AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Animal Learning and Cognition

ISSN: 2329-8456

2022, Vol. 48, No. 2, 105–122
https://doi.org/10.1037/xan0000317

Valence Generalization Across Nonrecurring Structures

Micah Amd
Department of Psychology, University of the South Pacific

Semantically meaningless letter strings correlated with affective attributes (US) can become evaluatively conditioned stimuli (CS). Jurchiș et al. (2020) recently demonstrated CS-US correlations may influence evaluations toward previously unseen strings when the latter are grammatically congruent with CS. We replicated those authors' findings in a modified extension (Experiment 1; N = 108), where emotional faces (US) were correlated with letter strings (CS) constructed from familiar (English) and unfamiliar (Phoenician) alphabets. CS-US trials were sandwiched by evaluations of strings that never appeared as CS but were constructed using similar grammar rules. Although CS and evaluated strings never overlapped, their individual elements (letters) recurred between phases. Element recurrence was controlled for in a second replication (Experiment 2; N = 140), where participants viewed Phoenician (/English) strings during conditioning and English (/Phoenician) strings during evaluations. We found credible evidence for valence generalization across strings from different alphabets but parallel grammars, suggesting the latter had been perceived as 'functionally equivalent' (Tonneau, 2004b). We provide support for this claim in a third study (Experiment 3; N = 79), where participants underwent a 'free selection' 2AFC discrimination task with sample and comparison strings taken from different alphabets. Increasing frequencies of grammar-congruent discriminations suggested strings were becoming functionally equated along overlapping grammar rules. We speculate how 'rules' which inform how elements are organized relative to each other can be abstracted and generalized across without specifying elemental properties (Spaulding, 1912).

Keywords: conditioning, learning theory, realism

Supplemental materials: https://doi.org/10.1037/xan0000317.supp

Unfamiliar symbols can come to elicit evaluative responses as "conditioned stimuli" (CS) following correlations with emotionally meaningful symbols (unconditioned stimuli, US), in what can be described as US-to-CS valence transfer/generalization (Mowrer, 1960). According to learning theory, if a pair of unfamiliar and relatively neutral symbols (call these CS1 and CS2) are respectively correlated with positively and negatively valenced US then CS1 should be positively evaluated relative to CS2, all else remaining equal (Staats & Staats, 1958). While the phenomenon of valence generalization following CS-US correlations has been investigated for some time (Mowrer, 2013; Staats, 1996), it remains unanswered whether the representational processes underlying transfer are exclusively propositional (De Houwer et al., 2021), or whether unqualified associations are necessary for describing the acquisition and expression of CS valence (Corneille & Mertens, 2020; Gawronski & Bodenhausen, 2006; Hofmann et al., 2010).

By "associations," we imply nonspecified mental links that incrementally form on the basis of regularities between spatio-temporally correlated stimuli and have no inherent truth value (Fazio, 2007; Gawronski & Bodenhausen, 2018). Propositions, on the other hand, describe specified relations that can be subjectively evaluated as true or false (De Houwer et al., 2021). According to some theorists, "simpler" associative architectures constitute and/or interact with propositional processes to influence evaluations following CS-US correlations (Gawronski & Bodenhausen, 2006, 2011; McLaren et al., 2019). Associations are *simpler* in the sense that they do not describe how representations are related beyond simple linkage—once activated, associations may be uncontrollably expressed and even supersede relational information in selected contexts (Mandelbaum, 2015; McConnell & Rydell, 2014). Because associative theory posits terms can be linked together without being explicitly specified, valences may be encoded and expressed with minimal deliberative influence (Gawronski & Bodenhausen, 2011). Conversely, propositional theorists question whether "mental associations" are conceptually useful for explaining evaluative effects (De Houwer et al., 2020; Mitchell et al., 2009). From this view, valence generalization is a product of contextually specified inferences informed by CS-US co-occurrences, which shifts evaluative beliefs through the relational specification of perceived contingencies (e.g., *CS co-occurs with US,*

Micah Amd https://orcid.org/0000-0003-4225-7026

All work reported was funded by a research grant from the University of South Pacific Research Office to the author and approved by the University ethics committee. The author thanks Li Mei Chew for helpful discussions on realism and Denise Passarelli for assistance with material preparation. All materials, procedures, data, and analysis scripts are openly available at https://doi.org/10.17605/OSF.IO/QDHMY. The author has no conflict of interest to declare.

Correspondence concerning this article should be addressed to Micah Amd, Department of Psychology, University of the South Pacific, CELT 205, USP Laucala, Suva, Fiji. Email: micah.amd.eab@hotmail.com

CS predicts US, CS is the same as US, etc.; De Houwer, 2018; De Houwer et al., 2020). The specification of CS-US relations as propositions with truth value is implied as both necessary and sufficient for generating evaluative effects (De Houwer et al., 2021).

We compared predictions derived from the above perspectives across the present study, which extends on an earlier report by Jurchiş et al. (2020). The highlight of that work was the demonstration of valence generalization across letter strings that were never presented during CS-US conditioning sequences but were structurally congruent with the former. We summarize those authors' study below, then highlight some limitations which the present study aimed to address.

In the work by Jurchiş et al. (2020), Romanian undergraduates underwent a simultaneous CS-US conditioning protocol where positively/negatively valenced images (US) selectively appeared with English letter strings (CS) from one of two artificial grammar categories (call these grammar A and grammar B). By "artificial grammars," we imply letter strings constructed following predetermined rules regarding how various elements are to be organized relative to one another (Norman et al., 2016; Reber, 1967; Scott & Dienes, 2010). For example, the same bigram *XM* could probabilistically precede the letters *X* or *V*, depending on whether the string was a member of grammar A or grammar B (Jurchiş et al., 2020, p. 1804). So, although artificial grammar strings were semantically meaningless as composites, they shared common organizational patterns. After CS-US pairings, participants evaluated strings that had never appeared as CS previously but were grammatically congruent with the latter. For brevity, suppose grammars A and B were exclusively associated with positive and negative US respectively. In this case, the expected conditioning effect would be A – B (positive – negative) > 0, which those authors reported moderate evidence for ($.17 < d's < .53$). Because conditioning effects were reported across exemplars never seen previously, but which were otherwise from the same grammar categories as CS, Jurchiş et al. (2020) reasoned that valences established through CS-US pairings had generalized across "nonconscious (associative) knowledge structures" (p. 1809).

We see six possible issues with this claim. First, strings constructed by Jurchiş et al. (2020) constituted of characters from a well-known alphabet, which can come with their own affective histories (Head et al., 2013; Scott & Dienes, 2008; Staats, 1996). If some elements were already salient, these may have controlled evaluative performances without requiring any grammar knowledge (Jurchiş et al., 2020, p. 1807). Second, those authors' conditioning sequences afforded extensive deliberation opportunities—each CS-US pair appeared for over 7 s and CS were repeated during blocks. Repeating CS for extended durations may have occasioned predictive (e.g., *I think those letters come before pleasant images*) and confirmatory (e.g., *those letters always predict pleasant images*) inferences within a conditioning block, which could influence evaluative beliefs (De Houwer, 2018). Third, since evaluation and conditioning trials repeated strings, participants may have generated evaluations based on familiarity with recurring letter combinations (Scott & Dienes, 2008). Fourth, Jurchiş et al. (2020) incorporated an awareness check after each string evaluation. The immediacy and sensitivity afforded by this trial-by-trial measure (Shanks & John, 1994) may have been offset by (potentially) priming participants' attention to specific letter configurations (e.g., letters corresponding to one's initials), or through induction of rules unrelated to the task contingencies (March et al., 2018). Fifth, Jurchiş et al.'s (2020) claims were based on postconditioning evaluations exclusively, meaning it is unknown whether artificial grammars had been affectively homogenous prior to conditioning (Silva, 2018). Seeing how non-neutral CS are less prone to evaluative conditioning effects (Cacioppo et al., 1992), it may have been the case that (some) stimulus evaluations simply reflected constituent (not acquired) valences. Finally, when comparing across performances between awareness categories, Jurchiş et al. (2020) reported participants who were "highly aware" of task contingencies produced more robust evaluative effects relative to participants classified as "less aware" (p. 1806). The positive relation between valence conditioning and higher strategy awareness, along with the limitations noted earlier, collectively question whether valence generalization across artificial grammars requires the assumption of unqualified mental links, seeing how the reported effects could be explained by propositional processes exclusively.

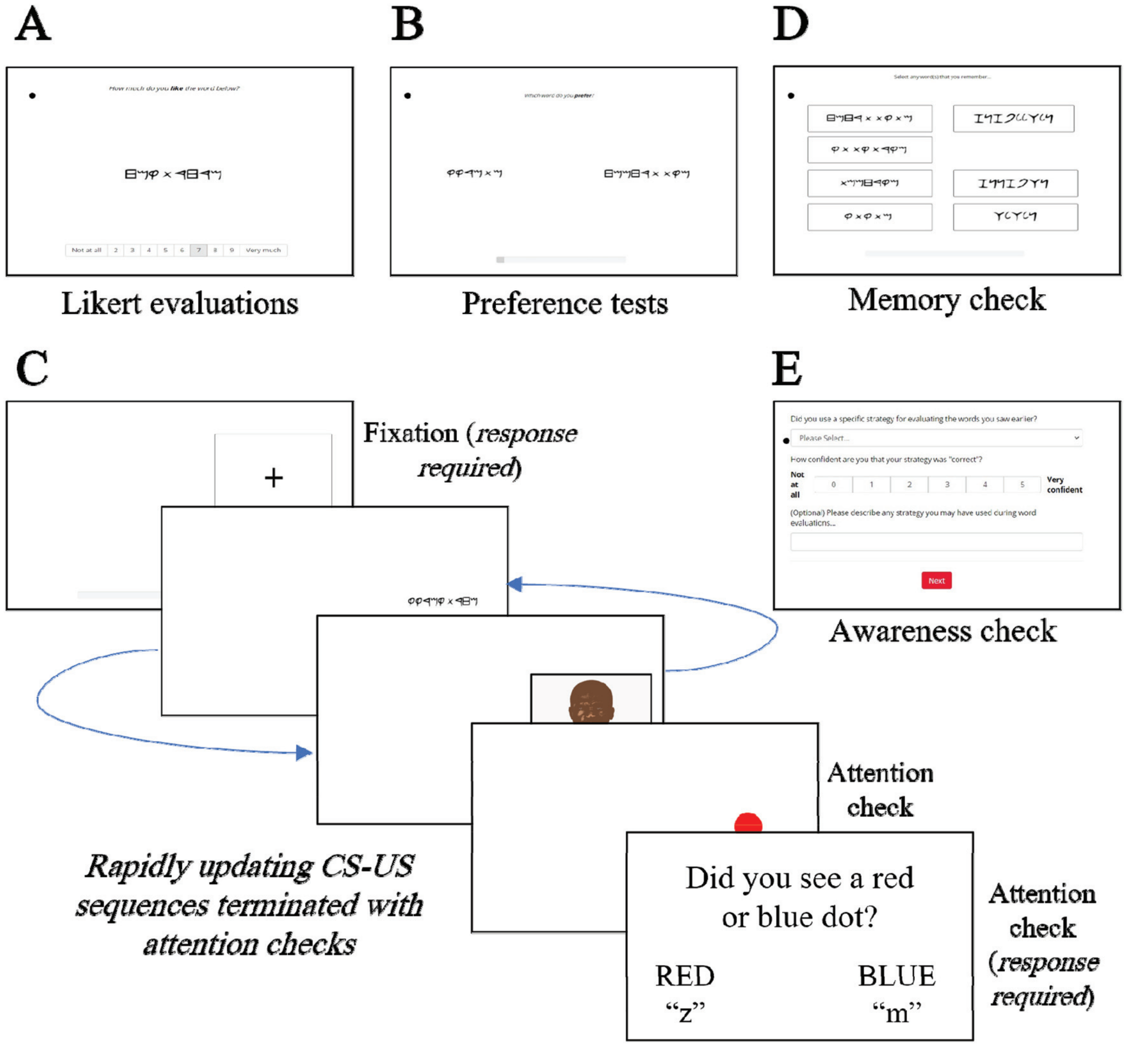
These limitations were addressed across the first pair of experiments reported here, where we tested whether valences generalize across novel strings from nonoverlapping grammars, similar to Jurchiş et al. (2020). Expanding on Jurchiş, we recorded preconditioning valences, incorporated 2-alternative forced choice (2AFC) preference checks, manipulated stimulus onset asynchronies (SOAs) between groups and alphabet familiarity within groups. We also included an awareness check at the end of evaluation trials (rather than on a trial-by-trial basis) to assess whether participants were cognizant of a global evaluative strategy—responses here were used to classify participants into *high*, *partial* and *least aware* subgroups during analyses (see Procedure section). The extent to which valence generalization could be explained by stimulus processing times and/or strategy awareness should inform the centrality of resource-intensive (propositional) processes during generalization.

In Experiment 1, strings constructed with familiar (English) and unfamiliar (Phoenician) alphabets were presented during conditioning and evaluation phases, respectively. Familiar and unfamiliar strings were constructed using matching grammar rules (see Materials section). Experiment 2 replicated the first study with a key procedural difference: strings presented during conditioning and evaluation phases were varied along alphabet. This ensured trained and tested sets contained distinct elements, reducing the likelihood of feature-mediated generalization. Finally, Experiment 3 explored whether English and Phoenician strings from congruent grammars become more likely to be matched together following increasing exposures to exemplars from both grammars. The final experiment tested whether regularities between grammar structures could be perceptually equated following multiple exposures to novel grammar exemplars. Additional details about individual experiments are provided later.

Experiment 1

In the first phase of the current study, participants completed Likert evaluations and two-alternative forced choice (2AFC) tasks to respectively record baseline evaluations and preferences toward English and Phoenician strings (Figure 1, Panels A and B). English strings were adopted from Jurchiş et al. (2020). These were transformed into a Phoenician script, a relatively unfamiliar alphabet for contemporary speakers (Rollston, 2020). Participants next underwent a CS-US conditioning protocol where English/Phoenician strings were correlated with happy/angry faces (Panel C), which was followed by a free-selection memory check (Panel D)

Figure 1
Phase Sequence Across Experiments 1 and 2



Note. Participants evaluated English/Phoenician strings before and after conditioning using 10-point Likert scales (A) and 2-alternative forced choice (2AFC) preference tests (B). Conditioning sequences initiated with a fixation point on the left or right sides of the screen (C). CS-US = conditioned stimulus-unconditioned-stimulus. A location-contingent keypress generated sequences of happy/angry faces (US) interspersed with English/Phoenician strings (CS) followed by attention checks. After conditioning and evaluation trials, participants completed memory checks (D) and indicated strategy awareness (E). See the online article for the color version of this figure.

and a second round of evaluation and preference tests. Near the end of the task, participants respectively indicated if they derived an evaluative strategy, then indicated their subjective confidence level in the accuracy of their strategy. Participants also had the option to elaborate on their subjective strategy near the end of the task (Panel E). Responses here were used to classify participants into *high*, *partial*, and *least aware* groups during analyses. Between groups, we manipulated stimulus processing times by varying CS and US onset asynchronies (SOAs) at 100 ms, 200 ms, and 400 ms. If resource-intensive deliberations are central to generalization, we assumed longer SOAs and/or higher strategy awareness would interact with evaluative effects. Alternatively, if evaluative effects are independent of processing times or strategy awareness, one could argue for the operation of an incremental (associative) learning process driven by exposure to CS-US frequencies, which were matched across participants.

Method

Participants

One-hundred and fifty-two psychology undergraduate students from the University of the South Pacific (USP) took part in exchange for course credit. A fixed-duration sampling strategy was followed for the months of October and November of 2020. The data of six participants were excluded for failing attention checks; 38 participants were excluded for slow Internet speeds (<8 Mbps), which produced SOA timing errors of ± 200 ms. The remaining $N = 108$ participants were randomly assigned to 100 ms ($n = 36$; 27.1 ± 7.8 years; 32 females), 200 ms ($n = 36$; 26.6 ± 7.4 years; 26 females), and 400 ms ($n = 36$; 24.8 ± 7.3 years; 28 females) SOA conditions. Sensitivity analyses for one-sided one-sample tests indicated samples of $n = 36$ could reliably detect moderate-to-large effects ($d > .42$) with 80% power when α error was set to 5% (Faul et al., 2009). All procedures reported were approved by the local IRB and comply with the Declaration of Helsinki. Participants completed all tasks within 30 min on average.

Materials

All tasks were designed and implemented on the Gorilla platform (Anwyl-Irvine et al., 2020) and are available online (https://app.gorilla.sc/openmaterials/120282). All participants completed demographic and personality surveys at the beginning of the experiment. These were unrelated to the present study and are not discussed further. English and Phoenician strings from two grammar categories were employed during conditioning and evaluation phases. English strings were taken from Jurchiș et al.'s (2020) open materials. These were directly converted to Phoenician characters using a freeware font package. Grammar structures for the two English categories were preserved across the two Phoenician categories. Each alphabet-grammar set contained 52 strings. From each set, 40 strings appeared as CS during conditioning trials, eight appeared during 2AFC preference tests, and four appeared during evaluation trials (see Table S4 in the online supplemental materials). Strings allocated to each condition varied between participants and never overlapped between phases. CS assignment to positive and negative US was counterbalanced between participants and alphabets. An unrelated set of English and Phoenician strings from unused grammars appeared as comparisons during memory checks (see Procedure). US consisted of 20 Black and 20 White attractiveness-matched male faces with happy (positive) and angry (negative) expressions from the Chicago Face Database (Ma et al., 2015). Normative valences for positive and negative faces, alongside additional face ratings data, are available in the online supplemental materials. All analyses were run on *RStudio* (RStudio Team, 2020) using the *tidyverse* (Wickham et al., 2019), *BayesianFirstAid* (Bååth, 2014), *ggdist* (Kay, 2021), *rstatix* (Kassambara, 2021), *forcats* (Wickham, 2021), *ggthemes* (Arnold, 2021), and *ggplot2* (Wickham, 2016) packages. The article was typeset on RMarkdown (Baumer & Udwin, 2015) on a *papaja* (Aust & Barth, 2020) generated template. Data and the markdown document are available in an online OSF repository (https://doi.org/10.17605/OSF.IO/QDHMY).

Procedure

Preconditioning Evaluations. Following consent and survey completion, participants evaluated 16 strings (four from each alphabet-grammar category) across 16 trials. Across each trial, participants indicated "how much they liked" the displayed string along 10-point scales (scored from 1 = *not at all* to 10 = *very much*). Interacting with the scale produced a blank 1,000-ms intertrial interval (ITI). If no response was detected within 10 s, the message *Timeout* would appear on screen for 1,000 ms before the onset of a new trial. Evaluation trials continued until 16 evaluations had been recorded.

Preconditioning Preferences. Participants viewed pairs of strings from the same alphabet but different grammars (e.g., PHOENICIAN-A|PHOENICIAN-B, or ENGLISH-A|ENGLISH-B). Participants were asked to select which "word they preferred more," and to provide "their best guess" when not sure of how to respond. Selection of either option produced a blank 1,000-ms ITI, followed by string pairs from the alternate alphabet. Similar to evaluation trials, a *Timeout* message appeared if no response was detected within 10 s. Preference trials continued until 16 responses were recorded.

Grammar Conditioning. Participants next underwent 80 conditioning trials, where 40 strings from each alphabet and grammar category (160 strings overall) were uniquely associated with happy or angry faces. Each conditioning trial commenced with a fixation point on the left/right sides of the screen. Clicking on the fixation with the mouse pointer "forced" participants to orient toward a rapidly updating stimulus sequence in the same location. Orienting toward stimulus sequences can facilitate CS-US acquisition (Amd & Baillet, 2019, 2017; Amd et al., 2018; Ribeiro et al., 2020; Sokolov, 1963). The display sequence contained 22 stimuli in randomized orders—this included 20 happy/20 angry faces, and a pair of English/Phoenician strings from the same grammar for a given participant. English and Phoenician strings never appeared in the same sequence. Grammar-valence assignment was randomized between participants.[1] Stimulus-onset asynchronies (SOAs) were varied between 100 ms, 200 ms, and 400 ms between participants. Display sequences terminated with a red/blue dot, or a triangle/square geometrical shape, for 500 ms followed by a 1,000-ms blank interval. Participants had to indicate dot color/figure shape from the preceding display. Failing this attention check three times consecutively dropped the participant from the study.

Memory Check. Completion of the conditioning task was followed by a 1,000-ms blank ITI, after which the following instructions appeared on screen:

> *You will now be shown a list of "words," some of which you may have seen before. Please select those words which appear familiar to you. Feel free to select as many (or few) words as you like. Each trial will automatically progress after 10 s.*

Upon pressing START, participants viewed eight strings across two columns, all from the same alphabet. This included four strings from familiar grammars (two from ENGLISH-A and two from ENGLISH-B, *or* two from PHOENICIAN-A and two from PHOENICIAN-B), and four strings from unfamiliar grammar structures (two from ENGLISH-C and two from ENGLISH-D, *or* two from PHOENICIAN-C and two from PHOENICIAN-D). Grammars C and D never appeared in any other phase. Across each trial, a text prompt appeared near the top of the screen reminding participants to *select any word(s) that* (they) *remembered from the earlier task.* Participants had 10 s to select any number of strings they wanted

---

[1] Some participants viewed English and Phoenician strings from grammar A with positive images (*US*+), and English and Phoenician strings from grammar B with negative images (*US*–). This can be summarily described as $\frac{US+(Eng{:}A|Pho{:}A)}{US-(Eng{:}B|Pho{:}B)}$. Other participants viewed $\frac{US+(Eng{:}B|Pho{:}A)}{US-(Eng{:}A|Pho{:}B)}$, or $\frac{US+(Eng{:}A|Pho{:}B)}{US-(Eng{:}B|Pho{:}A)}$, or $\frac{US+(Eng{:}B|Pho{:}B)}{US-(Eng{:}A|Pho{:}A)}$. During analysis, we categorized grammars based on whether they had been positively or negatively conditioned only.

using the mouse pointer. Clicking on a string immediately removed it from the screen. The location of strings were randomized between participants. Participants underwent eight memory check trials, four with English strings and four with Phoenician strings, in counterbalanced order (memory trials presenting the same alphabet never appeared concurrently). Each participant selected from 32 English and 32 Phoenician strings.

Postconditioning Evaluations and Preferences. Evaluation and preference phases described earlier were repeated, using the same strings presented during preconditioning but in different sequences.

Awareness Check. In the final phase, participants completed three survey items. First, participants selected from four options (yes/not sure/I think so/no) in response to the question *Did you use a specific strategy for evaluating the words you saw earlier?* Participants were classified into *least aware* (no), *partially aware* (not sure/I think so) and *highly aware* (yes) subgroups based on their response. Second, participants responded to *How confident are you that your strategy was "correct?"* using a 5-point scale (scored from 1 = *not at all* to 5 = *very confident*). Finally, participants had to option to describe "any strategy" they had used with written statements. Participant statements from each awareness[2] condition are provided in the online OSF file.

Results

Valence Evaluations

Evaluations collected before (*Pre*) and after (*Post*) conditioning were normalized ($\frac{Post-Pre}{Post+Pre}$) to reduce between-subjects variance and control for preconditioning differences. A 2 × 2 × 3 Type-2 ANOVA was run to explain variances across normalized valences. US valence (2) and string alphabet (2) were entered as repeated factors, with SOAs (3) as the between-subjects factor. Awareness was not included as a factor in our initial model to retain a balanced design and minimize chance detection of false positives. Additionally, a four-way interaction would be difficult to meaningfully interpret and could distract the reader from significant effects (Free, 2016). The homogeneity of variance assumption was not violated (p's $> .05$). None of the interaction terms reached significance (p's $> .1$). Significant main effects were observed for SOA, $F(2, 105) = 5.51$, $p = .005$, $\eta_p^2 = .09$; $MS_{Error} = .34$, and valence, $F(1, 105) = 12.63$, $p < .001$, $\eta_p^2 = .11$; $MS_{Error} = .24$.

Although a lack of significant interactions does not justify further tests, we had a priori reasons for investigating whether generalization was more likely during longer SOAs and/or higher strategy awareness. We reanalyzed our data in light of our mixed model's outcomes to investigate conditioning effects more thoroughly. We first estimated bias-corrected difference scores (Hedge's g) between positively and negatively conditioned grammars[3] for individual participants, combined across alphabets. A one-way Type-2 ANOVA with SOA as the between-subjects factor did not significantly explain valence effects, $F(2, 105) = 1.37$, $p = .260$, $\eta_p^2 = .03$, so no post hoc contrasts were run between SOAs.

Inspection of g-score distributions across SOAs (Figure 2, Panel A) suggested valence generalization may have been more likely during longer stimulus durations, motivating further tests. We ran series of frequentist and Bayesian one-sided single-sample tests to respectively estimate whether mean g-scores were significantly ($p < .05$) and/or credibly ($L^{\text{Bayes}} > 80\%$) greater than null estimates.[4] Across SOAs, we found credible evidence for generalization across participants exposed to 200-ms SOAs, $t(35) = 1.83$; $p = .057$; $L^{\text{Bayes}} = 87\%$; g [95%] = .24 [.11, .37], and across participants exposed to 400-ms SOAs, $t(35) = 2.69$; $p = .016$; $L^{\text{Bayes}} = 99\%$; g [95%] = .29 [.18, .4]. Next, we entered strategy awareness (3) and stimulus duration (3) into a 3 × 3 Type-2 ANOVA to explain g-score variance. No statistically significant outcomes were detected (p's $> .07$). One-sided Bayes tests produced credible ($L^{\text{Bayes}} > 83\%$) but statistically nonsignificant (p's $> .06$) evidence for generalization across *highly aware* participants exposed to 200-ms and 100-ms SOAs, as well as *partially aware* and *least aware* participants exposed to 400-ms SOAs.

2AFC Preferences

During 2AFC trials, participants selected between two strings from the same alphabet but alternate grammars (e.g., ENGLISH-A|ENGLISH-B, or PHOENICIAN-A|PHOENICIAN-B). Each participant completed 16 selections before and 16 selections after conditioning. Selecting strings from positively conditioned grammars were recorded as "hits." During analysis, we estimated hit proportions before and after conditioning for individual participants, then computed the difference in proportions ($\text{Hits}^{\text{After}} - \text{Hits}^{\text{Before}} = p^{diff}$). A positive (negative) p^{diff} indicates positively conditioned grammars were more (less) frequently selected after conditioning. A Type-2 mixed model with alphabet (2), SOA (3), and awareness (3) as factors produced a significant three-way interaction for explaining p^{diff} variance, $F(4, 99) = 3.05$, $p = .020$, $\eta_p^2 = .11$; $MS_{Error} = .125$. No other outcomes were significant (p's $> .2$).

Inspection of p^{diff} summaries faceted by SOA and awareness (see Figure 3) motivated one-sided tests to determine whether mean p^{diff}'s were greater than null estimates ($p^{diff} > 0$). We found credible evidence for $p^{diff} > 0$ across English grammars for *highly aware* participants exposed to 400-ms SOAs, $t(7) = 1.69$; $p = .27$; $L^{\text{Bayes}} =$

---

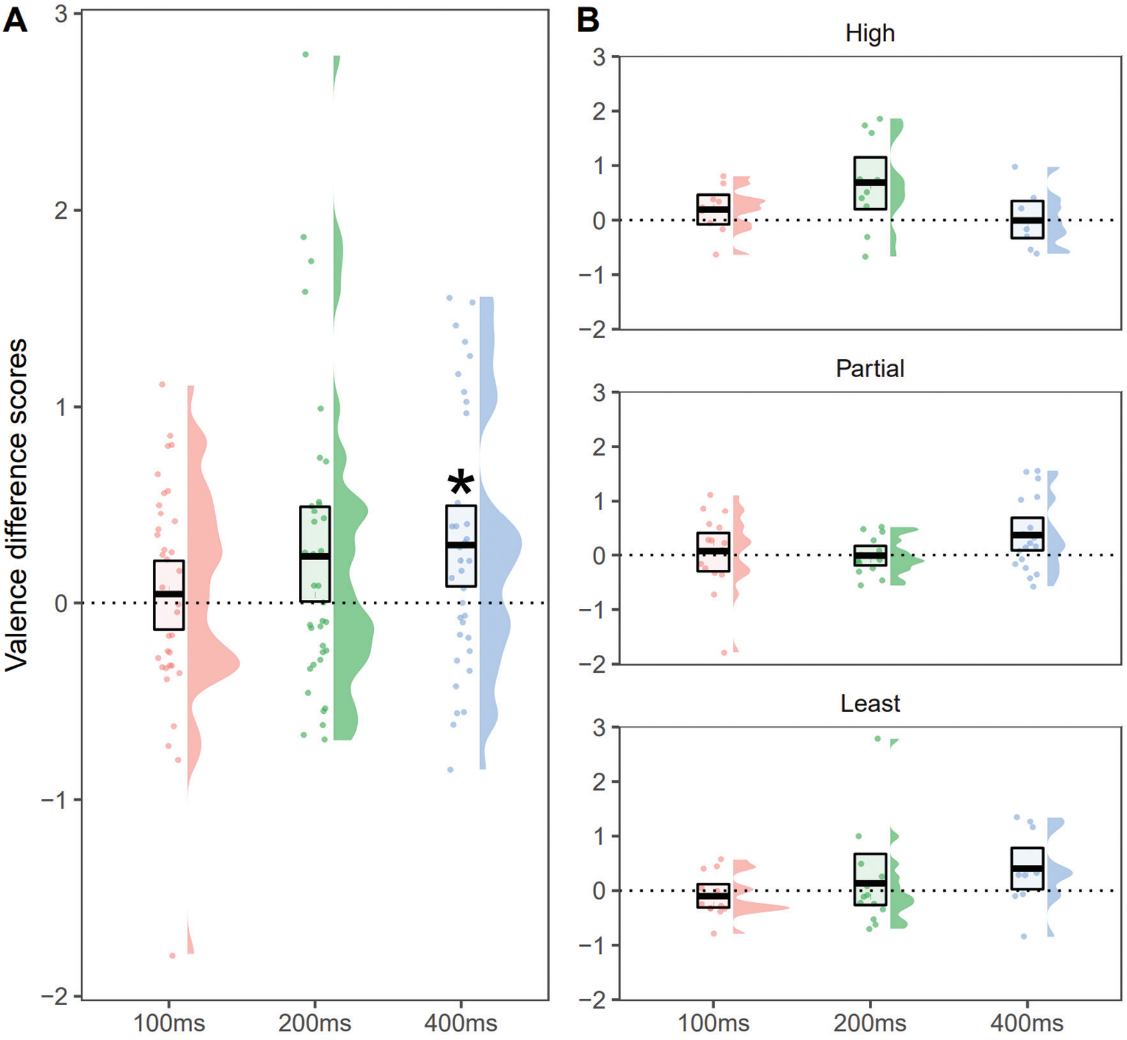
[2] Our awareness check varied from Jurchiș et al.'s (2020) in two important ways. First, the check appeared at the end of evaluation trials rather than after individual evaluations. Second, participants were asked to describe strategies without any strings being physically present, necessitating reliance on prior memories (which may not have been equally encoded across participants). These features collectively undermine our task's sensitivity and informational relevance as we could not estimate whether evaluative strategies varied across individual strings (Shanks & John, 1994). At best, our awareness levels represent differences in extant knowledge of global evaluative strategies. We return to this point in the General Discussion section.

[3] All g score estimates with 95% confidence intervals are available in Table S1 in the online supplemental materials.

[4] The null hypothesis was $g \leq 0$. All p-values were fdr-corrected to minimize detection of false positives and false negatives (Jafari & Ansari-Pour, 2019). Bayes likelihoods (L^{Bayes}) were estimated from posterior distributions derived from 10,000 Monte Carlo Markov chain (MCMC) simulations, which are sufficient to generate a normally distributed posterior (Kruschke, 2014, p. 184). L^{Bayes} describes whether posterior parameter distributions are somewhat ($L^{\text{Bayes}} > 70\%$), very ($L^{\text{Bayes}} > 80\%$) or extremely ($L^{\text{Bayes}} > 90\%$) likely to support the alternative hypothesis (relative to a null distribution) even when frequentist tests imply the null hypothesis cannot be statistically rejected (e.g., Amd & Passarelli, 2020). A continuous parameter of likelihood is more useful for assessing the credibility of alternative claims over binary claims of statistical significance (Kruschke, 2013; Kruschke & Liddell, 2018).

Figure 2

Distribution of Valence Difference Scores (y-Axes) Across Three Stimulus Onset Asynchrony (SOA) Groups (x-Axes) From Experiment 1



Note. Dots represent individual *g*-scores with half-violins illustrating their distribution. All scores were estimated as (Positive – Negative) differences. Crossbar plots indicate mean *g*-scores with 95% bootstrapped confidence intervals. Asterisks (*) illustrate significant (p's $< .05$) one-sided differences after correcting for multiple comparisons. Effect distributions faceted by strategy awareness (High, Partial, Least) are described in Panel B. See the online article for the color version of this figure.

92%; *partially aware* participants exposed to 400-ms SOAs, $t(17) = 2.15$; $p = .111$; $L^{\text{Bayes}} = 96\%$; and *least aware* participants exposed to 100-ms SOAs, $t(11) = 1.61$; $p = .233$; $L^{\text{Bayes}} = 93\%$ (Figure 3, top row). Credible evidence for $p^{diff} > 0$ across Phoenician grammars was observed for *highly aware* participants exposed to 100-ms SOAs only, $t(8) = 2.53$; $p = .106$; $L^{\text{Bayes}} = 99\%$ (bottom row). See Table S2 in the online supplemental materials for 2AFC performance summaries.

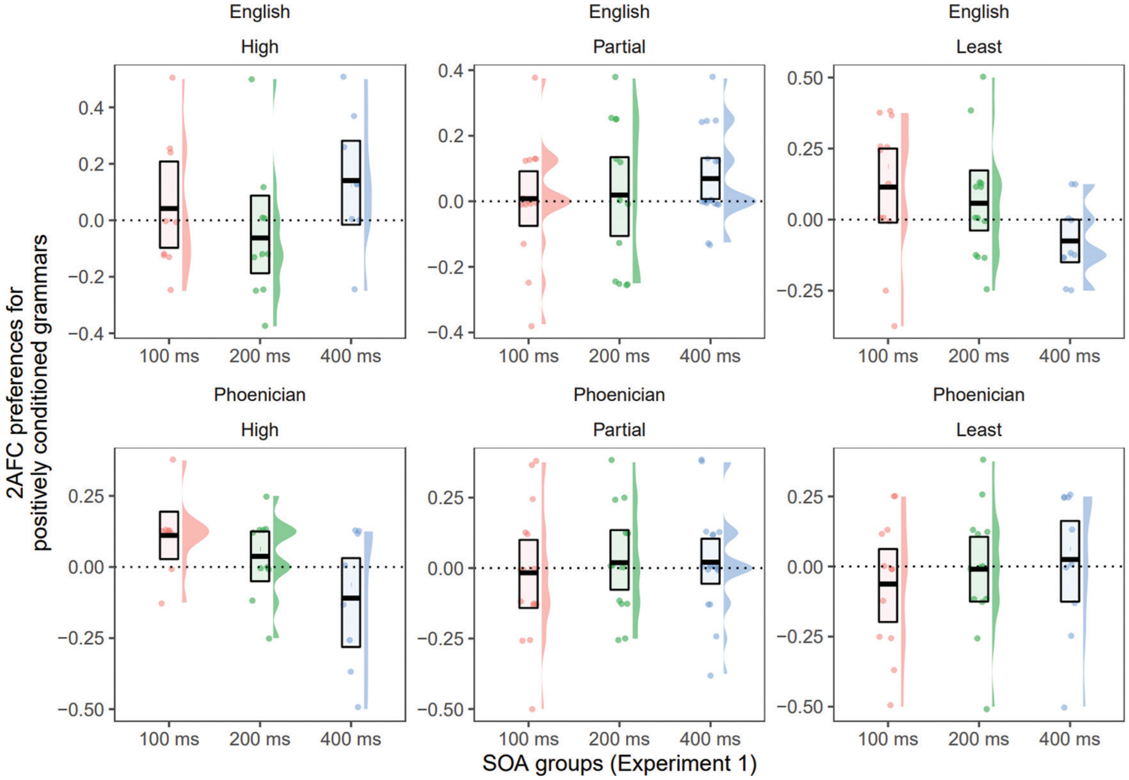
Memory Check

Participants viewed 32 English and 32 Phoenician strings across eight memory check trials. Across any given trial, participants viewed four familiar strings and four unfamiliar strings from the same alphabet. Familiar strings had appeared earlier as CS. Unfamiliar strings contained similar elements but were constructed using different grammar rules. Participants could select any number of strings within a fixed interval. Selection of familiar strings were scored as "hits"—selection of unfamiliar strings were scored as "false alarms." For each participant, we estimated proportions of hits and false alarms for strings from each alphabet, generating four outcome parameters (English hits, English false alarms, Phoenician hits, Phoenician false alarms; see Figure 4).

We ran four 3 × 3 Type-2 ANOVAs, with SOAs and awareness levels as between-subjects factors, to explain variance across each outcome parameter. Levene's tests were not violated for any model (p's $> .07$). No significant interactions or main effects for SOAs were detected. Main effects for awareness was significant across models explaining English false alarms, $F(2, 99) = 4.71$,

Figure 3

Dots Represent Individual 2-Alternative Forced Choice (2AFC) Proportion Differences Across Experiment 1



Note. Positive (>0) estimates along y-axes indicate positively conditioned grammars were selected more frequently after conditioning. Crossbar plots indicate mean proportion differences with 95% confidence intervals. stimulus onset asynchrony (SOA) conditions are described along x-axes. 2AFC performances are faceted by string alphabet (rows) and strategy awareness (columns). See the online article for the color version of this figure.

$p = .011$, $\eta_p^2 = .09$; $MS_{Error} = .126$, and Phoenician false alarms, $F(2, 99) = 1.83$, $p = .167$, $\eta_p^2 = .04$; $MS_{Error} = .052$. Overall, *least aware* participants ($M = .09$; $SE = .04$) misidentified unfamiliar English strings more frequently relative to *partially aware* ($M = .03$; $SE = .01$) and *highly aware* ($M = .04$; $SE = .03$) participants.

Evaluative Strategy Confidence

All participants indicated how confident they felt in the accuracy of their subjective evaluative strategies. A 3 × 3 Type-2 ANOVA explored whether confidence ratings were influenced by strategy awareness and SOA. The homogeneity of variance assumption was not violated ($p = .27$). The interaction was not significant ($p = .97$). Only a main effect of awareness was detected, $F(2, 99) = 5.59$, $p = .005$, $\eta_p^2 = .10$; $MS_{Error} = 9.73$. Pairwise tests indicated confidence ratings were statistically equivalent (p's $> .3$) across awareness levels during 100-ms and 200-ms SOAs. Across 400-ms SOAs, *highly aware* participants were significantly more confident relative to *partially* ($g = 1.02$ [.18 to 2.72]: $p = .06$) and *least aware* ($g = 1.14$ [.38 to 2.48]: $p = .02$) participants.

Discussion

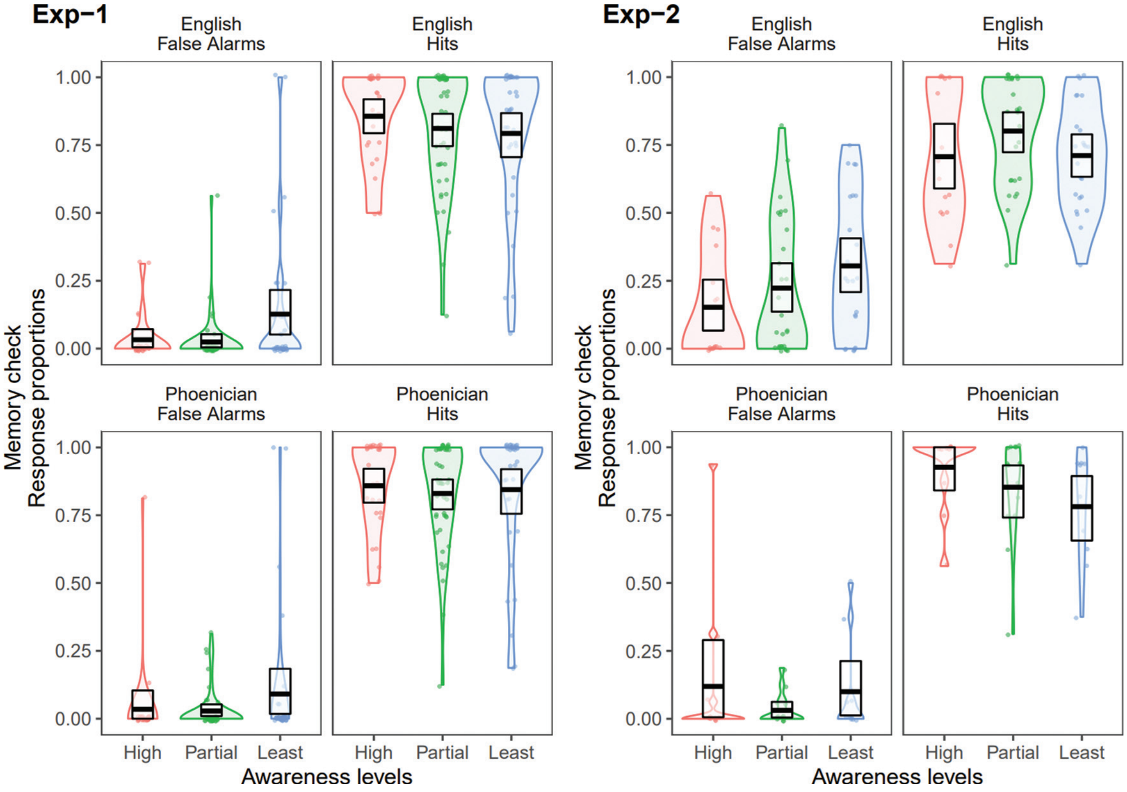
One-hundred and eight Fijian undergraduates viewed English and Phoenician letter strings (CS) with happy/angry faces (US). CS-US trials were sandwiched between evaluation and preference tests, which presented strings that never appeared as CS but were constructed using similar grammar rules. Stimulus processing times significantly influenced conditioning outcomes, with valence generalization becoming collectively more likely during longer SOAs. When parsed by awareness, MCMC-estimated likelihoods suggested *highly aware* participants exposed to 100 ms and 200 ms SOAs, as well as *least* and *partially aware* participants exposed to 400-ms SOAs, were very likely ($L^{\text{Bayes}} > 80\%$) to produce generalization.

Analysis of 2AFC preference tests were statistically inconclusive, although Bayes tests suggested English 2AFC performances were more likely to reflect conditioning effects relative to Phoenician performances. Across memory checks, all subgroups identified previously seen strings more frequently relative to strings that had never been seen. Memory check performances were not influenced by stimulus duration or alphabet familiarity. Parsed along awareness, *least aware* participants misidentified unfamiliar strings more frequently relative to remaining participants.

We replicated Jurchiș et al.'s (2020) findings and highlighted stimulus processing time as an important operating condition. Analysis of memory checks indicated *least aware* participants had misidentified unfamiliar strings more frequently relative to participants from other awareness levels. The same participants were also less confident in their evaluative strategy, suggesting they

Figure 4

Proportions of Hits and False Alarms Across English and Phoenician Strings (Facet Labels) Following Free-Selection Memory Checks Across Experiments 1 (Exp-1) and 2 (Exp-2)



Note. Dots and violins represent individual proportions and their distributions. Crossbars indicate mean proportions with 95% confidence intervals. Across Experiment 1 (left panel), all participants selected from English and Phoenician strings during memory check trials. Across Experiment 2 (right panel), participants selected from English or Phoenician strings if they had been assigned to PhoEval (top row) or EngEval (bottom row) groups, respectively. See the online article for the color version of this figure.

may have been least aware of task contingencies overall (Bar-Anan et al., 2010). On balance, because our awareness check appeared after evaluation trials and without any strings being physically present, our classifications likely reflected awareness of *global* over local evaluative strategies (recall Footnote 1).

Inspection of qualitative strategy descriptions {available in the online OSF file} suggests some participants had evaluated strings based on recurring and/or familiar elements. So, even though evaluated and conditioned strings never overlapped, structural features common to both sets may have contributed to generalization (e.g., trained and tested English strings sharing the trigram *VTR*). Combined with the observation that valence generalization was more likely during longer (≥ 200 ms) SOAs, a central role for resource-intensive deliberative processes can be inferred. On the other hand, uncertain outcomes across shorter SOAs may have been influenced by insufficient lags between stimulus presentations. A long history of associative learning research has demonstrated that a novel CS which appears "too soon" after an initial CS becomes less likely to be acquired since (inhibitory) elements of the latter may still be salient in active memory (Vogel et al., 2019). Applied to the present case, 100-ms intervals between CS may have been too brief for inhibitory traces from initial CS presentations to sufficiently decay, mitigating CS-US learning. Future works could explore this speculation more thoroughly and test for optimal CS-CS and CS-US intervals that facilitate evaluative conditioning of artificial grammars.

Our first experiment provides preliminary evidence that valences may generalize across strings with overlapping grammar structures, at least for participants exposed to SOAs of 200 ms or more. If generalization was mediated by abstracted "grammar rules" *viz* relations which specify how elements are organized relative to each other within a perceived structure (Spaulding, 1912), then mixed conditioning histories toward strings from different alphabets but the same grammar may have mitigated conditioning effects (recall Footnote 1). To see how, suppose a participant viewed English and Phoenician strings from grammar A with positive and negative US, respectively $\left(\frac{US+(Eng-A|Pho-B)}{US-(Eng-B|Pho-A)}\right)$, and another participant viewed both alphabets from grammar A with positive US exclusively $\left(\frac{US+(Eng-A|Pho-A)}{US-(Eng-B|Pho-B)}\right)$. If grammar rules are abstracted without term knowledge, then the former participant's mixed history would generate more variable evaluations relative to the latter participant. This limitation was controlled for in our second study, where string alphabets were varied between conditioning and evaluation phases. This ensured each alphabet-grammar combination was correlated with a unique valence category.

Experiment 2

Participants were randomly assigned to one of two groups (EngEval, PhoEval) in the present study. Participants assigned to *EngEval* viewed Phoenician strings during conditioning/memory trials and English strings during evaluation/2AFC trials. Alternatively participants assigned to *PhoEval* viewed English strings during conditioning/memory trials and Phoenician strings during evaluation/2AFC trials. Because conditioning trials presented strings from a single alphabet only, exemplars from individual grammars were exclusively associated with a single valence category. Additionally, because evaluated and conditioning strings were from different alphabets, familiarity with particular CS features mediating evaluations is less likely here relative to our first experiment. In sum, we tested whether valences could generalize across strings from different alphabets but overlapping grammars.

Method

Participants

One-hundred and sixty young adults from the United States were recruited from the academic site Prolific using a fixed-duration sampling strategy during November, 2020. Seven participants were excluded for failing attention checks, and 13 participants were excluded for variable Internet connection speeds, leaving a final sample of $N = 140$. These were randomly assigned to *EngEval* ($n = 70$; 24.2 ± 4.5 years, 26 females) and *PhoEval* ($n = 70$; 24.1 ± 4.4 years, 29 females) groups. Sensitivity analyses for one-sided pairwise tests indicated samples of $n = 70$ could reliably detect small-to-moderate effects (d's $> .3$) with 80% power. Participants were compensated at a rate of $8.50 per hour. All experimental procedures were completed within 30–40 min.

Materials

All materials from Experiment 1 were reused.

Procedure

Similar to the phase sequence in Experiment 1, participants completed 80 conditioning trials sandwiched by evaluation and preference tests. Across each conditioning trial, two strings from the same alphabet and grammar were randomly interleaved across sequences of happy or angry faces, as before. Different to our earlier sequence, each string was repeated once in varying combinations. To illustrate, suppose a participant initially viewed the string pairs *VTVTM-XXRVTM* and *VTTTVTM-XXRVTRTVM* (from ENGLISH-A; see online supplemental materials). Across later trials, these strings would be recombined in different sequences, such as *VTVTM-VTTTVTM* and *XXRVTRTVM-XXRVTM*, or some other configuration different to the initial sequence. Across *PhoEval* participants, English strings appeared during conditioning and memory checks, and Phoenician strings appeared during evaluation and preference tests. Assignment was reversed across *EngEval* participants, who viewed Phoenician CS and evaluated English strings. Grammar category assignment to happy and angry faces were counterbalanced across groups. All SOAs were held constant at 200 ms since our first experiment suggested shorter (100-ms) SOAs might be insufficient for serially presented strings to be equally encoded as CS (Vogel et al., 2019) whereas longer (400-ms) SOAs may facilitate deliberation by increasing processing times.

Results

Valence Evaluations

A 2 × 2 Type-2 ANOVA with valence (2) and group (2) respectively entered as repeated and between-subjects factors did not interact ($p = .36$) to explain variance across normalized evaluations. Main effects were observed for valence, $F(1, 138) = 7.14$, $p = .008$, $\eta_p^2 = .05$; $MS_{Error} = .05$, and group, $F(1, 138) = 7.79$, $p = .006$, $\eta_p^2 = .05$; $MS_{Error} = .15$.

Similar to Experiment 1, hedge-corrected difference scores were estimated for individual participants. One-sided tests produced credible evidence for generalization across *PhoEval* participants, $t(69) = 2.67$; $p = .009$; $L^{\text{Bayes}} = 99\%$; g [95%] = .16 [.1, .23], and *EngEval* participants, $t(69) = 1.38$; $p = .086$; $L^{\text{Bayes}} = 93\%$; g [95%] = .09 [.02, .15], with only the former (*PhoEval*) reaching statistical significance. Across awareness levels, credible ($L^{\text{Bayes}} > 89\%$) but statistically nonsignificant (p's $> .08$) evidence for generalization ($g > 0$) appeared across most subgroups, with the exception of "highly aware" *EngEval* participants ($L^{\text{Bayes}} = 30\%$). Individual valence effects and their central tendencies are illustrated in Figure 5.

2AFC Preferences

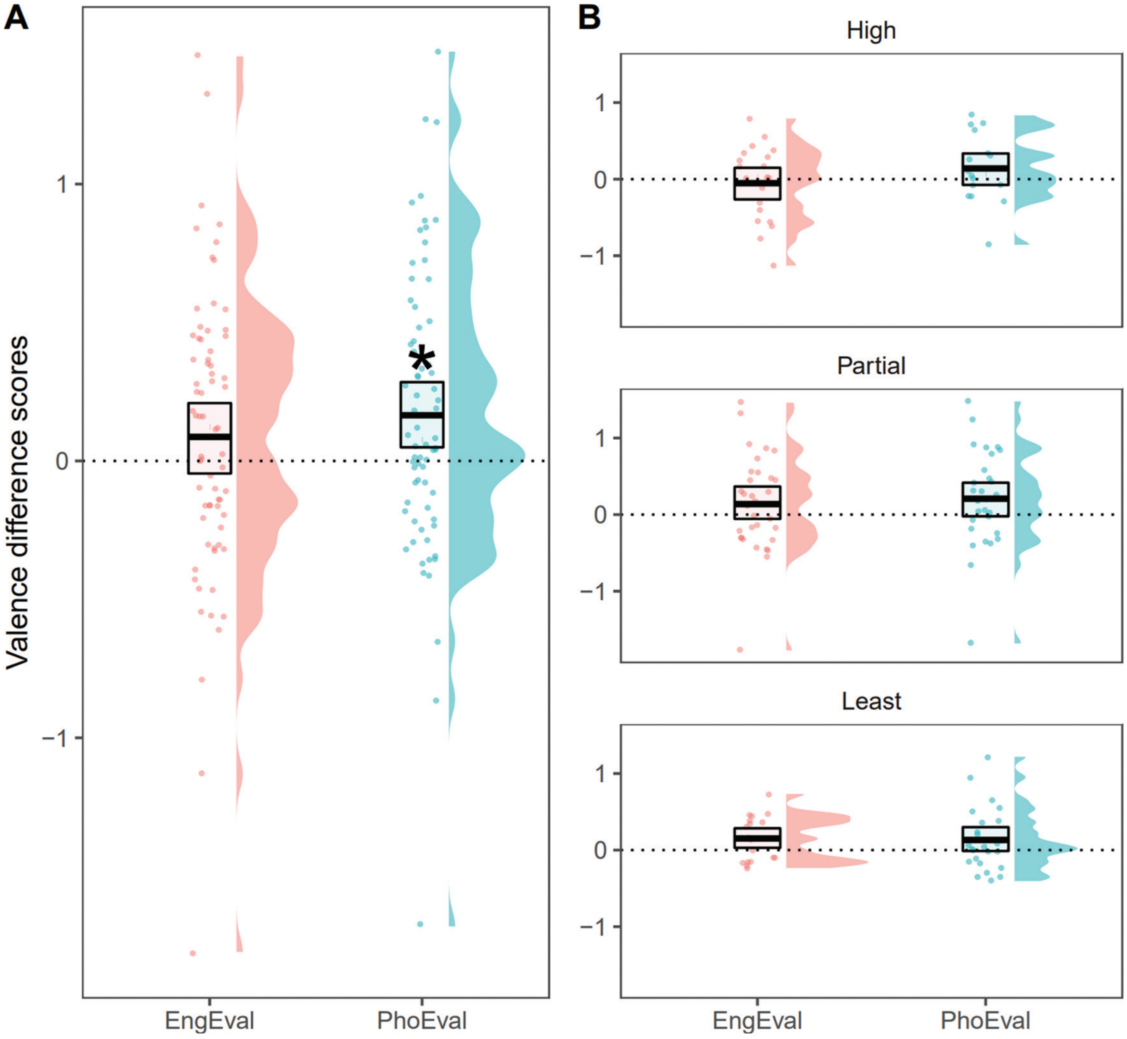
A 2 × 3 Type-2 ANOVA with group (2) and awareness (3) as between-subjects factors did not significantly interact ($p = .62$) when explaining variance in 2AFC proportion differences (p^{diff}). Only a main effect of group was found, $F(1, 134) = 8.03$, $p = .005$, $\eta_p^2 = .06$; $MS_{Error} = .302$. One-sided tests provided credible evidence for increased selections of positively conditioned grammars by *EngEval* participants, $t(69) = 1.27$; $p = .21$; $L^{\text{Bayes}} = 90\%$, and *PhoEval* participants, $t(69) = 5.57$; $p = .001$; $L^{\text{Bayes}} = 99\%$, with only the latter reaching significance. Parallel contrasts across awareness subgroups produced credible ($L^{\text{Bayes}} > 88\%$) but statistically nonsignificant (p's $> .11$) evidence for generalization for most subgroups, with the exception of "highly aware" *EngEval* participants ($L^{\text{Bayes}} = 37\%$). 2AFC performances are summarized in Figure 6 and Table S3 in the online supplemental materials.

Memory Check

We aggregated proportions of hits and false alarms for participants from *PhoEval* and *EngEval* groups as each group viewed strings from a single alphabet during memory checks. We ran two 2 × 3 Type-2 ANOVAs with group (2) and awareness (3) as between-subjects factors to explain variance across hits and false alarms respectively. Neither model produced significant interactions or main effects for awareness (p's $> .2$). Main effects for group were significant when explaining variance across hits, $F(1, 97) = 6.21$, $p = .014$, $\eta_p^2 = .06$; $MS_{Error} = .065$, and false alarms, $F(1, 97) = 10.15$, $p = .002$, $\eta_p^2 = .09$; $MS_{Error} = .079$. Inspection of Figure 4 (Exp-2) indicated *PhoEval* participants produced more false alarms and fewer hits relative to *EngEval* participants across most awareness levels, meaning Phoenician strings were more accurately detected than English strings.

Figure 5

Distribution of Valence Difference Scores (y-Axes) Across EngEval and PhoEval Groups (x-Axes) From Experiment 2, Faceted by Awareness Levels in Panel B



Note. Asterisks (*) illustrate significant (p =.009) one-sided differences relative to null (g = 0) estimates. See the online article for the color version of this figure.

Evaluative Strategy Confidence

A 2 × 2 Type-2 ANOVA with awareness and group entered as between-subjects factors, did not statistically interact (p = .70). Only a significant main effect of awareness was detected, $F(2, 134) = 16.30$, $p < .001$, $\eta_p^2 = .20$; $MS_{Error} = 25.93$. Across *EngEval*, pairwise tests confirmed *highly aware* participants were more confident then *partially aware* (g = .55 [–.01 to 1.2]: p = .05) and *least aware* (g = 1.39 [.65 to 2.63]: $p < .01$) participants. Within *PhoEval*, a parallel effect reached significance between *highly aware* and *least aware* participants only (g = 1.11 [.41 to 2.14]: $p < .01$).

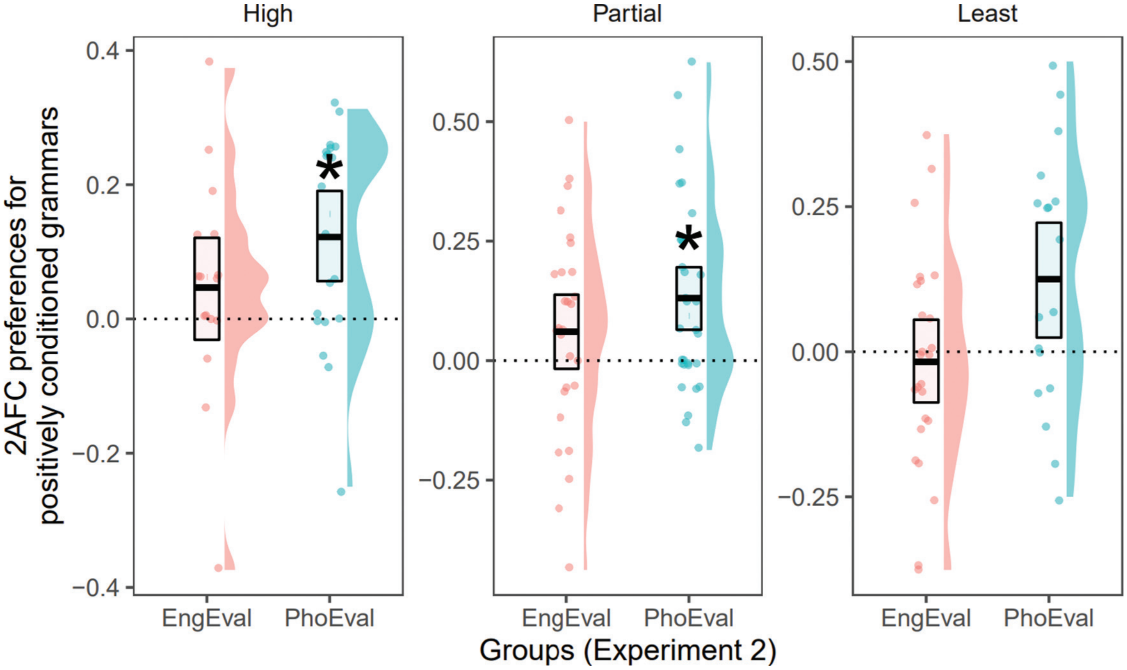
Discussion

One-hundred and forty American adults underwent grammar conditioning sandwiched by evaluation and preference tests. CS and evaluated strings were matched along grammar rules but varied along alphabet. Half our sample viewed Phoenician CS during conditioning and English strings during evaluations (*EngEval*). Remaining participants viewed English CS during conditioning and Phoenician strings during evaluations (*PhoEval*). Bayes tests revealed credible evidence for the alternative hypothesis (of valence generalization) for *EngEval* and *PhoEval* groups, with only the latter reaching statistical significance. Contrary to 2AFC outcomes reported across our first study, the majority of participants in the present work (5/6 awareness subgroups) produced credible evidence for a generalization effect. The variation in statistical significance rates may be considered weak evidence for a conditioning advantage across *PhoEval* participants.

Assuming valences generalize across structures perceived to be functionally equivalent (Amd et al., 2018; Tonneau, 2004b), we conjecture *PhoEval* participants were more likely to perceptually equate

Figure 6

Distribution of 2-Alternative Forced Choice (2AFC) Proportion Differences Across Experiment 2 Faceted By Awareness



Note. Participants who viewed English strings as conditioned stimulus and Phoenician strings during evaluations (PhoEval) selected positively conditioned grammars at significantly (*) higher frequencies after conditioning. See the online article for the color version of this figure.

congruent grammar structures relative to *EngEval* participants due to differential demands on their perceptual resources (Baddeley & Hitch, 2019). Recall that all participants in the second experiment had brief (200-ms) windows to process displayed CS. *PhoEval* participants, being already familiar with English characters, could be expected to have more perceptual resources available for detecting structural patterns across target strings and pay less attention to elemental features. Conversely, *EngEval* participants likely had to allocate more resources for discriminating characters from an unfamiliar dialect, leaving fewer resources for abstracting structure information. Analysis of memory check responses support this conjecture—*EngEval* participants correctly identified former CS more accurately relative to *PhoEval* participants, suggesting the former had dedicated more perceptual resources toward encoding string elements over abstracting grammar rules. It is also possible that *EngEval*'s familiarity with specific English letters/letter-combinations had interfered with information acquired during conditioning (Scott & Dienes, 2008; Scott & Dienes, 2010). Inspection of strategy descriptions provides partial evidence of evaluations being mediated by element attributes—for example, one participant reported *I remember seeing a lot of "x's," so I chose ones* (sic) *with this letter in*. Some/all of these possibilities may have contributed toward mitigating conditioning effects across *EngEval* participants.

Our second study provides further evidence that valences may generalize across overlapping grammars. Because trained and tested strings were constructed from different alphabets, familiarity with particular elements from earlier in the task sequence would be less likely to influence evaluations (Scott & Dienes, 2008). Note that for valences to generalize across grammars, the latter must have been perceived as "functionally equivalent" (Amd et al., 2013; Tonneau, 2004b). Stimuli can be functionally equivalent if they share overlapping structural and/or response properties (Berlyne, 1965; Tonneau, 2001). We had tacitly assumed that multiple exposures to strings from different grammars would suffice for abstracting and equating across grammar structures but had not empirically justified that claim. To address this oversight, we investigated whether strings from different alphabets became more likely to be perceived as functionally equivalent when they were from congruent grammar categories. Our final study tested the assumption that repeated exposures to multiple (grammar) exemplars can generate functionally equivalent relations between overlapping grammar rules.

Experiment 3

We adapted a 2AFC free-selection procedure (from Amd et al., 2017) and explored whether grammar-congruent discriminations become increasingly likely following exposures to sample and comparison strings varied along grammar and alphabet. During 2AFC trials, participants viewed English(/Phoenician) strings from grammars A/B as sample stimuli, followed by a pair of Phoenician (/English) comparison strings from grammars A and B. One of the presented comparisons were always from the same grammar category as the sample. Participants could freely select either comparison to progress the trial. Participant responses did not produce differential feedback, preventing strings from being functionally equated along common response properties (Berlyne, 1965, p. 49).

Each trial produced novel strings that were never repeated. We recorded whether freely selected comparisons were grammatically (in)congruent with the presented sample. We tested whether frequencies of grammar-congruent discriminations (GCDs) varied with increasing multiple exemplar exposures. If grammar structures are acquired following repeated exposures to grammar exemplars, as had been assumed across earlier experiments, we expected GCDs would become more frequent over time. Prior to analysis, test trials were binned into quartiles based on order of appearance. We estimated whether mean GCD frequencies ordinally varied between bins using two-sample Welch's tests. Increasing GCD frequencies would suggest grammar structures were perceived as functionally equivalent.

Method

Participants

Eighty-nine undergraduate USP students were recruited in exchange for course credit. A fixed-duration sampling strategy was followed from April to June 2021. Ten participants were dropped from analyses due to failing attention checks, leaving a final sample of $N = 79$ participants (21–35 years, 55 females). All participants were naïve to the experimental hypothesis and had not participated in any behavioral experiments previously.

Materials

All English and Phoenician strings from grammars A and B were interchangeably employed as samples and comparisons during test trials. The 2AFC discrimination task was designed and administered on E-Prime Go (Psychology Software Tools, n.d.), which participants accessed via online links. The task and associated data are available in the online supplemental materials.

Procedure

Participants viewed eight training trials followed by 80 test trials. Participants were instructed to categorize comparisons with samples based on which words "seem to go together." Across any given test trial, a sample (from grammar A or B) appeared near the top of the screen for 1,000 ms, followed by the onset of two comparisons (from grammars A and B) near the bottom of the screen. Sample and comparison strings were always from different alphabets. All three stimuli remained on screen until the participant pressed the letters "z" or "m" on the keyboard to select the left or right comparisons respectively. During training, participants viewed sample and comparison terms from the same alphabet (four trials with English strings, four trials with Phoenician strings), where one of the comparisons was a replica of the sample. Selecting the latter was followed by the feedback message "Correct!". Selecting the nonreplica produced a red X. The training phase served to orient participants to task demands while concurrently serving as an attention check: participants who produced more than one error during training (i.e., they did not match the sample with its replica despite receiving corrective feedback to do so) were dropped from analysis. After training, the program instructed participants to continue categorizing words that "seem to go together," while also being told that "no more corrective feedback" would be provided. Producing a keypress commenced the testing phase. This included 40 trials with Phoenician samples and English comparison pairs, and 40 trials with English samples and Phoenician comparison pairs. Trial sequences were randomized between participants. Selecting a comparison produced a blank 1,000-ms ITI with a fixation cross, followed by the subsequent trial. All stimuli remained on screen until a response was detected. After 80 trials, the task terminated and the program thanked participants for their time. No strings were repeated across trials.

Results

2AFC responses were binned into quartiles (*Bin-1/Bin-2/Bin-3/Bin-4*) by order of appearance. Each quartile represents 20 trials per participant, culminating to a total of 1,580 trials across 79 participants per quartile. Following removal of trials with response latencies $\leq$ 300 ms, the adjusted total trial counts for the four bins were 1,530 (*Bin-1*), 1,496 (*Bin-2*), 1,435 (*Bin-3*), and 1,442 (*Bin-4*). Counts of GCDs relative to bin-specific totals were 718/1,530; 712/1,496; 704/1,435, and 752/1,442. Mean GCD proportions across individual bins are illustrated in Figure 7, Panel A. Two-sample Welch's tests confirmed GCD frequencies were statistically higher across *Bin-4* relative to *Bin-1*, $t(2959.3) = -3$; $p = .008$, and *Bin-2*, t(2932.1) = –2.3; $p = .05$.

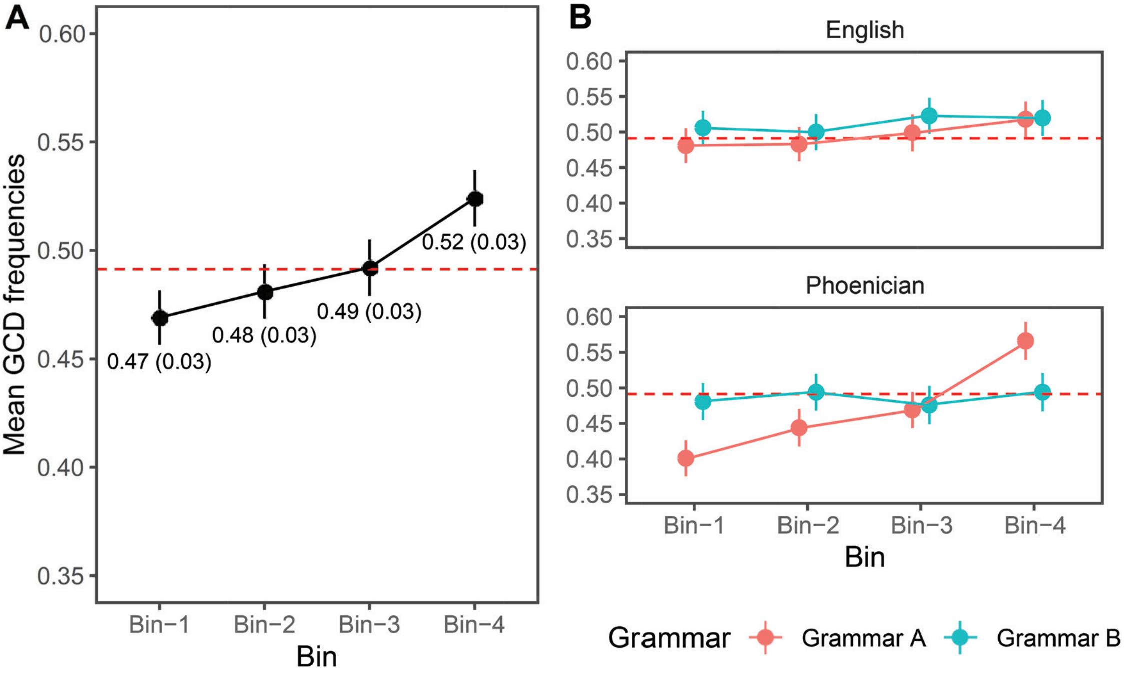
Given the asymmetrical evaluative effects generated by *PhoEval* and *EngEval* participants in Experiment 2, we decided to explore whether string alphabet and grammar structure had influenced GCD frequencies independent of bin sequence. GCD was logit-transformed to reliably estimate linear relationships with our categorical predictors (Peng et al., 2002). A two-predictor logistic regression showed that the predicted logit of (GCD) = $(-.02) + (-.10 \times \text{ALPHABET}^{sig}) + (.07 \times \text{GRAMMAR})$. According to the model, grammar-incongruent discriminations were more likely when Phoenician strings appeared as samples ($\beta = -.1$ [–.2 to 0]; $p = .05$). Mean GCDs parsed by alphabet and grammar are illustrated in Figure 7, Panel B. Across English samples, linear increases in GCD frequency were similar for strings from both grammars. Across Phoenician samples, a clear asymmetry was observed between grammars. On the one hand, GCD frequencies did not meaningfully vary when Phoenician samples were from grammar B. On the other hand, increasing GCD frequency was most evident when Phoenician samples were from grammar A. Kruskal-Wallis tests confirmed GCD frequencies had significantly increased for Phoenician samples from grammar A only, $\chi^2(3) = 20.3$; $p = .001$.

Discussion

Seventy-nine Fijian undergraduates underwent a 2AFC protocol where English (/Phoenician) strings from grammars A/B appeared as samples, followed by Phoenician (/English) strings from grammars A and B as comparisons. Participants were instructed to select comparisons that appeared to "go with" the presented sample. Over the course of 80 trials, GCD frequencies increased with multiple exposures to novel strings from both grammars, implying strings with overlapping grammar structures were increasingly being perceived as functionally equivalent (Tonneau, 2001). When parsed along alphabet and grammar, a statistical increase in GCD frequencies was noted across trials where Phoenician samples from grammar A (but not grammar B) had appeared. Across English samples, GCD frequencies increased at equivalent rates for both grammars.

Figure 7

Mean Frequencies (With Confidence Intervals in Parentheses) of Grammar-Congruent Discriminations (GCDs) During the 2-Alternative Forced Choice (2AFC) Free-Selection Task From Experiment 3



Note. Grammar-congruent discriminations (GCDs) are summarized along 20-trial bins, collapsed (A) and split (B) by alphabet and grammar categories. See the online article for the color version of this figure.

On the one hand, increased GCD frequencies in the presence of Phoenician samples corresponds with earlier claims of grammar structures being (more) readily acquired when structural elements are familiar (Experiment 2). If structural regularities are more readily discerned across comparisons constituting of familiar (English) instead of unfamiliar (Phoenician) elements, GCDs would be presumably more robust across the former. This seemed to be the case for GCD frequencies recorded for Phoenician samples from grammar A relative to English samples from both grammars. On the other hand, GCDs did *not* significantly increase when Phoenician samples from grammar B were presented. This is interesting as both grammars contained similar numbers of elements and element-combinations, leaving element familiarity as arguably the only distinguishing feature between string categories. Future investigations could explore the contribution of element familiarity to grammar conditioning by perhaps varying ratios of English and Phoenician elements within strings and across grammars, then comparing between GCD rates. It could be interesting to see if the collective familiarity of grammar elements influences structure acquisition, given the clear asymmetry in acquisition between Phoenician (but not English) grammars observed currently.

General Discussion

Our study expands on Jurchiș et al.'s (2020) report of valence generalization across artificial grammar structures in two important ways. First, across Experiment 1, we identified stimulus processing times (SOAs) as an important situational moderator—increasing CS/US durations predicted valence generalization for the majority of participants. The single exception were *highly aware* participants, who produced credible evidence for generalization during 100-ms SOAs.

Conditioning and evaluation phases across Experiment 1 and the procedure reported by Jurchiș et al. (2020) repeated elements between phases, raising the possibility that some participants may have generated evaluations based on idiosyncratic experiences with recurring/familiar letters (Scott & Dienes, 2008). This limitation was addressed in Experiment 2, where strings presented during conditioning and evaluation trials varied along alphabet. Results suggested strings from congruent grammar structures were evaluated in accordance with their conditioning histories, implying strings from different alphabets had become "functionally equivalent" on the basis of their overlapping grammars (Tonneau, 2004b). We found credible evidence for valence generalization across both groups in our second experiment, with statistically significant effects observed across *PhoEval* participants only. Analysis of memory checks across Experiment 2 revealed *PhoEval* participants generated more hits and fewer false alarms than *EngEval* participants—that is, English strings were recollected more accurately than Phoenician strings. A third study showed that repeated exposures to artificial grammar strings cause the latter to become functionally equated along overlapping grammar structures. Specifically, participants were observed to increasingly match grammatically congruent strings from different alphabets as "going together" following multiple exposures to novel strings from both grammars.

The first pair of experiments provide evidence for valence generalization across abstracted grammar structures. Because trained and tested strings contained no overlapping elements in Experiment 2, we had assumed there were no "common instances" that could mediate generalization (Brooks & Vokey, 1991). However, there remain other surface features that were not controlled for presently (e.g., letter sequences, Scott & Dienes, 2010; character fonts, Kuchinke et al., 2014), so claims of feature-mediated generalization cannot be dismissed just yet. It is even possible that feature-mediated generalization may have been more likely across Experiment 2 despite training and testing trials presenting different alphabets. To see how, recall that participants across Experiment 2 received more opportunities for acquiring grammar knowledge relative to participants from Experiment 1 (as strings were repeated during Experiment 2's training). If increased exposures facilitate perceptual learning of grammar structures (as "wholes"), the resulting knowledge could enable detection of a grammar's surface "parts," potentiating feature-mediated generalization. Just as the detection of a hypotenuse (a "part") is conditional on knowledge of right-angled triangles (it's "whole"; cf., Perry, 1912, pp. 107–109), the detection of overlapping surface features may be possible only after perceptual learning has taken place. A future work could investigate these speculations further by incorporating subjective knowledge checks across some of the tasks described here. This would help distinguish the extent to which (say) the 'free selection' performances reported presently were mediated by surface-level information or specified grammar knowledge.

Perceptual knowledge of grammar structures can be conceptualized as "organizing relations" which, unlike propositions, may be acquired and applied without specified term knowledge (Spaulding, 1912). A specified relation (e.g., *all swans are white*) can be subjectively evaluated as true/false but an organizing relation (*all __ are __*) may not. Assuming relations organize across perceptual regularities without detailed specification of perceived particularities allows for situational redundancy. For example, the organizing relation (*all __ are __*) can be redundantly applied across a practically infinite array of compatible symbols (e.g., *all swans are white, all leaves are green, all pigs are fat*) without assuming each propositional relation presupposes a "fragmented" representation.[5] By excluding requirements for a priori term knowledge, organizing relations are informationally simpler and ontologically prior to their situational applications (e.g., *all __ are __* must antecede the construction of the proposition *all swans are white*).

In the current study, grammar rules specified how elements were organized relative to one another across structures containing different elements. We propose "organizing relations" constituent to individual grammars had been acquired and functionally equated without assuming corresponding knowledge of particular elements. This would explain how novel strings could become perceived as functionally equivalent despite containing distinct elements (Experiment 3). Perceptions of functional equivalence, in turn, enables valence generalization across target structures (Amd et al., 2013, 2018; Amd & Roche, 2016, 2017). Presuming organizing relations can subsist independently of propositional relations can account for the present findings and related works (e.g., Jurchiș et al., 2020). Our presumption derives from a direct realist perspective, elaborated below.

Situating Organizing Relations

Within a direct realist framework, organizing relations are conceptualized as directly embodied across spatiotemporally extended "cross sections," which presume time and space as extended, nondiscrete, and phenomenally integrated (Holt, 1914; McMullen, 2018; Tonneau, 2013). These ontological assumptions enable cross sections to encompass a plurality of relations, including organizing (and affective; Boag, 2008) relations, which typically correspond with propositional/knowledge relations (Perry, 1912), but not always (Chew, 2016). We make no claims as to how "organizing relations" are mentally represented as, from a realist worldview, all experienced relations are environment-action couplings that become situationally activated in accordance with a context's affordances (Boodin, 1913; Tonneau et al., 2004). When activated relations come into conflict (Berlyne, 1965), their inadequate resolution may generate errors of memory (Chew, 2016) and perception (Tonneau, 2004a). By assuming organizing relations (grammar structures) can be abstracted and functionally equated "as is" (i.e., without detailed knowledge of their constituent elements), valences conditioned to members of a perceived structure can be predicted to generalize to other members of the same structure, as was noted here and by Jurchiș et al. (2020).

Our position does not dispute that evaluative effects are propositionally constructed—it is obvious that learning histories cannot be contextually reinstated without being relationally specified (e.g., Rantzen, 1993). We simply posit organizing (and affective) relations may not always overlap with propositional knowledge (e.g., Amd & Passarelli, 2020). Conceptualized as minimal informational units, organizing relations can be situationally redundant without positing truth criteria. By assuming organizing relations can be acquired independently of their terms, no assumptions are needed regarding the strengthening/weakening of unqualified mental links (Wills et al., 2019). These features render "organizing relations" conceptually distinct from propositional relations and mental associations (Spaulding, 1912). We conclude our discussion after addressing some limitations that potentially constrain the interpretations provided.

Limitations

Some concerns can be raised regarding our qualitative awareness check. First, the relevant response option was optional. Over half our sample chose not to respond, meaning no claims about those participants' subjective strategies can be made. We did not mandate responding at the end of the task to ensure participants did not feel "forced" to derive some arbitrary strategy postconditioning, which might have skewed reports (Hauser et al., 2018). Future replications could attempt to enforce mandatory awareness checks to explore whether the present response set correlates with their "forced" variants. Second, it is possible that the strategies participants reported awareness of may have been completely unrelated to task contingencies. Asking participants to affirm whether they were "aware of a specific strategy" does not inform us about the exact strategy applied. While our open-ended response option was designed to address this issue, responses were partially informative at best due to diagnosticity concerns (described below) and insufficient engagement.

---

[5] We are mostly in agreement with Porot and Mandelbaum's (2021) notion of "fragmented knowledge structures" in that organizing relations must be 'mandatorily' acquired (as relations cannot be incremental) and situationally redundant. We assume "redundancies" are embodied across directly experienced relations rather than their indirect representations (Tonneau et al., 2004).

The diagnosticity of our awareness check can be questioned on the grounds that it was implemented near the end of the task (rather than after each evaluation trial) and in the absence of any (previously evaluated) strings (Shanks et al., 2003; Shanks & John, 1994). Reports near the end of evaluation trials cannot inform whether strategies had varied across *individual* string evaluations. Asking participants if they "used a specific strategy" after all evaluations were complete is likely to reflect a global evaluative strategy at best. In response, we did not include a trial-by-trial awareness check to avoid priming attention to structural features over grammar knowledge, as this may have influenced valence attribution (e.g., March et al., 2018). Nevertheless, our inability to identify whether awareness levels varied across individual evaluations prevent us from claiming whether generalization was mediated by "unconscious" structures (Jurchiș et al., 2020). On balance, our conditioning tasks varied stimulus processing times, element familiarity and element recurrence. Procedurally, it seems unlikely that the current task would be *more* prone to induce contingency awareness relative to Jurchiș et al. (2020), though additional research would be required to (dis)confirm this claim.

Another concern may be the small-to-moderate effects reported presently, at least relative to those reported by Jurchiș et al. (2020). We had mentioned in the Introduction that Jurchiș et al. (2020) reported contrasts across postconditioning evaluations exclusively, whereas we ran tests across normalized values estimated from pre- and postconditioning evaluations. If we had run contrasts across postconditioning evaluations only, our effects would indeed have been more robust, but at the cost of concealing constituent (preconditioning) grammar valences. For example, the performances of *EngEval* participants from Experiment 2 produced a negligible difference (g = .08) following contrasts across normalized estimates. If we had contrasted across their postconditioning evaluations only, we would have reported a moderate effect (g = .21). Yet, inspection of preconditioning evaluations had indicated English grammars were differentially evaluated (g = .2) prior to any conditioning trials, raising concerns as to whether postconditioning evaluations reflected conditioned or constitutive valences (Silva, 2018). So, while normalizing across pre- and postconditioning values reduced effect magnitudes, the remaining effects can be interpreted with greater confidence as indicative of "genuine" valence generalization rather than preconditioning stimulus properties.

Finally, we had noted earlier that evaluative strategy confidence may reflect contingency awareness, in which case greater strategy confidence would be positively correlated with evaluative effects (Bar-Anan et al., 2010). Post hoc Spearman correlations between confidence ratings and valence difference scores found weak evidence for this claim across Experiment 1 (r_s = .17, p = .072) but not Experiment 2 (r_s = –.01, p = .895). A lack of statistical relations between confidence ratings and evaluative effects suggests grammar structures were "unconsciously" acquired, across which valences could have subsequently generalized (Jurchiș et al., 2020). Another possibility is that confidence ratings may have simply reflected dispositional differences in perceived confidence over actual contingency knowledge (Wolfe & Grosch, 1990). Perhaps (some) highly confident participants, by virtue of being more likely to "think they know" what the task required of them, had paid less attention to task contingencies and were less likely to acquire grammar knowledge. Alternatively, less (dispositionally) confident participants may have been less reliant on prior knowledge and paid more attention to task contingencies, increasing the likelihood of grammar knowledge acquisition. Future extensions could control for individual differences in dispositional confidence beforehand to increase the likelihood of confidence ratings reflecting contingency awareness over trait variables.

Conclusion

Jurchiș et al. (2020) claimed valence generalization across artificial grammars could be explained by "the unconscious formation of associations between elements of a grammar" (p. 8). The diagnostic limitations of our awareness check prevent us from claiming whether mediating knowledge structures were "unconscious"; however, because there were no common elements between trained and tested strings across Experiment 2, any interpretation requiring the formation of mental associations between common elements appears difficult to hold. On balance, Experiment 3's performances suggest perceptual knowledge of grammar structures (organizing relations) may have been incrementally (*associatively*) acquired. It is also possible that generalization could have been mediated by uncontrolled perceptual similarities, which we had speculated could have been *more* likely to be detected during Experiment 2. Future work will be required to fully rule out the possibility of perceptual/surface similarities mediating valence generalization across artificial grammars.

It could also be argued that the present performances were mediated by automated and/or deliberated propositional processes exclusively. Perhaps valences generalize across structures "only after a proposition" specifies their relationship, but what constitutes the "informational content" for the specifying proposition (De Houwer, 2018, p. 6)? Assuming causal chains constitute exclusively of propositional relations ultimately subsumes *all* psychological relations under propositional knowledge, but this can be questioned on logical and empirical grounds (Amd & Baillet, 2019; Amd et al., 2013; Amd & Passarelli, 2020; Kissler & Herbert, 2013; Morris, 2005; Spaulding, 1912). One could also assume combined operations of associative and propositional processes mediating performances (e.g., Gawronski & Bodenhausen, 2018) but this requires addressing "how" specified propositions interact with, and/or emerge from, unspecified associations in the first place (De Houwer et al., 2020; although see McLaren et al., 2019).

Within the realist perspective sketched out here, we assume a plurality of psychological (affective, organizing propositional) relations intersect across spatiotemporally extended cross sections to produce generalization (Holt, 1914; Tonneau, 2013). We propose relations "between elements" can be acquired without knowledge of their constituent elements (Spaulding, 1912), which enables relations specifying similar organizational patterns to be perceived as functionally equivalent despite constituting of dissimilar elements (Experiment 3). Inconsistent performances across string evaluations and strategy awareness may be due to conflicting overlaps across the plurality of available relations (Berlyne, 1965; Chew, 2016). There is some evidence that affective, motivational and organizing relations may overlap without being coupled to propositional knowledge (Amd & Baillet, 2019; Amd & Passarelli, 2020; Passarelli et al., 2022; Staats & Eifert, 1990). Future works will show whether so-called "organizing relations" can be acquired independently of (un)specified terms and/or behaviorally disassociated from other (affective, propositional) relations. In the meantime, we remain confident our findings can be accommodated, perhaps even challenged, by contemporary representationlist accounts (e.g., De Houwer et al., 2021; Gawronski & Bodenhausen, 2018). We welcome efforts to (dis)confirm and extend the present work, as they would only be conducive to our collective understanding of symbolic learning and valence generalization (Mowrer, 1960).

References

Amd, M., & Baillet, S. (2019). Neurophysiological effects associated with subliminal conditioning of appetite motivations. *Frontiers in Psychology*, *10*, Article 457. https://doi.org/10.3389/fpsyg.2019.00457

Amd, M., Barnes-Holmes, D., & Ivanoff, J. (2013). A derived transfer of eliciting emotional functions using differences among electroencephalograms as a dependent measure. *Journal of the Experimental Analysis of Behavior*, *99*(3), 318–334. https://doi.org/10.1002/jeab.19

Amd, M., de Almeida, J. H., de Rose, J. C., Silveira, C. C., & Pompermaier, H. M. (2017). Effects of orientation and differential reinforcement on transitive stimulus control. *Behavioural Processes*, *144*, 58–65. https://doi.org/10.1016/j.beproc.2017.08.014

Amd, M., de Oliveira, M. A., Passarelli, D. A., Balog, L. C., & de Rose, J. C. (2018). Effects of orientation and differential reinforcement II: Transitivity and transfer across five-member sets. *Behavioural Processes*, *150*, 8–16. https://doi.org/10.1016/j.beproc.2018.02.012

Amd, M., & Passarelli, D. A. (2020). Dissociating preferences from evaluations following subliminal conditioning. *Acta Psychologica*, *204*, Article 103023. https://doi.org/10.1016/j.actpsy.2020.103023

Amd, M., & Roche, B. (2016). A derived transformation of emotional functions using self-reports, implicit association tests, and frontal alpha asymmetries. *Learning & Behavior*, *44*(2), 175–190. https://doi.org/10.3758/s13420-015-0198-6

Amd, M., & Roche, B. (2017). Transforming valences through transitive inference: How are faces emotionally dissonant? *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *70*(12), 2478–2496. https://doi.org/10.1080/17470218.2016.1246576

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Arnold, J. B. (2021). *Ggthemes: Extra themes, scales and geoms for 'ggplot2'*. https://CRAN.R-project.org/package=ggthemes

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown*. https://github.com/crsh/papaja

Bååth, R. (2014). *Bayesian first aid: A package that implements Bayesian alternatives to the classical*. Test functions in R*. https://github.com/rasmusab/bayesian_first_aid

Baddeley, A. D., & Hitch, G. J. (2019). The phonological loop as a buffer store: An update. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *112*, 91–106. https://doi.org/10.1016/j.cortex.2018.05.015

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *The Quarterly Journal of Experimental Psychology*, *63*(12), 2313–2335. https://doi.org/10.1080/17470211003802442

Baumer, B., & Udwin, D. (2015). R markdown. *Wiley Interdisciplinary Reviews: Computational Statistics*, *7*(3), 167–177. https://doi.org/10.1002/wics.1348

Berlyne, D. E. (1965). *Structure and direction in thinking*. Wiley.

Boag, S. (2008). Mind as feeling or affective relations? *Theory & Psychology*, *18*(4), 505–525. https://doi.org/10.1177/0959354308091841

Boodin, J. E. (1913). Pragmatic realism–the five attributes. *Mind*, *22*(88), 509–525.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al.(1989). *Journal of Experimental Psychology: General*, *120*(3), 316–323. https://doi.org/10.1037/0096-3445.120.3.316

Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, L. G., & Petty, R. E. (1992). Rudimentary determinants of attitudes: Classical conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology*, *28*(3), 207–233. https://doi.org/10.1016/0022-1031(92)90053-M

Chew, L. M. (2016). *A direct realist understanding of memory errors* [PhD thesis]. Macquarie University.

Corneille, O., & Mertens, G. (2020). Behavioral and physiological evidence challenges the automatic acquisition of evaluations. *Current Directions in Psychological Science*, *29*(6), 569–574. https://doi.org/10.1177/0963721420964111

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, *13*(3), e28046. https://doi.org/10.5964/spb.v13i3.28046

De Houwer, J. D., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. *Advances in Experimental Social Psychology*, *61*, 127–183. https://doi.org/10.1016/bs.aesp.2019.09.004

De Houwer, J. D., Van Dessel, P., & Moran, T. (2021). Attitudes as propositional representations. *Trends in Cognitive Sciences*, *25*(10), 870–882. https://doi.org/10.1016/j.tics.2021.07.003

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*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition*, *25*(5), 603–637. https://doi.org/10.1521/soco.2007.25.5.603

Free, M. (2016). *What is a 4-way ANOVA?* https://www.quora.com/What-is-a-4-way-ANOVA

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. In J. M. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 59–127). Academic Press. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & Bodenhausen, G. V. (2018). Evaluative conditioning from the perspective of the associative-propositional evaluation model. *Social Psychological Bulletin*. *13*(3). https://doi.org/10.5964/spb.v13i3.28024

Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology*, *9*, Article 998. https://doi.org/10.3389/fpsyg.2018.00998

Head, J., Neumann, E., Helton, W. S., & Shears, C. (2013). Novel word processing. *The American Journal of Psychology*, *126*(3), 323–333. https://doi.org/10.5406/amerjpsyc.126.3.0323

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*, *136*(3), 390–421. https://doi.org/10.1037/a0018916

Holt, E. B. (1914). *The concept of consciousness*. The Macmillan Company.

Jafari, M., & Ansari-Pour, N. (2019). Why, when and how to adjust your p values? *Cell Journal*, *20*(4), 604–607. https://doi.org/10.22074/cellj.2019.5992

Jurchiș, R., Costea, A., Dienes, Z., Miclea, M., & Opre, A. (2020). Evaluative conditioning of artificial grammars: Evidence that subjectively-unconscious structures bias affective evaluations of novel stimuli. *Journal of Experimental Psychology: General*, *149*(9), 1800–1809. https://doi.org/10.1037/xge0000734

Kassambara, A. (2021). *Rstatix: Pipe-friendly framework for basic statistical tests*. https://CRAN.R-project.org/package=rstatix

Kay, M. (2021). *Ggdist: Visualizations of distributions and uncertainty*. https://doi.org/10.5281/zenodo.3879620

Kissler, J., & Herbert, C. (2013). Emotion, Etmnooi, or Emitoon?—Faster lexical access to emotional than to neutral words during reading. *Biological Psychology*, *92*(3), 464–479. https://doi.org/10.1016/j.biopsycho.2012.09.004

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Kruschke, J. K. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

Kuchinke, L., Krause, B., Fritsch, N., & Briesemeister, B. B. (2014). A familiar font drives early emotional effects in word recognition. *Brain and Language*, *137*, 142–147. https://doi.org/10.1016/j.bandl.2014.08.007

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Mandelbaum, E. (2015). *Associationist theories of thought*. Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/associationist-thought/

March, D. S., Olson, M. A., & Fazio, R. H. (2018). The implicit misattribution model of evaluative conditioning. *Social Psychological Bulletin*. *13*(3). https://doi.org/10.5964/spb.v13i3.27574

McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A dual-systems approach to attitudes. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 204–217). Guilford Press.

McLaren, I., McAndrew, A., Angerer, K., McLaren, R., Forrest, C., Bowditch, W., & Verbruggen, F. (2019). Mackintosh lecture: Association and cognition: Two processes, one system. *The Quarterly Journal of Experimental Psychology*, *72*(2), 98–117. https://doi.org/10.1177/1747021818766287

McMullen, T. (2018). Memory and its offspring: Realism and the "specious present." *Theory & Psychology*, *28* (2), 172–192. https://doi.org/10.1177/0959354318760741

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *The Behavioral and Brain Sciences*, *32*(2), 183–198. https://doi.org/10.1017/S0140525X09000855

Morris, M. (2005). Realism beyond correspondence. In H. Beebee & J. Dodd (Eds.), *Truthmakers: The contemporary debate* (pp. 49–65). Clarendon Press. https://doi.org/10.1093/acprof:oso/9780199283569.003.0004

Mowrer, O. H. (1960). *Learning theory and the symbolic processes*. Wiley. https://doi.org/10.1037/10772-000

Mowrer, O. H. (2013). *Psychology of language and learning*. Plenum Press.

Norman, E., Scott, R. B., Price, M. C., & Dienes, Z. (2016). The relationship between strategic control and conscious structural knowledge in artificial grammar learning. *Consciousness and Cognition: An International Journal*, *42*, 229–236. https://doi.org/10.1016/j.concog.2016.03.014

Passarelli, D. A., Amd, M., de Oliveira, M. A., & de Rose, J. C. (2022). Augmenting saliva, but not evaluations, through subliminal conditioning of eating-related words. *Behavioural Processes*, *194*, Article 104541. https://doi.org/10.1016/j.beproc.2021.104541

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, *96*(1), 3–14. https://doi.org/10.1080/00220670209598786

Perry, R. B. (1912). A realistic theory of independence. In E. B. Holt (Ed.), *The new realism: Cooperative studies in philosophy* (1st ed., pp. 99–151). The Macmillan Company.

Porot, N., & Mandelbaum, E. (2021). The science of belief: A progress report. *WIREs Cognitive Science*, *12*(2), e1539. https://doi.org/10.1002/wcs.1539

Psychology Software Tools. (n.d.). E-Prime. https://pstnet.com/products/e-prime/

Rantzen, A. J. (1993). Constructivism, direct realism and the nature of error. *Theory & Psychology*, *3*(2), 147–171. https://doi.org/10.1177/0959354393032001

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning & Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Ribeiro, G. W., Kawasaki, H. N., Menzori, L. R. F., Amd, M., de Rose, J. C., & de Souza, D. G. (2020). Emergent reading via stimulus pairing with orientation response. *The Psychological Record*, *70*(3), 397–410. https://doi.org/10.1007/s40732-020-00398-y

Rollston, C. (2020). The emergence of alphabetic scripts. In R. Hasselbach-Andee (Ed.), *A companion to ancient near eastern languages* (pp. 65–81). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119193814.ch4

RStudio Team. (2020). RStudio: Integrated development environment for R. RStudio, PBC. http://www.rstudio.com/

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1264–1288. https://doi.org/10.1037/a0012943

Scott, R. B., & Dienes, Z. (2010). Knowledge applied to new domains: The unconscious succeeds where the conscious fails. *Consciousness and Cognition: An International Journal*, *19*(1), 391–398. https://doi.org/10.1016/j.concog.2009.11.009

Shanks, D. R., & John, M. F. S. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, *17*(3), 367–447. https://doi.org/10.1017/S0140525X00035032

Shanks, D. R., Wilkinson, L., & Channon, S. (2003). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 248–261. https://doi.org/10.1037/0278-7393.29.2.248

Silva, F. J. (2018). The puzzling persistence of "neutral" conditioned stimuli. *Behavioural Processes*, *157*, 80–90. https://doi.org/10.1016/j.beproc.2018.07.004

Sokolov, E. N. (1963). Higher nervous functions; the orienting reflex. *Annual Review of Physiology*, *25*(1), 545–580. https://doi.org/10.1146/annurev.ph.25.030163.002553

Spaulding, E. G. (1912). A defense of analysis. In E. B. Holt (Eds.), *The new realism: Cooperative studies in philosophy* (1st ed., pp. 155–237). The Macmillan Company.

Staats, A. W. (1996). *Behavior and personality: Psychological behaviorism*. Springer.

Staats, A. W., & Eifert, G. H. (1990). The paradigmatic behaviorism theory of emotions: Basis for unification. *Clinical Psychology Review*, *10*(5), 539–566. https://doi.org/10.1016/0272-7358(90)90096-S

Staats, A. W., & Staats, C. K. (1958). Attitudes established by classical conditioning. *The Journal of Abnormal and Social Psychology*, *57*(1), 37–40. https://doi.org/10.1037/h0042782

Tonneau, F. (2001). Equivalence relations: A critical analysis. *European Journal of Behavior Analysis*, *2*(1), 1–33. https://doi.org/10.1080/15021149.2001.11434165

Tonneau, F. (2004a). Consciousness outside the head. *Behavior and Philosophy*, *32*(1), 97–123.

Tonneau, F. (2004b). Verbal understanding and Pavlovian processes. *The Behavior Analyst Today*, *5*(2), 158–169. https://doi.org/10.1037/h0100029

Tonneau, F. (2013). Neorealism: Unifying cognition and environment. *Review of General Psychology*, *17*(2), 237–242. https://doi.org/10.1037/a0032939

Tonneau, F., Abreu, N. K., & Cabrera, F. (2004). Sitting on the word “chair”: Behavioral support, contextual cues, and the literal use of symbols. *Learning and Motivation*, *35*(3), 262–273. https://doi.org/10.1016/j.lmot.2004.03.001

Vogel, E. H., Ponce, F. P., & Wagner, A. R. (2019). The development and present status of the SOP model of associative learning. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *72*(2), 346–374. https://doi.org/10.1177/1747021818777074

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H. (2021). *Forcats: Tools for working with categorical variables (factors)*. https://CRAN.R-project.org/package=forcats

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wills, A. J., Edmunds, C. E. R., Le Pelley, M. E., Milton, F., Newell, B. R., Dwyer, D. M., & Shanks, D. R. (2019). Dissociable learning processes, associative theory, and testimonial reviews: A comment on Smith and Church (2018). *Psychonomic Bulletin & Review*, *26*(6), 1988–1993. https://doi.org/10.3758/s13423-019-01644-3

Wolfe, R. N., & Grosch, J. W. (1990). Personality correlates of confidence in ones decisions. *Journal of Personality*, *58*(3), 515–534. https://doi.org/10.1111/j.1467-6494.1990.tb00241.xs

Received December 31, 2020
Revision received December 21, 2021
Accepted December 23, 2021 ■