A Comparison of Stimulus Fading and Stimulus Shaping on Perceptual Category Learning.

Journal:	*Quarterly Journal of Experimental Psychology*
Manuscript ID	QJE-STD-23-011.R3
Manuscript Type:	Standard Article
Date Submitted by the Author:	11-Sep-2023
Complete List of Authors:	Markham, Victoria; University of South Wales Faculty of Life Sciences and Education, Austin, Jennifer; Georgia State University Roderique-Davies, Gareth; University of South Wales Faculty of Life Sciences and Education May, Richard; University of South Wales Faculty of Life Sciences and Education
Keywords:	stimulus fading, stimulus shaping, errorless learning, trial-and-error learning, perceptual category learning

DOI: 10.1177/17470218231210446

Running Head: Comparing Stimulus Fading and Shaping

A Comparison of Stimulus Fading and Stimulus Shaping on Perceptual Category Learning

Victoria Markham
School of Psychology
University of South Wales
Pontypridd
CF37 1DL
United Kingdom
Telephone: +44(0)1443 483794
Victoria.Markham@southwales.ac.uk

Jennifer Austin
School of Psychology
University of South Wales
Pontypridd
CF37 1DL
United Kingdom
Telephone: +44(0)1443 483724
Jenn.Austin@southwales.ac.uk

Gareth Roderique-Davies
School of Psychology
University of South Wales
Pontypridd
CF37 1DL
United Kingdom
Telephone: +44(0)1443 482679
Gareth.rdavies@southwales.ac.uk

Richard May
School of Psychology
University of South Wales
Pontypridd
CF37 1DL
United Kingdom
Telephone: +44(0)1443 483671
Richard.May@southwales.ac.uk

DOI: 10.1177/17470218231210446

Author Note

Victoria A. Markham https://orcid.org/0000-0001-9790-1689

Jennifer Austin https://orcid.org/0000-0001-8751-4941

Gareth Roderique-Davies https://orcid.org/0000-0002-6446-749X

Richard May https://orcid.org/0000-0001-9198-4030

Jennifer Austin is now at Georgia State University.

The authors would like to thank Victoria Adshead and Josh Molina for their support and feedback on the study. We have no known conflicts of interest to disclose.

This study's design and its analysis were not pre-registered. All data, analysis code, and research materials have been made publicly available at the Open Science Framework and can be accessed at https://osf.io/rjy52/

Correspondence concerning this article should be addressed to Victoria A. Markham, School of Psychology, University of South Wales, Pontypridd, CF37 1DL, United Kingdom.
E-mail: victoria.markham@southwales.ac.uk

DOI: 10.1177/17470218231210446

Abstract

Errorless learning strategies such as stimulus fading and stimulus shaping are commonly used to teach complex skills like categorisation and concept formation. Despite widespread use, very few studies have explored the comparative effectiveness of these procedures in well controlled analyses. The vast majority of existing studies have been undertaken with clinical populations and have involved small numbers of participants (e.g., Single-case designs). The present study sought to compare stimulus fading, stimulus shaping and trial-and-error learning in a perceptual categorisation task. In Experiment 1, we found robust benefits of stimulus shaping when compared to stimulus fading or trial-and-error learning on measures of initial acquisition of discrimination and one measure of stimulus generalisation. These findings were replicated in a second experiment in which the dimension of fading/shaping was changed from a modification of the comparison stimuli (S-) to a modification of the target stimulus (S+). We discuss the implication our findings for the selection of errorless learning strategies in clinical settings.

Key words: stimulus fading, stimulus shaping, errorless learning, trial-and-error learning, perceptual category learning

DOI: 10.1177/17470218231210446

Discriminating among stimuli may be fundamental to acquiring functional skills (Fisher et al., 2021; Jones & Eayrs, 1992). Language acquisition relies heavily on the development of discrimination skills, which are essential for tasks such as vocabulary acquisition and categorization (Brown & Bebko, 2012; Donahoe & Palmer, 2004; Mazur & Odum, 2023). Methods to establish discrimination can include both trial-and-error (i.e., *errorful learning;* Middleton & Schwartz, 2012), or '*errorless learning*' strategies (Lancioni & Smeets, 1986). Trial-and-error strategies typically involve reinforcing responding in the presence of particular (target) stimuli and withholding reinforcement for responses to non-target stimuli (Fields, 2018). In contrast, errorless or 'easy-to-difficult' strategies (Amitay et al., 2006) involve tactics in which a task is manipulated to minimize the number of errors that occur (Green, 2001). For example, a target stimulus may be presented in an exaggerated form relative to its (non-target) comparators.

Most children and adults can acquire discrimination skills via trial-and-error methods (Lepper, 2013); however, in cases where learners may not readily acquire a discrimination, errorless learning (EL) strategies may be more effective (Etzel & LeBlanc, 1979). Errorless learning strategies have been used extensively in applied settings, including with individuals with autism and intellectual disabilities (Cooper et al., 2020), those with acquired brain injury (Donaghey et al., 2010), and memory impairment (Evans et al., 2000; de Werd et al., 2013). Errorless learning tactics feature in some professional practice guidelines (e.g., Behavior Analyst Certification Board, 2017) as well as in manualised behavioural interventions (Frost & Bondy, 2002).

Stimulus prompt fading is a commonly used EL method (Markham et al., 2020; Noell et al., 2021) which involves modifying teaching stimuli to increase the rate of correct responses at the outset of teaching. Over successive trials, tasks are then incrementally modified towards a terminal form (Noell et al., 2021). There are two primary types of stimulus prompt

DOI: 10.1177/17470218231210446

fading: stimulus fading and stimulus shaping (Etzel & LeBlanc, 1979). Stimulus fading, sometimes called “cross-dimensional” fading, involves the adaptation of a cue which is independent of the feature discrimination that the task is designed to teach. Thus, in a shape discrimination task, the target stimulus (e.g., an ellipse) might be presented in green with the comparators (e.g., circles) presented in black, and the color incrementally changed to black across trials.

In stimulus shaping, sometimes called “transfer-along-a-continuum” or “critical feature” fading, a feature of the to be-discriminated stimuli that is central to the discrimination is adapted. Using the shape discrimination example, stimulus shaping might involve the presentation of the ellipse in an exaggerated form (maximally different to the circles) at the outset of training and then adapted to be incrementally more like the non-target comparators. In a seminal study of stimulus shaping by Schilmoeller et al (1979), children were taught to discriminate between two shapes (circle and triangle) by initially presenting them with an apple and a tree. Across trials, modifications were made to the apple and tree until they ultimately led to the final stage of discrimination, where a circle and triangle were presented instead. Please refer to Markham et al. (2020) for a more comprehensive review of the literature around different within-stimulus EL methods. For the purpose of the present study, we have labelled those procedures which manipulate a feature salient to the final discrimination as stimulus shaping, and procedures which manipulate a non-critical feature as stimulus fading (Noell et al., 2021).

In a series of experiments, Pashler and Mozer (2013) compared stimulus shaping to trial-and-error learning in a perceptual categorisation task with neurotypical adults. In one experiment, participants categorised characters according to four features. One feature determined category membership (horn length) and the rest functioned as distractor (irrelevant) features (eye diameter, head color, and the presence of a nose). Participants were

DOI: 10.1177/17470218231210446

assigned to trial-and-error teaching or to a stimulus shaping condition. In the trial-and-error condition, the difference in horn length between category positive and category negative characters was presented in its terminal form from the outset. In the stimulus shaping condition, the horn length of the positive category members was exaggerated at the beginning of training. The length was then incrementally reduced (shaped) across training blocks. Participants taught via stimulus shaping were more accurate in a final test phase compared to those that had received trial-and-error training.

Stimulus fading and stimulus shaping procedures have been shown to offer some benefits relative to trial-and-error learning in applied contexts (Markham et al., 2020; Repp & Karsh, 1992; VanLaarhoven et al., 2003). Both procedures seek to minimize incorrect selections at the outset of training by manipulating the salience of some aspect of the stimulus array. As a result, the target stimuli are predictive of reinforcement to a relatively greater extent than during trial-and-error learning (Booth & Keenan, 2018). Minimising incorrect responses can be considered beneficial to the extent that it reduces the likelihood of some of the negative effects of extinction, such as learner frustration and other emotional responses. However, some studies are more equivocal on the benefits of EL. Some research has indicated that that trial- and-error based tactics produce comparable (Dunn & Clare, 2007; Voigt-Radloff et al., 2017), or even superior learning outcomes compared to EL methods, particularly with respect to the generalization of new skills (Ownsworth et al., 2017).

Relatively few studies have directly compared the effectiveness of different EL tactics in applied settings (Markham et al., 2020). Most of the comparative studies have used single case experimental designs (SCED) which, although powerful for investigating effectiveness at the level of the individual, may be less helpful for determining efficacy for a given population (e.g., external validity). Of this literature, there is some evidence that stimulus shaping may be more effective, particularly for learners with intellectual disabilities. For

DOI: 10.1177/17470218231210446

example, Strand and Morris (1988), compared the effects of stimulus shaping, stimulus fading and trial and error teaching on size and intensity discriminations in young learners with intellectual disabilities. The authors found that both EL conditions were superior to trial-and-error. In addition, there was some evidence that stimulus shaping produced better outcomes than stimulus fading; however, these latter effects were not statistically robust. Similarly, Schreibman (1975) failed to teach young autistic learners using extra-stimulus prompts (non-critical feature manipulations), but were successful with within-stimulus prompts (critical feature manipulations; see also Koegal and Rincover (1976)). These findings are consistent with the non-human animal literature on discrimination learning where stimulus fading (intensity fading) has been shown to be less effective than stimulus shaping (transfer-along-a-continuum). For example, Ploog and Williams (1995) showed that pigeons were able to acquire a difficult discrimination (flicker frequency) only when they were trained with progressively more difficult flicker frequencies, and not when the correct stimulus was predicted by change in colour that was gradually faded out across trials or via trial and error.

From a theoretical perspective, these findings are consistent with research on cue competition. Cue competition is a well studied effect in the learning literature in which one stimulus feature which is predictive of stimulus class membership, interferes with the extent to which other stimuli come to evoke responding. Research on cue competition effects include work on stimulus blocking (Johnson & Cumming, 1968) and overshadowing (Lau et al., 2020). In blocking, a stimulus that predicts reinforcement can interfere with the ability of a novel stimulus (presented concurrently) to evoke later responding. For example, Johnson and Cumming (1968) found that pigeons trained to respond to a color or line orientation followed by training with a combination of both, responded only to the stimulus they initially encountered when the stimuli where presented in isolation in a final test phase.

DOI: 10.1177/17470218231210446

In overshadowing, the intensity of one stimulus impacts the evocative effect of a second stimulus; the salient stimulus in a two-stimulus compound interferes with what is learned about the second (less salient) stimulus. A recent series of experiments reported by Lau et al. (2020) illustrated the overshadowing effect in the context of perceptual category learning. In one experiment, participants were presented with two categories of stimuli with which they were required to determine stimulus class membership. In a control condition, category membership was determined by the number of dots on the stimuli. In the overshadowing condition, category membership could be determined by both dot number and the orientation of the stimulus (tilted left or right). In a later test phase, in which category membership was predicted solely by dot number, participants in the overshadowing group fared much worse than the control group in the categorisation task.

Overshadowing rather than blocking might explain the relative benefits observed in the effectiveness of stimulus shaping relative to stimulus fading (Lau et al., 2020). Specifically, in stimulus fading the salient (non-critical) dimension and the critical dimension are both present for the first time from the beginning of training. The salient (non-critical) feature effectively impedes learning about the critical feature when the salient dimension is eventually removed (i.e., faded). The (albeit limited) literature directly comparing the effectiveness of stimulus fading and stimulus shaping is certainly supportive of this interpretation.

The aforementioned study by Pashler and Mozer (2013) provided some valuable insights for understanding under what conditions stimulus shaping might be beneficial in teaching new discriminations. On the other hand, a number of questions remain unanswered. For example, the authors did not take any measures of stimulus generalization. Stimulus generalization refers to the extent that responding acquired under the control of stimuli in the training and testing sets, extends to novel stimuli that share similar features (Stokes & Baer,

DOI: 10.1177/17470218231210446

1977). Stimulus generalization is often overlooked in studies on errorless learning; Markham et al (2020), found that as little as a third of studies reported such measures. We argue that this is a problematic feature of the literature given the role of stimulus generalization in the development of functional skills. There may be limited benefit for an individual if a new skill occurs only with the exemplars, contexts or materials that have been directly taught (Stokes & Baer, 1977; Schroeder et al., 1998 Wunderlich et al., 2014). This is particularly important consideration in applied contexts as generalization is often not a guaranteed outcome (Bailey, 1981; Wunderlich & Vollmer, 2017). Indeed, EL procedures have been shown to *impair* generalization of skills relative to trial and error in some contexts (Clare & Jones, 2008; Jones et al., 2010; Ownsworth et al., 2017). Accordingly, we argue that generalization tests should be a key feature of analyses of EL procedures.

In summary, while stimulus fading and shaping strategies are commonly used in clinical practice, there is little robust evidence from well-powered studies in support of the relative effectiveness of any particular variation. There is some evidence which is suggestive of a benefit of stimulus fading in the context of teaching individuals with IDD (Lancioni & Smeets, 1986; Strand & Morris, 1988); however, this evidence is limited by methodological concerns and studies involving small numbers of participants. Furthermore, few studies have explored the impact of stimulus fading or shaping methods beyond the stimuli targeted for discrimination (i.e., generalization outcomes). Finally, we currently know very little about the extent to which particular teaching procedures are more or less well suited to particular target discriminations. Elucidating the effects of variations of stimulus prompt fading with neurotypical adult participants may provide some useful insights with which to further explore these phenomena in clinical populations (e.g., proof of concept). Moreover, such evidence may prove useful for informing 'default' or proscribed learning strategies such as

DOI: 10.1177/17470218231210446

those in clinical manuals targeting discrimination learning (Thompson, 2011; Frost & Bondy, 2002).

In the present study, we undertook an extension of the work described by Pashler and Mozer (2013). We made a number of adaptations to the procedures. First, we employed a task in which the to-be-discriminated target stimulus (i.e., category member) was presented concurrently with the (non-target) comparator stimuli (i.e., simultaneous discrimination training). In the original study, all stimuli were presented individually and successively. We made this adaptation both because it enabled us to reduce chance-level responding from 50% (i.e., yes or no) in the original study to 16.7% (i.e., a choice of one in six), but also because in educational contexts simultaneous discrimination tasks are more commonly used than successive tasks (Halbur et al., 2021). Second, whereas Pashler and Mozer compared stimulus shaping to trial and error, we included stimulus fading as a second experimental condition. Thus, we undertook a three-arm randomised trial of two varieties of stimulus prompt fading and a trial-and-error teaching condition. Third, we incorporated two further tests phases to assess two types of generalization. Our research questions were as follows:

RQ1: Do the benefits of stimulus shaping observed in Pashler and Mozer (2013) replicate in a simultaneous discrimination task?

RQ2: How does stimulus fading compare with both stimulus shaping and trial and error-in a perceptual category discrimination task.

RQ3: What is the impact of stimulus shaping and stimulus fading on measures of generalization compared to trial-and-error learning?

DOI: 10.1177/17470218231210446

Method

Participants and Setting

The present study was conducted in compliance with the American Psychological Association's Ethical Principles (2017) and approved by the institution's ethics committee as being of low risk to prospective participants. Criteria for inclusion consisted of being aged 18 or over with access to the internet via a computer. We employed opportunity sampling to recruit participants. We sent an invitation to participate via email to students and staff in the Department of Psychology and Therapeutic Studies at the (*deleted for anonymity)*. In addition, we posted an invitation on listservs (e.g., those supporting psychology postgraduate students) and across a range of social media sites.

In planning our sample size, we estimated the anticipated effect size by examining the results obtained by Pashler and Mozer (2013). The authors reported mean differences in the test phase of 20% (Experiment 4) and 30% (Experiment 5) in accuracy between the fading (EL condition) and trial and error conditions. These differences are indicative of a large effect of fading on test performance using Cohen's (1988) criteria. In addition, we also performed an a priori power analyses to determine the required sample size assuming both a large (F =0.4) and medium (F=0.25) effect. As such, we planned to run three separate ANOVAs corresponding to each of the three test phases (Test, Generalisation 1 and Generalisation 2), we applied a Bonferroni correction which reduced our alpha level from 0.05 to 0.016 for the purpose of each calculation. The power calculations were performed using GPower (Faul et al., 2009). We calculated that with power of .80, the minimum sample size needed assuming a large effect size was 84, and the minimum required assuming a medium effect was 207. Finally, we also took into account resource constraints in planning our sample size. We therefore aimed to recruit 40 participants to each group (n=120).

DOI: 10.1177/17470218231210446

Participants could access the experiment at any time and in any location. As part of the pre-experiment information sheet, we asked participants to undertake the experiment in one sitting using either a Personal Computer (PC) or laptop rather than a smartphone or tablet. Once participants had read the information and had indicated their consent to proceed, the programme randomly allocated them to one of three experimental conditions: Stimulus shaping (n = 41), stimulus fading (n = 39), or a control condition (n = 40). Only those participants that completed the full experiment were included in our data analysis (i.e., complete case analysis). Two hundred and seven participants started the experiment of which 120 completed all the phases (dropout rate of 42%). The number of participants that did not complete all the stages was higher in the control and broadly consistent across stimulus shaping and stimulus fading: control (n= 37; 48%); stimulus shaping (n=25; 38%); stimulus fading (n=25; 39%).

Materials and Stimuli

During all trial presentations a conditional discrimination task was presented on screen in which participants selected one stimulus from an array of six. The stimuli were generated from a bank of animal-like creatures called "Fribbles" designed for behavioural and cognitive research (Barry et al., 2014; Williams, 1997). We used the image editing software *Paint.net,* to adapt the critical feature and opacity of the stimuli according to the target discriminations. Figure 1 shows examples from the stimulus sets. We undertook pilot work with different adaptations of the critical feature and opacity levels in order to avoid ceiling and floor effects during testing. This was achieved by both testing discrimination accuracy in each condition and soliciting qualitative feedback (e.g., asking participants to rate how easy they found identifying the S+ during the test phases where no feedback was provided).

DOI: 10.1177/17470218231210446

The experiment was programmed in PHP: Hypertext Pre-processor and Structured Query Language. On each trial presentation, the S+ was programmed to appear according to the allocated fading or shaping level for that particular trial and the S- stimuli were programmed to be randomly selected from a pool of 27 stimuli (experimental stimuli are available at https://osf.io/rjy52/ on the Open Science Framework (OSF). The random selection from the stimuli pools and the number of stimuli in each reduced the likelihood of two consecutive presentations being identical. Stimuli sets within the Fribbles vary along several parameters; for example, features such as the 'head' or 'leg' type on the creature. We also employed the use of two arbitrary (printed) category names during training and testing: '*Sogi*' and '*Keza*'. These were names that had been employed in a previous study on adult category learning by Behrmann and Williams (2007).

General Procedure

Upon randomization to one of the three conditions, participants were presented with a series of training and testing blocks. Each block consisted of 12 trials and participants received a total of six blocks in total (i.e., 72 screens) corresponding to the six phases of the experiment. During a trial, participants were asked to select a stimulus from an array of six (the S+ and five S- stimuli). A correct response was defined as selecting the stimulus in which the target feature appeared larger relative to the other five stimuli. During the Phases 1-5 the target feature was the creature's tail. In the final phase (Generalization 2), the selected feature were the creature's legs. A schematic diagram of the experiment is available in the supplementary materials provided on the OSF link shown above.

Phases 1 – 3: Training

Stimuli from Set 1 were used during the first three phases. Each trial began with the six stimuli presented on the screen under which was printed: "*Click on the Sogi*." Participants

DOI: 10.1177/17470218231210446

selected a stimulus by clicking on one of the images. This selection generated a green square which appeared around the corresponding stimulus along with a blue button which read: "*select*" appeared on the screen. Participants confirmed their selection by clicking on the "*select*" button, upon which the word "*incorrect*" in red text or "*correct*" in green text, appeared. The position of the correct stimulus (S+) was counterbalanced such that it occupied each position in the array of six, twice during each 12- trial block. The sequence for the position it appeared in was determined at random for each trial. A blue bar depicting progress through the experiment appeared at the bottom of the screen. Participants were unable to return to previously completed trials.

Within-Stimulus Conditions. In the stimulus fading and the stimulus shaping conditions, the appearance of the incorrect stimuli (S-) stimuli was systematically altered during Phases 1-3. The incorrect stimuli on each trial were "faded-in" over the course of the three phases, while the S+ stimulus remained the same throughout. In Phase 1, the S+ was clearly differentiated from the five S- stimuli. In Phase 2, the contrast between the S+ stimulus and S- stimuli was marginally less apparent compared to Phase 1. In Phase 3, the contrast was reduced further still.

Stimulus shaping. In this condition, the target feature (i.e., the size of the tail) on the incorrect stimuli was reduced in size. In Phase 1, the feature on all of the S- stimuli were 150% smaller than the terminal size. In Phase 2, the feature was 100% smaller. In Phase 3, the feature was 75% smaller (see Figure 2).

Stimulus fading. In this condition, the size of the target feature was identical across all of the training phases (i.e., S+ tail was 30% smaller than the S- tails); however, the saturation and color intensity of the S- stimuli was systematically increased across the training phases (e.g., Trial block 1-3). During Phase 1, the intensity of all five of the S- stimuli on each

DOI: 10.1177/17470218231210446

screen presentation was reduced by 75% and the S+ remained at 100% intensity. During Phase 2, the intensity of the S- stimuli was 50% lower than the S+. Finally in Phase 3, the intensity of the S- stimuli was only 25% lower than the S+. See Figure 3 for an example.

Control condition. In this condition, no manipulation of the S- stimuli was employed. Stimulus presentations during Phases 1-3 were identical. As such, the tail on each of the S- stimuli was 30% smaller than the tail on the S+ stimulus in every trial and all stimuli were presented at full intensity throughout. The provision of feedback was identical to both stimulus fading conditions; informational feedback was provided during all three phases.

Phase 4: Test

Following completion of Phase 3 in all conditions, participants were advised they were starting a test stage and would no longer receive feedback on their performance. During this phase, participants were presented with a block of 12 trials in a similar fashion to the training phases but with two exceptions. First, no within-stimulus manipulations were present; the tail on each of the S- stimuli was 30% smaller than the tail on the S+ stimulus in for every trial and all stimuli were presented at full intensity throughout. Second, no feedback was provided.

Generalization

Following the completion of Trial blocks 1-4 with Stimulus Set 1, generalization tests were arranged using a further two stimulus sets. In both cases, a size discrimination determined the correct response (i.e., a feature on the S+ was always slightly larger compared to the S-).

Phase 5: Generalization 1. During the first generalization test, Stimulus Set 2 was used. The critical feature was the same as Set 1 (i.e., tail). The category name was also the same (i.e., *Sogi*). All other trial presentations were identical to Phase 4.

DOI: 10.1177/17470218231210446

Phase 6: Generalization 2. During the second generalization test, Stimulus Set 3 was used. The critical feature for this set was the legs, which were arranged to be bigger on the S+. During this condition, a different category name '*Keza*' was used. The purpose of changing the feature and the name was to examine whether participants responded in accordance with a generalized rule that the relative size difference of *any* feature determined category membership (so-called 'rule-based generalization', Maes et al., 2015). All other trial presentations were identical to Phase 5. Following completion of Phase 6, participants were presented with a debrief screen explaining the nature of the experiment and thanking them for their participation.

Sample screenshots of all the conditions are available on the OSF. The average time for participants to complete the six trial blocks in the stimulus shaping condition was 7 mins and 23 seconds, 7 minutes and 6 seconds in the stimulus fading condition, and 8 minutes and 25 seconds in the control condition.

Data Analysis

We assessed whether the training condition (i.e., control, stimulus fading, stimulus shaping) affected accuracy across the three tests of acquisition (i.e., same set stimuli, different set same feature stimuli, and different set different feature stimuli). We had initially planned to perform a series of ANOVAs to compare the effect of experimental condition; however, the data did not meet the assumption of normally distributed residuals. One possible reason for this is that our dependent variable involved a bounded scale (0-12) which can push the variance towards zero as the mean approaches the upper and lower bounds. As a result, we decided to use a logistic regression modeling strategy which is more well suited to such data (see Barr, 2008, 2021; Kieschnick & McCullough; 2003). We used a Bayesian multilevel modelling approach for each test. In a Bayesian framework, prior information

DOI: 10.1177/17470218231210446

about the parameters in the model (e.g., priors) are specified before the observed data is incorporated. Combining these the prior parameter distributions with the observed data generates a 'posterior distribution' estimate of the parameter values. Statistical inferences can then be made by examining the central tendency and spread of the posterior distributions. In the current study, we generated posterior distributions which estimated the mean probability of a correct response (along with corresponding 95% credibility intervals (CI)) for each training condition. We then used these posterior distributions to calculate pairwise contrasts between the training types. We interpreted these contrasts as suggesting a statistically robust effect of training type if the 95% CI interval did not cross zero.

One of the benefits of specifying priors in a model is that it allows the exclusion of unreasonable values before the model sees the data (McElreath, 2020). Accordingly, we specified mildly informative, so-called 'regularizing', priors. We used a varying intercepts model with two-level structure in which repeated observations (e.g., trials in a test block) were nested within participants. Experimental condition was included as a fixed effect in the model. All models incorporated a Bernoulli likelihood function with a logit link. All Bayesian models were created in Stan computational framework (Carpenter et al., 2017) and accessed using the Brms package (Bürkner, 2018). Model estimation was performed using Markov Chain Monte Carlo via the No-U-Turn Sampler (Hoffman & Gelman, 2014). See supplementary materials for full model specification and details.

Transparency and openness

We report how we obtained participants, the number of participants who completed the experiment in each condition as well as attrition in each condition. We also report all measures in the study and have followed the American Psychological Association's Journal Article Reporting Standards (Kazak, 2018) for reporting data. Data were analysed using R,

DOI: 10.1177/17470218231210446

version 4.0.0 (R core Team, 2020). The study design and analysis plan were not pre-registered; however, all data, analysis code, and research materials have been made available on the OSF.

Results and Discussion

Training

Figure 4 depicts participant responding (raw data) across all phases of the experiment for each of the three conditions. In the control condition, mean percentage correct increased gradually during the training phases (Phases 1-3); however, remained within a range of 20-30%. In contrast, participants in the stimulus shaping and fading conditions responded with a high degree of accuracy during these phases.

Tests

Same Stimuli Test

In the first test participants obtained mean (raw) accuracy scores of 64.4%, 21.4% and 28.3% in the stimulus shaping, stimulus fading and the control conditions respectively. Our model estimations indicated a statistically robust advantage of stimulus shaping relative to the two other conditions. The posterior contrasts estimated that those participants that received training via stimulus shaping were 36.3% (95% CI [21.4 – 50.4]) more likely to be accurate relative to the participants in the control condition, and 44.1% (95% CI [29.3 – 56.5]) more accurate than participants that had received training via stimulus fading. We estimated that participants who received stimulus fading were 7.7% less likely to respond accurately than those who received training in the control condition; however, this difference was not statistically robust (95% CI [-24.1 – 6.7]).

DOI: 10.1177/17470218231210446

Generalization Test 1

In the first generalization test, participants responded to test trials involving a novel stimulus set, but one in which the same target feature determined category membership (i.e., size of the tail). Participants obtained mean (raw) accuracy scores of 56.9%, 26.5% and 29.6% in the stimulus shaping, stimulus fading and the control conditions respectively. Our model estimations indicated a statistically robust advantage of the stimulus shaping training relative to the two other conditions. The posterior contrasts estimated that those participants that received training via stimulus shaping were 28.5% (95% CI [11.1 – 45.3]) more likely to respond correctly relative to the participants in the control condition, and 31% (95% CI [13 – 47.9]) more likely to respond correctly than participants that had received training via stimulus fading. We estimated that participants who received stimulus fading were 2.5% (95% CI [-20.9 – 15.4) less likely to respond accurately than those who received training in the control condition; however, again this was not a statistically robust difference with the posterior distribution credibility interval ranging on the contrast ranging from a 20.9% reduction in accuracy to a 15.4% improvement in accuracy.

Generalization Test 2

In the second generalization test, participants responded to test trials involving another novel stimulus set. This time the target feature was a different stimulus feature than that which determined category membership in training (i.e., leg length). The raw scores indicated that participants tended to perform better in the control condition to either of the within-stimulus training conditions obtaining mean accuracy scores in this test of 16.7%, 16.9% and 27.3% in the stimulus shaping, stimulus fading and the control conditions respectively. The posterior contrasts indicated that most credible score was a 13.3% (95% CI [1- 26.8]) improvement in accuracy having received training in the control condition relative

DOI: 10.1177/17470218231210446

to the stimulus shaping training, and a 13.1% (95% CI [1- 26.8]) accuracy benefit having received training in the control condition relative to the stimulus fading training. When comparing the scores of participants who received the EL training conditions, we estimated zero benefit of either training condition relative to the other (95% CI [-13.8 – 13.8]).

Taken together, the findings from Experiment 1 suggest that stimulus shaping was more effective than stimulus fading or trial-and-error (no EL) in producing perceptual category learning in a simultaneous discrimination task. The relative benefits observed in the stimulus shaping condition extended to novel stimuli that were not used during the initial training, but that shared the target feature (e.g., size of the tail) determining category membership. Neither stimulus shaping or fading of the strategies were effective as measured by the final generalization test (Generalization 2 test) in which the target discrimination was the relative size difference of a new feature (different to that taught during training). Interestingly, participants in the control condition appeared to respond more accurately in this second generalization test than participants in either of the EL conditions.

In Experiment 2 we sought to replicate the findings from Experiment 1 while adjusting some aspects of the task that have been shown to be important in the context of within-stimulus EL procedures. In Experiment 1 we manipulated the S- as part of the adaptations across training trials; however, previous research has suggested that manipulation of the S+ may result in fewer errors during training ultimately impacting on acquisition (Stella & Etzel, 1986). Therefore, it is possible that manipulating the S+ rather than the S- stimuli might alter effectiveness of the EL procedures examined here. Given the large differences in the effectiveness of the procedures we observed in Experiment 1 and examination of the impact of adapting these parameters seems salutary.

DOI: 10.1177/17470218231210446

Experiment 2

In Experiment 1 we found that stimulus fading was no more effective than trial-and-error in establishing a new discrimination (or indeed chance-level responding). In Experiment 2 we explored a variation of fading which is commonly employed in the errorless learning literature. A key consideration in designing a stimulus fading protocol is whether the S+ or the S- stimuli are manipulated. In the earliest demonstrations of stimulus fading, pigeons acquired a simple discrimination when the S- was faded in and the S+ was held constant (Terrace; 1963a, 1963b). More recently, researchers have employed fading along both the S+ and S- in teaching discrimination skills to learners with intellectual disabilities (Strand ,1989; Repp & Karsh, 1992); however, there is no consensus as to which method is more effective (e.g., Zawlocki & Walls, 1983). Thus, in Experiment 2 we attempted a systematic replication of the first experiment but we altered the dimension of the stimuli that was manipulated across trials. In Experiment 1, we had adapted the S- across trials, in Experiment 2 we manipulated a dimension of the S+.

Method

All aspects of the methodology were identical to Experiment 1 with the exception of the modifications made to the fading and shaping dimensions as detailed below.

Participants and Setting

Participant recruitment, sample size determination and allocation to groups was identical to Experiment 1. Only those participants that completed the full experiment were included in our data analysis (i.e., complete case analysis). Two hundred and sixty-one participants started the experiment of which 131 completed all six phases (dropout rate of 50%). The number of participants that completed all the phases was as follows: control (n= 45); stimulus shaping (n=43), and stimulus fading n=43). The number of participants who

DOI: 10.1177/17470218231210446

failed to complete the experiment was similar in the stimulus shaping condition and the control: 54% during the stimulus shaping condition (n=50), and 54% (n=52) in the trial-and-error condition. However, fewer dropped out of the stimulus fading condition (39%; n=28).

Materials and Stimuli

The stimuli employed in Experiment 2 were identical to those used in Experiment 1 with the exception that all modifications were made to the S+ stimuli as opposed to the S- stimuli.

Procedures

The procedures were identical to those described in Experiment 1, with the exception that a correct response involved the selection of the stimulus that had one target feature that appeared smaller relative to that feature on the other five stimuli on the screen. This adaptation was selected so that we were able to maintain consistency across the two Experiments in terms of the dimension by which the fading took place in the stimulus shaping condition. During Experiment 1, the critical feature on the S- stimuli were altered from small to large and at the final discrimination the S- stimuli were 30% smaller than the S+. In Experiment 2, the critical feature of the S+ stimulus was also altered from small to large using the same increments as Experiment 1.

Within-Stimulus Conditions. For both stimulus fading and stimulus shaping conditions, the saliency of the S+ was manipulated during training Phases 1-3. While the S+ stimuli were changed over the course of the three training phases, the S- stimuli remained the same throughout. Consistent with Experiment 1, in the first phase, the S+ was clearly differentiated from the five S- stimuli. In Phase 2, the contrast between the S+ stimulus and S- stimuli was not as apparent as during the first level. In Phase 3, the difference was reduced further still.

DOI: 10.1177/17470218231210446

Stimulus shaping. During this condition, the target feature (i.e., the size of the tail) was manipulated. In Phase 1, the tail on the S+ stimuli appeared 150% smaller than the terminal size. In Phase 2, the tail was 100% smaller. In Phase 3, the tail was 75% smaller.

Stimulus fading. During this condition, the size of the target feature was identical across all of the training phases (i.e., S+ tail was 30% smaller than the S- tails). However, the saturation and color intensity of the S+ stimuli was systematically increased across the training phases (i.e., Phases 1-3) until all the stimuli were the same color intensity at the final discrimination. During Phase 1, the intensity of the S+ stimuli on each screen presentation was reduced by 75% and the S- remained at 100%. During Phase 2, the intensity of the S+ image was 50% lower than the S- stimuli. Finally in Phase 3, the intensity of the S+ stimuli was 25% lower than the S- stimuli.

Control

During this condition, no manipulation of the S- stimuli was employed. Stimulus presentations during Phases 1-3 were identical. As such, the tail on the S+ stimulus was 30% smaller than the tail on the S- stimuli in every trial and all stimuli were presented at full color intensity and saturation throughout. Informational feedback was provided during all three phases.

Generalization

Generalization 1 and Generalization 2 were identical to those described in Experiment 1 (with the exception that the tail on S+ was 30% smaller rather than larger). Screenshots of all the conditions in Experiment 2 are available on the OSF.

DOI: 10.1177/17470218231210446

Data analysis

We analysed the data using an identical multilevel logistic regression modelling strategy to that that for Experiment 1.

Results and Discussion

Training (Phases 1-3)

Figure 6 depicts participant responding across all phases for each of the three conditions in Experiment 2. In the control condition, mean percentage correct increased gradually during the training phases (Phases 1-3); however, remained within a range of 20-30 % during the final training phase (3). Participants in the two within-stimulus conditions responded relatively accurately during the training phases, although accurate responding during training in the stimulus shaping condition was more variable compared to Experiment 1.

Tests

Same Stimuli Test

In the first test in which participants responded to trials involving the stimulus set they were trained with they obtained mean accuracy scores of 60.4%, 26.4% and 33.9% in the stimulus shaping, stimulus fading and the control conditions respectively. As with Experiment 1, the model suggested a statistically robust advantage of stimulus shaping relative to the two other conditions. The posterior contrasts estimated that those participants that received stimulus shaping were 25.7% (95% CI [5.2 – 45.4]) more likely to be accurate relative to the participants in the control condition, and 31% (95% CI [11.2 – 49.5]) more accurate than participants that had received stimulus fading. Our analysis estimated that participants who received stimulus fading were 5.3% less likely to respond accurately than

DOI: 10.1177/17470218231210446

those who received training in the control condition; however, the range of the posterior distribution indicated that this was not a statistically robust difference (95% CI [-24.8 – 15.1]).

Generalization Test 1

In the first generalization test, participants responded to test trials involving a new stimulus set than they were trained with, but in which the same stimulus feature determined category membership (e.g., size of the tail). Participants obtained mean accuracy scores of 47.9%, 23.1% and 34.1% in the stimulus shaping, stimulus fading and the control conditions respectively. The model indicated a statistically robust advantage of the shaping training when compared to the fading condition. The posterior contrast estimated that the most credible difference was a 25.1% (95% CI [6.2 – 42]) advantage for those participants that received stimulus shaping relative to the participants who had received stimulus fading. The model estimated a 13.7% (95% CI [-5.5% – 32.2%]) improvement in accuracy having received stimulus shaping compared to training with no stimulus manipulation (control), but again this difference was not statistically robust. The posterior contrast between participants who had received training in the control condition and the stimulus fading condition was suggestive of an advantage of training in the control condition; the most credible value was an 11.5% (95% CI [-7.1 – 29.9]) improvement, however, the 95% credible interval contained zero.

Generalization Test 2

In the second generalization test, participants responded to test trials involving a new stimulus set than they were trained with but that also involved a discrimination that involved a different stimulus feature than that which determined category membership in training. Consistent with the first experiment, participants tended to perform better in the control

DOI: 10.1177/17470218231210446

condition to either of the within-stimulus conditions. Participants' mean accuracy scores in the second test of generalization were 24%, 17.6% and 32.6% in the stimulus shaping, stimulus fading and the control conditions respectively. The posterior contrasts indicated that most credible value was a 11.6% (95% CI [-4.8%- 26.8]) benefit in accuracy for participants in the control condition compared to those that received stimulus shaping, and a 17.8% (95% CI [1- 33.6]) benefit relative to those in the stimulus fading condition. When comparing the scores of participants who received stimulus shaping, the model estimated zero benefit of either training condition relative to the other (95% CI[-13.8 – 13.8]). There was no statistically robust benefit of receiving stimulus shaping relative to stimulus fading. The model estimated that participants in the stimulus shaping condition were 6.2% more likely to respond correctly in these tests (95% CI [-10.9 – 22.9).

General Discussion

The purpose of the present study was to extend the work of Pashler and Mozer (2013) by comparing stimulus shaping, stimulus fading and trial-and-error teaching in a perceptual categorisation task. We also explored the impact of the procedures on measures of generalization. Taken together, the findings we have obtained from Experiments 1 and 2 suggest that stimulus shaping is more effective relative to trial-and-error and when compared to stimulus fading. Moreover, these benefits extended to the categorisation of stimuli beyond those directly targeted in the training (i.e., generalization outcomes), which is often neglected in studies on EL discrimination training. Our results also indicate that the benefits were apparent irrespective of whether the S+ or S- was manipulated as part of the EL procedure.

The finding that shaping (i.e., manipulating a critical dimensions of a target) was relatively more effective than fading or trial-and-error is consistent with the little existing research directly comparing these approaches (e.g., Lancioni & Smeets, 1986; Strand &

DOI: 10.1177/17470218231210446

Morris, 1988). Our study does contribute to the literature in that it is the first systematic comparison of these approaches in a large neurotypical sample. Previous comparative studies have generally involved small numbers of participants and looked at atypical populations (Markham et al., 2020). For example, while Strand and Morris (1988) reported no statistically significant differences between the shaping and fading training conditions at a group-level, this was likely to have been in-part an artefact of the low power in the study; their analyses were indicative of a benefit of shaping when analysis was undertaken at the individual participant level. The findings reported here represent the first demonstrations of a comparative study exploring these procedures using a large group-based, randomised design.

A notable feature of the results was the marked decrease in accuracy from training to tests following training in the stimulus fading condition. While participants responded with a high degree of accuracy in the training phases (i.e., Phases 1-3), upon testing (Phase 4) they responded at levels of accuracy comparable to the trial-and-error group, which approximated chance-level responding. This abrupt decrease in accuracy suggests that participants' responding was under the sole control of the (irrelevant) features of the stimuli that were being manipulated during training (i.e., the opacity of the image). This finding is consistent with the theoretical interpretation and existing research arguing that the failure of stimulus fading might be indicative of overshadowing (e.g., Lau et al., 2020). This is a potentially important finding given that fading procedures are a tactic that are recommended and used as part of EL strategies in clinical settings. One implication is that in the early stages of teaching, these procedures might give the impression they are effective when in fact they are not. Maximizing success from the outset of training procedures can be advantageous in clinical settings (e.g., access to positive reinforcement, reducing frustrations); however, there is little value in establishing correct responding under faulty stimulus control, particularly if these teaching procedures absorb valuable time and resources. To our knowledge, there is

DOI: 10.1177/17470218231210446

scant research exploring whether the success (or failure) of these procedures is dependent on the particular discrimination targeted.

Our findings suggest that further exploration of the specific conditions under which fading (and indeed shaping) is and is not effective is warranted. The current study showed that when category membership was based on a subtle difference between the perceptual features of the candidate stimuli, fading in of the intensity of the entire image was an inferior tactic relative to directly exaggerating the feature central to the discrimination. If overshadowing was responsible for the relative failure of stimulus fading compared to stimulus shaping, we might expect stimulus fading to perform worse than trial-and-error given that the latter did not involve the manipulation of the non-critical feature. We did not observe robust differences in the performance of participants who had received stimulus fading or trial-and-error. We suspect, however, that this is most probably due to floor effects. We programmed the differences in the trial-and-error condition to maximise the difficulty of the discrimination so that we would be able to clearly detect any benefits of the EL procedures. While stimulus fading has proved to be inferior compared to trial-and-error in previous research (Koegal and Rincover, 1976) the current experiments were not designed to test this prediction. Future studies might seek to explore this research question with a trial-and-error condition more likely to generate stimulus control.

We observed some features of participant responding that were inconsistent with previous research on shaping and fading. For instance, in Experiment 2, where the S+ was manipulated, accurate responding was lower in the stimulus shaping condition across the test phases, compared to participant performance in Experiment 1 where we had modified the S-. This finding is not in line with previous comparisons of manipulating the S+ and S-, as Stella and Etzel (1986) discovered lower accuracy in their stimulus shaping condition when the S- was manipulated. We did not design the study to explicitly test for these differences and

DOI: 10.1177/17470218231210446

therefore this observation may simply reflect differences in the sample that are not related to the training conditions. Nevertheless, these questions could represent a promising avenue for future research and underscore the benefits of conducting well-powered trials capable of identifying subtle but important differences.

Our findings are also useful to the extent that we explored generalization outcomes. As highlighted in Markham et al. (2020), measures of generalization are seldom reported in studies on EL procedures. The benefits in accuracy observed in the stimulus shaping condition extended to stimuli beyond that directly targeted in the training. Participants who received training in the shaping condition were more likely to respond correctly to a categorisation task in which they had to identify characters in accordance with a specific target feature (e.g., a tail). This finding suggests that stimulus shaping via the S+ or S- stimuli may promote some degree of stimulus generalization. Conversely, the fading and trial-and-error conditions did not lead to stimulus generalization. This latter point should be treated with some caution given that these procedures did not produce robust learning with the targeted stimuli therefore generalization would not be expected. Neither of the errorless learning conditions led to relatively better generalization as measured by the second test. This assessment evaluated whether participants learnt to respond to a novel target discrimination in which the relative size of a different feature (the legs of the character rather than the tail) was the basis for categorization (i.e., rule-based generalization; Maes et al., 2015). There was no evidence that participants in either EL condition responded more accurately than chance in these tests in either experiment. The finding that accuracy was lower during this second generalization test is also consistent with experiments examining the difference between predictive and non-predictive features in determining the learner's attention to stimuli during discrimination training (e.g., Mackintosh, 1975). For example, when there is a predictable transfer from those critical features learned during training to generalization stimuli (i.e., an

DOI: 10.1177/17470218231210446

intradimensional shift) then participants respond with higher accuracy compared to generalization tests whereby stimuli do not have those specific critical features (i.e., an *extradimensional* shift; Eimas, 1966; Rittle & Baron, 1969). Interestingly, we found a marginal benefit of trial-and-error teaching in these tests when responding was compared to what would be expected by chance level responding alone. One possible explanation is that trial-and-error teaching likely produced greater variability in responding during the training phases, which may promote generalization (Raviv et al., 2022). This finding certainly warrants future examination, particularly given the findings from research in clinical settings which have found relatively better generalization following trial-and-error learning compared to error-controlled learning strategies (e.g., Ownsworth et al., 2017).

One limitation of the present study was the lack of control over the size of screen, and therefore stimulus presentations that participants experienced across the conditions. While efforts were made to minimize the magnitude of the difference in screen size by asking participants to avoid the use of portable devices for example, we were not able to verify that this occurred. While it is possible participant experience of the experiment differed unsystematically across the groups, our randomized design may have mitigated the impact of this confound to some extent. Future research on these questions might consider standardising the presentations of stimuli by conducting the experiment under strict laboratory conditions.

We evaluated the effects of stimulus fading and stimulus shaping with a neurotypical adult sample. It might be argued that examining errorless learning tactics with verbally sophisticated learners is a limitation given that these strategies are more likely to be used in applied settings. There are a number of reasons why we believe this work to be informative. First, this research will further our understanding as to whether errorless learning tactics have benefits that are generalisable across different groups. Second, there are contexts in which

DOI: 10.1177/17470218231210446

reducing learner error might be useful even for those with sophisticated verbally repertoires; for example, where the discriminations are not easily verbalized (e.g., radiology, dermatology), or in scenarios where relying on verbal explanations or extensive training is an impractical solution (e.g., safety information/warnings, road signage). Finally, this work is notoriously difficult to power adequately in applied settings (see the paucity of group-based analyses reported in Markham, 2020), thus work with bigger more readily accessible participant pools can provide useful (albeit imperfect) insights.

A further limitation relates to way in which shaping or fading was implemented across the phases. Unlike some other studies on errorless learning and stimulus discrimination more generally (e.g., Stella & Etzel, 1986; Graff & Green, 2004) the participants' progress through the experiment was not determined by accuracy or mastery. For example, all participants progressed through to the final generalization phase independent of response accuracy during the earlier training phases. It is possible that participants may have performed better in the later tests had progression been related to performance during training. For example, in previous research participants have either remained on a particular fading level and received additional training until a criterion was achieved (e.g., Wolfe & Cuvo, 1978; Allen & Fuqua, 1985). As a counterpoint to this, (and as can be seen in Figures 4 and 6) participants in both the stimulus fading and stimulus shaping conditions performed with a high degree of accuracy in the final training phase, which suggests that participants had mastered the discrimination. Furthermore, our interest was in the effects of training where these parameters are fixed. Adapting experimental procedures to account for the differences in the performances of individual learners introduces additional sources of control over responding which then makes it difficult to evaluate the contribution of an independent variable to behaviour change. Moreover, in clinical settings, arranging bespoke and response-dependent procedures is often unfeasible. Nevertheless, future research may further examine the

DOI: 10.1177/17470218231210446

importance of tailoring the number of fading levels on the relative effectiveness of these procedures.

While stimulus prompt fading strategies are commonly recommended in clinical practice, very few studies have explored the relative effectiveness of these techniques in well controlled studies or evaluated generalization outcomes. The findings we have reported here suggest that stimulus shaping was, on average, a relatively more effective teaching procedure than stimulus fading when the discrimination involved a subtle difference in the perceptual features of the stimuli. This benefit persisted in a generalization test in which participants selected novel (unseen) stimuli from within the same category (e.g., generalization test). Accordingly, our findings suggest that stimulus shaping might be a candidate for use as the default strategy (compared to stimulus fading) in settings in which within-stimulus errorless learning strategies are used and the target discrimination involves marginal perceptual differences between the stimuli (e.g., category or concept learning). Given that we found very little benefit of stimulus fading, future studies should seek to explore the conditions under which stimulus fading tactics are (and are not) effective. We consider this to be a pressing issue given that that both stimulus shaping and stimulus fading are used widely in clinical and applied settings.

Data Accessibility Statement

The data and materials from the present experiment are publicly available at the Open Science Framework website:https://osf.io/rjy52/

DOI: 10.1177/17470218231210446

References

Allen, K. D., & Fuqua, R. W. (1985). Eliminating selective stimulus control: A comparison of two procedures for teaching mentally retarded children to respond to compound stimuli. *Journal of Experimental Child Psychology, 33,* 57-71. https://doi.org/10.1016/0022-0965(85)90029-3

American Psychological Association (2017). Ethical principles of psychologists and code of conduct. https://www.apa.org/ethics/code

Amitay, S., Irwin, A., & Moore, D. R. (2006). Discrimination learning induced by training with identical stimuli. *Nature Neuroscience, 9,* 1446-1448. https://doi.org/10.1038/nn1787

Bailey, S. L. (1981). Stimulus over-selectivity in learning disabled children. *Journal of Applied Behavior Analysis, 14,* 239-248. https://doi.org/10.1901/jaba.1981.14-239

Barry T.J., Griffith J.W., De Rossi S., & Hermans, D. (2014). Meet the Fribbles: novel stimuli for use within behavioural research. *Frontiers in Psycholy*. 5:103. doi: 10.3389/fpsyg.2014.00103

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barr, D. J. (2021). *Learning statistical models through simulation in R: An interactive textbook.* Version 1.0.0. Retrieved from https://psyteachr.github.io/stat-models-v1.

DOI: 10.1177/17470218231210446

Behavior Analyst Certification Board. (2017). BCBA task list (5th ed.). https://www.bacb.com/wp-content/uploads/2020/08/BCBA-task-list-5th-ed-211019.pdf

Behrmann, M., & Williams, P. (2007). Impairments in part-whole representations of objects in two cases of integrative visual agnosia. *Cognitive Neuropsychology, 24*, 701-730. https://doi.org/10.1080/02643290701672764

Booth, N. & Keenan, M. (2018). Discrete Trial Teaching: A study on the comparison of three training strategies. *Interdisciplinary Education and Psychology, 2*(2), 1-13. https://doi.org/10.31532/InterdiscipEducPsychol.2.2.003

Brown, S. M., & Bebko, J. M. (2012). Generalization, overselectivity, and discrimination in the autism phenotype: A review. *Research in Autism Spectrum Disorders, 6*, 733 – 740. https://doi.org/10.1016/j.rasd.2011.10.012

Bürkner, P. (2018). *Bayesian item response modeling in R with brms and Stan*. arXiv. https://doi.org/10.48550/arXiv.1905.09501

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76* (1), 1-32. htpps://doi.org/10.18637/jss.v076.i01

Clare, L., & Jones, R.S. (2008). Errorless learning in the rehabilitation of memory impairment: A critical review. *Neuropsychology Review,18* (1),1-23. https://doi.org/10.1007/s11065-008-9051-4.

DOI: 10.1177/17470218231210446

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied Behavior Analysis* (3rd ed.). Pearson Education, Inc.

de Werd, M. M., Boelen, D., Rikkert, M. G., & Kessels, R. P. (2013). Errorless learning of everyday tasks in people with dementia. *Clinical Interventions in Aging*, *8*, 1177–1190. https://doi.org/10.2147/CIA.S46809

Donaghey, C.L., McMillan, T. M., & O'Neill, B. (2010). Errorless learning is superior to trial and error when learning a practical skill in rehabilitation: A randomized controlled trial. *Clinical Rehabilitation, 24*. 195-201. https://doi.org/10.1177/0269215509353270

Donahoe, J. W., Palmer, D. C., & Dorsel, V. P. (2004). *Learning and Complex Behavior*. Ledgetop Pub.

Dunn, J., & Clare, L. (2007). Learning face-name associations in early-stage dementia: Comparing the effects of errorless learning and effortful processing. *Neuropsychological Rehabilitation,17*, 735-54. https://doi.org/10.1080/09602010701218317

Eimas, P. D. (1966). Effects of overtraining and age on intradimensional and extra-dimensional shifts in children. *Journal of Experimental Child Psychology, 3*(4), 348 – 355. https://doi.org/10.1016/0022-0965(66)90079-8

Etzel, B. C. & LeBlanc, J. M. (1979). The simplest treatment alternative: The law of parsimony applied to choosing appropriate instructional control and errorless-learning procedures for the difficult-to-teach child. *Journal of Autism & Developmental Disorders, 9,* 361-382. https://doi.org/10.1007/BF01531445

Evans, J. J., Wilson, B. A., Schuri, U., Andrade, J., Baddeley, A., Bruna, O., Canavan, T., Del Sala, S., Green, R., Laaksonen., R., Lorenzi, L., & Taussik, I. (2000). A comparison of "errorless" and "trial-and-error" learning methods for teaching

DOI: 10.1177/17470218231210446

individuals with acquired memory deficits. *Neuropsychological Rehabilitation, 10*, 67-101. https://doi.org/10.1080/096020100389309

Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods,* 41, 1149-1160. doi:10.3758/BRM.41.4.1149.

Fields, L. (2018). Transfer of discriminative control during stimulus fading conducted without reinforcement. *Learning & Behavior, 46,* 79–88. https://doi.org/10.3758/s13420-017-0294-x

Fisher, W. W., Piazza, C. C., & Roane, H. S. (2021). *Handbook of applied behavior analysis* (2nd Ed). Guilford Press.

Frost, L. A., & Bondy, A. S. (2002). *PECS: The Picture Exchange Communication System training manual* (2nd ed.). Pyramid Educational Consultants, Inc.

Graff, R. B., & Green, G. (2004). Two methods for teaching simple visual discriminations to learners with severe disabilities. *Research in Developmental Disabilities, 25,* 295-307. https://doi.org/10.1016/j.ridd.2003.08.002

Green, G. (2001). Behavior analytic instruction for learners with autism: Advances in stimulus control technology. *Focus on Autism and Other Developmental Disabilities*, *16* (2), 72-85. https://doi.org/10.1177/108835760101600203

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15,* 1593-1623. https://doi.org/10.48550/arXiv.1111.4246

Johnson, D. F., & Cumming, W. W. (1968). Some determiners of attention. *Journal of the Experimental Analysis of Behavior*, *11*(2), 157-166.

DOI: 10.1177/17470218231210446

Jones, R. S. P. & Eayrs, C. B. (1992). The use of errorless learning procedures in teaching people with a learning disability: A critical review. *Mental Handicap Research, 5,* 204-212. https://doi.org/10.12691/education-4-4-1

Jones, R. S. P., Clare, L., MacPartlin, C., & Murphy, O. (2010). The effectiveness of trial-and-error and errorless learning in promoting the transfer of training. *European Journal of Behavior Analysis*, *11*(1), 29–36. https://doi.org/10.1080/15021149.2010.11434332

Kazak, A. E. (2018). Editorial: Journal article reporting standards. *American Psychologist, 73*(1), 1–2. https://doi.org/10.1037/amp0000263

Kieschnick, R., & McCullough, B. D. (2003). Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions. *Statistical Modelling*, *3*(3), 193–213. https://doi.org/10.1191/1471082X03st053oa

Koegel, R. L., & Rincover, A. (1976). Some detrimental effects of using extra stimuli to guide learning in normal and autistic children. *Journal of Abnormal Child Psychology*, 4, 59-71.

Lancioni, G. E., & Smeets, P. M. (1986). Procedures and parameters of errorless discrimination training with developmentally impaired individuals. *International Review of Research in Mental Retardation, 14,* 135-164. https://doi.org/10.1016/S0074-7750(08)60077-4

Lau, J. S.-H., Casale, M. B., & Pashler, H. (2020). Mitigating cue competition effects in human category learning. *Quarterly Journal of Experimental Psychology*, 73(7), 983–1003. https://doi.org/10.1177/1747021820915151

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan, 2nd Edition* (2 ed.). CRC Press.

DOI: 10.1177/17470218231210446

Maes, E., DeFilippo, G., Inkster, A. B., Lea, S. E. G., De Houwer, J., D'Hooge, R., Beckers, T., & Wills, A. J. (2015). Feature- versus rule-based generalization in rats, pigeons and humans. *Animal Cognition, 18*, 1267-1284. https://doi10.1007/s10071-015-0895-8

Markham, V. A., Giles, A. F., Roderique-Davies, G., Adshead, V., Tamiaki, G., & May, R. J. (2020). Applications of within-stimulus errorless learning methods for teaching discrimination skills to individuals with intellectual and developmental disabilities: A systematic review. *Research in Developmental Disabilities, 97.* https://doi.org/10.1016/j.ridd.2019.103521

Mazur, J.E., & Odum, A.L. (2023). Learning and Behavior (9th ed.). Routledge. https://doi.org/10.4324/9781003215950

Middleton, E. L., & Schwartz, M. F. (2012). Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation, 22* (2),138-68. https://doi.org/10.1080/09602011.2011.639619

Noell, G. H., Call, N. A., Ardoin, S. P., & Miller, S. J. (2021). Building complex repertoires from discrete behaviors: Establishing stimulus control, behavioral chains, and strategic behavior. In W.W. Fisher, C.C. Piazza, & H.S. Roane, Eds.), *Handbook of Applied Behavior Analysis* (2nd ed., pp. 252-270). The Guildford Press.

Ownsworth, T., Fleming, J., Tate, R., Beadle, E., Griffin, J., Kendall, M., Schmidt, J., Lane-Brown, A., Chevignard, M., & Shum, D.H.K. (2017). Do people with severe traumatic brain injury benefit from making errors? A randomized controlled trial of error-based and errorless learning. Neurorehabilitation and Neural Repair, 31,1072-1082. https://doi.org10.1177/1545968317740635

DOI: 10.1177/17470218231210446

Pashler, H., & Mozer, M. C. (2013). When does fading enhance perceptual category learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1162-1173. https://doi.org/10.1037/a0031679

Ploog, B. O., & Williams, B. A. (1995). Two methods of stimulus fading applied to a simultaneous flicker rate discrimination in pigeons. *Learning and Motivation*, *26*(2), 161–182. https://doi.org/10.1016/0023-9690(95)90003-9

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raviv, L., Lupyan, G., Green, S. C. (2022). How variability shapes learning and generalization. *Trends in Cognitive Sciences, 26,* 462-483. https://doi.org/10.1016/j.tics.2022.03.007

Repp, A. C., & Karsh, K. G. (1992). An analysis of a group teaching procedure for persons with developmental disabilities. *Journal of Applied Behavior Analysis, 25,* 701-712. https://doi.org/10.1901/jaba.1992.25-701

Rittle, R. H., & Baron, M. R. (1969). Selective attention and rate of discrimination learning as a function of intradimensional variability. *Journal of Experimental Psychology, 82*(2), 327–331. https://doi.org/10.1037/h0028124

Schilmoeller, G. L., Schilmoeller, K. J., Etzel, B. C., & LeBlanc, J. M. (1979). Conditional discrimination after errorless and trial-and-error training. *Journal of the Experimental Analysis of Behavior*, *31*, 405–420. https://doi.org/10.1901/jeab.1979.31-405

Schreibman, L. (1975). Effects of within-stimulus and extra-stimulus prompting on discrimination learning in autistic children. *Journal of Applied Behavior Analysis, 8*(1), 91–112. https://doi.org/10.1901/jaba.1975.8-91

DOI: 10.1177/17470218231210446

Stella, M E., & Etzel, B. C. (1986). Stimulus control of eye orientations: Shaping S+ only versus shaping S- only. *Analysis and Intervention in Developmental Disabilities, 6*, 137-153, https://doi.org/10.1016/0270-4684(86)90011-X

Schroeder, C. M., Schuster, J. W., & Hemmeter, M. L. (1998). Efficiency in programming for generalization: Comparison of two methods for teaching expressive labeling to pre-schoolers with developmental delays. *Journal of Developmental and Physical Disabilities, 10*, 109-131. https://doi.org/ 10.1023/A:1022803631613

Stella, M E., & Etzel, B. C. (1986). Stimulus control of eye orientations: Shaping S+ only versus shaping S- only. *Analysis and Intervention in Developmental Disabilities, 6*, 137-153, https://doi.org/10.1016/0270-4684(86)90011-X

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349-367. https://doi.org/10.1901/jaba.1977.10-349

Strand, S. C. (1989). S+ versus S- fading in teaching visual discriminations to severely mentally handicapped children. *Journal of Mental Deficiency Research, 33*, 283-299. https://doi.org/10.1111/j.1365-2788.1989.tb01478.x

Strand, S. C., & Morris, R. C. (1988). Criterion-related versus non-criterion-related prompt training with severely mentally handicapped children. *Journal of Mental Deficiency Research, 32,* 137-151. https://doi.org/10.1111/j.1365-2788.1988.tb01399.x

Thompson, T. (2011). *Individualized autism intervention*. Paul H. Brookes Publishing Co.

VanLaarhoven, T., Johnson, J. W., Repp, A. C., Karsh, K. G., & Lenz, M. W. (2003). Discrimination training: A comparison of two procedures for presenting multiple examples within a fading and non-fading paradigm. *Research in Developmental Disabilities, 24*, 1-18. https://doi.org/10.1016/S0891-4222(02)00171-3

DOI: 10.1177/17470218231210446

Voigt-Radloff, S., de Werd, M.M.E., Leonhart, R., Boelen, D.H.E., Olde Rikkert, M.G.M., Fliessbach, K., Kloppel, S., Heimbach, B., Fellgiebel, A., Dodel, R., Eschweiler, G.W., Hausner, L., Kessels, R.P.C., & Hull, M. *(2017)*. Structured relearning of activities of daily living in dementia: The randomized controlled REDALI-DEM trial on errorless learning. *Alzheimer's Research & Therapy, 9,* 1-11. https://doi.org/10.1186/s13195-017-0247-9

Williams, P. (1997). *Prototypes, exemplars, and object recognition*. Yale University.

Wolfe, V. F., & Cuvo, A. J. (1978). Effects of within-stimulus and extra-stimulus prompting of letter discrimination by mentally retarded persons. *American Journal of Mental Deficiency, 83,* 297-303. PMID: 717444

Wunderlich, K. L., Vollmer, T. R., Donaldson, J. M., & Phillips, C. L. (2014). The effects of serial and concurrent training on acquisition and generalization. *Journal of Applied Behavior Analysis, 47*, 723-737. https://doi.org/10.1002/jaba.154

Wunderlich, K. L., & Vollmer, T. R. (2017). Effects of serial and concurrent training on receptive identification tasks: A systematic replication. *Journal of Applied Behavior Analysis, 50,* 641-652. https://doi.org/10.1002/jaba.401

Zawlocki, R. J., & Walls, R. T. (1983). Fading on the S+, the S-, both, or neither. *American Journal of Mental Deficiency, 87,* 462-464. PMID: 6829626.

DOI: 10.1177/17470218231210446

Figure Captions

Figure 1. Examples of stimuli used in sets 1,2, and 3 and their associated category names (*Sogi* and *Keza*). The red circle has been used to highlight the critical feature for each set.

Figure 2. Examples of S- stimuli during the three levels of shaping. The S+ stimulus remained the same size throughout.

Figure 3. Examples of S- stimuli during the three levels of stimulus fading. The S+ stimulus remained at 100% color intensity throughout.

Figure 4. Participant responses during the training and test phases of Experiment 1 for each condition. Open circles indicate mean level. Horizontal lines indicate medians. Boxplot hinges show the inter-quartile range (IQR). Boxplot whiskers extend from the boxplot hinges to the largest (upper) and smallest (lower) observed value no further than 1.5 x the IQR. 1 = Phase 1, 2 = Phase 2, 3 = Phase 3, T = Same Stimuli Test, G = Generalization test 1, G2 = Generalization test 2.

Figure 5. Same Stimuli Test (left), Same Stimuli Different Feature (middle), and Different Stimul Different Feature (right) point range visualizations of the model's estimates in each condition (mean and 95% credibility intervals) in Experiment 1. Dashed line indicates chance level responding (16.7% accuracy).

Figure 6. Participant responses during the training and test phases of Experiment 2 for each condition. Open circles indicate mean level. Horizontal lines indicate medians. Boxplot hinges show the inter-quartile range (IQR). Boxplot whiskers extend from the boxplot hinges to the largest (upper) and smallest (lower) observed value no further than 1.5 x the IQR. 1 = Phase 1, 2 = Phase 2, 3 = Phase 3, T = Same Stimuli Test, G = Generalization test 1, G2 = Generalization text 2.

DOI: 10.1177/17470218231210446

Figure 7. Same Stimuli Test (left), Same Stimuli Different Feature (middle), and Different Stimuli Different Feature (right) point range visualizations of the models' estimates in each condition (mean and 95% credibility intervals) in Experiment 2. Dashed line indicates chance level responding (16.7% accuracy).

DOI: 10.1177/17470218231210446

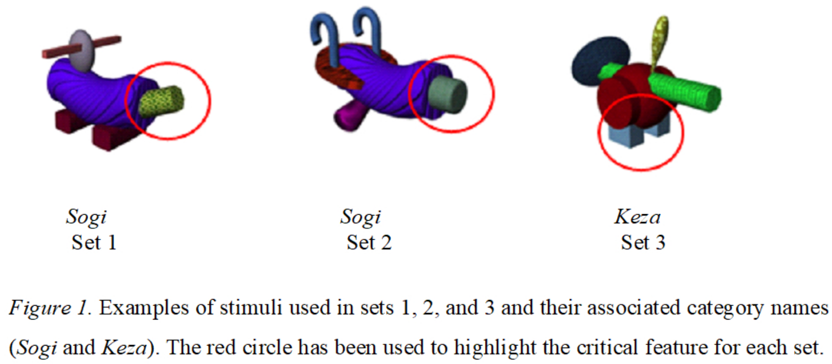

Figure 1. Examples of stimuli used in sets 1,2, and 3 and their associated category names (Sogi and Keza). The red circle has been used to highlight the critical feature for each set.

203x114mm (144 x 144 DPI)

DOI: 10.1177/17470218231210446

Figure 2. Examples of S- stimuli during the three levels of shaping. The S+ stimulus remained the same size throughout.

Figure 2. Examples of S- stimuli during the three levels of shaping. The S+ stimulus remained the same size throughout.

203x141mm (144 x 144 DPI)

DOI: 10.1177/17470218231210446

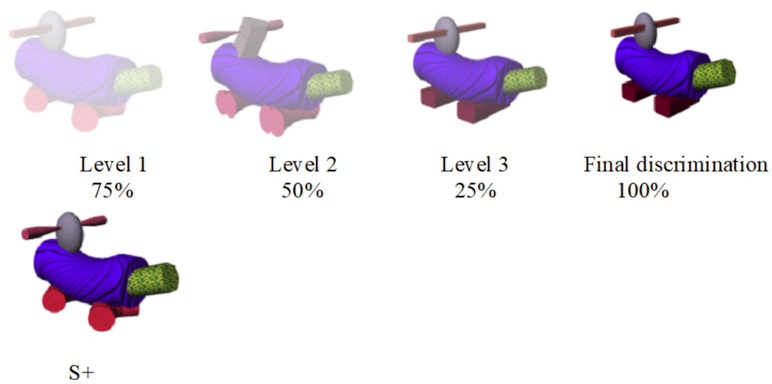

Figure 3. Examples of S- stimuli during the three levels of stimulus fading. The S+ stimulus remained at 100% color intensity throughout.

Figure 3. Examples of S- stimuli during the three levels of stimulus fading. The S+ stimulus remained at 100% color intensity throughout.

203x141mm (144 x 144 DPI)

DOI: 10.1177/17470218231210446

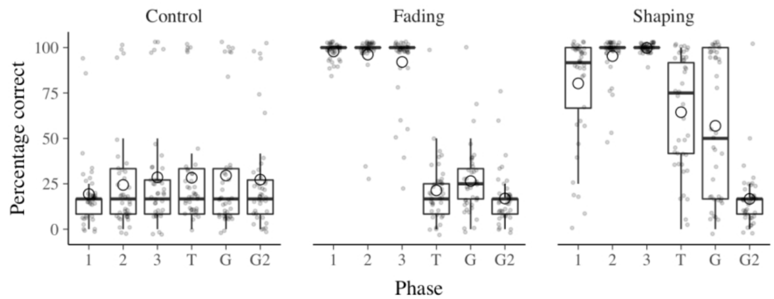

Figure 4. Participant responses during the training and test phases of Experiment 1 for each condition. Open circles indicate mean level. Horizontal lines indicate medians. Boxplot hinges show the inter-quartile range (IQR). Boxplot whiskers extend from the boxplot hinges to the largest (upper) and smallest (lower) observed value no further than 1.5 x the IQR. 1 = Phase 1, 2 = Phase 2, 3 = Phase 3, T = Same Stimuli Test, G = Generalization test 1, G2 = Generalization test 2.

Figure 4. Participant responses during the training and test phases of Experiment 1 for each condition. Open circles indicate mean level. Horizontal lines indicate medians. Boxplot hinges show the inter-quartile range (IQR). Boxplot whiskers extend from the boxplot hinges to the largest (upper) and smallest (lower) observed value no further than 1.5 x the IQR. 1 = Phase 1, 2 = Phase 2, 3 = Phase 3, T = Same Stimuli Test, G = Generalization test 1, G2 = Generalization test 2.

223x144mm (192 x 192 DPI)

DOI: 10.1177/17470218231210446

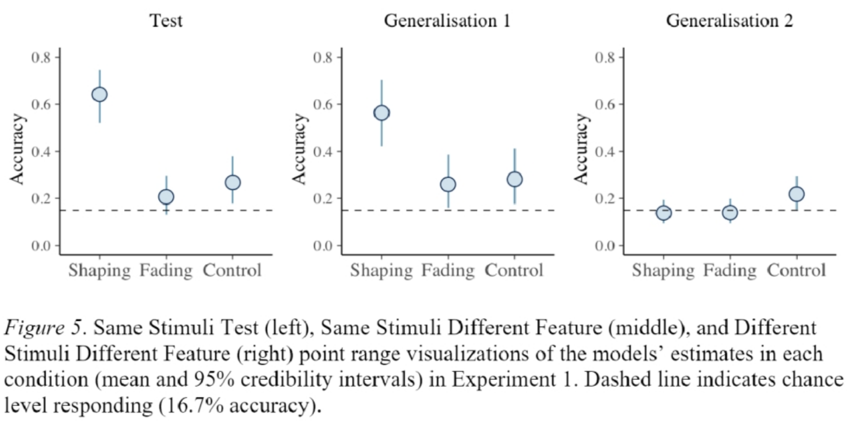

Figure 5. Same Stimuli Test (left), Same Stimuli Different Feature (middle), and Different Stimul Different Feature (right) point range visualizations of the model's estimates in each condition (mean and 95% credibility intervals) in Experiment 1. Dashed line indicates chance level responding (16.7% accuracy).

217x144mm (192 x 192 DPI)

DOI: 10.1177/17470218231210446

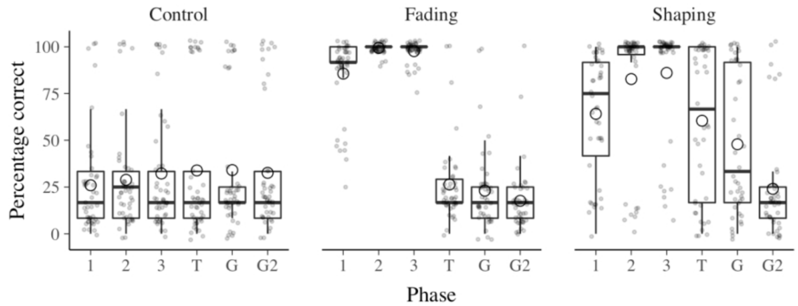

Figure 6. Participant responses during the training and test phases of Experiment 2 for each condition. Open circles indicate mean level. Horizontal lines indicate medians. Boxplot hinges show the inter-quartile range (IQR). Boxplot whiskers extend from the boxplot hinges to the largest (upper) and smallest (lower) observed value no further than 1.5 x the IQR. 1 = Phase 1, 2 = Phase 2, 3 = Phase 3, T = Same Stimuli Test, G = Generalization test 1, G2 = Generalization test 2.

Figure 6. Participant responses during the training and test phases of Experiment 2 for each condition. Open circles indicate mean level. Horizontal lines indicate medians. Boxplot hinges show the inter-quartile range (IQR). Boxplot whiskers extend from the boxplot hinges to the largest (upper) and smallest (lower) observed value no further than 1.5 x the IQR. 1 = Phase 1, 2 = Phase 2, 3 = Phase 3, T = Same Stimuli Test, G = Generalization test 1, G2 = Generalization text 2.

217x144mm (192 x 192 DPI)

DOI: 10.1177/17470218231210446

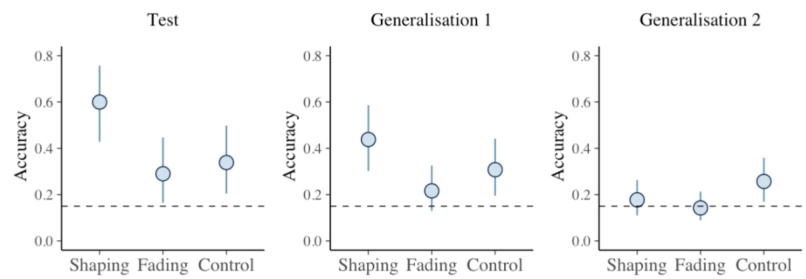

Figure 7. Same Stimuli Test (left), Same Stimuli Different Feature (middle), and Different Stimuli Different Feature (right) point range visualizations of the models' estimates in each condition (mean and 95% credibility intervals) in Experiment 2. Dashed line indicates chance level responding (16.7% accuracy).

Figure 7. Same Stimuli Test (left), Same Stimuli Different Feature (middle), and Different Stimuli Different Feature (right) point range visualizations of the models' estimates in each condition (mean and 95% credibility intervals) in Experiment 2. Dashed line indicates chance level responding (16.7% accuracy).

217x144mm (192 x 192 DPI)

DOI: 10.1177/17470218231210446

Table 1. *Posterior contrasts for experimental conditions in test phases of Experiment 1*

Experimental Phase	Contrast	Point estimate	$CI_{95\%}$ Lower	$CI_{95\%}$ Upper
Test	Shaping vs Control	36.3%	21.4%	50.4%
	Shaping vs Fading	44.1%	30.1%	56.5%
	Fading vs Control	-7.7%	-23.1%	7.6%
Generalization 1	Shaping vs Control	28.5%	11.1%	45.3%
	Shaping vs Fading	31%	13%	47.9%
	Fading vs Control	-2.5%	-20.9%	15.4%
Generalization 2	Shaping vs Control	-13.3%	-26.8%	1%
	Shaping vs Fading	0%	-13.8%	13.8%
	Fading vs Control	-13.1%	-26.8%	1%

DOI: 10.1177/17470218231210446

Table 2. *Posterior contrasts for experimental conditions in test phases of Experiment 2*

			$CI_{95\%}$	
Experimental Phase	Contrast	Point estimate	Lower	Upper
Test	Shaping vs Control	25.7%	5.2%	45.4%
	Shaping vs Fading	31%	11.2%	49.5%
	Fading vs Control	-5.3%	-24.8%	15.1%
Generalization 1	Shaping vs Control	13.7%	-5.5%	32.2%
	Shaping vs Fading	25.1%	6.2%	42%
	Fading vs Control	-11.5%	-29.9%	7.1%
Generalization 2	Shaping vs Control	-11.6%	-28.2%	4.8%
	Shaping vs Fading	6.2%	-10.9%	22.9%
	Fading vs Control	-17.8%	-33.6%	-1%

DOI: 10.1177/17470218231210446