Received: 23 August 2022 | Accepted: 1 April 2023

DOI: 10.1002/jeab.850

RESEARCH ARTICLE

Journal of the Experimental Analysis of Behavior SEAB

Revaluation of overselected stimuli: Emergence of control by underselected stimuli depends on degree of overselectivity

Stephanie Gomes-Ng[1,2] | **Peter B. C. Kim**[2] | **Sarah Cowie**[2] | **Douglas Elliffe**[2]

[1]Auckland University of Technology, Auckland, New Zealand

[2]The University of Auckland, Auckland, New Zealand

Correspondence
Stephanie Gomes-Ng, Department of Psychology, Auckland University of Technology, Private Bag 92006, Auckland, 1142, New Zealand.
Email: stef.gomes-ng@aut.ac.nz

Editor-in-Chief: Mark Galizio
Handling Editor: Manish Vaidya

Abstract

Stimulus overselectivity describes strong control by one stimulus element at the expense of other equally relevant elements. Research suggests that control by underselected stimuli emerges following extinction of the overselected stimulus ("revaluation") and the emergence is larger when overselectivity is greater. We compared such revaluation effects with a control compound or condition in two experiments. Human participants chose between compound S+ and S- stimuli. Then, to assess control by compound-stimulus elements, participants chose between individual elements in a testing phase without feedback. The S+ element chosen most often (the overselected element) underwent revaluation, during which choice of that element was extinguished and choice of a novel element reinforced. Thereafter, participants completed a retesting phase. Revaluation reduced choice of the overselected element. Choice of the underselected element decreased for participants with low overselectivity but increased for participants with high overselectivity. This was not the case for a control compound that did not undergo revaluation (Experiments 1 and 2) or in a control condition in which the overselected element continued to be reinforced during revaluation (Experiment 2). These findings suggest that overselectivity levels may modulate revaluation effects, and they also highlight the importance of the contingency change in post-revaluation changes in stimulus control.

KEYWORDS
compound stimulus, humans, retrospective revaluation, stimulus control, stimulus overselectivity

When multiple stimuli differentially signal reinforcer availability, stimulus control may be *divided* between the stimuli (e.g., Blough, 1969; Chase, 1968; Cowie et al., 2017; Davison, 2018; Davison & Elliffe, 2010; Shahan & Podlesnik, 2006, 2007). The presence of multiple differential stimuli is necessary but not sufficient for divided stimulus control to occur. Under some circumstances, behavior may be controlled by a single stimulus or small subset of stimuli at the expense of control by other equally relevant stimuli (Lovaas et al., 1971; Ploog, 2010). Such *stimulus overselectivity* can lead to difficulties with acquisition, maintenance, and/or generalization of discrimination learning. Overselectivity has been observed in individuals with autism spectrum disorder (ASD; e.g., Dube et al., 2010; Kelly et al., 2015; Leader et al., 2009; see Ploog, 2010, for a review) and also in preschool-aged children (Kelly & Reed, 2021), older adults aged 60 to 89 (Kelly et al., 2016; McHugh & Reed, 2007; McHugh et al., 2010), and typically developing humans under high cognitive load (Broomfield et al., 2010; Reed & Gibson, 2005; Reed et al., 2011; Reed, Altweck, et al., 2012).

Two mechanisms have been proposed to underlie stimulus overselectivity. The *attention-deficit* view posits that overselectivity occurs due to a failure to observe and/or attend to all stimuli. As a result, only those stimuli that are observed and attended to control behavior (Dube, 2009; Dube & McIlvane, 1999; Dube et al., 1999, 2003, 2010; Lovaas et al., 1971; Reed, 2019). In support of this, eye-tracking data from individuals with intellectual disabilities suggest that overselected stimuli

(i.e., those exerting selective control) are observed more often and for longer durations than underselected stimuli (i.e., those exerting little to no control). Additionally, interventions that differentially reinforce observing of underselected stimuli appear to remediate overselectivity, although their effects may not necessarily maintain over time (Dube & McIlvane, 1999; Dube et al., 2003, 2010; Farber et al., 2017; Walpole et al., 2007; but see also Reed, Altweck, et al., 2012).

Alternatively, *comparator* theories suggest that stimulus overselectivity arises due to a performance deficit. According to this view, participants attend to and learn about all stimuli, but such learning is not always expressed behaviorally due to competition for control between stimuli. Such competition occurs via a "comparator" mechanism, which effectively selects one stimulus, or a subset of stimuli, to control behavior (Reed, 2011; see also Miller & Matzel, 1988). Thus, behavioral control by the selected stimulus or stimuli is apparent, whereas control by other stimuli is masked. The implication of this explanation of stimulus overselectivity is that control by underselected stimuli may *emerge* under suitable conditions.

Findings from studies of cue competition in associative learning provide evidence for such emergence (see Miller & Witnauer, 2016, for a review). Following conditioning with a compound cue (e.g., AX), the element supporting a stronger conditioned response (e.g., Cue A) is "revalued" by placing it on extinction, resulting in an *increase* in responding to the other cue (e.g., Cue X). The magnitude of this *retrospective revaluation* effect depends on several parameters. For example, longer durations of compound-cue exposure or of revaluation produce a stronger effect (e.g., Blaisdell et al., 1999; Larkin et al., 1998; Matzel et al., 1985). The physical or functional similarity between cues also appears to determine the direction of post-revaluation change; more distinctive cues produce the revaluation effect, whereas more similar cues result in reduced responding to all cues (*mediated extinction*; e.g., Balleine et al., 2005; Liljeholm & Balleine, 2009).These retrospective revaluation effects are thought to arise due to within-compound stimulus associations; associations between overselected and underselected cues form during discrimination training, so when the overselected cue is extinguished in revaluation, a representation of the underselected cue may be "activated" associatively, resulting in changes in control by that cue (e.g., Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994; see Miller & Witnauer, 2016, for a brief overview of associative accounts of retrospective revaluation).

Revaluation effects have also been observed in studies of stimulus overselectivity, demonstrating that these effects are not isolated to associative-learning paradigms and providing support for comparator theories of overselectivity (e.g., Broomfield et al., 2010; Kelly et al., 2015; Reed et al., 2009; Reed, Reynolds, et al., 2012; Reynolds & Reed, 2011a, 2018; Reynolds et al., 2012). In these studies, participants learn to choose a compound stimulus, AB, over another compound, CD. Then, in a testing phase, the overselected and underselected elements of the AB compound are identified by presenting one element from each compound (e.g., A vs. C) and recording the frequencies with which participants choose each element. The element, A or B, that was chosen more frequently by each participant (the overselected element) undergoes revaluation training, during which choice of that element is extinguished and participants are reinforced for choosing a novel element instead. This is followed by another testing phase, to determine how revaluation affected control by elements A and B. Typically, after revaluation, control by the overselected element decreases and control by the underselected element emerges. No such systematic changes are evident for a control compound whose elements do not undergo revaluation.

Thus, stimulus overselectivity may be remediated by extinguishing choice of the overselected stimulus, which results in an emergence of control by underselected stimuli. This revaluation effect is stronger when revaluation training lasts for longer and when a larger number of novel stimuli is used in revaluation training (Reed, Reynolds, et al., 2012). In addition to procedural factors, another variable that may be important is baseline levels of overselectivity. That is, the effects of revaluation on control by overselected and underselected elements may differ depending on whether initial control by these elements is more unequal (i.e., overselectivity levels are higher) or more equal (i.e., overselectivity levels are lower). Indeed, Broomfield et al. (2010) found that participants displaying greater overselectivity (i.e., larger differences in the strength of control by the overselected and underselected elements) showed a larger reduction in control by the overselected element and greater emergence of control by the underselected element than participants with lower overselectivity).

Given that revaluation training may be used to remediate overselectivity, it is important to consider how such training may differentially influence participants with higher or lower levels of overselectivity as well as the mechanisms that may underlie any such differences. Presently, little is known about how levels of overselectivity affect revaluation effects. Only Broomfield et al. (2010) have analyzed data separately for participants showing lower and higher levels of overselectivity; other studies aggregated data across participants, thus masking any potential differences based on overselectivity level. As a result, the replicability and generality of the graded effects of revaluation training reported by Broomfield et al. are presently unclear. Furthermore, Broomfield et al. did not include a control compound or condition against which to compare the effects of revaluation. Thus, the extent to which post-revaluation changes in control represented

real effects arising from the contingency change (extinction of the overselected element) during revaluation is unclear. This is particularly applicable to participants displaying lower overselectivity, for whom changes in control were smaller and thus may have reflected regression to the mean effects rather than genuine changes. Although other studies have found little systematic change for a control compound that does not undergo revaluation (e.g., Reed et al., 2009; Reed, Reynolds, et al., 2012)—providing some evidence that post-revaluation changes in control are not simply due to regression to the mean—it is unclear whether this is true across different levels of overselectivity.

The present study investigated further the effects of overselectivity level on revaluation effects in a systematic replication of Broomfield et al.'s (2010). In Experiment 1, neurotypical participants chose between compound stimuli during discrimination training, and then choice of the overselected element was extinguished and choice of a novel element reinforced in revaluation training. This occurred under high cognitive load to induce overselectivity (see Broomfield et al., 2010; Reed & Gibson, 2005). Unlike Broomfield et al., we also included a control compound whose elements did not undergo revaluation. Thus, we compared changes in stimulus control between the revalued and control compounds separately for participants displaying different levels of overselectivity. Experiment 2 replicated Experiment 1 with the addition of a novel control condition in which choice of the overselected element continued to be reinforced. This allowed us to compare stimulus control in the typical revaluation procedure with a control procedure in which no change in contingency occurred.

EXPERIMENT 1

Method

Participants

Fifty-nine participants without a diagnosis of ASD were recruited, 44 through advertisement posters placed around a large university campus in New Zealand and 15 through the university's SONA research participation system. The exclusion of participants with ASD replicates Broomfield et al. (2010) and other recent studies that have investigated overselectivity in nonclinical populations (e.g., Reed, Reynolds, et al., 2012; Reynolds & Reed, 2018) and was further justified because some research suggests that overselectivity is more prevalent and may be governed by different processes in ASD compared with nonclinical populations (e.g., Kelly et al., 2015; Leader et al., 2009; Reed, 2019). Participants were fluent in English (as per the university's admissions criteria, and this was also evident to the experimenter while providing verbal instructions to participants before the session).

Participants recruited through advertisements were recruited in two phases. In the first phase, which ran for the duration of one academic semester, participants were compensated with entries into a drawing to win a $250 shopping voucher ($N = 33$). In the second phase, participants were compensated with a $20 petrol voucher after the session ($N = 11$). Participants recruited through SONA were also recruited during the second phase and compensated with experiential learning course credit. Preliminary analyses indicated that these different recruitment and compensation methods had no effect on participants' performance (see Supplementary Figure 1).

The ages of 43 of the participants ranged from 17 to 38 years ($M = 22.34$ years, $SD = 4.32$); the remaining 16 participants declined to provide age data. At the beginning of the experiment, a computer program randomly assigned each participant a unique four-digit identifier ranging from 1000 to 9999 and also randomly assigned participants to Group 1 or Group 2. The group number determined which stimuli were used during the experiment.

All participants completed the Autism Quotient (AQ) Questionnaire (Baron-Cohen et al., 2001) as part of the experiment. The AQ is a questionnaire measuring symptoms of ASD in the general population. Participants with scores of 32 or greater ($N = 4$) were excluded from analyses because such scores suggest clinically significant symptoms of ASD (see e.g., Kelly et al., 2015; Reed, 2011, 2019; Reed & Gibson, 2005, for further discussion).

Apparatus and materials

The experiment was conducted in a quiet room located at the university. Participants were seated at a table facing a Dell P2214H computer monitor with a screen size of 22 inches (~66 cm, 1920 × 1080 desktop resolution), which was connected to a computer running Windows 10. The experiment was programmed using the Pygame 1.9.6 package in Python 3.7. This program recorded all experimental events and the time at which each event occurred. The monitor background remained white throughout the experiment.

Figure 1 shows the stimuli used in the experiment. Stimuli were 20 abstract symbols from the fonts Wingdings, Wingdings 2, Stylebats, and Larasukma. They were separated into two groups, one comprising Stimuli 1 to 10 (Group 1) and the other comprising Stimuli 11 to 20 (Group 2). Eight stimuli from one group (Group 1 for participants in Group 1 and Group 2 for participants in Group 2) were randomly selected by the computer as the training stimuli (Stimuli A to H), and four stimuli from the other group were randomly selected as novel stimuli (Stimuli I to L) for the revaluation-training phase. Supplementary Table 1 provides details of the training and novel stimuli for each participant.

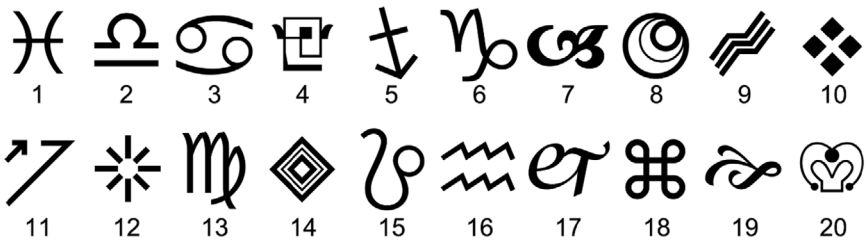

FIGURE 1 Stimuli used in the present experiment. Group 1 consisted of Stimuli 1 to 10, and Group 2 consisted of Stimuli 11 to 20.

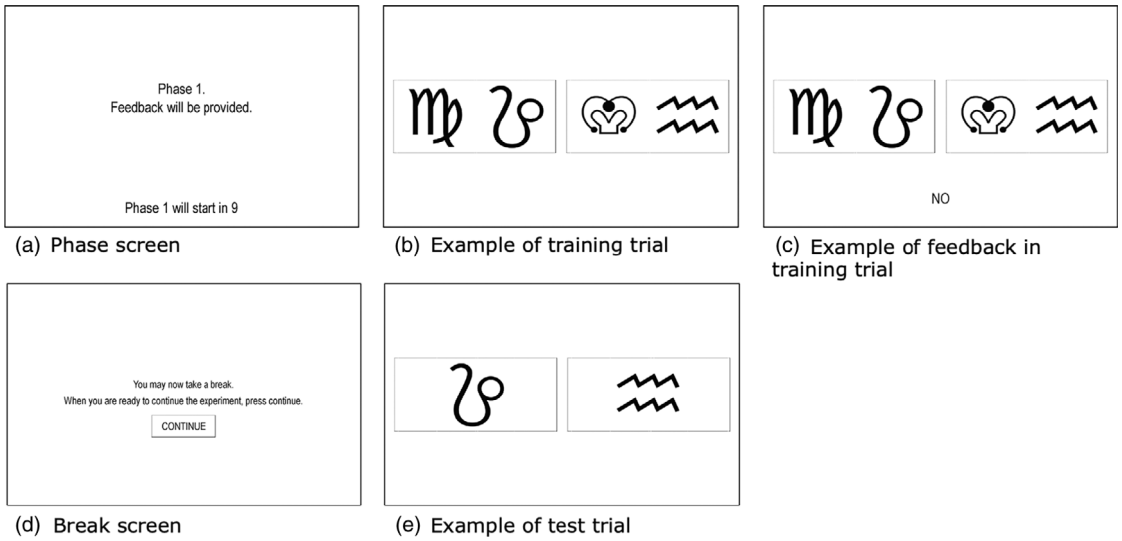

FIGURE 2 Screenshots of the experimental program. (a) Beginning of each phase, (b) Example of training trials, (c) Example of feedback in training trial, (d) Break screen, and (e) Example of test trial.

Procedure

Instructions to participants

Before beginning the experiment, participants were informed (by the first or second author) that they would be completing a computer task consisting of four phases, that in each phase they would choose between symbols presented in rectangles on the computer screen, and that choices would sometimes result in feedback from the computer. Participants were told that their goal in the experiment was to make correct choices. Additionally, they were instructed to perform a concurrent-load task in which they counted backwards in sevens starting from a five-digit number throughout the experiment.

A general script of the verbal instructions provided to participants is in the Supplementary Material. Participants were asked to verify verbally that they understood the instructions. If necessary or requested by the participant, the experimenter repeated the instructions to provide clarification. After participants verified that they understood the instructions, the experimental session began.

Experimental sessions

Each participant completed one experimental session, which was split into four phases: training (Phase 1), testing (Phase 2), revaluation (Phase 3), and retesting (Phase 4). Each phase began with a screen informing participants of the number of the phase they were in and whether the computer would provide feedback (Figure 2a). A countdown at the bottom of the screen informed participants of the time (in seconds) until the start of the phase. The countdown timer started at 30 s in Phase 1 and at 10 s for the other phases. Throughout the session, participants simultaneously completed a verbal-subtraction concurrent-load task. The experimenter remained seated in a corner of the experiment room to ensure that participants completed the concurrent-load task.

Training

During training, participants were presented with two compound stimuli, either AB and CD or EF and GH, in each trial (Figure 2b). Each compound stimulus comprised two symbols presented side by side within a containing rectangle. Symbols AB and EF were the S+ compound stimuli, and CD and GH were the

S- compound stimuli. The positions of the S+ and S- stimuli were counterbalanced across trials, and trials occurred in a random order. Participants selected a compound stimulus by clicking anywhere inside its containing rectangle using the computer mouse. A response to the S + compound was defined as correct and resulted in positive feedback ("YES" presented on the bottom of the screen), whereas a response to the S- compound was defined as incorrect and resulted in negative feedback ("NO" presented on the bottom of the screen; Figure 2c). Feedback lasted for 1 s and was followed by a 500-ms intertrial interval during which the screen was blank.

To minimize participant fatigue, the program allowed participants to take a break every 13 trials. Participants could resume the experiment at any time by clicking a "CONTINUE" button (Figure 2d). Participants took an average of 2.24 breaks ($SEM = 0.29$) lasting an average of 0.33 min ($SEM = 0.04$) during training. Training continued until 10 consecutive correct responses were made for each pair of compound stimuli. This criterion was based on previous published work on overselectivity (e.g., Broomfield et al., 2010; Kelly et al., 2015; Reed et al., 2009; Reed, Reynolds, et al., 2012; Reynolds & Reed, 2011a, 2018; Reynolds et al., 2012). If participants did not reach the criterion within 45 min, the experiment terminated.

Testing

Test trials were similar to training trials, except that only one element of each compound stimulus was presented (Figure 2e). Participants were given a choice between one of the S+ elements from training (A, B, E, or F) and one of the S- elements from training (C, D, G, or H). The eight pairs of elements used in test trials were A vs. C, A vs. D, B vs. C, B vs. D, E vs. G, E vs. H, F vs. G, and F vs. H. Five of each pair were presented in a random order, and the locations of the S+ and S- elements were counterbalanced across trials. No feedback or breaks occurred during the testing phase.

Revaluation

The computer program randomly selected one of the S+ compound stimuli from training (i.e., AB or EF) and identified the element of that compound that was selected more often during the testing phase. This element was the overselected S+ element for revaluation training. If participants chose both elements of the selected compound equally often during testing, the program evaluated the other compound stimulus and used the most selected element from that compound for revaluation. If the elements of the other compound were also selected equally often (i.e., the participant displayed no overselectivity at all), the program randomly selected one element from either S+ compound ($N = 9$ participants did not show any overselectivity). This process ensured that whenever possible, the element selected for revaluation was indeed an overselected element.

Revaluation was similar to training except that participants were presented with a choice between two single-element stimuli: the overselected element selected for revaluation and one of four novel stimuli (Stimuli I to L). The locations of the stimuli were counterbalanced across trials. Choice of the novel stimulus was defined as correct and resulted in positive feedback, whereas choice of the overselected element was incorrect and resulted in negative feedback. The program allowed participants to take a break every 13 trials. Participants took an average of 4.08 breaks ($SEM = 0.36$) lasting an average of 0.29 min ($SEM = 0.04$) in the revaluation phase. Thus, overall, breaks were short (typically, less than a minute). The greater number of breaks in the revaluation phase relative to training reflects the larger number of stimulus pairs (four, as there were four novel stimuli, I, J, K, and L) in this phase. As in training, revaluation continued until 10 consecutive correct responses were made for each overselected/novel stimulus pair.

Retesting

The retesting phase was identical to the testing phase.

Concurrent-load task

In typically developing individuals, overselectivity is more likely under conditions of high cognitive load (Reed & Gibson, 2005). Thus, to increase the likelihood of overselectivity occurring, participants completed a concurrent-load task at the same time as the computer task. This occurred throughout the entire experimental session. Participants were given a randomly generated five-digit number below 20,000 and were instructed to count backwards verbally in sevens throughout the session, excluding breaks and in between phases. Participants were verbally prompted by the experimenter ("Please continue the subtraction task") if they paused for more than 7 s. All participants completed the subtraction task, with no obvious attempts to avoid the task (e.g., by consistently reporting incorrect numbers or by pausing for long periods).

Preliminary data analysis

Supplementary Table 1 shows data for individual participants. In addition to the four participants with AQ scores of 32 or greater whose data were excluded, five participants did not meet the criterion of 10 successive correct responses within 45 min during the training phase (i.e., these participants never progressed to the testing, revaluation, and retesting phases), so their data were excluded from analyses. Thus, the final sample size consisted of 50 participants. For the 50 participants who completed all four phases, the mean times to complete training, testing, revaluation, and retesting were 5.93 ($SEM = 0.89$), 3.97 ($SEM = 0.22$), 7.01 ($SEM = 0.85$), and 3.50 min ($SEM = 0.18$), respectively. The mean total

time to complete the session was 20.42 min ($SEM = 1.63$).

Because we counterbalanced stimuli across participants (see Supplementary Table 1), stimuli were grouped according to whether they were the overselected (i.e., more selected) or underselected (i.e., less selected) element of the revalued compound (i.e., the compound whose overselected element underwent revaluation training) or of the other, control, compound. We then used responses during the testing and retesting phases to calculate the percentage of trials in which each stimulus element was chosen.

In the analyses that follow, we used *JASP* (JASP Team, 2018) to conduct Bayesian hypothesis tests (e.g., to compare choice between elements, see Wagenmakers, Love, et al., 2018; Wagenmakers, Marsman, et al., 2018). Bayesian tests are advantageous because they quantify the strength of the evidence for the null or alternative hypothesis with a Bayes factor (BF_{10}) rather than just providing a p value. The Bayes factor reflects the likelihood of H_1 (e.g., differences between elements or groups) relative to H_0 (e.g., no differences between elements or groups) given the obtained data. Additionally, given our relatively small sample size, Bayesian analyses provide greater confidence in our analytic results. We interpreted the Bayes factor using Lee and Wagenmakers's (2013) classification. Specifically, a Bayes factor of 1 indicates no evidence for either hypothesis. A Bayes factor greater than 1 indicates stronger support for the alternative hypothesis, with values above 10 indicating strong support and above 100 indicating extreme support. Similarly, a Bayes factor less than 1 indicates stronger support for the null hypothesis, with values less than 0.1 or 0.001 indicating strong or extreme support, respectively, for the null hypothesis.

We ran preliminary analyses to determine whether there were systematic differences between participants recruited using different methods (advertisements or SONA) and compensated differently (entries into a drawing to win a shopping voucher, a petrol voucher, or experiential learning course credit). Participants were split into three groups based on recruitment method and compensation, and we used one-way between-groups Bayesian analyses of variance (ANOVAs) to compare percentage of choice of each element during pre-revaluation testing (Phase 2) and the percentage of change in choice from testing to retesting. Supplementary Figure 1 shows percentage of choice for each recruitment group, and Supplementary Table 2 shows the results of the Bayesian ANOVAs. The Bayes factors ranged from 0.18 to 1.44, indicating little to no support for differences between participants recruited and compensated differently. Therefore, we combined data across recruitment and compensation methods for all subsequent analyses.

Results

Mean trials to reach criterion

The mean number of trials to reach the criterion of 10 consecutive correct responses in training was 17.97 ($SEM = 2.04$), averaged across participants and both training stimulus pairs (AB vs. CD and EF vs. GH). In the revaluation phase, the mean number of trials to reach criterion was 14.70 ($SEM = 1.40$).

Stimulus control before revaluation

The leftmost two panels in Figure 3 show the mean (averaged across participants) percentage of trials in which each S+ element of the revalued and control compounds was chosen during the testing (*pre*-revaluation, Phase 2) and retesting (*post*-revaluation, Phase 4) phases. Of the 50 participants, 41 demonstrated some degree of overselectivity during pre-revaluation testing. The remaining nine participants chose both elements of both compounds in 100% of test trials, so no overselectivity was apparent for these participants. The size of the difference in choice between the overselected and underselected elements was larger for the revalued compound than for the control compound due to the method used to select the element to revalue (which ensured that whenever possible the selected element was from a compound for which overselectivity was apparent). A Bayesian repeated-measures ANOVA on pre-revaluation percentages provided extreme support for differences in percentage of choice between elements ($BF_{10} = 1.05 \times 10^{11}$, error $\% = 0.96$). Post hoc comparisons indicated extreme support for differences in choice between the overselected and underselected elements of the revalued compound ($BF_{10} = 7.01 \times 10^{6}$, error $\% < 10^{-11}$) and the control compound ($BF_{10} = 2{,}616.95$, error $\% < 10^{-5}$).

To determine whether levels of overselectivity for the revalued compound were significantly different from chance—that is, from the level of overselectivity that would be expected if both stimulus elements had the same probability of being chosen—we conducted an analysis based on the binomial distribution (Reynolds & Reed, 2011a, 2011b; Reed, Reynolds, et al., 2012). First, we calculated the mean probability, p, of choosing the overselected and underselected elements of the revalued compound during testing for each participant. Using this obtained probability, we calculated the probability of choosing the overselected element x times in 10 trials and the underselected element y times in 10 trials, $P(X \cap Y)$, where $X \sim Binomial(x_{\in [0,\ 10]},\ p)$ and $Y \sim Binomial(y_{\in [0,\ 10]},\ p)$. These binomial probabilities were entered into an 11×11 contingency table, and the contents of this table were multiplied by another 11×11 contingency table containing the absolute $|x - y|$ difference for each combination of x and y.

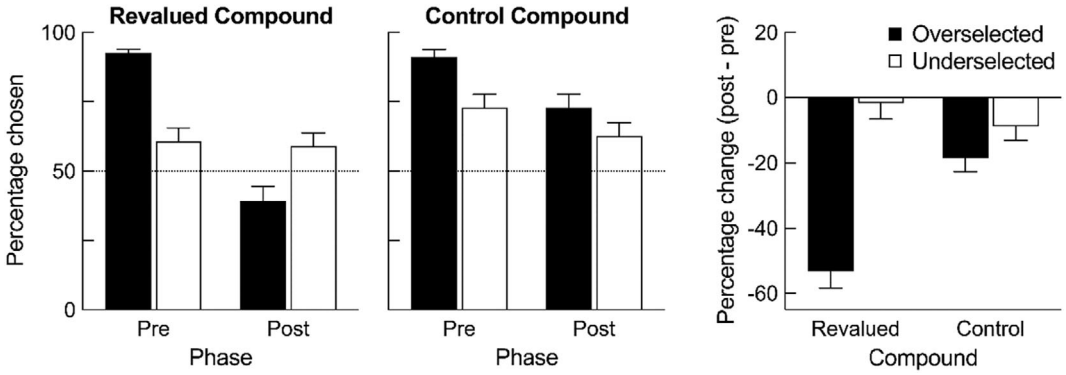

FIGURE 3 Mean percentage choice of each S+ element before and after revaluation training. The left and middle panels show mean (averaged across participants) percentage of choice for the revalued and control compounds, respectively, before (pre) and after (post) revaluation training. The right panel shows the percentage of change in choice pre- to post-revaluation. The error bars represent the *SEM*.

Summing all cells of the resulting table and then multiplying by 10 produced the expected difference between percentage of choice of the overselected and underselected elements for the revalued compound.

The mean (averaged across participants) expected difference between choice of the overselected and underselected elements of the revalued compound was 11.30% ($SEM = 0.94$). A Bayesian one-tailed paired-samples *t* test indicated substantial support for the alternative hypothesis that obtained differences were greater than expected differences ($BF_{10} = 153{,}385.90$). Thus, levels of overselectivity for the revalued compound were higher than the level expected by random variation of selection of the overselected and underselected elements. This means that the concurrent-load task successfully induced overselectivity.

Effects of revaluation on stimulus control

Differences between choice in the testing phase (Phase 2) and retesting phase (Phase 4) provide insight into the effects of revaluation on control by the overselected and underselected elements of both compound stimuli. The right panel of Figure 3 shows the mean percentage of change in choice, calculated by subtracting percentage of choice before revaluation from percentage of choice after revaluation. We used Bayesian two-tailed paired-samples *t* tests to compare pre- and post-revaluation choice.

For the revalued compound, choice of the overselected element decreased for almost all participants after revaluation training (strong support for a difference; $BF_{10} = 9.80 \times 10^{10}$, error % $< 10^{-13}$), whereas there was little systematic change across all participants for the underselected element (moderate support for no difference; $BF_{10} = 0.17$, error % = 0.07). For the control compound, choice of both elements was lower after revaluation; there was strong support for this difference for the overselected element ($BF_{10} = 254.80$, error % $< 10^{-9}$), whereas support for a difference was weak and anecdotal for the underselected element ($BF_{10} = 1.66$, error % = 0.02). Thus, the main effect of revaluation training was to decrease choice of the overselected element of both compounds and to a greater extent for the revalued compound (Figure 3).

Relation between overselectivity levels and postrevaluation choice

One potential explanation for why revaluation training appeared to have little effect on choice of the underselected element when the data were combined across all participants (Figure 3) is that the effects of revaluation may have differed for participants with different levels of overselectivity (Broomfield et al., 2010). Therefore, we examined the relation between initial levels of overselectivity and changes in choice after revaluation training. We calculated the level of overselectivity for each participant as the percentage of difference in choice between the overselected and underselected elements of the revalued compound during pre-revaluation testing (Phase 2). We then grouped participants into three groups. Participants displaying no overselectivity (i.e., a 0% difference) were grouped together (Group None: $N = 9$). For the remaining participants, all of whom showed some degree of overselectivity, we followed Broomfield et al. (2010) and calculated the mean level of overselectivity (38% difference) and split the sample at this mean. Therefore, participants showing overselectivity below the mean level were grouped together (Group Low: $N = 20$) and participants showing overselectivity above the mean level were grouped together (Group High: $N = 21$).

As suggested by a reviewer, we explored potential differences between participants with different overselectivity levels by examining correlations between overselectivity level, or percentage of change in choice, and age, AQ score, total session time, and the number of trials or breaks in the training and revaluation phases. Bayes factors ranged from 0.21 to 1.30, indicating that

there was little evidence for a correlation between overselectivity levels or percentage of change in choice and any of these participant variables. Thus, participants' overselectivity levels and post-revaluation changes in choice appeared not to relate to age, AQ score, or overall performance measures or learning-history differences (e.g., session time, number of trials to criterion). Figure 4 shows the mean percentage of change in choice from testing (Phase 2, pre-revaluation) to retesting (Phase 4, post-revaluation) separately for participants displaying no, low, or high overselectivity. A Bayesian mixed ANOVA with element as a within-subjects factor and overselectivity level as a between-subjects factor indicated that the model with the strongest support included element, overselectivity level, and the interaction term ($BF_{10} = 3.21 \times 10^{19}$, error % = 2.44). To probe the interaction, we conducted one-way ANOVAs comparing groups for each element. Table 1 summarizes the results

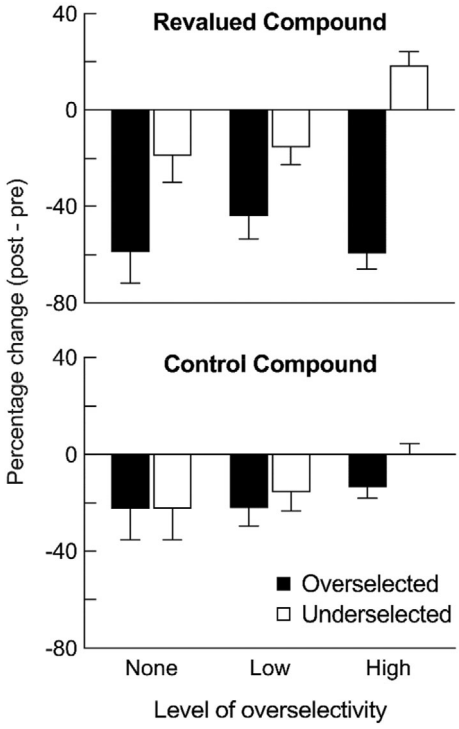

FIGURE 4 Mean percentage of change in choice pre- to post-revaluation for participants displaying no, low, or high overselectivity in Experiment 1. The error bars represent the *SEM*.

TABLE 1 Results of Bayesian ANOVAs comparing participants with no, low, and high overselectivity levels

	Overselected element		Underselected element	
Compound	**BF_{10}**	**Error %**	**BF_{10}**	**Error %**
Revalued	0.35	0.04	79.59*	0.01
Control	0.23	0.03	0.73	0.03

*indicates strong support for differences depending on overselectivity level.

of these ANOVAs. The only element for which there was support for differences depending on overselectivity level was the underselected element of the revalued compound. For this element, choice decreased after revaluation for participants with no or low overselectivity, whereas it increased for participants with high overselectivity (Figure 4). Post hoc comparisons indicated strong support for differences between participants with high overselectivity and those with no or low overselectivity ($BF_{10} = 15.59$ and 75.52, respectively, both error % $< 10^{-5}$), whereas there was no support for differences between participants with no or low overselectivity ($BF_{10} = 0.38$, error % = 0.002).

Thus, in summary, the main effect of overselectivity level was to modulate the effects of revaluation training on choice of the underselected element of the revalued compound. Choice of that element slightly decreased, which may suggest a slight decrease in control, for participants showing little to no overselectivity. In contrast, choice of the underselected element of the revalued compound increased, indicating an emergence of control, for participants showing high overselectivity. These differences between participants with lower and higher overselectivity levels explain why revaluation appeared to have little effect on choice of the underselected element of the revalued compound when the data were aggregated across all participants (Figure 3).

Discussion

The present experiment replicated and extended previous research examining the effects of postlearning "revaluation" training on control by overselected and underselected stimuli. Specifically, we investigated the relation between initial levels of overselectivity and the effects of revaluation on stimulus control in typically developing humans. Stimulus overselectivity was successfully induced in this population using a concurrent-load task. This adds to the growing body of literature demonstrating that overselectivity occurs in humans without a diagnosis of ASD under high cognitive load (e.g., Broomfield et al., 2010; Reed & Gibson, 2005; Reed, Reynolds, et al., 2012; Reynolds & Reed, 2011a, 2011b, 2018). Revaluation training with the overselected element of the revalued compound appeared to have little effect on control by the underselected element when the data were aggregated across all participants (Figure 3), but this was because the effects of revaluation differed depending on initial overselectivity levels (Figure 4): Choice of the underselected element of the revalued compound *decreased* slightly for participants with no or low overselectivity, whereas it *increased* for those with high overselectivity. Thus, extinguishing overselected stimuli produced changes in control by underselected stimuli, in line with comparator theories of overselectivity.

However, in contrast to previous research, the revaluation effect (emergence of control by the underselected element) was only observed for participants with high overselectivity.

The present results partially replicate Broomfield et al. (2010); we found post-revaluation differences in stimulus control between participants displaying lower and higher levels of overselectivity. However, unlike Broomfield et al., we found differences in the *direction* of change in choice of the underselected element for participants with lower and higher overselectivity. Specifically, for participants with *lower* overselectivity, choice of the underselected element decreased in our experiment but increased in Broomfield et al.'s. The reasons for the difference are unclear, as few studies have examined the relation between overselectivity levels and post-revaluation changes in stimulus control.

One possibility is that the difference between our findings and Broomfield et al.'s (2010) is more apparent than real. Unlike Broomfield et al., we included a control compound that did not undergo revaluation, allowing us to assess more thoroughly the extent to which post-revaluation changes in choice were directly related to revaluation. For participants displaying no or low overselectivity, changes in choice were similar for the revalued and control compounds, particularly for the underselected elements (Figure 4). This may suggest that at least a portion of the post-revaluation changes for participants with lower overselectivity may be artifactual. That is, they may represent regression to the mean effects rather than real changes caused by revaluation. It is possible that the same was true in Broomfield et al.'s experiment, as changes in control by the underselected element were small for their participants with low overselectivity. If so, this would reduce the apparent difference between the present findings and Broomfield et al.'s findings given that differences were mainly observed for participants with low overselectivity. Note that this line of reasoning does not invalidate the general conclusion that the effects of revaluation depend on initial overselectivity levels. Instead, it may be the case that extinguishing overselected stimuli only results in the emergence of control by underselected stimuli if initial overselectivity is substantial (here, when the percentage of difference in choice between the overselected and underselected elements was at least 40%).

One question that remains unanswered is which aspect(s) of revaluation training were responsible for the post-revaluation changes in stimulus control for participants with high overselectivity. It is generally thought that the change in contingency for the overselected element of the revalued compound is responsible for the emergence of control by the underselected element (Broomfield et al., 2010; Kelly et al., 2015; Leader et al., 2009; Reed et al., 2009; Reed, Reynolds, et al., 2012; Reynolds et al., 2012). However, participants also experience additional discrimination training with, and thus exposure to, the overselected element of the revalued compound. This is not the case for elements of the control compound. As a result, it is unclear how these procedural factors may contribute to post-revaluation changes in choice by underselected elements. That is, differences in results between the revalued and control compounds (e.g., Figures 3 and 4) may be related to the change in contingency, additional discrimination training, and/or exposure to the overselected element of the revalued compound. To explore which of these may contribute to post-revaluation changes in choice as well as to investigate whether changes in choice after revaluation were artifactual, we conducted a second experiment which directly assessed the role of the contingency change.

EXPERIMENT 2

The purpose of Experiment 2 was to investigate the extent to which extinction of the overselected element contributed to post-revaluation changes in choice. To that end, we replicated Experiment 1, with the addition of a control condition in which choice of the overselected element continued to be reinforced during the revaluation phase. That is, no contingency change occurred in the control condition, but participants still experienced additional discrimination training with and exposure to the overselected element. To the extent that the contingency change is responsible for changes in choice of the underselected element after revaluation, no systematic changes in choice should be apparent for participants in the control condition.

Including the control condition also provided another opportunity to investigate whether post-revaluation changes in choice are real or artifactual for participants displaying little to no overselectivity. If such changes in choice in Experiment 1 were indeed artifactual, then similar changes should be apparent for participants with lower overselectivity in the control condition of Experiment 2. In other words, if regression to the mean effects were responsible for the post-revaluation decreases in choice for participants with lower overselectivity in Experiment 1 (Figure 4), similar regression to the mean should occur in Experiment 2. In contrast, if there are differences between the results of the revaluation (i.e., replication of Experiment 1) and control conditions, this would suggest that post-revaluation changes in choice may represent real effects arising due to extinction of the overselected element.

Method

Participants

Thirty-one participants without a diagnosis of ASD were recruited. Five participants were recruited through

advertisements around a large university campus and compensated with a $20 petrol voucher after the session. The remaining 26 participants were recruited through the university's SONA research participation system and compensated with experiential learning course credit. The ages of 30 of the participants ranged from 19 to 58 years ($M = 22.20$ years, $SD = 7.13$).

At the beginning of the experiment, the computer program randomly assigned each participant a unique four-digit identifier ranging from 1000 to 9999 and also randomly assigned participants to a group (Group 1 or 2) and condition (Condition 1 or 2). Fifteen participants were allocated to Condition 1 (*revalue* condition), and 16 participants to Condition 2 (*control* condition). Four participants (two from each condition) had AQ scores of 32 or greater, so their data were excluded.

Apparatus, materials, and procedure

The same apparatus, materials, and procedure as Experiment 1 were used, except for the contingencies arranged during Phase 3 in Condition 2 (the control condition). In the control condition, choice of the overselected element was followed by positive feedback and choice of the novel element was followed by negative feedback in the third phase. Participants in both conditions were given the same instructions as in Experiment 1 by either the first author or a student researcher trained and supervised by the first author. For consistency, we continue to refer to Phase 3 as the revaluation phase here, even though no change in contingency occurred in the control condition. Supplementary Table 3 provides details of the training and novel stimuli used for individual participants in Experiment 2.

Participants in the revalue condition took an average of 5.62 min ($SEM = 0.64$) to complete training, 3.69 min ($SEM = 0.36$) to complete testing, 4.97 min ($SEM = 0.57$) to complete revaluation, and 3.30 min ($SEM = 0.27$) to complete retesting. In this condition, participants took an average of 3.0 breaks ($SEM = 0.41$) in training and 3.46 breaks ($SEM = 0.18$) in revaluation, with a mean break duration of 0.21 min ($SEM = 0.04$) in training and 0.14 min ($SEM = 0.04$) in revaluation. In the control condition, the mean time to complete the training, testing, revaluation, and retesting phases was 5.71 ($SEM = 0.96$), 4.02 ($SEM = 0.50$), 5.01 ($SEM = 0.84$), and 3.17 min ($SEM = 0.33$), respectively. The mean number of breaks was 2.5 ($SEM = 0.49$) in training and 3.14 ($SEM = 0.10$) in revaluation, with a mean break duration of 0.23 ($SEM = 0.06$) and 0.24 min ($SEM = 0.07$), respectively. The total average time to complete the session was 17.59 min ($SEM = 1.50$) in the revalue condition and 17.92 min ($SEM = 2.34$) in the control condition. Thus, the average time to complete the session was similar, and breaks were short and infrequent in both conditions.

Results

Mean trials to reach criterion

The mean number of trials, averaged across participants and both training stimulus pairs, to reach the criterion of 10 consecutive correct responses in training was 22.04 ($SEM = 2.79$) in the revalue condition (Condition 1) and 19.36 ($SEM = 2.35$) in the control condition (Condition 2). There was moderate support for no difference in trials to criterion between conditions during the training phase (independent samples t test: $BF_{10} = 0.42$, error % = 0.002). In the revaluation phase (Phase 3), the mean number of trials to reach criterion was 12.29 ($SEM = 0.54$) in the revalue condition and 10.88 ($SEM = 0.41$) in the control condition. There was only anecdotal evidence for a difference in trials to criterion between conditions in the revaluation phase ($BF_{10} = 1.61$, error % = 0.01). Therefore, overall, the number of trials to reach criterion differed little between participants in the revalue and control conditions.

Stimulus control before and after revaluation

In this experiment, the first two phases (training and testing) were identical for participants in the revalue and control conditions. Thus, we expected participants in both conditions to behave similarly before the revaluation phase. The left two panels of Figure 5 show the mean (averaged across participants) percentage of trials in which each S+ element of the revalued and control compounds was chosen during testing (*pre*-revaluation; Phase 2) and retesting (*post*-revaluation; Phase 4) in each condition. Four participants in the revalue condition and one participant in the control condition chose both elements from both compounds in 100% of test trials (i.e., no overselectivity). The remaining participants displayed some degree of overselectivity for the revalued compound, and most also showed overselectivity for the control compound but to an overall smaller extent (due to the method used to select the element to revalue).

A Bayesian mixed ANOVA with element as a within-subjects factor and condition as a between-subjects factor indicated strongest support for the model containing only the main effect of element ($BF_{10} = 1{,}001.87$, error % = 0.40). Analysis of effects, which summarizes the extent to which each effect should be included in the model, provided extremely strong support for inclusion of the main effect of element in the model ($BF_{incl} = 778.42$), whereas there was no support for including the main effect of condition ($BF_{incl} = 0.62$) or the interaction term ($BF_{incl} = 0.47$). Thus, there was little evidence for differences in pre-revaluation percentage of choice between conditions. In contrast, there was notably strong evidence for differences in choice depending on the element. Specifically, choice of the overselected element was higher than of the underselected element of the revalued

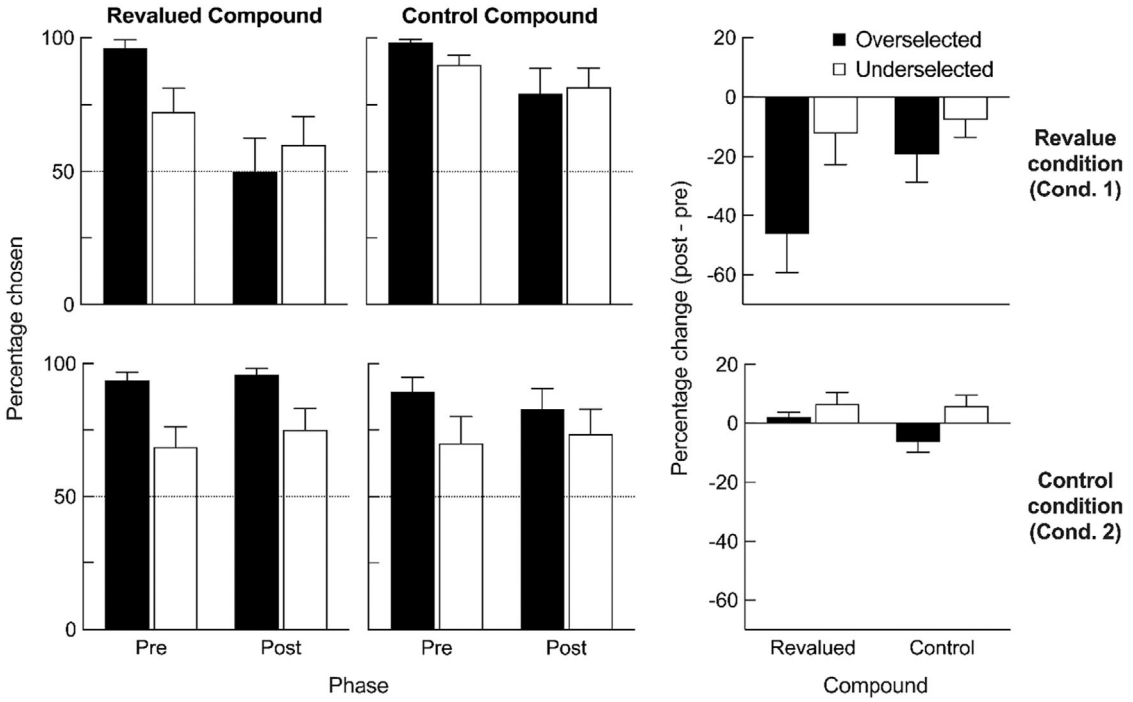

FIGURE 5 Mean percentage of choice of each S+ stimulus element before and after the revaluation phase for participants in the revalue and control conditions in Experiment 2. The left and middle panels show the mean (averaged across participants) percentage of choice for the revalued and control compounds, respectively, before (pre) and after (post) revaluation. The right panels show the percentage of change in choice pre- to post-revaluation. The error bars represent the *SEM*.

compound ($BF_{10} = 405.91$, error % < 10^{-7}). Differences in choice between the elements of the control compound were smaller, and there was moderate support for such differences ($BF_{10} = 9.28$, error % < 10^{-7}).

Analyses based on binomial theory indicated that the mean expected difference between choice of the overselected and underselected elements of the revalued compound was 8.92% (*SEM* = 1.83) in the revalue condition and 10.93% (*SEM* = 1.35) in the control condition. There was strong evidence that expected differences were smaller than obtained differences ($BF_{10} = 28.30$, error % < 0.001), indicating that the concurrent-load task induced overselectivity.

In the revaluation phase (Phase 3), choice of the overselected element of the revalued compound was extinguished in the revalue condition but not in the control condition. The right panel of Figure 5 shows the percentage of change in choice after this phase, averaged across participants in each condition. A Bayesian mixed ANOVA indicated substantial support for a model containing the main effects of element and condition and the interaction term ($BF_{10} = 574.22$, error % = 1.21). Thus, the effects of revaluation differed between elements and between conditions.

In the revalue condition (Figure 5), overall choice was similar to that for Experiment 1 (Figure 3). For the revalued compound, choice of both elements decreased. There was strong support for this decrease for the overselected element ($BF_{10} = 12.16$, error % < 0.001), whereas there was anecdotal evidence for no difference between pre- and post-revaluation choice for the underselected element ($BF_{10} = 0.50$, error % = 0.02). Choice of the elements of the control compound also decreased, though there was little to no support for such changes ($BF_{10} = 1.36$, error % < 10^{-5} for the overselected element; $BF_{10} = 0.57$, error % = 0.02 for the underselected element). In the control condition, choice changed little after revaluation, although it is worth noting that percentage of choice of the underselected element of the revalued compound increased for most (nine out of 13) participants who showed some degree of overselectivity in the control condition. Nevertheless, there was little to no evidence for differences between pre- and post-revaluation choice in the control condition ($BF_{10} = 0.60$ and 0.76 for the overselected and underselected elements of the revalued compound, respectively, and $BF_{10} = 1.08$ and 0.70 for the overselected and underselected elements of the control compound, respectively; all error % = 0.02). Therefore, in summary choice changed differently between conditions and such changes were consistent with the contingencies arranged in the revaluation phase in each condition.

Relation between overselectivity levels and post-revaluation choice

There was little change in choice after revaluation in the control condition (Figure 5), so we focus on the revalue

condition for analyses of the relation between overselectivity levels and post-revaluation choice. Because the sample size was small in the revalue condition, splitting the data based on initial overselectivity level, as we did in Experiment 1, produced very small groups: Three participants displayed no overselectivity, seven displayed low levels of overselectivity (i.e., below the mean), and three displayed high overselectivity. Visual inspection indicated that patterns of post-revaluation choice for participants with no, low, or high overselectivity were similar to those in Experiment 1 (see Figure 4). Thus, given the small sample sizes in Experiment 2, we chose to combine data from Experiment 1 (which was identical to the revalue condition) with the revalue condition of Experiment 2 and ran a linear regression analysis with overselectivity level as the predictor and percentage of change in choice as the dependent variable (total combined $N = 63$).

Figure 6 shows percentage of change in choice for each element, plotted as a function of overselectivity level. There was no evidence for a relation between overselectivity level and percentage of change in choice for the overselected element of the revalued compound ($BF_{10} = 0.37$) or for either element of the control compound ($BF_{10} = 0.27$ and 2.22 for the overselected and underselected elements, respectively). In contrast, there was extremely strong support for a positive relation between overselectivity level and percentage of change in choice of the underselected element of the revalued compound, $\beta = 0.55$, 95% CI = [0.31, 0.79]; $BF_{10} = 1{,}792.30$; $R^2 = .28$).[1] Thus, overselectivity level predicted changes in choice of the underselected element of the revalued compound, with choice of that element decreasing for participants with lower levels of overselectivity and increasing for participants with higher overselectivity.

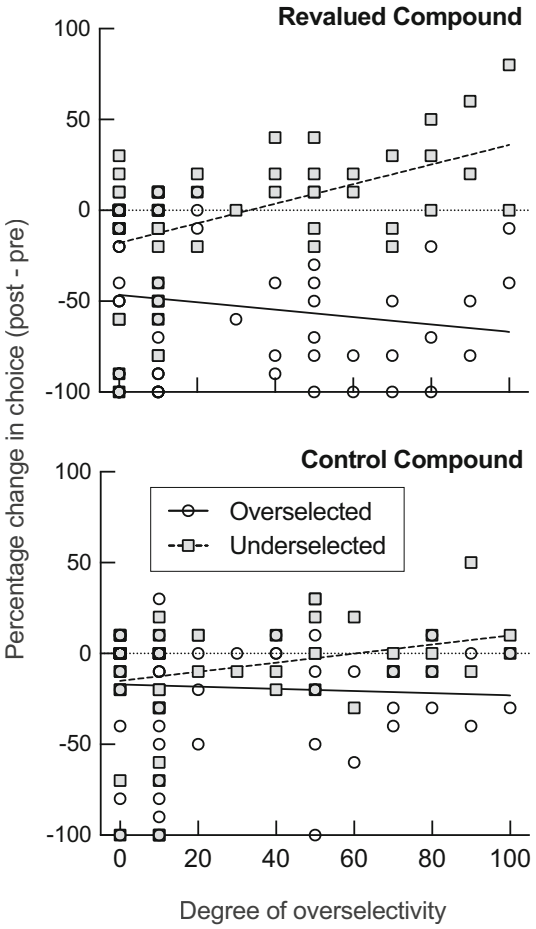

FIGURE 6 Percentage of change in choice as a function of overselectivity level in the typical revaluation procedure. Data from Experiment 1 and from Condition 1 of Experiment 2.

Discussion

This experiment was conducted as a follow up to Experiment 1 to investigate further the role of contingency change in post-revaluation changes in choice. We did this by comparing changes in choice in the typical revaluation procedure (revalue condition) with a control condition in which choice of the overselected element continued to be reinforced in the revaluation phase.

In contrast to the revalue condition and to Experiment 1 (Figure 4), choice changed little in the control condition (Figure 5). This suggests that post-revaluation changes in choice in the typical revaluation procedure are primarily related to extinction of the overselected element; if other factors, such as additional exposure to the overselected element or regression to the mean effects, were responsible for changes in choice, then similar changes should have been apparent in the control condition here. This is an important finding for participants with no or low overselectivity, for whom post-revaluation changes in choice in Experiment 1 (Figure 4) were small and similar for both the revalued and control compounds, tentatively suggesting that they may have been artifactual. The present results suggest otherwise; differences between the revalue and control conditions here highlight the importance of the contingency change in the typical revaluation procedure. Indeed, even though there was a small revaluation effect for the underselected element of the revalued compound in the control condition, this effect was likely more apparent than real.

Thus, our findings suggest that only a small (if any) portion of post-revaluation changes in choice in the typical revaluation procedure is likely to be artifactual. Nevertheless, given our small sample size and the small post-revaluation changes in choice that were seen in the control condition, further investigation into the mechanisms underlying the post-revaluation emergence in

[1] As suggested by a reviewer, we also ran this regression including primary participant variables (age, AQ score, total session time, and number of trials to criterion in training and revaluation), but there was no evidence for a model including any of these variables (all $BF_{inclusion} < 1$), whereas the evidence supporting inclusion of overselectivity level remained very strong ($BF_{inclusion} = 59.31$).

control by underselected stimuli is required. As a starting point, replicating the current experiment with a larger sample would provide further insight into the role of the contingency change in the typical revaluation procedure.

GENERAL DISCUSSION

Previous studies suggest that extinction of overselected stimuli (revaluation) results in the emergence of behavioral control by underselected stimuli (e.g., Broomfield et al., 2010; Kelly et al., 2015; Reed et al., 2009; Reed, Reynolds, et al., 2012; Reynolds & Reed, 2018; Reynolds et al., 2012). Limited evidence also suggests that the extent of such emergence is larger for individuals with greater overselectivity (Broomfield et al., 2010). Across two experiments, we extended previous research investigating how levels of overselectivity modulate post-revaluation changes in stimulus control. A novel aspect of our experiments was the inclusion of a control compound (Experiments 1 and 2) or condition (Experiment 2) against which to compare the effects of revaluation across overselectivity levels. Our main finding was that the effects of revaluation depended on overselectivity level; choice of the underselected element decreased for participants with lower overselectivity (mediated extinction), whereas it increased for those with higher overselectivity (retrospective revaluation; Figures 4 and 6). Such post-revaluation changes were probably related to the change in contingency during revaluation, as changes were generally not apparent when choice of the overselected element continued to be reinforced (control condition; Figure 5).

Why might revaluation effects depend on initial levels of overselectivity? One possibility is that the degree of overselectivity and the effects of revaluation depend on the strength of associations between stimulus elements. In a recent study, and drawing on the associative-learning literature, Reynolds and Reed (2018) suggested that stronger within-compound associations reduce cue competition and thus result in lower overselectivity and mediated extinction, whereas weaker associations increase cue competition and produce higher overselectivity and retrospective revaluation (see also Liljeholm & Balleine, 2009; Reed & Quigley, 2019; Shevill & Hall, 2004; Sissons et al., 2009; Westbrook et al., 1983). In support of this, Reynolds and Reed showed that shorter durations of compound-stimulus presentation during discrimination training—a manipulation previously shown to weaken within-compound associations in associative-learning paradigms—produced higher overselectivity and retrospective revaluation, whereas longer durations produced lower overselectivity and mediated extinction. Although we did not manipulate the strength of within-compound associations in the current study, some participants may have formed such associations more (or less) readily than others, resulting in lower (or higher) overselectivity and mediated extinction (or retrospective revaluation).

Learned associations between stimulus elements may also explain why choice of the elements of the control compound decreased for participants with lower overselectivity levels in the present experiment (Figures 4 to 6). Although the elements of the revalued and control compounds never appeared together, they may have become associated with each other due to pairing with a common outcome (e.g., the elements of both S+ compounds were associated with the same positive feedback). Indeed, separately trained stimuli may become functionally equivalent if paired with a common outcome (e.g., Minster et al., 2006; Schenk, 1994), and postlearning revaluation results in mediated extinction for such functionally equivalent stimuli (Liljeholm & Balleine, 2009). Thus, to the extent that associations were formed between the revalued and control compounds in the present study, the effects of revaluation would be expected to generalize across compounds. That this was the case for participants with lower overselectivity lends further support to our suggestion that these participants formed associations between stimuli more readily than participants with higher overselectivity.

The above line of reasoning is post hoc, as we did not experimentally manipulate levels of overselectivity or within-compound associations. Nevertheless, there was strong evidence that overselectivity levels predicted changes in choice for the underselected element of the revalued compound (Figure 6). To be sure, and to explore other potential variables that may modulate revaluation effects, future studies should manipulate variables that have been shown to influence overselectivity levels such as stimulus complexity (e.g., Reed et al., 2011), stimulus duration (e.g., Reynolds & Reed, 2018), or concurrent-task difficulty (e.g., Reed & Gibson, 2005). If studies manipulating such variables replicate the present results, this would provide stronger evidence that revaluation effects depend on overselectivity levels. Furthermore, the larger increases in choice of the underselected element for participants with higher overselectivity in the present study may partly reflect the fact that there is greater room for such increases to occur, relative to participants for whom choice is already close to 100% (see Broomfield et al., 2010, for a similar argument). Thus, future studies should consider arranging a procedure not limited by a ceiling (e.g., a free-operant procedure), which will help to clarify further the extent to which overselectivity levels influence post-revaluation changes in stimulus control.

Additionally, investigating the role of within- and between-compound associations may be a promising avenue to understanding stimulus overselectivity and changes in control by underselected stimuli following revaluation (see also Reed & Quigley, 2019; Reynolds & Reed, 2018). This is interesting both theoretically and practically. Theoretically, it suggests that research on

stimulus equivalence may be relevant to stimulus overselectivity; for example, the behavioral mechanisms underlying stimulus overselectivity may be similar to those underlying poor equivalence-class formation (an equivalence class is a group of stimuli that is functionally equivalent due to pairing with each other; e.g., Minster et al., 2006). Practically, individuals with ASD appear to have difficulties forming within-compound associations (e.g., Plaisted et al., 1998; Reed, 2011), and stimulus overselectivity is thought to be prevalent in this population (Ploog, 2010). If associations between stimuli underlie overselectivity, then interventions directed at establishing or strengthening such associations—such as training equivalence relations between overselected and underselected stimuli—may help to reduce overselectivity.

Taken together, our and Broomfield et al.'s (2010) findings call for a thorough analysis of individual-level data in future studies of stimulus overselectivity. Our findings illustrate that conclusions based solely on group-level analyses may be incomplete to the extent that group data are not representative of individual data. Even if retrospective revaluation is evident in group-mean and individual data, the extent of such revaluation may still be correlated with overselectivity level (e.g., as in Broomfield et al., 2010). Clinically, this implies that the efficacy of revaluation-based interventions may depend, to some extent, on the individual's initial level of overselectivity. However, as both Broomfield et al. and we used nonclinical populations, it remains to be seen whether similar relations between levels of overselectivity and post-revaluation changes are apparent in clinical populations. Additionally, it is unclear why we found differences in the direction of changes in choice for participants with lower and higher overselectivity, whereas Broomfield et al. found differences only in the extent of retrospective revaluation. Future replications of the current study with clinical and nonclinical populations and with larger sample sizes will help to clarify the generality of our results. Varying the extent of contingency change (e.g., by varying the reinforcer ratio for responses to the overselected and novel elements in the revaluation phase) may also provide a further test of the role of contingency change in retrospective revaluation.

In conclusion, the present experiments extended previous research examining the effects of postlearning extinction of overselected stimuli on control by underselected stimuli. Our findings suggest a correlation between initial levels of overselectivity and the effects of revaluation and demonstrate the role of the contingency change in producing such changes in choice. The present findings also highlight several interesting avenues for future research such as experimental manipulations to examine the effects of overselectivity level more thoroughly, investigations of the role of within- and between-compound associations, and further control conditions, which will provide greater insight into the mechanisms underlying the emergence in control by underselected stimuli after revaluation.

ACKNOWLEDGMENTS

The authors thank Linda Kim, Niyanta Nand, Francisca Rahjardja, and Reema Khan for their help in running the experiment. Open access publishing facilitated by Auckland University of Technology, as part of the Wiley - Auckland University of Technology agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

ETHICS APPROVAL

This experiment was conducted under Protocol 022074, approved by the University of Auckland Human Ethics Committee.

ORCID

Douglas Elliffe https://orcid.org/0000-0001-9449-9540

REFERENCES

Balleine, B. W., Espinet, A., & González, F. (2005). Perceptual learning enhances retrospective revaluation of conditioned flavor preferences in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(3), 341–450. https://doi.org/10.1037/0097-7403.31.3.341

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*, 5–17. https://doi.org/10.1023/A:1005653411471

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior*, *27*, 63–76. https://doi.org/10.3758/BF03199432

Blough, D. S. (1969). Attention shifts in a maintained discrimination. *Science*, *166*(3901), 125–126. https://doi.org/10.1126/science.166.3901.125

Broomfield, L., McHugh, L., & Reed, P. (2010). Factors impacting emergence of behavioral control by underselected stimuli in humans after reduction in control by overselected stimuli. *Journal of the Experimental Analysis of Behavior*, *94*(2), 125–133. https://doi.org/10.1901/jeab.2010.94-125

Chase, S. (1968). Selectivity in multidimensional stimulus control. *Journal of Comparative and Physiological Psychology*, *66*(3), 787–792. https://doi.org/10.1037/h0026542

Cowie, S., Davison, M., & Elliffe, D. (2017). Control by past and present stimuli depends on the discriminated reinforcer differential. *Journal of the Experimental Analysis of Behavior*, *108*(2), 184–203. https://doi.org/10.1002/jeab.268

Davison, M. (2018). Divided stimulus control: Which key did you peck, or what color was it? *Journal of the Experimental Analysis of Behavior*, *109*(1), 107–124. https://doi.org/10.1002/jeab.295

Davison, M., & Elliffe, D. (2010). Divided stimulus control: A replication and a quantitative model. *Journal of the Experimental Analysis of Behavior*, *94*(1), 13–23. https://doi.org/10.1901/jeab.2010.94-13

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, *49*(1), 60–80. https://doi.org/10.1080/713932614

Dube, W. V. (2009). Stimulus overselectivity in discrimination learning. In P. Reed (Ed.), *Behavioral theories and interventions for autism* (pp. 23–46). Nova Science Publishers.

Dube, W. V., Dickson, C. A., Balsamo, L. M., O'Donnell, K. L., Tomanari, G. Y., Farren, K. M., Wheeler, E. E., & McIlvane, W. J. (2010). Observing behavior and atypically restricted stimulus control. *Journal of the Experimental Analysis of Behavior*, *94*(3), 197–313. https://doi.org/10.1901/jeab.2010.94-297

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D., Balsamo, L. M., & Fowler, T. R. (1999). Eye tracking assessment of stimulus overselectivity in individuals with mental retardation. *Experimental Analysis of Human Behaviour Bulletin*, *17*, 267–271.

Dube, W. V., Lombard, K. M., Farren, K. M., Flusser, D. S., Balsamo, L. M., Fowler, T. R., & Tomanari, G. Y. (2003). Stimulus overselectivity and observing behavior in individuals with mental retardation. In S. Soraci & K. Murata-Soraci (Eds.), *Visual information processing* (pp. 109–123). Praeger.

Dube, W. V., & McIlvane, W. J. (1999). Reduction of stimulus overselectivity with nonverbal differential observing responses. *Journal of Applied Behavior Analysis*, *32*(1), 25–33. https://doi.org/10.1901/jaba.1999.32-25

Farber, R. S., Dickson, C. A., & Dube, W. V. (2017). Reducing overselective stimulus control with differential observing responses. *Journal of Applied Behavior Analysis*, *50*(1), 87–105. https://doi.org/10.1002/jaba.363

JASP Team. (2018). *JASP* (Version 0.16.4) [Computer software]. https://jasp-stats.org

Kelly, M. P., Leader, G., & Reed, P. (2015). Stimulus over-selectivity and extinction-induced recovery of performance as a product of intellectual impairment and autism severity. *Journal of Autism and Developmental Disorders*, *45*, 3098–3106. https://doi.org/10.1007/s10803-015-2466-x

Kelly, M. P., Leader, G., & Reed, P. (2016). Factors producing over-selectivity in older individuals. *Age*, *38*, Article 63. https://doi.org/10.1007/s11357-016-0026-x

Kelly, M. P., & Reed, P. (2021). Age effects on the development of stimulus over-selectivity are mediated by cognitive flexibility and selective attention. *International Journal of Behavioral Development*, *45*(1), 89–96. https://doi.org/10.1177/0165025420949702

Larkin, M. J. W., Aitken, M. R. F., & Dickinson, A. (1998). Retrospective revaluation of causal judgments under positive and negative contingencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1331–1352. https://doi.org/10.1037/0278-7393.24.6.1331

Leader, G., Loughnane, A., McMoreland, C., & Reed, P. (2009). The effect of stimulus salience on over-selectivity. *Journal of Autism and Developmental Disorders*, *39*, 330–338. https://doi.org/10.1007/s10803-008-0626-y

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Liljeholm, M., & Balleine, B. W. (2009). Mediated conditioning versus retrospective revaluation in humans: The influence of physical and functional similarity of cues. *Quarterly Journal of Experimental Psychology*, *62*(3), 470–482. https://doi.org/10.1080/17470210802008805

Lovaas, O. I., Schreibman, L., Koegel, R., & Rehm, R. (1971). Selective responding by autistic children to multiple sensory input. *Journal of Abnormal Psychology*, *77*(3), 211–222. https://doi.org/10.1037/h0031015

Matzel, L. D., Schachtman, T. R., & Miller, R. R. (1985). Recovery of an overshadowed association achieved by extinction of the overshadowing stimulus. *Learning and Motivation*, *16*(4), 398–412. https://doi.org/10.1016/0023-9690(85)90023-2

McHugh, L., & Reed, P. (2007). Age trends in stimulus overselectivity. *Journal of the Experimental Analysis of Behavior*, *88*(3), 369–380. https://doi.org/10.1901/jeab.2007.88-369

McHugh, L., Simpson, A., & Reed, P. (2010). Mindfulness as a potential intervention for stimulus over-selectivity in older adults. *Research in Developmental Disabilities*, *31*(1), 178–184. https://doi.org/10.1016/j.ridd.2009.08.009

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. *Psychology of Learning and Motivation*, *22*, 51–92. https://doi.org/10.1016/S0079-7421(08)60038-9

Miller, R. R., & Witnauer, J. E. (2016). Retrospective revaluation: The phenomenon and its theoretical implications. *Behavioural Processes*, *123*, 15–25. https://doi.org/10.1016/j.beproc.2015.09.001

Minster, S. T., Jones, M., Elliffe, D., & Muthukumaraswamy, S. D. (2006). Stimulus equivalence: Testing Sidman's (2000) theory. *Journal of the Experimental Analysis of Behavior*, *85*(3), 371–391. https://doi.org/10.1901/jeab.2006.15-05

Plaisted, K., O'Riordan, M., & Baron-Cohen, S. (1998). Enhanced visual search for a conjunctive target in autism: A research note. *Journal of Child Psychology and Psychiatry*, *39*(5), 777–783. https://doi.org/10.1111/1469-7610.00376

Ploog, B. O. (2010). Stimulus overselectivity four decades later: A review of the literature and its implications for current research in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*, 1332–1349. https://doi.org/10.1007/s10803-010-0990-2

Reed, P. (2011). Comparator mechanisms and autistic spectrum conditions. In T. R. Schachtman & S. R. Reilly (Eds.), *Associative learning and conditioning: Human and animal applications* (pp. 168–190). Oxford University Press.

Reed, P. (2019). Unpredictability reduces over-selective responding of individuals with ASD who have language impairments. *Research in Autism Spectrum Disorders*, *57*, 35–45. https://doi.org/10.1016/j.rasd.2018.10.006

Reed, P., Altweck, L., Broomfield, L., Simpson, A., & McHugh, L. (2012). Effect of observing-response procedures on overselectivity in individuals with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *27*(4), 237–246. https://doi.org/10.1177/1088357612457986

Reed, P., Broomfield, L., McHugh, L., McCausland, A., & Leader, G. (2009). Extinction of over-selected stimuli causes emergence of under-selected cues in higher-functioning children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*, 290–298. https://doi.org/10.1007/s10803-008-0629-8

Reed, P., & Gibson, E. (2005). The effect of concurrent task load on stimulus over-selectivity. *Journal of Autism and Developmental Disorders*, *35*, 601–614. https://doi.org/10.1007/s10803-005-0004-y

Reed, P., Petrina, N., & McHugh, L. (2011). Over-selectivity as a learned response. *Research in Developmental Disabilities*, *32*(1), 201–206. https://doi.org/10.1016/j.ridd.2010.09-016

Reed, P., & Quigley, M. (2019). Over-selectivity decreases with increased training: A role for within-compound associations. *Acta Psychologica*, *198*, Article 102868. https://doi.org/10.1016/j.actpsy.2019.102868

Reed, P., Reynolds, G., & Fermandel, L. (2012). Revaluation manipulations produce emergence of underselected stimuli following simultaneous discrimination in humans. *The Quarterly Journal of Experimental Psychology*, *65*(7), 1345–1360. https://doi.org/10.1080/17470218.2012.656663

Reynolds, G., & Reed, P. (2011a). Effects of schedule of reinforcement on over-selectivity. *Research in Developmental Disabilities*, *32*(6), 2489–2501. https://doi.org/10.1016/j.ridd.2011.07.011

Reynolds, G., & Reed, P. (2011b). The strength and generality of stimulus over-selectivity in simultaneous discrimination procedures. *Learning and Motivation*, *42*(2), 113–122. https://doi.org/10.1016/j.lmot.2010.12.001

Reynolds, G., & Reed, P. (2018). The effect of stimulus duration on over-selectivity: Evidence for the role of within-compound associations. *Journal of Experimental Psychology: Animal Learning and Cognition*, *44*(3), 293–308. https://doi.org/10.1037/xan0000175

Reynolds, G., Watts, J., & Reed, P. (2012). Lack of evidence for inhibitory processes in over-selectivity. *Behavioural Processes*, *89*(1), 14–22. https://doi.org/10.1016/j.beproc.2011.09.008

Schenk, J. J. (1994). Emergent relations of equivalence generated by outcome-specific consequences in conditional discrimination. *The Psychological Record*, *44*(4), 537–558. https://doi.org/10.1007/BF03395143

Shahan, T. A., & Podlesnik, C. A. (2006). Divided attention performance and the matching law. *Learning & Behavior*, *34*, 255–261. https://doi.org/10.3758/BF03192881

Shahan, T. A., & Podlesnik, C. A. (2007). Divided attention and the matching law: Sample duration affects sensitivity to reinforcement allocation. *Learning & Behavior*, *35*, 141–148. https://doi.org/10.3758/BF03193049

Shevill, I., & Hall, G. (2004). Retrospective revaluation effects in the conditioned suppression procedure. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, *57*(4), 331–347. https://doi.org/10.1080/02724990344000178

Sissons, H. T., Urcelay, G. P., & Miller, R. R. (2009). Overshadowing and CS duration: Counteraction and a reexamination of the role of within-compound associations in cue competition. *Learning & Behavior*, *37*, 254–268. https://doi.org/10.3758/LB.37.3.254

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and Motivation*, *25*(2), 127–151. https://doi.org/10.1006/lmot.1994.1008

Wagenmakers, E.-J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouger, J. N., & Morey, R. D. (2018). Bayesian inference for psychology Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, *25*, 35–57. https://doi.org/10.3758/s13423-017-1343-3

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., … Morey, R. D. (2018b). Bayesian inference for psychology Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*, 58–76. https://doi.org/10.3758/s13423-017-1323-7

Walpole, C. W., Roscoe, E. M., & Dube, W. V. (2007). Use of a differential observing response to expand restricted stimulus control. *Journal of Applied Behavior Analysis*, *40*(4), 707–712. https://doi.org/10.1901/jaba.2007.707-712

Westbrook, R. F., Homewood, J., Horn, K., & Clarke, J. C. (1983). Flavour-odour compound conditioning: Odour-potentiation and flavor-attenuation. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, *35*(1), 13–33. https://doi.org/10.1080/14640748308400911

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gomes-Ng, S., Kim, P. B. C., Cowie, S., & Elliffe, D. (2023). Revaluation of overselected stimuli: Emergence of control by underselected stimuli depends on degree of overselectivity. *Journal of the Experimental Analysis of Behavior*, 1–16. https://doi.org/10.1002/jeab.850