.64	48
		Total	20.36	4.90	96	17.65	6.97	96	20.50	1.30	96
2	back	greater	19.88	5.08	24	20.46	7.08	24	20.00	1.64	24
		lesser	13.17	2.73	24	17.92	9.02	24	20.75	0.74	24
		Total	16.52	5.27	48	19.19	8.12	48	20.38	1.32	48
	forw	greater	23.38	3.16	24	14.25	4.87	24	20.38	1.72	24
		lesser	19.96	4.67	24	12.63	5.84	24	20.46	1.67	24
		Total	21.67	4.30	48	13.44	5.38	48	20.42	1.67	48
	Total	greater	21.63	4.54	48	17.35	6.78	48	20.19	1.67	48
		lesser	16.56	5.11	48	15.27	7.98	48	20.60	1.28	48
		Total	19.09	5.44	96	16.31	7.44	96	20.40	1.50	96
3	back	greater	19.54	5.28	24	19.92	7.31	24	19.96	1.81	24
		lesser	12.58	2.89	24	18.96	8.27	24	20.67	0.64	24
		Total	16.06	5.49	48	19.44	7.74	48	20.31	1.39	48
	forw	greater	23.71	3.64	24	14.58	4.59	24	20.63	1.01	24
		lesser	19.46	4.47	24	12.25	4.71	24	20.67	1.24	24
		Total	21.58	4.57	48	13.42	4.75	48	20.65	1.12	48
	Total	greater	21.63	4.96	48	17.25	6.61	48	20.29	1.49	48
		lesser	16.02	5.09	48	15.60	7.47	48	20.67	0.98	48
		Total	18.82	5.74	96	16.43	7.07	96	20.48	1.27	96
Total	back	greater	20.46	4.77	24	20.38	7.20	24	19.99	1.72	24
		lesser	13.33	2.78	24	18.54	8.60	24	20.78	0.59	24
		Total	16.90	5.29	48	19.46	7.96	48	20.38	1.34	48
	forw	greater	23.82	3.12	24	14.99	4.60	24	20.46	1.45	24
		lesser	20.10	4.31	24	13.28	5.32	24	20.61	1.28	24
		Total	21.96	4.19	48	14.13	5.03	48	20.54	1.37	48
Total	Total	greater	22.14	4.36	48	17.68	6.60	48	20.22	1.60	48
		lesser	16.72	4.96	48	15.91	7.60	48	20.69	1.00	48

Note: *AO* and *Leg* represent the Assessment Order and Leg Position factors, respectively. *Total* signals the collapsing of levels for the respective factor. The leg level identifies the leg position that was used during training. The at-risk classes and safe class had a total of 27 and 21 trials, respectively.

TABLE 2. Summary of the Similarity x Leg Position x Assessment Order Factorial ANOVAs for Experiment 1, Generalization data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Trained At-risk Class						
A (Similarity)	2117.92	1	2117.92	54.843	< 0.001**	0.373
B (Leg Position)	1845.281	1	1845.28	47.783	< 0.001**	0.342
AB	208.42	1	208.42	5.397	0.022*	0.055
Error	3552.847	92	38.618			
C (Assessment Order)	130.083	2	65.042	25.457	< 0.001**	0.217
AC	4.694	2	2.347	0.919	0.401	0.01
BC	12.25	2	6.125	2.397	0.094	0.025
ABC	6.861	2	3.431	1.343	0.264	0.014
Error	470.111	184	2.555			
Untrained At-risk Class						
A (Similarity)	225.781	1	225.781	1.837	0.179	0.02
B (Leg Position)	2042.67	1	2042.67	16.621	< 0.001**	0.153
AB	0.281	1	0.281	0.002	0.962	0
Error	11306.181	92	122.893			
C (Assessment Order)	104.84	2	52.42	10.011	< 0.001**	0.098
AC	3.563	2	1.781	0.34	0.712	0.004
BC	45.882	2	22.941	4.381	0.014*	0.045
ABC	20.271	2	10.135	1.936	0.147	0.021
Error	963.444	184	5.236			
Safe Class						
A (Similarity)	16.056	1	16.056	3.519	0.064	0.037
B (Leg Position)	1.681	1	1.681	0.368	0.545	0.004
AB	7.347	1	7.347	1.61	0.208	0.017
Error	419.75	92	4.563			
C (Assessment Order)	0.583	2	0.292	0.688	0.504	0.007
AC	0.861	2	0.431	1.016	0.364	0.011
BC	1.194	2	0.597	1.409	0.247	0.015
ABC	0.028	2	0.014	0.033	0.968	0
Error	78	184	0.424			

Assessment Order Post Hoc Test: LSD

Source	M Difference	*p*
Trained At-risk Class		
1 vs. 2	1.27	< 0.001**
1 vs. 3	1.54	< 0.001**
2 vs. 3	0.27	0.167
Untrained At-risk Class		
1 vs. 2	1.33	< 0.001**
1 vs. 3	1.22	0.002**
2 vs. 3	-0.12	0.735

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007). In the post hoc test, the numbers 1, 2, and 3 refer levels within Assessment Order.
*p< 0.05; **p< 0.01

TABLE 3. Number of participants per group, and group mean and SD of percentage of correctly classified trials for the Type of Training (Groups 1, 2, and ND training) x Leg Angle Factorial ANOVAs for Experiment 1, Generalization data.

	Trained At-risk Class			Untrained At-risk Class			Safe Class			
Type of Training	Leg Angle	M	SD	Leg Angle	M	SD	Leg Angle	M	SD	n
Greater Similarity	101-125	78.21	19.41	27-53	88.69	13.12	80-90	99.87	0.62	24
	127-153	97.52	5.59	55-79	19.77	23.49	91-100	94.72	13.93	24
	Total	88.22	11.56	Total	55.52	17.04	Total	97.43	6.90	24
Lesser Similarity	101-125	50.75	26.75	27-53	79.17	22.26	80-90	99.12	3.74	24
	127-153	96.43	6.66	55-79	16.88	19.05	91-100	97.08	8.42	24
	Total	74.44	15.96	Total	49.19	19.70	Total	98.14	6.10	24
ND Training	101-125	34.51	29.24	27-53	66.87	37.63	80-90	97.35	9.84	24
	127-153	78.13	32.53	55-79	21.15	28.97	91-100	97.92	4.38	24
	Total	57.04	29.19	Total	44.81	30.30	Total	97.62	7.57	24

Note: *Total* signals the collapsing of levels for the respective factor. Group 1 and 2 were trained with at-risk stimuli in the forward position.

TABLE 4. Summary of the Type of Training (Groups 1, 2, and ND training) x Leg Angle Factorial ANOVAs for Experiment 1, Generalization data. Summary of the Type of Training (Groups 1, 2, and ND training) One-way ANOVA for Experiment 1, Generalization data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Forward Trained At-risk Class						
A (Type of Training)	314629.92	2	157314.958	15.329	< 0.001**	0.308
Error	708123.71	69	10262.663			
B (Leg Angle)	1166384.03	26	44860.924	109.571	< 0.001**	0.614
AB	131025.95	52	2519.73	6.154	< 0.001**	0.151
Error	734503.60	1794	409.422			
Backward Untrained At-risk Class						
A (Type of Training)	37170.78	2	18585.391	1.38	0.258	0.038
Error	929038.07	69	13464.32			
B (Leg Angle)	2057932.10	26	79151.235	147.801	< 0.001**	0.682
AB	63971.19	52	1230.215	2.297	< 0.001**	0.062
Error	960730.45	1794	535.524			
Safe Class						
A (Type of Training)	142.56	2	71.282	0.081	0.922	0.002
Error	60368.17	69	874.901			
B (Leg Angle)	2620.52	20	131.026	1.496	0.073	0.021
AB	4857.44	40	121.436	1.386	0.056	0.039
Error	120881.83	1380	87.596			

Main Effect Type of Training Post Hoc Test: Tukey's HSD

Source	M Difference	*p*
Forward Trained At-risk Class		
Greater vs. Lesser Similarity	13.79	0.044*
ND Training vs. Greater Similarity	31.10	< 0.001**
ND Training vs. Lesser Similarity	17.31	0.008**

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).
*$p < 0.05$; **$p < 0.01$

TABLE 4 CONTINUED. Summary of the Type of Training (Groups 1, 2, and ND training) x Leg Angle Factorial ANOVAs for Experiment 1, Generalization data. Summary of the Type of Training (Groups 1, 2, and ND training) One-way ANOVA for Experiment 1, Generalization data.

AB Interaction Post Hoc Test: One-way ANOVA/Tukey's HSD

Sources	SS	df	MS	*F*	*p*
Backward Untrained At-risk Class:Type of Training x Leg Angles 27-53					
A (Type of Training)	5747.04	2	2873.52	4.137	0.02*
Error	47929.89	69	694.64		
Source			M Difference	*p*	
Greater vs. Lesser Similarity			9.52	0.427	
ND Training vs. Greater Similarity			21.83	0.015*	
ND Training vs. Lesser Similarity			12.30	0.245	
Sources	SS	df	MS	*F*	*p*
Backward Untrained At-risk Class:Type of Training x Leg Angles 55-79					
A (Type of Training)	228.10	2	114.05	0.195	0.823
Error	40350.37	69	584.79		

Note: *p< 0.05; **p< 0.01

TABLE 5. Number of participants per group, and group mean and SD of percentage of correctly classified trials for the Type of Training (Groups 3, 4, and ND training) x Leg Angle Factorial ANOVAs for Experiment 1, Generalization data.

	Trained At-risk Class			Untrained At-risk Class			Safe Class			
Type of Training	Leg Angle	M	SD	Leg Angle	M	SD	Leg Angle	M	SD	n
Greater Similarity	27-53	94.74	10.15	101-125	60.58	29.39	80-89	96.94	7.08	24
	55-79	55.34	29.44	127-153	89.29	28.48	90-100	93.56	10.92	24
	Total	75.77	17.68	Total	75.46	26.66	Total	95.17	8.17	24
Lesser Similarity	27-53	86.71	11.47	101-125	54.17	29.58	80-89	99.44	1.27	24
	55-79	9.19	8.42	127-153	82.14	34.71	90-100	98.48	3.90	24
	Total	49.38	10.31	Total	68.67	31.86	Total	98.94	2.79	24
ND Training	27-53	66.87	37.63	101-125	34.51	29.24	80-89	97.08	10.83	24
	55-79	21.15	28.97	127-153	78.13	32.53	90-100	98.11	3.98	24
	Total	44.80	30.29	Total	57.05	29.17	Total	97.62	7.58	24

Note: *Total* signals the collapsing of levels for the respective factor. Group 3 and 4 were trained with at-risk stimuli in the backward position.

TABLE 6. Summary of the Type of Training (Groups 3, 4, and ND training) x Leg Angle Factorial ANOVAs for Experiment 1, Generalization data. Summary of the Type of Training (Groups 3, 4, and ND training) One-way ANOVA for Experiment 1, Generalization data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Backward Trained At-risk Class						
A (Type of Training)	361290.58	2	180645.29	15.887	< 0.001**	0.315
Error	784566.19	69	11370.524			
B (Leg Angle)	1766366.03	26	67937.155	132.515	< 0.001**	0.658
AB	228215.59	52	4388.761	8.561	< 0.001**	0.199
Error	919739.37	1794	512.675			
Forward Untrained At-risk Class						
A (Type of Training)	111406.32	2	55703.161	2.527	0.087	0.068
Error	1520853.48	69	22041.355			
B (Leg Angle)	1058225.59	26	40700.984	91.909	< 0.001**	0.571
AB	56101.40	52	1078.873	2.436	< 0.001**	0.066
Error	794459.02	1794	442.842			
Safe Class						
A (Type of Training)	3687.54	2	1843.768	2.468	0.092	0.067
Error	51549.82	69	747.099			
B (Leg Angle)	3224.57	20	161.229	1.469	0.083	0.021
AB	6374.19	40	159.355	1.452	0.035*	0.04
Error	151459.44	1380	109.753			

Type of Training Post Hoc Test: Tukey's HSD

Source	M Difference	*p*
Backward Trained At-risk Class		
Greater vs. Lesser Similarity	26.39	< 0.001**
ND Training vs. Greater Similarity	30.92	< 0.001**
ND Training vs. Lesser Similarity	4.53	0.726

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).

**p* < 0.05; ***p* < 0.01

TABLE 6 CONTINUED. Summary of the Type of Training (Groups 3, 4, and ND training) x Leg Angle Factorial ANOVAs for Experiment 1, Generalization data. Summary of the Type of Training (Groups 3, 4, and ND training) One-way ANOVA for Experiment 1, Generalization data.

AB Interaction Post Hoc Test: One-way ANOVA/Tukey's HSD

Sources	SS	df	MS	*F*	*p*
Forward Untrained At-risk Class:Type of Training x Leg Angles 101-125					
A (Type of Training)	8856.746	2	4428.373	5.123	0.008**
Error	59646.34	69	864.44		
Source			M Difference	*p*	
Greater vs. Lesser Similarity			6.41	0.731	
ND Training vs. Greater Similarity			26.07	0.008**	
ND Training vs. Lesser Similarity			19.66	0.06	
Sources	SS	df	MS	*F*	*p*
Forward Untrained At-risk Class:Type of Training x Leg Angles 127-153					
A (Type of Training)	1533.801	2	766.901	0.748	0.477
Error	70705.251	69	1024.714		

Sources	SS	df	MS	*F*	*p*
Safe Class:Type of Training x Leg Angles 80-89					
A (Type of Training)	94.753	2	47.377	0.841	0.436
Error	3886.574	69	56.327		
Sources	SS	df	MS	*F*	*p*
Safe Class:Type of Training x Leg Angles 90-100					
A (Type of Training)	360.422	2	180.211	3.599	0.033*
Error	3455.005	69	50.073		
Source			M Difference	*p*	
Greater vs. Lesser Similarity			-4.92	0.048*	
ND Training vs. Greater Similarity			-4.55	0.074	
ND Training vs. Lesser Similarity			0.38	0.981	

Note: *p< 0.05; **p< 0.01

TABLE 7. Number of participants per group, and group mean and SD of number of degrees the leg was extended beyond safe limits for the Assessment Order x Leg Position x Similarity Factorial ANOVAs for Experiment 1, Demonstration data.

			Trained At-risk Direction			Untrained At-risk Direction		
AO	Leg	Similarity	M	SD	n	M	SD	n
1	back	greater	9.68	11.18	22	15.02	13.16	22
		lesser	15.34	8.65	24	16.54	8.50	21
		Total	12.63	10.23	46	15.76	11.03	43
	forw	greater	11.65	6.12	22	15.41	8.36	24
		lesser	16.75	9.10	24	15.56	7.01	23
		Total	14.31	8.15	46	15.48	7.65	47
	Total	greater	10.66	8.96	44	15.22	10.79	46
		lesser	16.05	8.81	48	16.03	7.68	44
		Total	13.47	9.24	92	15.62	9.36	90
2	back	greater	7.23	9.97	22	12.00	10.42	22
		lesser	14.08	8.08	24	14.93	7.66	21
		Total	10.80	9.58	46	13.43	9.19	43
	forw	greater	12.22	8.14	22	14.11	7.50	24
		lesser	17.58	10.86	24	15.70	8.56	23
		Total	15.01	9.92	46	14.89	7.99	47
	Total	greater	9.72	9.34	44	13.10	8.98	46
		lesser	15.83	9.63	48	15.33	8.06	44
		Total	12.91	9.93	92	14.19	8.57	90
3	back	greater	8.13	11.15	22	13.36	11.69	22
		lesser	13.91	8.79	24	14.20	7.28	21
		Total	11.15	10.30	46	13.77	9.68	43
	forw	greater	11.94	8.47	22	13.87	8.28	24
		lesser	17.53	10.70	24	15.20	9.80	23
		Total	14.86	10.00	46	14.52	8.98	47
	Total	greater	10.04	9.97	44	13.63	9.95	46
		lesser	15.72	9.86	48	14.72	8.61	44
		Total	13.00	10.26	92	14.16	9.28	90
Total	back	greater	8.35	11.13	22	13.46	12.66	22
		lesser	14.44	8.42	24	15.22	8.27	21
		Total	11.40	10.14	46	14.34	10.73	43
	forw	greater	11.94	8.57	22	14.46	7.97	24
		lesser	17.29	10.11	24	15.49	8.41	23
		Total	14.61	9.60	46	14.97	8.18	47
Total	Total	greater	10.14	10.07	44	13.96	10.46	46
		lesser	15.87	9.38	48	15.35	8.31	44

Note: *AO* and *Leg* represent the Assessment Order and Leg Position factors, respectively. *Total* signals the collapsing of levels for the respective factor. The leg level identifies the leg position that was used during training.

TABLE 8. Summary of the Similarity x Leg Position x Assessment Order Factorial ANOVAs for Experiment 1, Demonstration data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Trained At-risk Direction						
A (Similarity)	2256.516	1	2256.52	9.442	0.003**	0.097
B (Leg Position)	712.634	1	712.634	2.982	0.088	0.033
AB	9.547	1	9.547	0.04	0.842	0
Error	21030.867	88	238.987			
C (Assessment Order)	17.626	2	8.813	0.697	0.499	0.008
AC	5.955	2	2.978	0.235	0.79	0.003
BC	83.459	2	41.73	3.3	0.039*	0.036
ABC	5.155	2	2.578	0.204	0.816	0.002
Error	2225.598	176	12.645			
Untrained At-risk Direction						
A (Similarity)	130.648	1	130.648	0.608	0.438	0.007
B (Leg Position)	27.001	1	27.001	0.126	0.724	0.001
AB	9.034	1	9.034	0.042	0.838	0
Error	18473.688	86	214.81			
C (Assessment Order)	127.797	2	63.898	3.351	0.037*	0.038
AC	26.096	2	13.048	0.684	0.506	0.008
BC	34.137	2	17.068	0.895	0.41	0.01
ABC	12.742	2	6.371	0.334	0.716	0.004
Error	3279.604	172	19.067			

Assessment Order Post Hoc Test: LSD

Source	M Difference	*p*
Untrained At-risk Direction		
1 vs. 2	1.45	0.028*
1 vs. 3	1.47	0.072
2 vs. 3	0.03	0.953

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007). In the post hoc test, the numbers 1, 2, and 3 refer levels within Assessment Order.
*p< 0.05; **p< 0.01

TABLE 9. Number of participants per group, and group mean and SD of number degrees the leg was extended beyond safe limits for the Assessment Order x Type of Training (Groups 1, 2, and ND training) Factorial ANOVAs for Experiment 1, Demonstration data.

		Trained At-risk Direction			Untrained At-risk Direction		
AO	Type of Training	M	SD	n	M	SD	n
1	greater	11.65	6.12	22	15.41	8.36	24
	lesser	16.75	9.10	24	15.56	7.01	23
	ND Training	23.82	15.83	22	16.01	12.55	24
	Total	17.39	11.97	68	15.66	9.50	71
2	greater	12.22	8.14	22	14.11	7.50	24
	lesser	17.58	10.86	24	15.70	8.56	23
	ND Training	22.32	16.85	22	15.90	15.35	24
	Total	17.38	12.92	68	15.23	10.94	71
3	greater	11.94	8.47	22	13.87	8.28	24
	lesser	17.53	10.70	24	15.20	9.80	23
	ND Training	21.63	17.35	22	15.29	13.84	24
	Total	17.05	13.10	68	14.78	10.77	71
Total	greater	11.94	8.57	22	14.46	7.97	24
	lesser	17.29	10.11	24	15.49	8.41	23
	ND Training	22.59	17.62	22	15.74	13.77	24

Note: *AO* represents the Assessment Order factor. *Total* signals the collapsing of levels for the respective factor. Group 1 and 2 were trained with at-risk stimuli in the forward position.

TABLE 10. Summary of the Type of Training (Groups 1, 2, and ND training) x Assessment Order Factorial ANOVAs for Experiment 1, Demonstration data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Forward Trained At-risk Direction						
A (Type of Training)	3747.124	2	1873.562	4.522	0.015*	0.122
Error	26933.59	65	414.363			
B (Assessment Order)	5.7	2	2.85	0.24	0.787	0.004
AB	63.963	4	15.991	1.346	0.256	0.04
Error	1544.626	130	11.882			
Backward Untrained At-risk Direction						
A (Type of Training)	65.731	2	32.865	0.107	0.898	0.003
Error	20845.87	68	306.557			
B (Assessment Order)	27.127	2	13.564	0.977	0.379	0.014
AB	15.716	4	3.929	0.283	0.889	0.008
Error	1888.216	136	13.884			

Type of Training Post Hoc Test: Tukey's HSD

Source	M Difference	*p*
Forward Trained At-risk Direction		
ND Training vs. Greater Similarity	10.66	0.01*
ND Training vs. Lesser Similarity	5.30	0.284

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).
*$p < 0.05$; **$p < 0.01$

TABLE 11. Number of participants per group, and group mean and SD of number degrees the leg was extended beyond safe limits for the Assessment Order x Type of Training (Groups 3, 4, and ND training) Factorial ANOVAs for Experiment 1, Demonstration data.

		Trained At-risk Direction			Untrained At-risk Direction		
AO	Type of Training	M	SD	n	M	SD	n
1	greater	9.68	11.18	22	15.02	13.16	22
	lesser	15.34	8.65	24	16.54	8.50	21
	ND Training	16.01	12.55	24	23.82	15.83	22
	Total	13.79	11.11	70	18.49	13.30	65
2	greater	7.23	9.97	22	12.00	10.42	22
	lesser	14.08	8.08	24	14.93	7.66	21
	ND Training	15.90	15.35	24	22.32	16.85	22
	Total	12.55	12.01	70	16.44	12.91	65
3	greater	8.13	11.15	22	13.36	11.69	22
	lesser	13.91	8.79	24	14.20	7.28	21
	ND Training	15.29	13.84	24	21.63	17.35	22
	Total	12.57	11.70	70	16.43	13.20	65
Total	greater	8.35	11.13	22	13.46	12.66	22
	lesser	14.44	8.42	24	15.22	8.27	21
	ND Training	15.74	13.77	24	22.59	17.62	22

Note: *AO* represents the Assessment Order factor. *Total* signals the collapsing of levels for the respective factor. Group 3 and 4 were trained with at-risk stimuli in the backward position.

TABLE 12. Summary of the Type of Training (Groups 3, 4, and ND training) x Assessment Order Factorial ANOVAs for Experiment 1, Demonstration data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Backward Trained At-risk Direction						
A (Type of Training)	2118.345	2	1059.173	2.95	0.059	0.081
Error	24055.35	67	359.035			
B (Assessment Order)	73.336	2	36.668	2.863	0.061	0.041
AB	33.293	4	8.323	0.65	0.628	0.019
Error	1716.386	134	12.809			
Forward Untrained At-risk Direction						
A (Type of Training)	3087.158	2	1543.579	3.469	0.037*	0.101
Error	27585.89	62	444.934			
B (Assessment Order)	182.324	2	91.162	4.658	0.011*	0.07
AB	33.273	4	8.318	0.425	0.79	0.014
Error	2426.801	124	19.571			

Type of Training Post Hoc Test: Tukey's HSD

Source	M Difference	*p*
Forward Untrained At-risk Direction		
ND Training vs. Greater Similarity	9.13	0.041*
ND Training vs. Lesser Similarity	7.37	0.125

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).
*p < 0.05; **p < 0.01

TABLE 13. Number of participants per group, and group mean and SD of number of correctly classified trials for the Assessment Order x Absolute Frequency x Relative Frequency Factorial ANOVAs for Experiment 2, Generalization data.

			Trained At-risk Class			Untrained At-risk Class			Safe Class		
AO	Abslt	Rltv	M	SD	n	M	SD	n	M	SD	n
1	10	1:4	25.33	1.31	24	17.04	3.92	24	18.71	3.16	24
		4:1	23.58	3.15	24	15.88	4.79	24	20.46	1.87	24
		Total	24.46	2.54	48	16.46	4.37	48	19.58	2.71	48
	40	1:4	25.67	1.24	24	17.75	3.80	24	19.88	2.47	24
		4:1	23.62	2.20	24	16.92	3.30	24	20.38	1.25	24
		Total	24.65	2.05	48	17.33	3.55	48	20.13	1.95	48
	Total	1:4	25.50	1.27	48	17.40	3.84	48	19.29	2.87	48
		4:1	23.60	2.69	48	16.40	4.10	48	20.42	1.57	48
		Total	24.55	2.30	96	16.90	3.98	96	19.85	2.37	96
2	10	1:4	24.67	1.88	24	15.42	4.51	24	18.79	3.26	24
		4:1	22.83	3.63	24	14.46	4.22	24	20.46	1.10	24
		Total	23.75	3.01	48	14.94	4.35	48	19.63	2.55	48
	40	1:4	25.17	1.76	24	16.67	4.07	24	19.17	2.58	24
		4:1	23.29	2.55	24	14.42	3.80	24	20.42	1.72	24
		Total	24.23	2.36	48	15.54	4.06	48	19.79	2.26	48
	Total	1:4	24.92	1.82	48	16.04	4.30	48	18.98	2.91	48
		4:1	23.06	3.11	48	14.44	3.97	48	20.44	1.43	48
		Total	23.99	2.70	96	15.24	4.20	96	19.71	2.40	96
3	10	1:4	24.67	2.39	24	15.96	4.49	24	19.21	3.08	24
		4:1	22.92	4.18	24	13.92	4.72	24	20.50	1.32	24
		Total	23.79	3.48	48	14.94	4.67	48	19.85	2.43	48
	40	1:4	25.33	2.26	24	16.67	4.50	24	19.63	2.46	24
		4:1	22.92	2.59	24	14.54	4.14	24	20.38	1.61	24
		Total	24.13	2.70	48	15.60	4.41	48	20.00	2.09	48
	Total	1:4	25.00	2.33	48	16.31	4.46	48	19.42	2.77	48
		4:1	22.92	3.44	48	14.23	4.40	48	20.44	1.46	48
		Total	23.96	3.10	96	15.27	4.53	96	19.93	2.26	96
Total	10	1:4	24.89	1.91	24	16.14	4.31	24	18.90	3.13	24
		4:1	23.11	3.64	24	14.75	4.59	24	20.47	1.44	24
		Total	24.00	3.03	48	15.44	4.49	48	19.69	2.55	48
	40	1:4	25.39	1.79	24	17.03	4.11	24	19.56	2.49	24
		4:1	23.28	2.43	24	15.29	3.88	24	20.39	1.52	24
		Total	24.33	2.38	48	16.16	4.08	48	19.97	2.10	48
Total	Total	1:4	25.14	1.86	48	16.58	4.22	48	19.23	2.84	48
		4:1	23.19	3.09	48	15.02	4.25	48	20.43	1.48	48

Note: *AO, Abslt,* and *Rltv* represent the Assessment Order, Absolute Frequency, and Relative Frequency factors, respectively. *Total* signals the collapsing of levels for the respective factor.

TABLE 14. Summary of the Similarity x Absolute Frequency x Relative Frequency Factorial ANOVAs for Experiment 2, Generalization data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Trained At-risk Class						
A (Relative Frequency)	272.222	1	272.222	15.893	< 0.001**	0.147
B (Absolute Frequency)	8	1	8	0.467	0.496	0.005
AB	2	1	2	0.117	0.733	0.001
Error	1575.778	92	17.128			
C (Assessment Order)	21.437	2	10.719	7.964	< 0.001**	0.08
AC	0.715	2	0.358	0.266	0.767	0.003
BC	1.021	2	0.51	0.379	0.685	0.004
ABC	1.187	2	0.594	0.441	0.644	0.005
Error	247.639	184	1.346			
Untrained At-risk Class						
A (Relative Frequency)	175.781	1	175.781	3.765	0.055	0.039
B (Absolute Frequency)	36.837	1	36.837	0.789	0.377	0.009
AB	2.17	1	2.17	0.046	0.83	0.001
Error	4294.931	92	46.684			
C (Assessment Order)	172.312	2	86.156	26.777	< 0.001**	0.225
AC	14.146	2	7.073	2.198	0.114	0.023
BC	0.965	2	0.483	0.15	0.861	0.002
ABC	8.549	2	4.274	1.328	0.267	0.014
Error	592.028	184	3.218			
Safe Class						
A (Relative Frequency)	103.92	1	103.92	8.044	0.006**	0.08
B (Absolute Frequency)	5.837	1	5.837	0.452	0.503	0.005
AB	9.753	1	9.753	0.755	0.387	0.008
Error	1188.486	92	12.918			
C (Assessment Order)	2.382	2	1.191	0.887	0.414	0.01
AC	2.507	2	1.253	0.934	0.395	0.01
BC	2.382	2	1.191	0.887	0.414	0.01
ABC	2.424	2	1.212	0.903	0.407	0.01
Error	246.972	184	1.342			

Assessment Order Post Hoc Test: LSD

Source	M Difference	*p*
Trained At-risk Class		
1 vs. 2	0.56	< 0.001**
1 vs. 3	0.59	0.005**
2 vs. 3	0.03	0.816
Untrained At-risk Class		
1 vs. 2	1.66	< 0.001**
1 vs. 3	1.63	< 0.001**
2 vs. 3	-0.03	0.858

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007). In the post hoc test, the numbers 1, 2, and 3 refer levels within Assessment Order.

*p< 0.05; **p< 0.01

TABLE 15. Number of participants per group, and group mean and SD of number of correctly classified trials for the Assessment Order x Group Factorial ANOVAs for Experiment 1, Groups 1 vs. Experiment 2, Group 3: Generalization data.

		Trained At-risk Class			Untrained At-risk Class			Safe Class		
AO	Group	M	SD	n	M	SD	n	M	SD	n
1	1	24.38	2.52	24	16.13	4.32	24	20.38	1.58	24
	3	23.62	2.20	24	16.92	3.30	24	20.38	1.25	24
	Total	24.00	2.37	48	16.52	3.82	48	20.38	1.41	48
2	1	23.38	3.16	24	14.25	4.87	24	20.38	1.72	24
	3	23.29	2.55	24	14.42	3.80	24	20.42	1.72	24
	Total	23.33	2.84	48	14.33	4.32	48	20.40	1.70	48
3	1	23.71	3.64	24	14.58	4.59	24	20.63	1.01	24
	3	22.92	2.59	24	14.54	4.14	24	20.38	1.61	24
	Total	23.31	3.15	48	14.56	4.32	48	20.50	1.34	48
Total	1	23.82	3.12	24	14.99	4.60	24	20.46	1.45	24
	3	23.28	2.43	24	15.29	3.88	24	20.39	1.52	24

Note: *AO* represents the Assessment Order factor. *Total* signals the collapsing of levels for the respective factor.

TABLE 16. Summary of the Group x Assessment Order Factorial ANOVAs for Experiment 1, Groups 1 vs. Experiment 2, Group 3: Generalization data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Trained At-risk Class						
A (Group)	10.563	1	10.563	0.504	0.481	0.011
Error	963.097	46	20.937			
B (Assessment Order)	14.681	2	7.34	5.134	0.008**	0.1
AB	3.792	2	1.896	1.326	0.271	0.028
Error	131.528	92	1.43			
Untrained At-risk Class						
A (Group)	3.361	1	3.361	0.07	0.793	0.002
Error	2217.861	46	48.214			
B (Assessment Order)	138.764	2	69.382	29.727	< 0.001**	0.393
AB	4.514	2	2.257	0.967	0.384	0.021
Error	214.722	92	2.334			
Safe Class						
A (Group)	0.174	1	0.174	0.031	0.86	0.001
Error	254.986	46	5.543			
B (Assessment Order)	0.431	2	0.215	0.348	0.707	0.008
AB	0.597	2	0.299	0.482	0.619	0.01
Error	56.972	92	0.619			

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).
*$p < 0.05$; **$p < 0.01$

TABLE 17. Number of participants per group, and group mean and SD of percentage of correctly classified trials for the Leg Angle x Type of Training (Groups 1, 3, and ND training) Factorial ANOVAs for Experiment 2, Generalization data.

	Trained At-risk Class			Untrained At-risk Class			Safe Class			
Type of Training	Leg Angle	M	SD	Leg Angle	M	SD	Leg Angle	M	SD	n
1:4	101-125	87.93	12.37	27-53	95.04	10.00	80-90	97.35	6.77	24
	127-153	99.70	0.80	55-79	28.63	22.21	91-100	88.47	18.68	24
	Total	94.03	6.62	Total	63.07	15.21	Total	93.12	11.85	24
4:1	101-125	72.97	17.75	27-53	90.77	10.84	80-90	99.12	2.44	24
	127-153	98.51	3.20	55-79	19.87	19.88	91-100	94.86	11.63	24
	Total	86.21	9.01	Total	56.64	14.38	Total	97.09	7.22	24
ND Training	101-125	34.51	29.24	27-53	66.87	37.63	80-90	97.35	9.84	24
	127-153	78.13	32.53	55-79	21.15	28.97	91-100	97.92	4.38	24
	Total	57.05	29.17	Total	44.80	30.29	Total	97.62	7.58	24

Note: *Total* signals the collapsing of levels for the respective factor. Group 1 and 3 were trained with at-risk stimuli in the forward position.

TABLE 18. Summary of the Type of Training (Groups 1, 3, and ND training) x Leg Angle Factorial ANOVAs for Experiment 2, Generalization data. Summary of the Type of Training (Groups 1, 3, and ND training) One-way ANOVA for Experiment 2, Generalization data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Forward Trained At-risk Class						
A (Type of Training)	490226.62	2	245113.312	30.084	< 0.001**	0.466
Error	562187.50	69	8147.645			
B (Leg Angle)	823271.89	26	31664.303	87.986	< 0.001**	0.56
AB	151740.97	52	2918.096	8.109	< 0.001**	0.19
Error	645625.00	1794	359.88			
Backward Untrained At-risk Class						
A (Type of Training)	110529.26	2	55264.632	4.833	0.011*	0.123
Error	788962.62	69	11434.241			
B (Leg Angle)	2224795.38	26	85569.053	173.283	< 0.001**	0.715
AB	95149.75	52	1829.803	3.705	< 0.001**	0.097
Error	885898.49	1794	493.812			
Safe Class						
A (Type of Training)	6090.54	2	3045.267	2.248	0.113	0.061
Error	93474.43	69	1354.702			
B (Leg Angle)	11478.54	20	573.927	4.149	< 0.001**	0.057
AB	15421.81	40	385.545	2.787	< 0.001**	0.075
Error	190877.43	1380	138.317			

Main Effect Type of Training Post Hoc Test: Tukey's HSD

Source	M Difference	*p*
Forward Trained At-risk Class		
1:4 vs. 4:1	7.82	0.27
ND Training vs. 1:4	36.91	< 0.001**
ND Training vs. 4:1	29.09	< 0.001**
Backward Untrained At-risk Class		
1:4 vs. 4:1	6.43	0.528
ND Training vs. 1:4	18.21	0.009**
ND Training vs. 4:1	11.78	0.124

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).

*$p < 0.05$; **$p < 0.01$

TABLE 18 CONTINUED. Summary of the Type of Training (Groups 1, 3, and ND training) x Leg Angle Factorial ANOVAs for Experiment 2, Generalization data. Summary of the Type of Training (Groups 1, 3, and ND training) One-way ANOVA for Experiment 2, Generalization data.

AB Interaction Post Hoc Test: One-way ANOVA/Tukey's HSD

Sources	SS	df	MS	*F*	*p*
Forward Trained At-risk Class:Type of Training x Leg Angles 101-125					
A (Type of Training)	36452.63	2	18226.31	41.342	< 0.001**
Error	30419.95	69	440.87		
Source			M Difference	*p*	
1:4 vs. 4:1			14.96	0.042*	
ND Training vs. 1:4			53.42	< 0.001**	
ND Training vs. 4:1			38.46	< 0.001**	

Sources	SS	df	MS	*F*	*p*
Forward Trained At-risk Class:Type of Training x Leg Angles 127-153					
A (Type of Training)	7061.01	2	3530.51	9.904	< 0.001**
Error	24595.56	69	356.46		
Source			M Difference	*p*	
1:4 vs. 4:1			1.19	0.974	
ND Training vs. 1:4			21.58	0.001**	
ND Training vs. 4:1			20.39	0.001**	

Sources	SS	df	MS	*F*	*p*
Backward Untrained At-risk Class:Type of Training x Leg Angles 27-53					
A (Type of Training)	11069.07	2	5534.53	10.163	< 0.001**
Error	37575.35	69	544.57		
Source			M Difference	*p*	
1:4 vs. 4:1			4.27	0.802	
ND Training vs. 1:4			28.18	< 0.001**	
ND Training vs. 4:1			23.91	0.002**	

Sources	SS	df	MS	*F*	*p*
Backward Untrained At-risk Class:Type of Training x Leg Angles 55-79					
A (Type of Training)	1074.59	2	537.29	0.933	0.398
Error	39739.21	69	575.93		

Sources	SS	df	MS	*F*	*p*
Safe Class:Type of Training x Leg Angles 80-90					
A (Type of Training)	50.00	2	25.00	0.505	0.606
Error	3417.89	69	49.54		

Sources	SS	df	MS	*F*	*p*
Safe Class:Type of Training x Leg Angles 91-100					
A (Type of Training)	1114.82	2	557.41	3.323	0.042*
Error	11572.69	69	167.72		
Source			M Difference	*p*	
1:4 vs. 4:1			-6.39	0.209	
ND Training vs. 1:4			-9.44	0.036*	
ND Training vs. 4:1			-3.06	0.694	

Note: *p< 0.05; **p< 0.01

TABLE 19. Number of participants per group, and group mean and SD of number of degrees the leg was extended beyond safe limits for the Assessment Order x Absolute Frequency x Relative Frequency Factorial ANOVAs for Experiment 2, Demonstration data.

			Trained At-risk Direction			Untrained At-risk Direction		
AO	Abslt	Rltv	M	SD	n	M	SD	n
1	10	1:4	9.75	6.83	23	11.51	8.52	24
		4:1	13.44	8.01	23	12.67	9.31	24
		Total	11.60	7.60	46	12.09	8.85	48
	40	1:4	8.23	6.59	24	8.41	5.17	24
		4:1	12.48	6.00	23	12.00	7.46	23
		Total	10.31	6.60	47	10.16	6.57	47
	Total	1:4	8.98	6.68	47	9.96	7.15	48
		4:1	12.96	7.02	46	12.34	8.37	47
		Total	10.95	7.10	93	11.14	7.83	95
2	10	1:4	7.88	7.84	23	10.96	10.14	24
		4:1	12.65	9.14	23	13.59	10.35	24
		Total	10.27	8.76	46	12.27	10.22	48
	40	1:4	6.80	7.31	24	7.63	6.70	24
		4:1	13.13	6.20	23	11.38	6.93	23
		Total	9.90	7.44	47	9.46	7.00	47
	Total	1:4	7.33	7.51	47	9.30	8.66	48
		4:1	12.89	7.73	46	12.50	8.82	47
		Total	10.08	8.08	93	10.88	8.84	95
3	10	1:4	8.17	7.59	23	10.33	9.58	24
		4:1	12.63	9.60	23	12.68	10.60	24
		Total	10.40	8.85	46	11.50	10.07	48
	40	1:4	5.82	6.41	24	8.49	8.20	24
		4:1	12.22	5.68	23	10.44	7.04	23
		Total	8.95	6.81	47	9.44	7.64	47
	Total	1:4	6.97	7.04	47	9.41	8.87	48
		4:1	12.42	7.80	46	11.58	9.01	47
		Total	9.67	7.88	93	10.48	8.96	95
Total	10	1:4	8.60	7.37	23	10.93	9.32	24
		4:1	12.91	9.13	23	12.98	9.97	24
		Total	10.75	8.59	46	11.96	9.67	48
	40	1:4	6.95	6.76	24	8.18	6.72	24
		4:1	12.61	5.82	23	11.27	7.41	23
		Total	9.78	6.90	47	9.72	7.26	47
Total	Total	1:4	7.78	7.09	47	9.56	8.21	48
		4:1	12.76	7.62	46	12.12	8.79	47

Note: *AO*, *Abslt*, and *Rltv* represent the Assessment Order, Absolute Frequency, and Relative Frequency factors, respectively. *Total* signals the collapsing of levels for the respective factor.

TABLE 20. Summary of the Relative Frequency x Absolute Frequency x Assessment Order Factorial ANOVAs for Experiment 2, Demonstration data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Trained At-risk Direction						
A (Relative Frequency)	1730.014	1	1730.01	11.695	0.001**	0.116
B (Absolute Frequency)	66.031	1	66.031	0.446	0.506	0.005
AB	31.758	1	31.758	0.215	0.644	0.002
Error	13165.341	89	147.925			
C (Assessment Order)	77.839	2	38.92	5.393	0.005**	0.057
AC	36.148	2	18.074	2.505	0.085	0.027
BC	15.883	2	7.942	1.101	0.335	0.012
ABC	5.772	2	2.886	0.4	0.671	0.004
Error	1284.519	178	7.216			
Untrained At-risk Direction						
A (Relative Frequency)	470.082	1	470.082	2.398	0.125	0.026
B (Absolute Frequency)	354.774	1	354.774	1.81	0.182	0.02
AB	19.548	1	19.548	0.1	0.753	0.001
Error	17836.833	91	196.009			
C (Assessment Order)	21.248	2	10.624	1.01	0.366	0.011
AC	14.146	2	7.073	0.673	0.512	0.007
BC	10.493	2	5.247	0.499	0.608	0.005
ABC	23.862	2	11.931	1.135	0.324	0.012
Error	1913.548	182	10.514			

Assessment Order Post Hoc Test: LSD

Source	M Difference	*p*
Trained At-risk Direction		
1 vs. 2	0.86	0.026*
1 vs. 3	1.27	0.005**
2 vs. 3	0.41	0.265

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007). In the post hoc test, the numbers 1, 2, and 3 refer levels within Assessment Order.
*p< 0.05; **p< 0.01

TABLE 21. Number of participants per group, and group mean and SD of number of degrees the leg was extended beyond safe limits for the Assessment Order x Group Factorial ANOVAs for Experiment 1, Groups 1 vs. Experiment 2, Group 3: Demonstration data.

		Trained At-risk Direction			Untrained At-risk Direction		
AO	Group	M	SD	n	M	SD	n
1	1	11.65	6.12	22	15.41	8.36	24
	3	12.48	6.00	23	12.00	7.46	23
	Total	12.07	6.01	45	13.74	8.03	47
2	1	12.22	8.14	22	14.11	7.50	24
	3	13.13	6.20	23	11.38	6.93	23
	Total	12.68	7.15	45	12.77	7.28	47
3	1	11.94	8.47	22	13.87	8.28	24
	3	12.22	5.68	23	10.44	7.04	23
	Total	12.08	7.10	45	12.19	7.81	47
Total	1	11.94	8.57	22	14.46	7.97	24
	3	12.61	5.82	23	11.27	7.41	23

Note: *AO* represents the Assessment Order factor. *Total* signals the collapsing of levels for the respective factor.

TABLE 22. Summary of the Group x Assessment Order Factorial ANOVAs for Experiment 1, Groups 1 vs. Experiment 2, Group 3: Demonstration data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Trained At-risk Direction						
A (Group)	15.217	1	15.217	0.126	0.724	0.003
Error	5195.924	43	120.835			
B (Assessment Order)	10.789	2	5.395	0.555	0.576	0.013
AB	2.728	2	1.364	0.14	0.869	0.003
Error	836.172	86	9.723			
Untrained At-risk Direction						
A (Group)	358.609	1	358.609	2.311	0.135	0.049
Error	6981.491	45	155.144			
B (Assessment Order)	57.43	2	28.715	2.977	0.056	0.062
AB	3.752	2	1.876	0.194	0.824	0.004
Error	868.1	90	9.646			

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).
$*p < 0.05$; $**p < 0.01$

TABLE 23. Number of participants per group, and group mean and SD of number of degrees the leg was extended beyond safe limits for the Assessment Order x Type of Training (Groups 1, 3, and ND training) Factorial ANOVAs for Experiment 2, Demonstration data.

		Trained At-risk Direction			Untrained At-risk Direction		
AO	Type of Training	M	SD	n	M	SD	n
1	1:4	8.23	6.59	24	8.41	5.17	24
	4:1	12.48	6.00	23	12.00	7.46	23
	ND Training	23.82	15.83	22	16.01	12.55	24
	Total	14.62	12.13	69	12.14	9.38	71
2	1:4	6.80	7.31	24	7.63	6.70	24
	4:1	13.13	6.20	23	11.38	6.93	23
	ND Training	22.32	16.85	22	15.90	15.35	24
	Total	13.86	12.62	69	11.64	10.91	71
3	1:4	5.82	6.41	24	8.49	8.20	24
	4:1	12.22	5.68	23	10.44	7.04	23
	ND Training	21.63	17.35	22	15.29	13.84	24
	Total	12.99	12.64	69	11.42	10.44	71
Total	1:4	6.95	6.76	24	8.18	6.72	24
	4:1	12.61	5.82	23	11.27	7.41	23
	ND Training	22.59	17.62	22	15.74	13.77	24

Note: *AO* represents the Assessment Order factor. *Total* signals the collapsing of levels for the respective factor. Group 1 and 3 were trained with at-risk stimuli in the forward position.

TABLE 24. Summary of the Type of Training (Groups 1, 3, and ND training) x Assessment Order Factorial ANOVAs for Experiment 2, Demonstration data.

Sources	SS	df	MS	*F*	*p*	partial η^2
Forward Trained At-risk Direction						
A (Type of Training)	8576.427	2	4288.213	12.846	< 0.001**	0.28
Error	22031.87	66	333.816			
B (Assessment Order)	90.592	2	45.296	5.744	0.004**	0.08
AB	44.554	4	11.139	1.413	0.233	0.041
Error	1040.886	132	7.885			
Backward Untrained At-risk Direction						
A (Type of Training)	2079.632	2	1039.816	3.809	0.027*	0.101
Error	18562.41	68	272.977			
B (Assessment Order)	19.874	2	9.937	0.934	0.395	0.014
AB	26.928	4	6.732	0.633	0.64	0.018
Error	1446.425	136	10.635			

Type of Training Post Hoc Test: Tukey's HSD

Source	M Difference	*p*
Forward Trained At-risk Direction		
ND Training vs. 1:4	15.64	< 0.001**
ND Training vs. 4:1	9.98	0.006**
Backward Untrained At-risk Direction		
ND Training vs. 1:4	7.56	0.021*
ND Training vs. 4:1	4.47	0.251

Note: The categorizations of effect size for η^2 are the following: small = .01, medium = .06, and large = .14 (Kittler, Menard, & Phillips, 2007).
*p < 0.05; **p < 0.01

Figure 1. Each data point shows the groups' percent of trials classified as safe for each leg angle. The data are shown as a generalization gradient. The S+ value is 90 degrees. The S- value is 120 degrees.

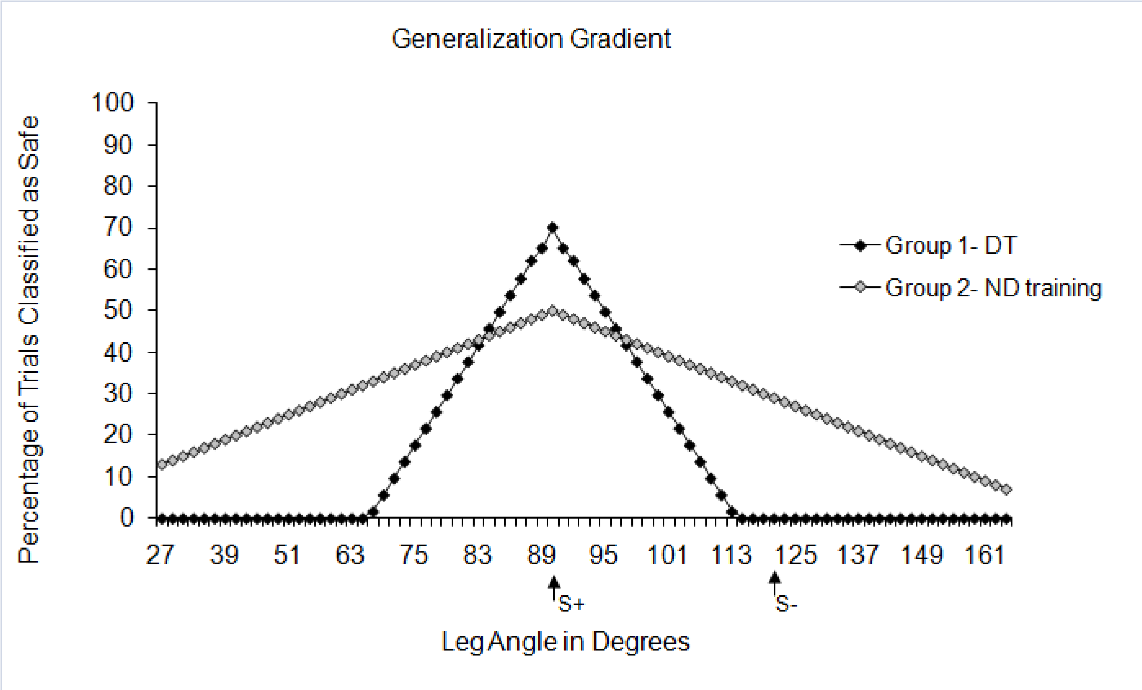


Figure 2. Each data point shows the groups' percent of trials classified as safe for each leg angle. The data are shown as a generalization gradient. The S+ value is 90 degrees. The S- value is 120 degrees.

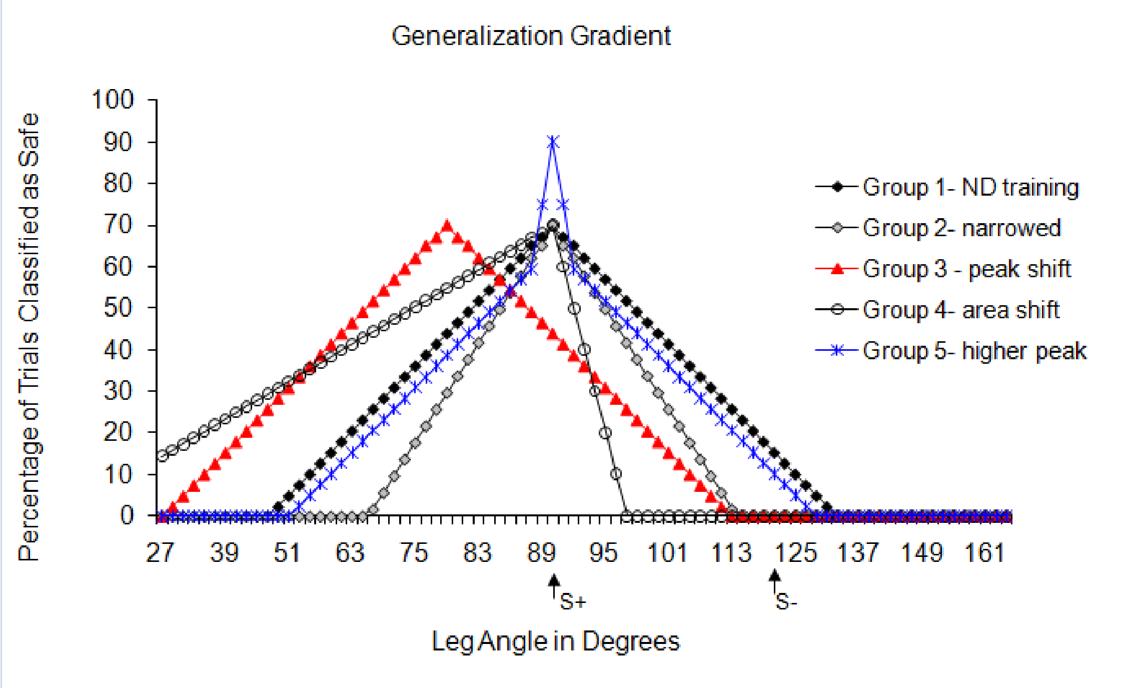


Figure 3. The data are the groups' percent of trials classified as safe for each leg angle. The data are shown as a generalization gradient across the three stimulus classes: safe, trained at risk, and untrained at risk. The arrows signify the value of the S- stimuli used during discrimination training for groups 1 and 2. The two vertical dashed lines represent the S+ stimuli used during training.

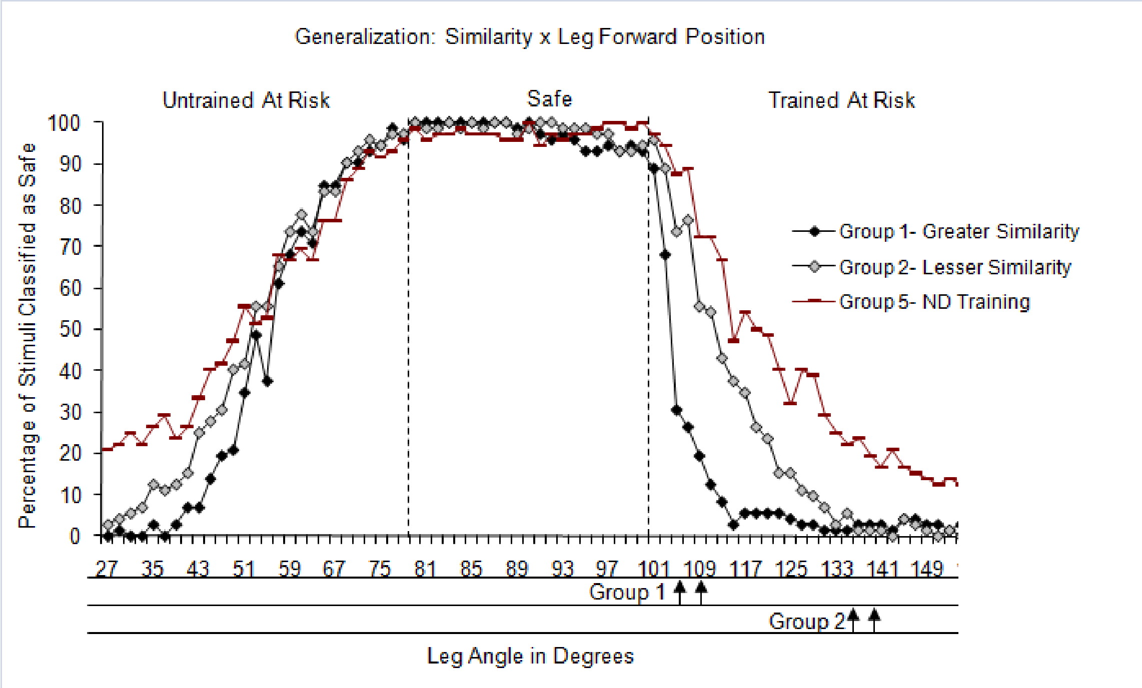


Figure 4. The data are the groups' percent of trials classified as safe for each leg angle. The data are shown as a generalization gradient across the three stimulus classes: safe, trained at risk, and untrained at risk. The arrows signify the value of the S- stimuli used during discrimination training for groups 3 and 4. The two vertical dashed lines represent the S+ stimuli used during training.

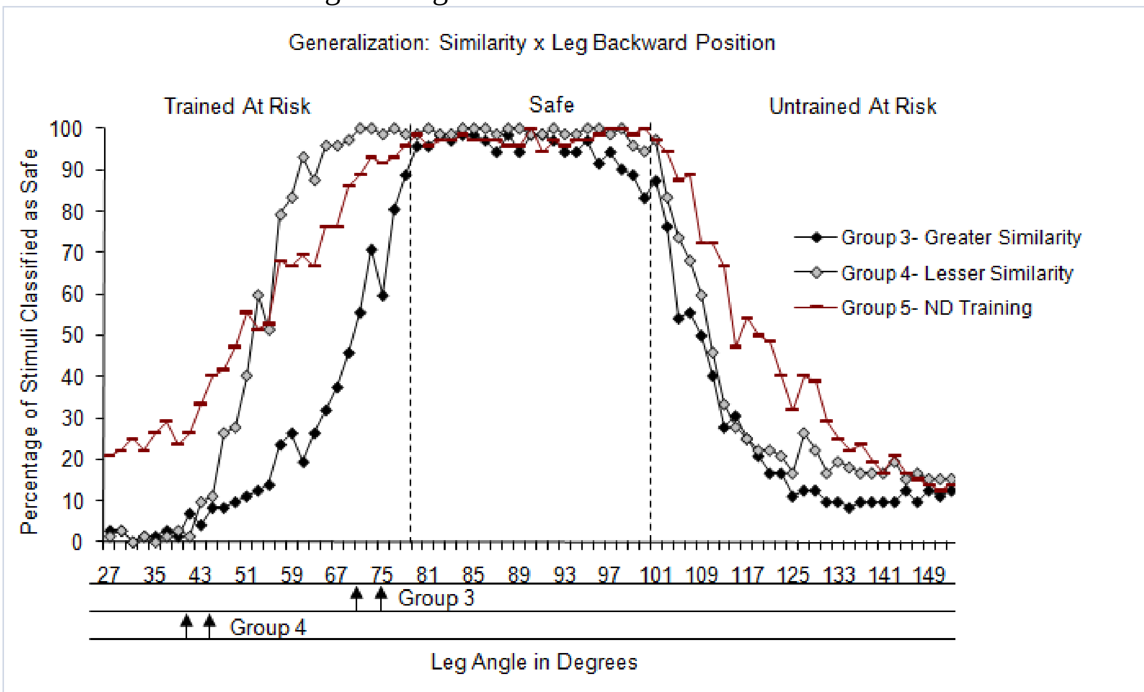


Figure 5. The bars represent the percentage of trained at-risk picture stimuli classified correctly according to the Similarity and Leg Position factors. The leg position label signifies the location of the leg within each at-risk stimulus during training. The error bars show the SEM.

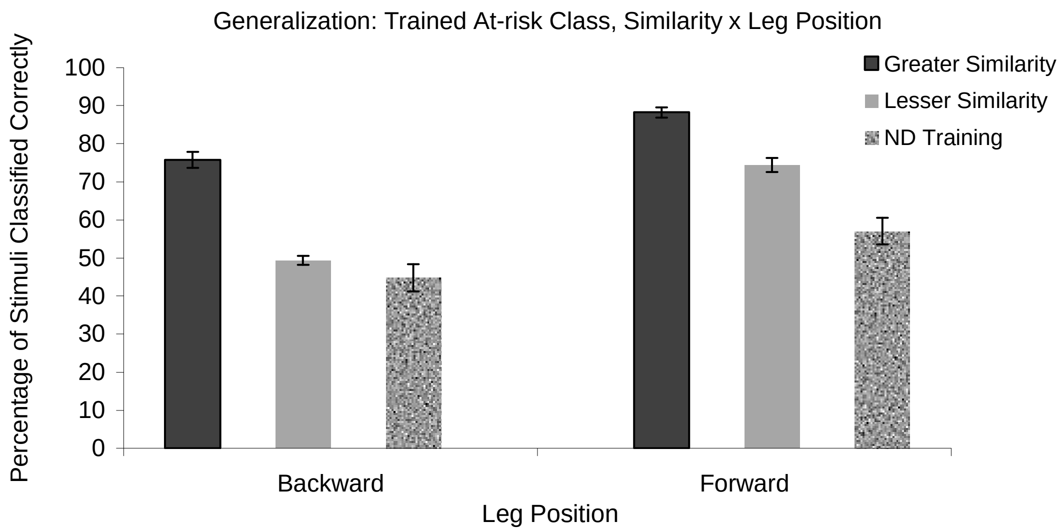


Figure 6. The bars represent the percentage of untrained at-risk picture stimuli classified correctly according to the Similarity and Leg Position factors. The leg position label signifies the location of the leg within each at-risk stimulus during training. The error bars show the SEM.

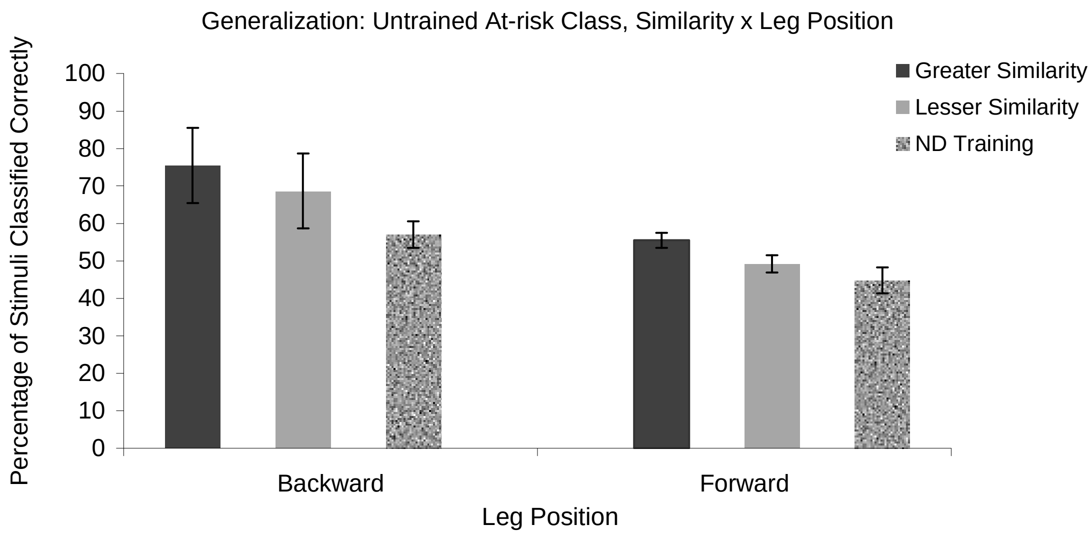


Figure 7. The bars represent the percentage of safe picture stimuli classified correctly according to the Similarity and Leg Position factors. The leg position label signifies the location of the leg within each at-risk stimulus during training. The error bars show the SEM.

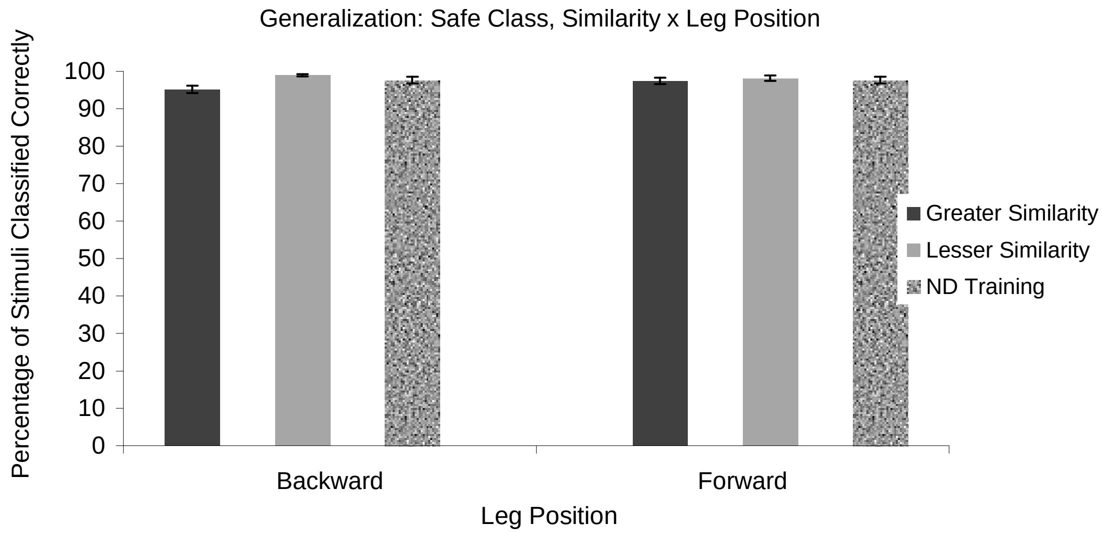


Figure 8. The data for the demonstration variable are the groups' mean number of degrees the leg was over-extended in the trained at-risk direction. The error bars are the SEM.

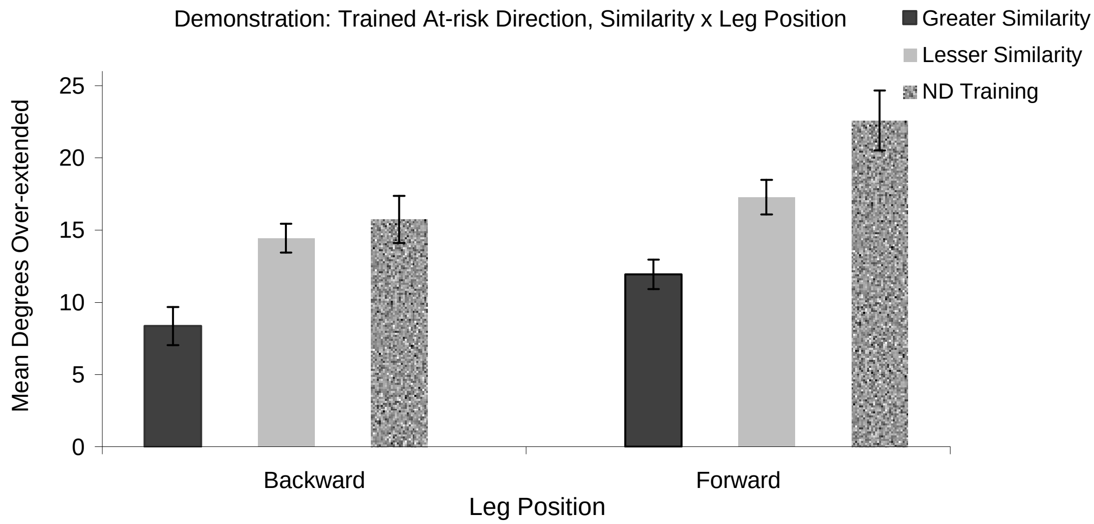


Figure 9. The data for the demonstration variable are the groups' mean number of degrees the leg was over-extended in the untrained at-risk direction. The error bars are the SEM.

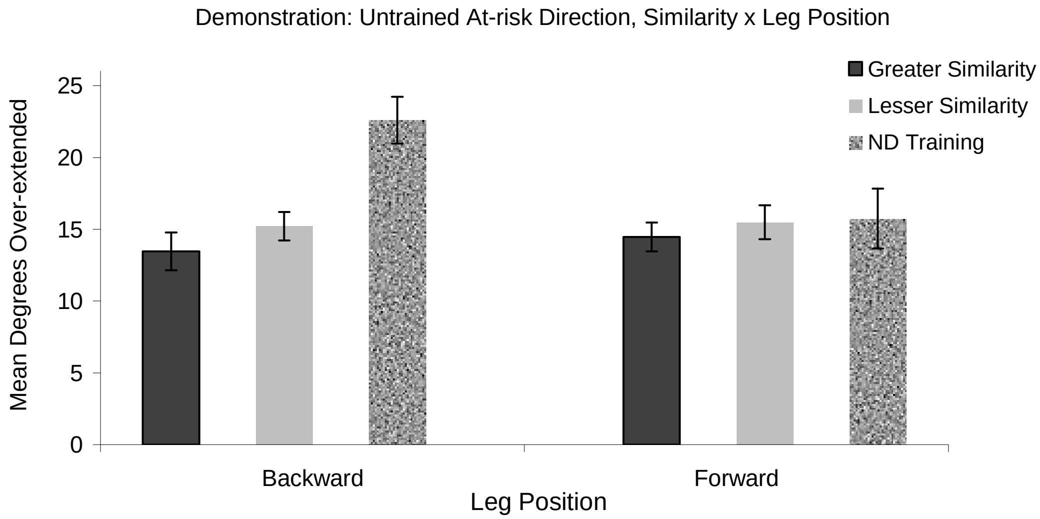


Figure 10. The data are the groups' percent of trials classified as safe for each leg angle. The data are shown as a generalization gradient across the three stimulus classes: safe, trained at risk, and untrained at risk. The arrows signify the value of the S- stimuli used during discrimination training for groups 2 and 4. The two vertical dashed lines represent the S+ stimuli used during training.

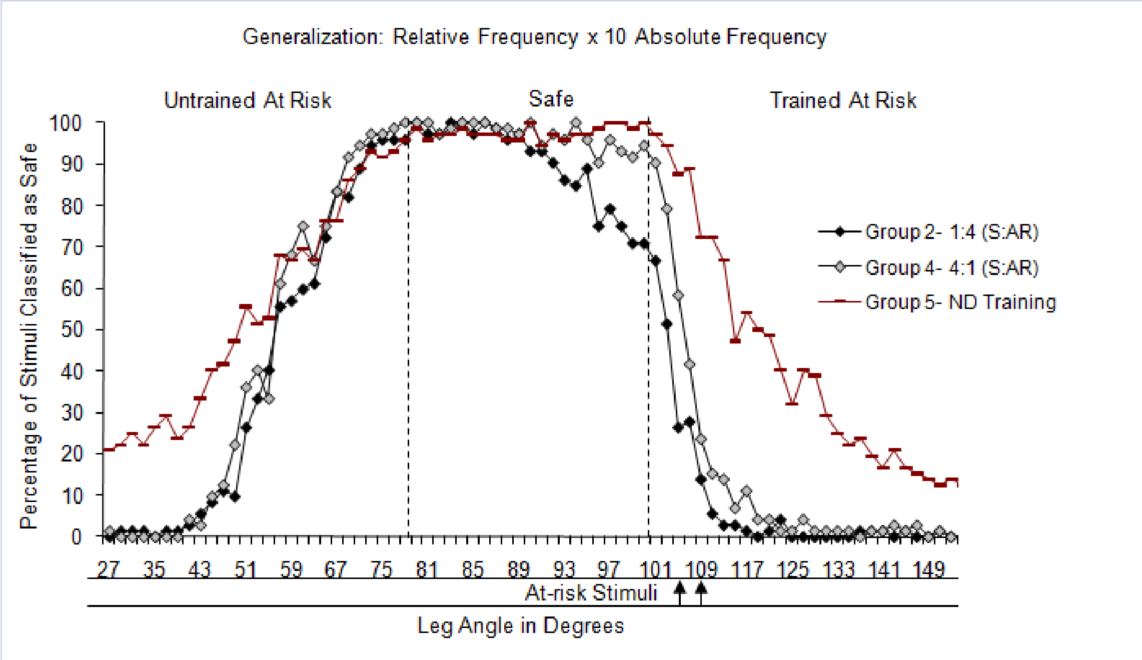


Figure 11. The data are the groups' percent of trials classified as safe for each leg angle. The data are shown as a generalization gradient across the three stimulus classes: safe, trained at risk, and untrained at risk. The arrows signify the value of the S- stimuli used during discrimination training for groups 1 and 3. The two vertical dashed lines represent the S+ stimuli used during training.

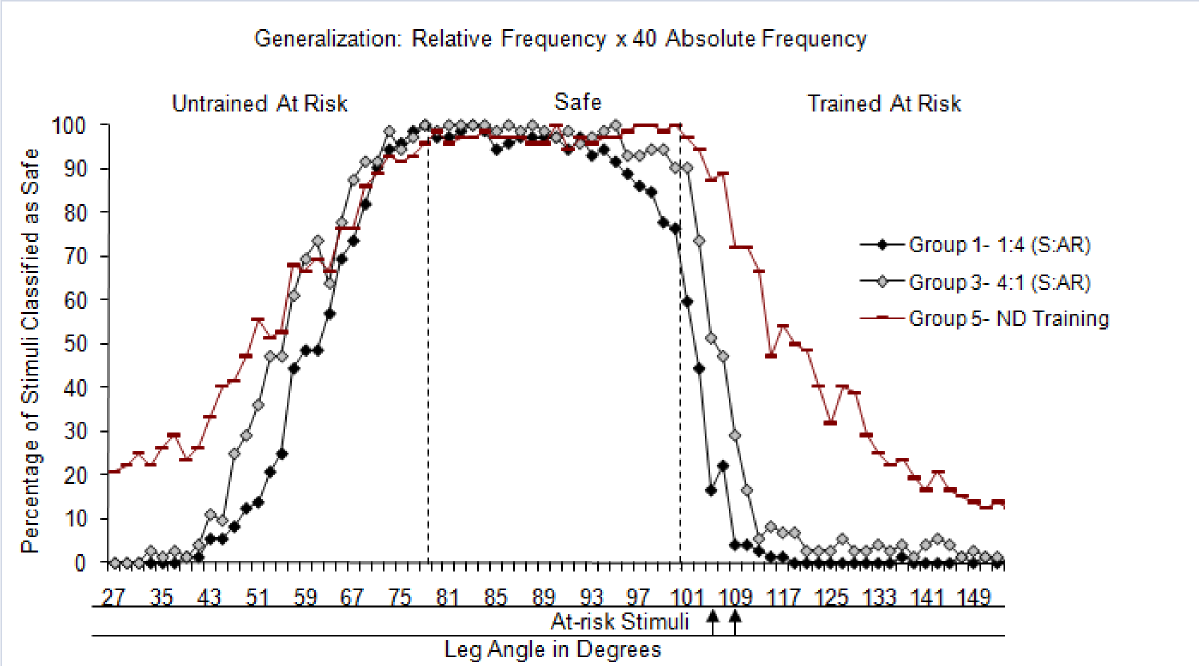


Figure 12. The data points represent the percentage of trained at-risk picture stimuli classified correctly according to the Relative Frequency and Absolute Frequency factors. The error bars show the SEM. The ND training group was trained with 32 safe trials.

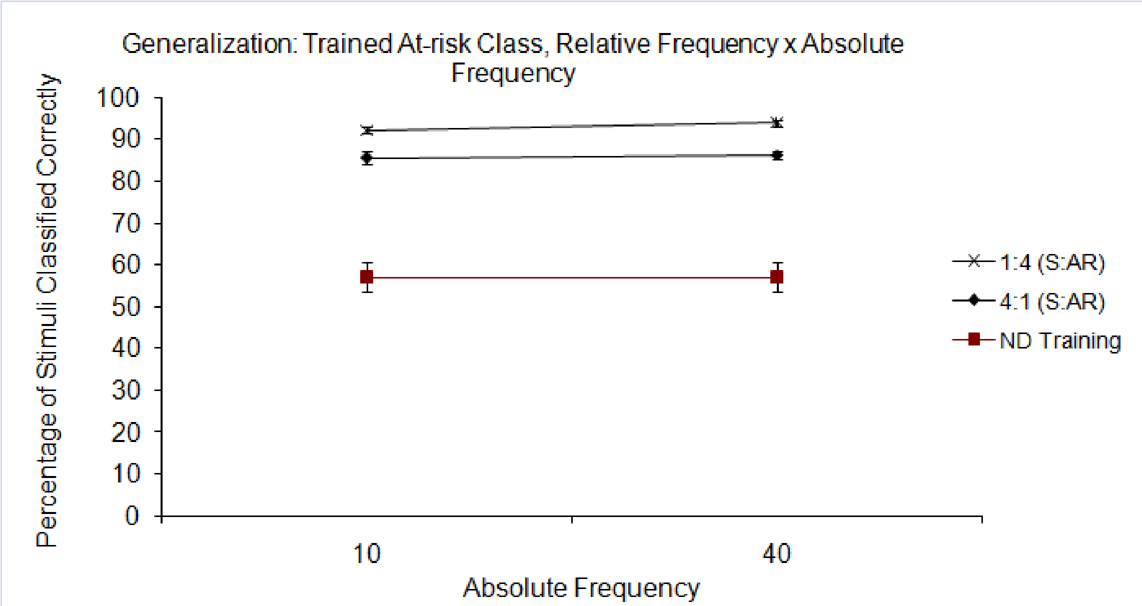


Figure 13. The data points represent the percentage of untrained at-risk picture stimuli classified correctly according to the Relative Frequency and Absolute Frequency factors. The error bars show the SEM. The ND training group was trained with 32 safe trials.

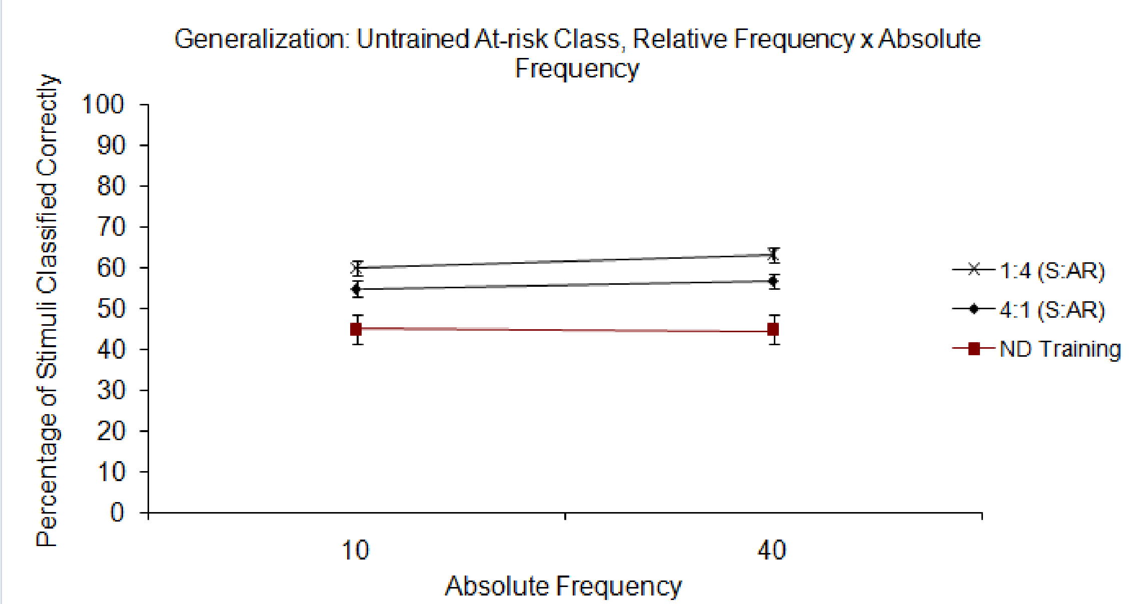


Figure 14. The data points represent the percentage of safe picture stimuli classified correctly according to the Relative Frequency and Absolute Frequency factors. The error bars show the SEM. The ND training group was trained with 32 safe trials.

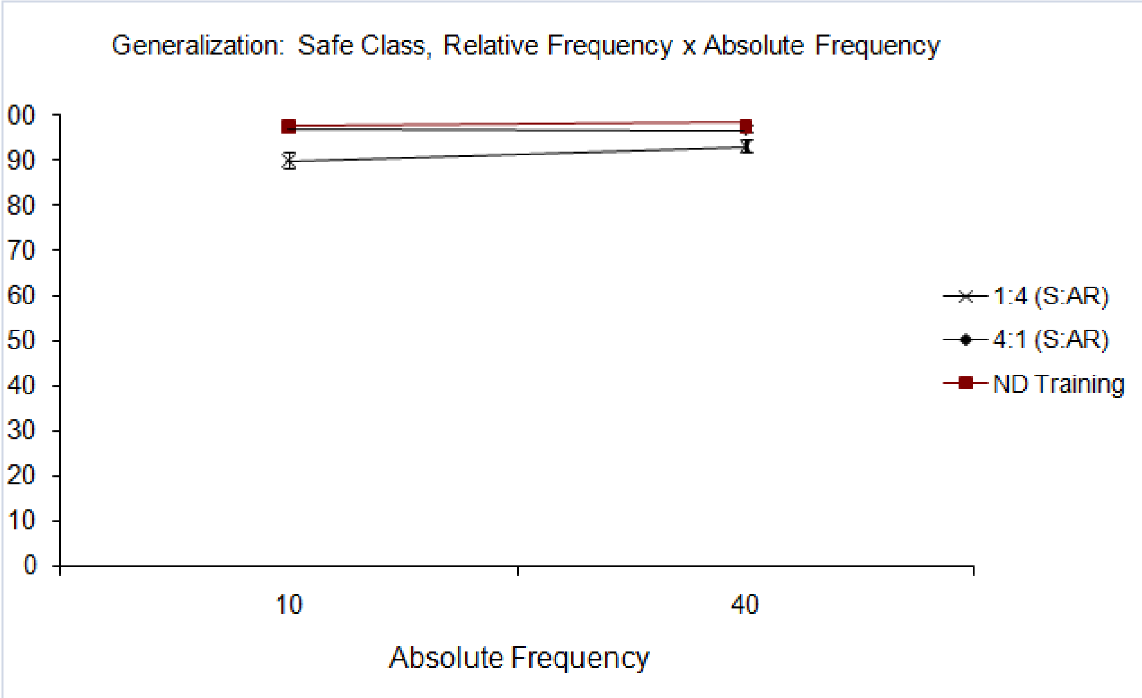


Figure 15. The data for the demonstration variable are the groups' mean number of degrees the leg was over-extended in the trained at-risk direction. The error bars are the SEM. The ND training group was trained with 32 safe trials.

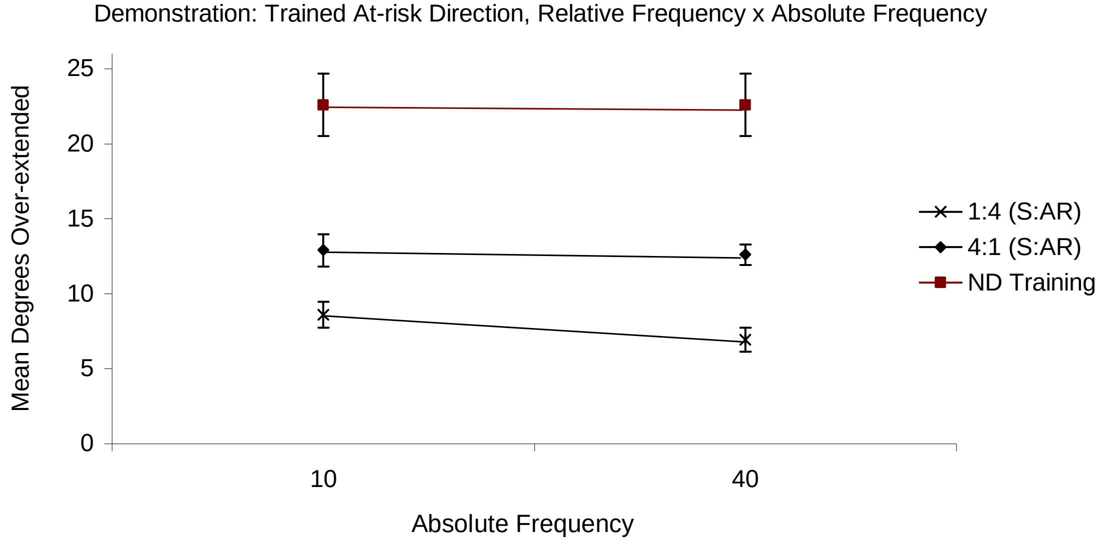
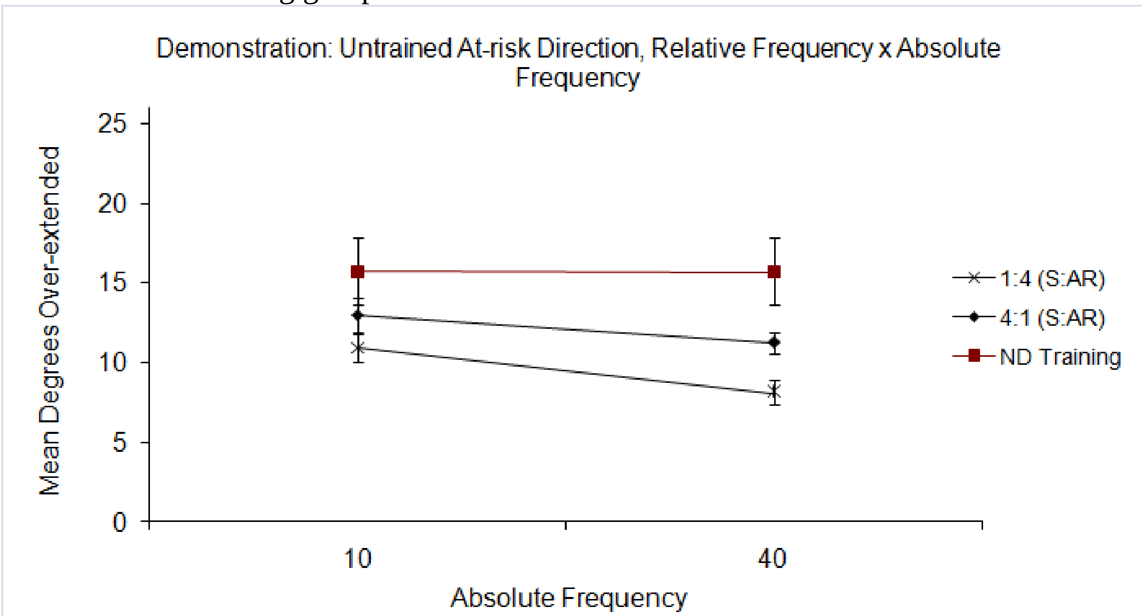


Figure 16. The data for the demonstration variable are the groups' mean number of degrees the leg was over-extended in the trained at-risk direction. The error bars are the SEM. The ND training group was trained with 32 safe trials.



Appendix A

"Are you ready to begin the session?" *If not, wait until the subject is ready.*

"Your participation may be terminated by the research investigator if your performance provides a clear indication of inattention to the tasks presented (e.g., talking on phone, using internet, engaging in activities other than those instructed during the experiment, and not following discrete instructions.) If your participation is terminated, you will not receive any money."

"Please turn off and do not use your cell phone during the session. If you use your cell phone during the session, then you will not receive money for your participation." *Wait until the subject turns off their cell phone and is sitting still before continuing with this script.*

"I am going to show you a slideshow. I want you to observe all of the slides in detail. Everything in the slideshow will run automatically. In other words, you don't need to touch the computer."

"Some slides will take longer than others. Again you must pay close attention to every slide and please do not touch the computer."

Point at clipboard to show them where it is. "There are data collection sheets on the clipboard next to the computer which you will be instructed to use during the slideshow. In order to receive the full amount of money, you must answer all questions on data collection sheets. Even if you are unsure about a question, you must give an answer to receive the full amount of money."

Point at clipboard. "Do not touch the data collection sheets until the slideshow prompts you to use them. Again, do not touch the data collection sheets until the slideshow prompts you to use them."

"Once I begin the session I am unable to answer any questions. "Would you like me to repeat anything that I've read so far?" *If yes, then reread the appropriate part.*

If they ask a question that is not covered in the script, then say "I'm sorry but I can't answer that question. Please do the best you can with the information that I've provided."

"Alright, I'm going to begin the slideshow."
Start PowerPoint slideshow.

Appendix B

Order 1: FORWARD>BACKWARD

Picture 1:

"Move your foot <u>(forward or backward)</u> until it meets the yard stick (yard stick marks a vertical position)."

-Take PICTURE 1 of vertical leg.

"During this picture taking process I'm going to ask you to move your leg. If you are moving your leg and you feel that the chair is in the way, then please move your leg around the piece of the chair. Again, if you are moving your leg and you feel that the chair is in the way, then please move your leg around the piece of the chair."

Picture 2:

"Please listen to my instructions. My instructions will tell you to move your lower-leg. Please do not move until I POINT at you signaling that it is okay for you to move your lower leg." *[Exclude these two sentences for sets B and C].*

READ SLOWLY

"You will now use the leg safety information that you learned during the training slideshow. I'm going to ask you to **show me the range** of **safe forward** and **safe backward** positions." *[Exclude these two sentences for sets B and C].*

"Based on the training, show me the **FARTHEST** that you can place your right leg **FORWARD**, while keeping it SAFE. Please look down at your leg when you move it."

"Move your right leg now." *POINT at the subject.* "Hold this position while I take a picture."

-Take PICTURE 2.

Picture 3:

"Move your foot <u>(forward or backward)</u> until it meets the yard stick (yard stick marks a vertical position)."

-Make sure marker board has arrow pointing at floor.

-Move the subject's leg so it is vertical again. Take PICTURE 3.

Picture 4:

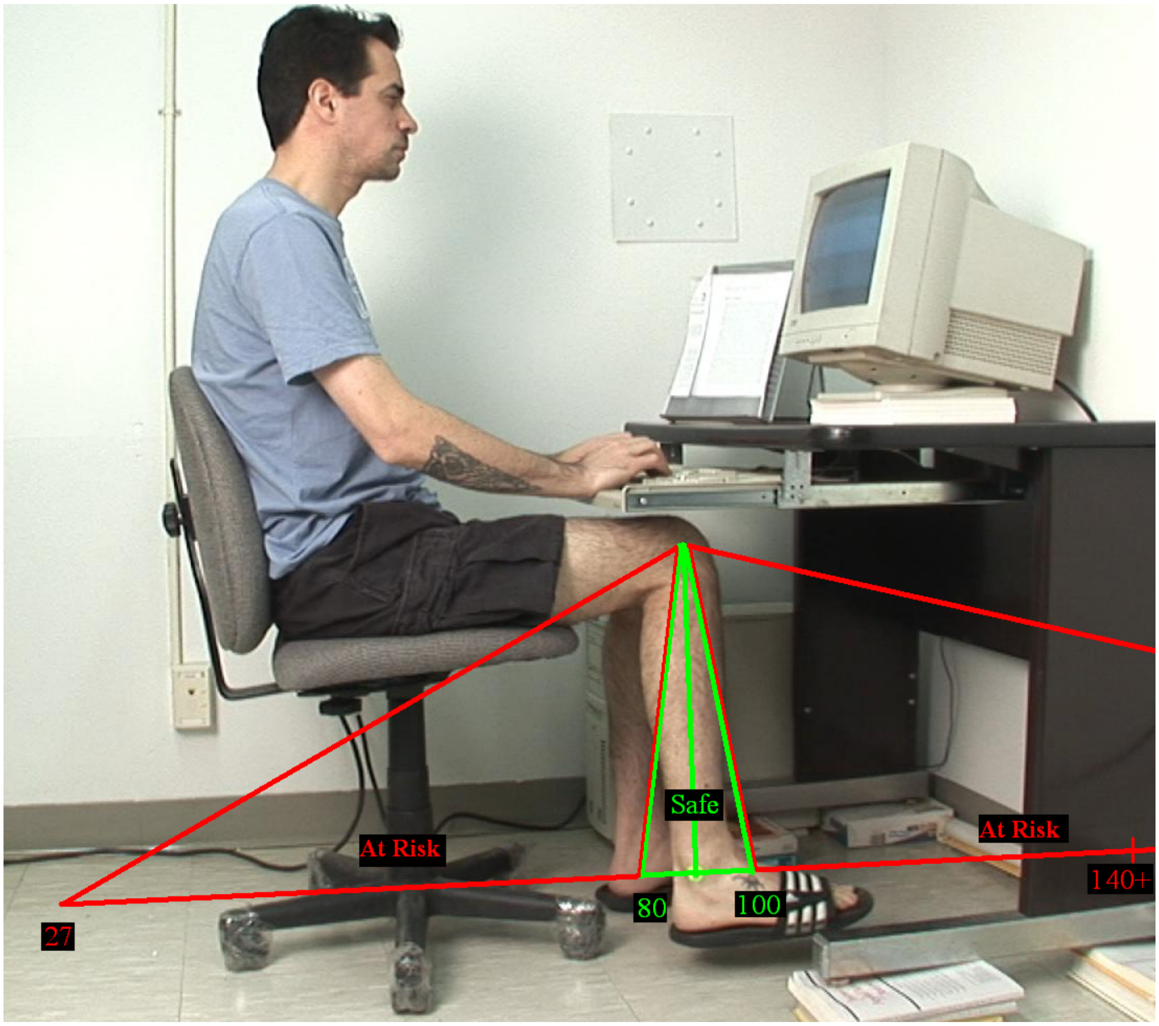
"Again, please listen to my instructions. My instructions will tell you to move your lower-leg. Please do not move until I POINT at you signaling that it is okay for you to move your lower leg." *[Exclude these two sentences for sets B and C].*

"Again, you will now use the leg safety information that you learned during the training slideshow." *[Exclude these two sentences for sets B and C].*

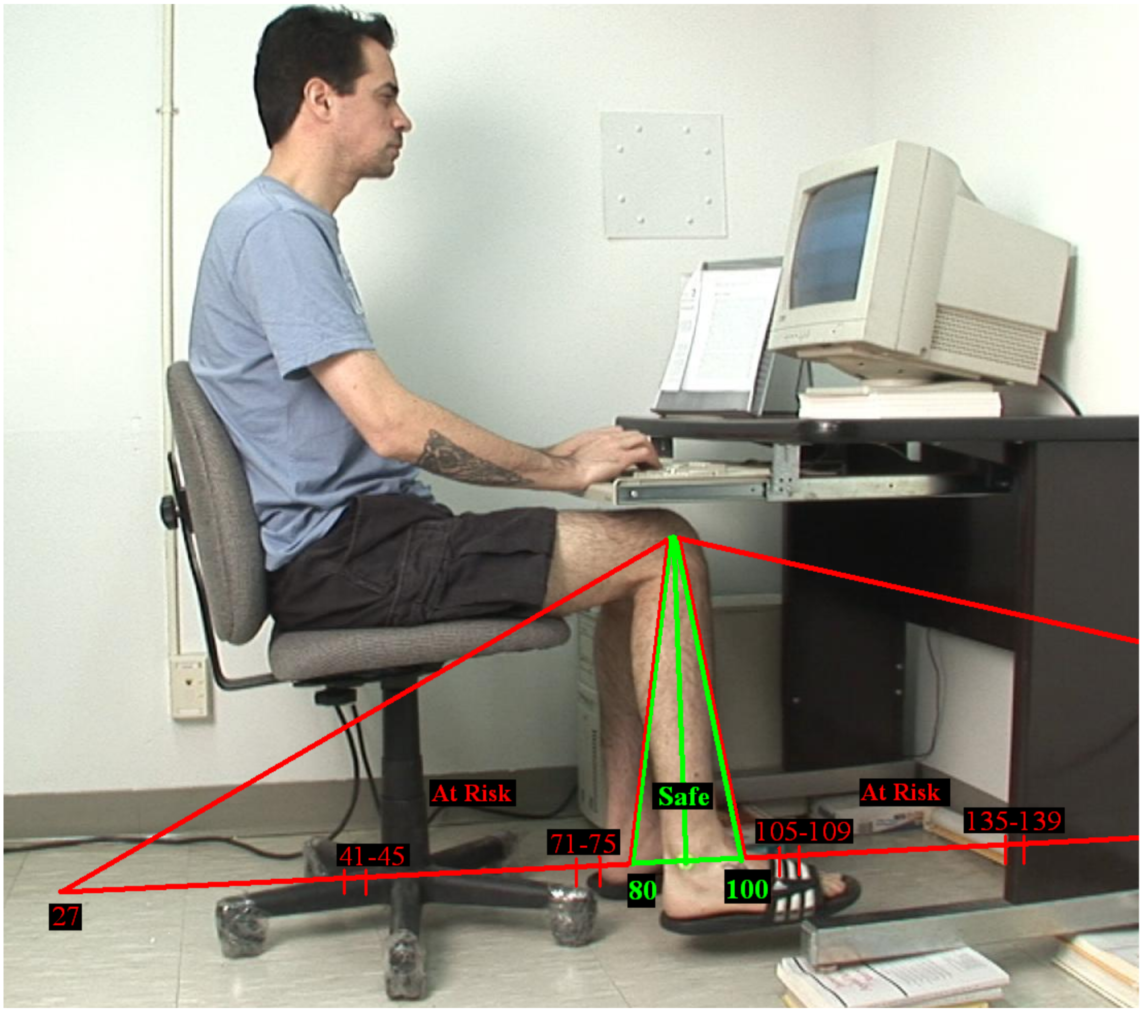
"Based on the training, show me the **FARTHEST** that you can place your right leg **BACKWARD**, while keeping it SAFE. Please look down at your leg when you move it."

"Move your right leg now." *POINT at the subject.* "Hold this position while I take a picture."
-Take PICTURE 4.

Appendix C. The picture shows the safe and at-risk classes of stimuli.



Appendix D. The picture shows the levels within the similarity and leg position factors.



Bibliography

Anger, K., Stupfel, J., Ammerman, T., Tamulinas, A., Bodner, T., & Rohlman, D. S. (2006). The suitability of computer-based training for workers with limited formal education: A case study from the US agricultural sector. *International Journal of Training and Development, 10*, 26-284.

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

Bornstein, M. H. (1974). Perceptual generalization: A note on the peak shift. *Psychological Bulletin, 81*, 804–808.

Burke, M. J., Sarpy, S. A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R. O., & Islam, G. (2006). Relative effectiveness of worker safety and health training methods. *American Journal of Public Health. 96*, 315-324.

Catania, A. C. (2007). *Learning (Interim 4th Edition)*. New York: Sloan Publishing.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). Applied behavior analysis *(2nd edition).* New Jersey: Pearson.

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

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

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

Hearst, E., & Koresko, M. B. (1968). Stimulus generalization and amount of prior training on variable interval reinforcement. *Journal of Comparative and Physiological Psychology, 66***,** 133-138.

Jenkins, H. M., & Harrison, R. H. (1960). Effect of discrimination training on auditory generalization. *Journal of Experimental Psychology, 59*, 246–253.

Kittler, J.E., Menard, W., & Phillips, K.A. (2007). Weight concerns in individuals with body dysmorphic disorder. *Eating Behaviors*, *8*, 115-120.

Komaki, J., Heinzmann, A. T., & Lawson, L. (1980). Effects of training and feedback: Component analysis of a behavioral safety program. *Journal of Applied Psychology*, *65*, 261-270.

Krause, T. R. (1997). *The behavior-based safety process (2nd edition).* New York: Van Nostrand Reinhold.

Lyons, J., & Thomas, D. R. (1967). Effects of interdimensional training on stimulus generalization: II. Within-subjects design. *Journal of Experimental Psychology, 75*, 572-574.

Malott, R. W. (1992). A theory of rule-governed behavior and organizational management. *Journal of Organizational Behavior*, *12*, 45−65.

McCann, K. B., & Sulzer-Azaroff, B. (1996). Cumulative trauma disorders: Behavioral injury prevention at work. *Journal of Applied Behavioral Science, 32*, 277-291.

McSween, T. M. (2003). *The values-based safety process: Improving your safety culture with behavior-based safety (2nd edition).* New Jersey: John Wiley & Sons, Inc.

National Institute for Occupational Safety and Health [NIOSH]. (2001). The national occupational research agenda for musculoskeletal disorders. *DHHS (NIOSH) Publication No. 2001-117*. Washington, DC: Author

National Research Council and the Institute of Medicine [NRC]. (2001). *Musculoskeletal disorders and the workplace: Low back and upper extremities.* Panel on musculoskeletal disorders and the workplace. Commission on behavioral and social sciences and education. Washington, DC: National Academy Press.

Perkins, C. C. Jr., Hershberger, W. A., & Weyant, R. G. (1959). Difficulty of a discrimination as a determiner of subsequent generalization along another dimension. *Journal of Experimental Psychology, 57*, 181-186.

Pierrel, R., & Sherman, J. G. (1962). Generalization and discrimination as a function of the S-D - S-Delta intensity difference. *Journal of the Experimental Analysis of Behavior*, *5*, 67–71.

Schwartz, B., Wasserman, E. A., & Robbins, S. J. (2002). *Psychology of learning and behavior (5th edition)*. New York: W. W. Norton & Company, Inc.

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

Taylor, M. A., Olvina, M., & Alvero, A. M. (2008). Assessment of safety discrimination and demonstration following training of one or two stimulus classes. Manuscript in preparation.

Thomas, D. R. (1962). The effects of drive and discrimination training on stimulus generalization. *Journal of Experimental Psychology*, *64*, 24-28.

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

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

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

Thomas, D. R., & Vogt, M. B. (1983). Stimulus generalization in humans viewed from a signal detection perspective. *New Zealand Journal of Psychology, 12*, 1-6.

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

Tomie, A., Davitt, G. A., & Thomas, D. R. (1975). Effects of stimulus similarity in discrimination training upon wavelength generalization in pigeons. *Journal of Comparative and Physiological Psychology, 101***,** 146-150.

U.S. Department of Labor: Occupational Safety & Health Administration. (1999). Ergonomics Program. Retrieved January 18, 2008 from http://www.osha.gov/pls/oshaweb/owadisp.show_document?p_table=FEDERAL_REGISTER&p_id=16305

Weinberg, S. G. (1973). Some determinants of the peak shift in stimulus generalization. *Psychological Reports, 32*, 47–58.

Wilcox, R. R. (2001). *Fundamentals of modern statistical methods: Substantially improving power and accuracy*. New York: Springer-Verlag.

Wirth, O., & Sigurdsson, S. O. (2008). When workplace safety depends on behavior change: Topics for behavioral safety research. *Journal of Safety Research, 39*, 589-598.