Learning and Motivation 95 (2026) 102316

Contents lists available at ScienceDirect

Learning and Motivation

journal homepage: www.elsevier.com/locate/l&m

Implicit generalization of space-valence associations across artificial grammar categories

Micah Amd [1]

Department of Psychology, University of the South Pacific, Suva, Fiji

ARTICLE INFO

Keywords:
Polarity correspondence principle
Space-valence associations
Evaluative conditioning

ABSTRACT

Previous research on space-valence associations (SVAs) has demonstrated how emotional valences are differentially embodied across spatial dimensions. However, the extent to which SVAs can generalize to abstract categories, along with the extent to which transferred SVAs rely on conscious inferential processes, remains unaddressed. The present work addresses these gaps by linking natural relational poles spanning vertical, horizontal, temporal, and affective dimensions with non-recurring, meaningless strings from Artificial Grammar Categories (AGCs) using a matching-to-sample (MTS) task. Natural pole evaluations were highly differentiated before and after MTS, confirming established SVAs. Natural pole evaluations were also strongly and positively related to explicit MTS performance and strategy awareness. In contrast, novel AGC strings linked with affective and vertical poles evoked valence differentials comparable to their natural counterparts after MTS. However, this transfer was not observed for AGCs linked to horizontal or temporal poles. Importantly, successful AGC evaluative shifts were unrelated to explicit MTS categorization accuracy or strategy awareness. These findings suggest that while embodied valences can implicitly generalize to abstracted structures, the efficacy of this transfer depends heavily on the inherent cognitive stability of the spatial anchor.

A core feature of human cognition is the tendency to map abstract affective concepts onto concrete physical space (Pohl & Miklashevsky, 2025; Proctor & Cho, 2006). This phenomenon, commonly described as 'space-valence associations' (SVAs), illustrates how spatial coordinates are not affectively neutral. In a comprehensive meta-analysis on SVAs, Pohl and Miklashevsky (2025) reported robust evidence that individuals map positive and negative valence onto spatial quadrants without explicit training. Along a vertical axis for instance, positive and negative stimuli are typically associated with upper space and lower space respectively. Relatedly, along the horizontal axis, positive and negative stimuli are typically mapped onto right and left spaces respectively, although horizontal SVAs can be more variable due to inter-individual variability in culturally-assigned reading directions (Román et al., 2015), context-dependent motor consequences (Brouillet et al., 2015), body-specific handedness effects (Casasanto, 2009; Kong, 2013), and cumulative language experience (Tsaregorodtseva, Rinaldi & Marelli, 2025). Irrespective of individual differences in particular valence-coordinate mappings (e.g. 'left' being classified as more positive than 'right' – cf. Kong, 2013), the underlying observation that differential valences are mapped to opposite relational poles remains a well-established finding (Pohl & Miklashevsky, 2025). The phenomenon of SVAs can be explained as an expression of the Polarity Correspondence Principle, or PCP, proposed by Proctor and Cho (2006), who posited that binary oppositions, such as those defining opposite spatial poles (e.g., 'above' vs. 'below') are 'implicitly coded' with positive (+) and negative (-) polarity values. According to PCP, this polarity coding is a fundamental aspect of how stimuli

E-mail address: micah.amd@proton.me.

[1] This study was not pre-registered.

https://doi.org/10.1016/j.lmot.2026.102316
Received 24 March 2026; Received in revised form 23 June 2026; Accepted 14 July 2026

and responses are 'mentally represented', with response selection typically being faster and more accurate when the polarity of the stimulus corresponds with the polarity of the required response. While the binary positive and negative 'poles' implied by the PCP framework can suggest discrete, categorical mappings, evidence indicates that space-valence associations often operate along a continuous analog scale. Under this view, relative spatial gradients continuously track incremental changes in valence and magnitude, rather than reflecting an abrupt, binary switch in polarity (Freddi et al., 2015; see also Koch et al., 2023).

If opposite relational poles are systematically differentiated along valence, then linguistic terms denoting the former (e.g., the words 'above' vs. 'below', or 'left' vs. 'right') can be predicted to elicit comparable SVAs, as has been shown previously (Meier & Robinson, 2004; Pohl & Miklashevsky, 2025; for other physical dimensions, see Amd, 2024). On balance, differential evaluations of familiar poles cannot, by themselves, establish whether such effects arise *because* said terms denote relational poles, *or* because they carry socially learned connotations unrelated to relations under investigation. For example, in a study by Amd (2024), participants who explicitly associated the relational poles 'brighter' and 'darker' with semantically meaningless strings subsequently evaluated those strings in opposite directions, seemingly supporting an SVA account along the physical dimension of luminosity. We note 'seemingly' since the observed valence effects might have reflected appraisals of relational poles along a physical dimension of brightness, but they could have just as easily reflected a transfer of culturally entrenched connotations, given how 'brighter' can be synonymous with 'smarter' or 'more intelligent' in social settings, which would imbue positive valences independent of luminosity. This concern applies, in principle, to 'any' familiar stimulus category (e.g. arrows, hands, words) that can have socially established connotations independent of their reference to physical relational poles.

A central aim of the present study was to mitigate familiarity-related confounds during the assessment of SVAs by contingently linking novel, non-recurring, semantically meaningless strings from two Artificial Grammar Categories (AGCs). Briefly, AGCs comprise fixed sets of letters arranged according to category-specific structural rules, such as probabilistic letter transitions, that do not resemble natural language structures (Reber, 1967). By systematically rearranging the same set of letters according to different AGC rules, it is possible to generate multiple superficially novel, non-repeating strings (see Supplementary Table S1). In the present study, non-recurring strings from two distinct categories (AGC1 and AGC2) were assigned to opposing relational poles using a matching-to-sample (MTS) task (e.g., AGC1 strings were linked with 'above' and AGC2 strings with 'below' for one participant group – see Procedure). Critically, because each string was novel at the time of presentation, and all strings were composed of the same letters, any subsequent evaluative differences could not readily be attributed to familiarity or element-specific learning histories. If AGCs exhibit valence effects comparable to their naturally linked counterparts, where (say) the evaluative differences between AGC1 and AGC2 mirror those elicited by 'above' and 'below', we can claim SVA-related valences had generalized from familiar poles to abstracted grammar categories.

Alongside vertical or horizontal spatial poles, participants in the present study were trained along temporal ('before' versus 'after') or affective ('happy' versus 'unhappy') relational poles. The inclusion of temporal poles was motivated by prior demonstrations of words like 'before' and 'after' being inherently mapped along left/right or back/front axes, revealing a spatial correspondence (von Sobbe et al., 2019). According to the PCP, temporal poles should elicit valence differentials similar to spatial poles, as previously demonstrated (de la Fuente et al., 2014; von Sobbe et al., 2019; Janczyk et al., 2025). Lastly, including explicitly affective poles ('happy' versus 'unhappy') served as a baseline control. Affective terms are inherently differentiated due to direct verbal conditioning histories (Staats, 1996). By comparing the differentiation across spatial and temporal poles against this explicit affective baseline, we explored whether embodied spatial mappings generalize with the same robustness as inherently emotional terms (Amd, 2022).

An additional feature of the present study was the inclusion of trial-by-trial strategy-awareness checks following each evaluation trial, adapting a paradigm used by Jurchiș et al. (2020). In that study, meaningless AGC strings acquired valence after being paired with positive or negative stimuli following a Pavlovian conditioning protocol. Importantly, analyses of post-evaluation awareness checks identified a subset of participants who showed valence generalization despite reporting little or no awareness of any explicitly inferred strategy, suggesting the valences had generalized from affective stimuli to abstract grammatical structures 'implicitly', that is, in the relative absence of explicit strategic reasoning. Amd (2022) subsequently replicated and extended Jurchiș et al.'s findings across AGC strings composed of unfamiliar Phoenician letters (Experiment 2), showcasing how affective and structural properties may be appraised independently of consciously articulated strategies. Consequently, trial-by-trial awareness checks alongside evaluations of non-recurring AGC strings in the present study will reveal whether (any) transfer of valence from natural relational poles to abstracted AGC structures relies on explicitly derived strategies, versus relatively 'implicit' mappings.

A demonstration of SVA generalization to abstract grammatical structures would illustrate that spatial and temporal coordinates with inherent valences can directly influence the emotional appraisal of abstract linguistic configurations that carry no prior spatio-temporal or socially learned connotations (Amd, 2022; Amd, 2024; Jurchiș et al., 2020). Crucially, if this emotional loading can be observed 'implicitly' viz., independently of coherent attributional processes, we could claim that spatio-temporally embodied valences can impact novel symbolic structures without accompanying propositional attributions (Amd, 2023).

These research questions were investigated using two confirmatory (CH1, CH2) and two exploratory (EH1, EH2) hypotheses. First, consistent with the SVA literature, we predicted that natural relational poles (e.g., the words 'above' and 'below') would be significantly and non-negligibly different along valence (CH1). Because participants were already semantically familiar with these natural poles, we expected their evaluations would co-vary with relatively high levels of self-reported strategy awareness (CH2). Upon confirming these established phenomena, we tested our first exploratory hypothesis (EH1). Following previous works showcasing the generalization of valences from emotional terms to AGC structures (Amd, 2022; Jurchiș et al., 2020), we tentatively predicted that novel, non-recurring AGCs would become differentiated along valence after being contingently linked to the natural poles, demonstrating generalization to abstract structural categories (EH1). Furthermore, to test whether such transfer can occur in the relative absence of explicit awareness, our second exploratory hypothesis predicted that evaluative differentiation of AGCs would still be

observed among participant subsets showing chance-level MTS performance and low strategy awareness (EH2). Support for these exploratory hypotheses would suggest that SVAs can generalize to novel structural categories implicitly, without requiring explicit articulation.

1. Method

1.1. Data availability and ethics statement

All reported procedures were approved by the University of the South Pacific Research Ethics Committee (REC). Data, analysis scripts, and the REC approval letter have been included with the manuscript. This study was not pre-registered.

1.2. Participants

Two hundred and fifty-nine undergraduate students from the University of the South Pacific volunteered for the current study in exchange for course credit, or a cash payment of FJ$10.00, between the months of March 2023 and April 2024. Participants were recruited through an online participant pool and word-of-mouth. Thirty-eight participants were omitted from the analysis for not meeting the pre-training screening threshold (minimum 13/16 correct; see Procedure). The remaining sample of two hundred and twenty-one participants were distributed across four groups, classified as AFFECTIVE ($N = 54$), comprising thirty females ($M = 25.9$, $SD = 5.3$ years), eighteen males ($M = 27.5$, $SD = 5.2$ years), and six participants identifying as other genders ($M = 21.8$, $SD = 4.2$ years); VERTICAL ($N = 59$), comprising thirty-six females ($M = 27.3$, $SD = 5.3$ years), nineteen males ($M = 27.3$, $SD = 5.4$ years), and four participants identifying as other genders ($M = 25.5$, $SD = 5.7$ years); HORIZONTAL ($N = 53$), comprising thirty-nine females ($M =$

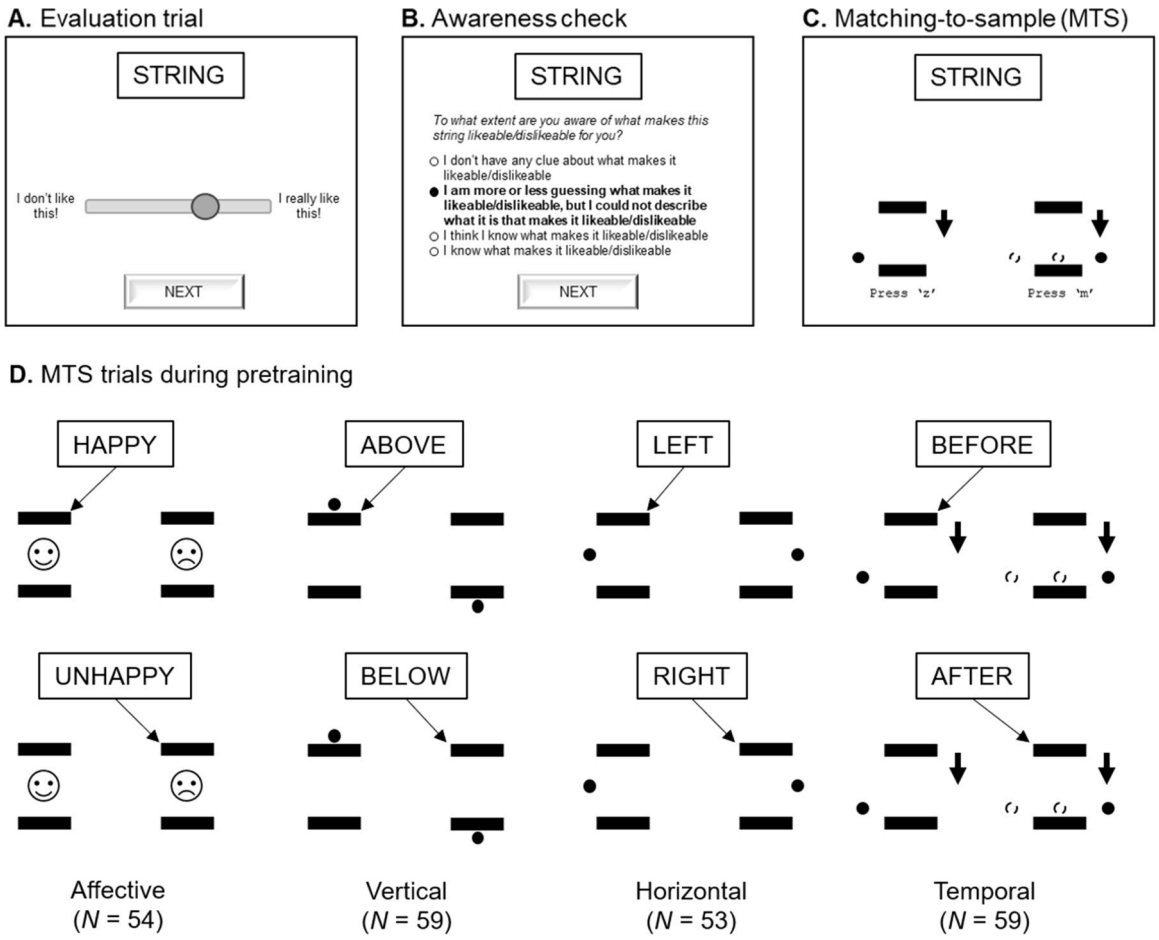

Fig. 1. Strings were evaluated along 9-point slider scales (Panel A), coupled with awareness checks comprising four response options (Panel B). Sandwiched between evaluation phases, participants underwent instrumental matching-to-sample (MTS) trials that presented natural poles or nonrecurring AGC strings as samples alongside comparisons depicting opposite relational poles (Panel C). All participants completed the same MTS pretraining trials, which required natural poles to be linked with their respective comparison type (Panel D).

26.6, *SD* = 5.5 years), eleven males (*M* = 28.3, *SD* = 4.8 years), and three participants identifying as other genders (*M* = 22.0, *SD* = 6.1 years); and TEMPORAL (*N* = 59), comprising thirty-two females (*M* = 26.3, *SD* = 5.6 years), sixteen males (*M* = 27.4, *SD* = 5.1 years), and eleven participants identifying as other genders (*M* = 28.4, *SD* = 4.8 years).

A 4 (*Group*: AFFECTIVE, VERTICAL, HORIZONTAL, TEMPORAL) x 2 (*Stimulus*: Natural, Artificial) x 2 (*Time*: Pre-MTS, Post-MTS) mixed analysis of variance (ANOVA), with group entered as a between-subjects factor, and with stimulus category (artificial, natural), and time (before, after) entered as within-subjects factors, was planned to explain variances across outcome parameters. A sensitivity power analysis was run on G*Power (Faul et al., 2007) to determine the minimum effect size detectable by the ANOVA design. The 'ANOVA: Fixed effects, special, main effects and interactions' test option was selected, with input parameters set to α= .05, power = .80, total sample size = 221, number of groups = 4 (the between-subjects factor levels), and numerator df = 15 (reflecting all main effects and interactions in a 4 ×2×2 design). The analysis indicated that the study was sufficiently powered to detect moderate-to-large effects ($\eta_p^2 = 0.083$). Participant classifications based on awareness checks were derived after data collection and not considered during power analyses. Participants completed all tasks in under 30 min on average.

1.3. Materials

Eight natural English words specifying opposite relational poles (ABOVE, BELOW, LEFT, RIGHT, BEFORE, AFTER, HAPPY, and UNHAPPY) appeared during evaluation and instrumental MTS trials. Additionally, one hundred letter strings, fifty per AGC, were deployed across evaluation and instrumental trials (see Supplementary Table S1). These were adapted from Amd (2022)'s materials, who presented the two categories as AGC-A and AGC-B. The mean number of characters across AGC-A (7.7) and AGC-B (7.66) were statistically non-distinguishable, t (98) = .15; p = 0.88.

In the present study, the assignment of AGC-A and AGC-B was counter-balanced between AGC1 and AGC2 categories. So, for approximately half of all participants, AGC-A and AGC-B were classified as AGC1 and AGC2 respectively. For the remaining participants, AGC-B and AGC-A were classified as AGC1 and AGC2 respectively. Because this assignment was fully counterbalanced, any item-specific variance was distributed evenly across conditions and was not modeled as an independent factor. Note that baseline evaluations of all AGC strings prior to the MTS task did not significantly differ from statistical noise thresholds (see Results), empirically confirming that neither category possessed an inherent valence that could systematically confound the learning outcomes.

Alongside natural and artificial terms, four pairs of comparison images, comprising eight achromatic images illustrating opposite relational poles, appeared during MTS trials (Fig. 1, Panel D). Comparison effectiveness for indicating relation poles was validated in an unpublished pilot study and during pretraining trials in the current study. All tasks were programmed on E-Prime 3 (Verdonschot et al., 2019), and distributed as self-contained executable files through the E-Prime Go website (https://pstnet.com/eprime-go/). Participants received links to download the task onto a Windows device, which instructed them to complete the task in a quiet environment and have a working internet connection. After participants completed all task phases, the program automatically uploaded the encrypted data to a secure server on the E-Prime website. Data analysis and manuscript preparation was conducted using R via the Posit interface (https://posit.co/), utilizing the following packages: *rprime* (Mahr, 2020), *emmeans* (Lenth, 2024), *glue* (Hester & Bryan, 2024), *purrr* (Wickham, Vaughan, et al., 2023), *rstatix* (Kassambara, 2023), *forcats* (Wickham, 2023), and *ARTool* for aligned rank transform analyses (Elkin et al., 2021; Wobbrock et al., 2011). Manuscript preparation was aided by the *apa* (Gromer, 2023), *kableExtra* (Zhu, 2024), and *knitr* (Xie et al., 2024) packages. All materials, data, and analysis scripts have been included as a supplementary file with the manuscript.

1.4. Procedure

The study began with participants evaluating eight natural words, eight AGC1 strings, and eight AGC2 strings in a randomized order. Each evaluation trial presented a target string near the top-center of the screen for 500 ms, followed by a 9-point slider near the bottom of the screen. The slider was anchored by the labels *I don't like this* (1) and *I really like this* (9) on the left and right sides, respectively (Fig. 1, Panel A). The slider tab was located at the scale midpoint at the start of each trial. Participants interacted with the slider to generate a *Next* button. Clicking this button immediately replaced the slider scale with the awareness check item: 'To what extent are you aware of what makes this string likeable/dislikeable for you?' The previous target string remained on screen (Panel B), and participants selected one of four response options: (a) 'I don't have any clue about what makes it likeable/dislikeable', (b) 'I am more or less guessing what makes it likeable/dislikeable, but I could not describe what it is', (c) 'I think I know what makes it likeable/dislikeable', or (d) 'I know what makes it likeable/dislikeable'. Selecting an option generated another *Next* button, which produced a blank inter-trial interval (ITI) of 500 ms before the subsequent trial. After completing all twenty-four evaluation and awareness trials, the program instructed participants of the upcoming matching-to-sample (MTS) task.

Each MTS trial commenced with a sample stimulus displayed near the top half of the screen for 500 ms, followed by two comparison stimuli near the bottom left and right sides of the screen. Samples comprised natural relational poles across the first 16 MTS trials, and non-recurring strings from AGC1 or AGC2 across the remaining 28 MTS trials. Comparison pairs comprised stimuli denoting opposite relational poles along spatial, temporal, or affective dimensions (Fig. 1, Panel C). Sample and comparison stimuli remained on screen until an acceptable keypress (the letters 'z' or 'm' for left or right comparisons, respectively) was detected.

Across the first 36 trials (all 16 natural sample trials and the first 20 AGC sample trials), keypresses produced the feedback message *Correct* in a green font for 1000 ms if the response was consistent with the target contingency, or *X* in a red font if inconsistent. Across the remaining 8 AGC trials, all responses produced blank ITIs of 1000 ms with no feedback. Only participants who mapped natural words to their corresponding relational pole comparisons correctly on at least 13 out of the 16 trials (>80% accuracy) were included in

the final dataset. This criterion ensured the inclusion of participants who adequately understood the relations indicated by the visual comparisons.

During the AGC trials, comparison pairs remained consistent with the participant's assigned group. For example, participants assigned to the TEMPORAL group viewed AGC samples alongside comparisons indicating *Before* and *After* relations only, exposing them to AGC1 → *Before* and AGC2 → *After* contingencies. Similarly, participants assigned to the VERTICAL (AGC1 → *Above*; AGC2 → *Below*), HORIZONTAL (AGC1 → *Left*; AGC2 → *Right*), or AFFECTIVE (AGC1 → *Happy*; AGC2 → *Unhappy*) conditions viewed visual comparisons indicating *Above* and *Below* positions, *Left* and *Right* positions, or *Happy* and *Unhappy* cartoon faces, respectively. After completing the MTS task, participants underwent a second round of 24 evaluation and awareness trials. This post-MTS phase comprised the exact eight natural words participants had seen previously, alongside eight entirely *novel*, non-recurring strings from each grammar category (AGC1 and AGC2). Completing this second round of evaluations and awareness checks marked the end of the experiment.

2. Results

2.1. Evaluative differences between natural and artificial poles

Evaluations produced for natural and AGC poles along 9-point scales are summarized in Table 1. String ratings were transformed into absolute normalized difference scores across individual participants to quantify the magnitude of valence differentiation between opposite relational poles for each stimulus type and time point. This transformation aimed to capture the degree of valence differentiation between poles while mitigating for individual variability in evaluative directionality, reducing scale-use bias (Amd, 2022; Amd, 2024). Scores were calculated using the following steps: first, for each participant, the mean evaluation rating was calculated for stimuli representing each relational 'pole' p within a stimulus category (e.g. $P1 = Above$, $P2 = Below$) at each time point t ($t1$ = Pre-MTS, $t2$ = Post-MTS). Let $_{P1,t}$ and $_{P2,t}$ represent the mean ratings for Poles 1 and 2 stimuli, respectively, within either the Natural (NAT) or Artificial (AGC) categories at time t. An absolute normalized difference score, reflecting the magnitude of differentiation between poles, can be estimated for each domain and time point using the formula:

$$d_{diff,t} = \left| \frac{\widehat{i}_{P2,t} - \widehat{i}_{P1,t}}{\widehat{i}_{P2,t} + \widehat{i}_{P1,t}} \right|$$

This yielded four difference scores per participant (D-NAT^{t1}, D-NAT^{t2}, D-AGC1^{t1}, and D-AGC2^{t2}). These corresponded to the levels of the within-subjects factors in the subsequent ANOVA: Natural Pre-MTS (D-NAT^{t1}), Artificial Pre-MTS (D-AGC^{t1}), Natural Post-MTS (D-NAT^{t2}), and Artificial Post-MTS (D-AGC^{t2}). Fig. 2 depicts individual participant d-scores alongside group means (Panel A), as well as aggregated means across groups (Panel B), with error bars representing 95% confidence intervals. Horizontal intercepts across panels indicate arbitrary noise thresholds (d =.1), which were deemed more conservative than conventional null estimates.

False-discovery rate (FDR) corrected one-sided contrasts exploring the null hypotheses that pole differences did not exceed statistical 'noise' thresholds (.1) were rejected across all natural pole contrasts before and after MTS (all p's < .026; all Hedge's g >.61 - see Rows 1 and 2 across Table 3). These outcomes conform to the SVA principle, demonstrating that linguistic terms denoting opposite horizontal, vertical, affective and temporal relational poles are differentially evaluated. Following contrasts of AGC poles, the null hypothesis was not rejected for any group before MTS trials (p's > .14; g's < .39 - see Row 3, Table 2). Following MTS however, AGCs linked with affective, vertical and temporal poles were significantly differentiated beyond a noise threshold (p's < .013; g's > .67 – Row 4), with no significant differences being observed for AGCs linked with horizontal poles (p > .9, g <.1).

Table 1
Descriptive summaries of evaluations before and after matching-to-sample (MTS) trials.

Group	Relational Direction	Stimulus Category	Mean (SD) before MTS	Mean (SD) after MTS
VERTICAL	ABOVE	Artificial	3.22 (2.16)	4.63 (2.41)
		Natural	7.51 (1.69)	7.03 (2.17)
	BELOW	Artificial	3.23 (2.2)	3.06 (2.12)
		Natural	4 (2.04)	3.77 (1.87)
HORIZONTAL	LEFT	Artificial	3.8 (2.79)	3.5 (2.81)
		Natural	4.83 (2.11)	4.58 (2.2)
	RIGHT	Artificial	3.69 (2.76)	3.94 (2.87)
		Natural	6.81 (1.65)	6.82 (2.03)
TEMPORAL	BEFORE	Artificial	3.08 (2.26)	3.42 (2.41)
		Natural	4.67 (2.1)	4.59 (2.17)
	AFTER	Artificial	2.92 (2.22)	4.27 (2.82)
		Natural	7.07 (2.2)	6.85 (1.67)
AFFECTIVE	HAPPY	Artificial	3.16 (2.42)	4.42 (2.38)
		Natural	8.32 (1.16)	8.05 (1.45)
	UNHAPPY	Artificial	3.02 (2.32)	2.24 (1.31)
		Natural	1.78 (1.03)	1.79 (1.09)

Note. Mean evaluations, with standard deviations, across natural and artificial grammar strings provided before and after matching-to-sample (MTS) trials.

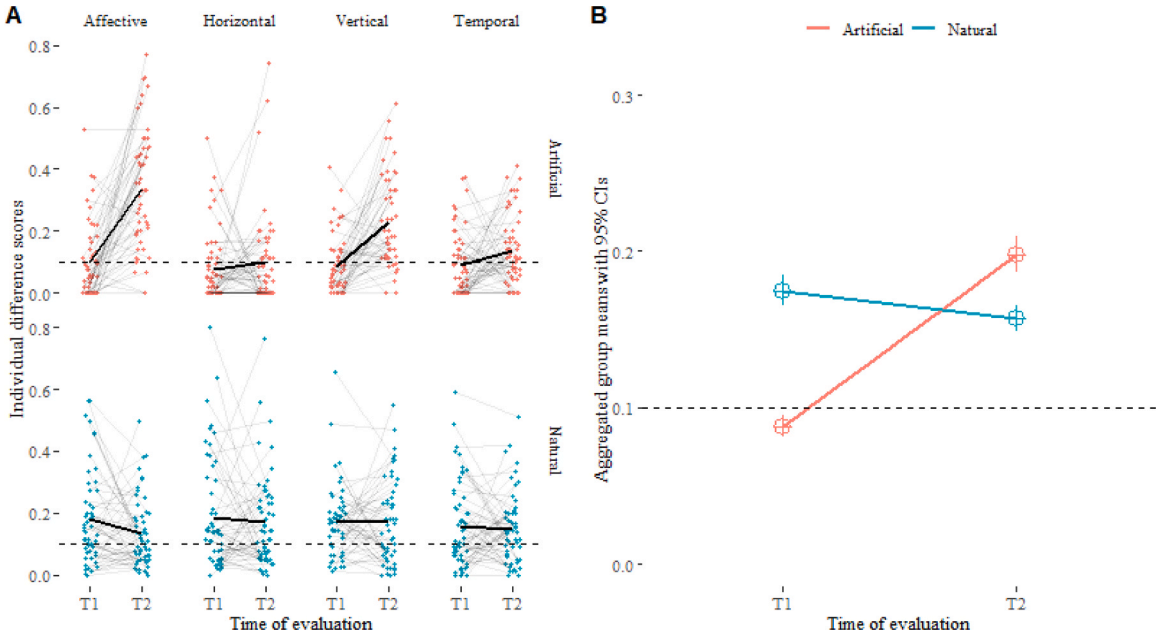

Fig. 2. Points representative of individual *d*-scores and group-aggregated grand means with 95% confidence intervals collected before and after MTS trials (*x*-axis) are depicted across Panels A and B respectively. Across Panel A, solid lines indicate evaluative trends before and after MTS. Dashed intercepts across both panels illustrate an arbitrary noise threshold for differentiating negligible ($d \leq .1$) from meaningful ($>.1$) effects.

Table 2
Awareness response frequencies by group, stimulus category, and time of evaluation.

Group	Category	Time	Minimally Aware	Partially Aware	Moderately Aware	Maximally Aware
AFFECTIVE	Artificial	T1	155 (33.4%)	84 (18.1%)	112 (24.1%)	113 (24.4%)
		T2	92 (19.8%)	101 (21.8%)	132 (28.4%)	139 (30%)
	Natural	T1	39 (8.4%)	60 (13%)	113 (24.4%)	251 (54.2%)
		T2	23 (5%)	40 (8.6%)	139 (30%)	262 (56.5%)
HORIZONTAL	Artificial	T1	203 (43%)	82 (17.4%)	78 (16.5%)	109 (23.1%)
		T2	178 (37.7%)	78 (16.5%)	88 (18.6%)	128 (27.1%)
	Natural	T1	52 (11%)	55 (11.7%)	143 (30.4%)	221 (46.9%)
		T2	63 (13.4%)	52 (11%)	136 (28.9%)	220 (46.7%)
VERTICAL	Artificial	T1	106 (24.5%)	88 (20.4%)	115 (26.6%)	123 (28.5%)
		T2	84 (19.4%)	89 (20.6%)	122 (28.2%)	137 (31.7%)
	Natural	T1	35 (8.1%)	66 (15.3%)	143 (33.2%)	187 (43.4%)
		T2	36 (8.4%)	54 (12.6%)	104 (24.2%)	236 (54.9%)
TEMPORAL	Artificial	T1	176 (37.3%)	132 (28%)	73 (15.5%)	91 (19.3%)
		T2	150 (31.8%)	124 (26.3%)	85 (18%)	113 (23.9%)
	Natural	T1	39 (8.3%)	75 (15.9%)	115 (24.4%)	243 (51.5%)
		T2	38 (8.1%)	60 (12.7%)	107 (22.7%)	267 (56.6%)

Note. Values represent frequencies of awareness responses before (T1) and after (T2) instrumental MTS trials. Proportions of awareness responses relative to the total is displayed in parentheses.

2.2. Omnibus tests

A 4 (*AFFECTIVE*, *VERTICAL*, *HORIZONTAL*, *TEMPORAL*) x 2 (Natural, Artificial) x 2 (Pre-MTS, Post-MTS) mixed ANOVA using Type III sums of squares (given the unbalanced design) was planned to explain variance across *d*-scores. Levene's test indicated homogeneity of variances across groups was violated ($p < 0.01$). Relatedly, Shapiro-Wilk tests suggested the data was not normally distributed for any group (all p's $< .05$). Additionally, *d*-scores were significantly autocorrelated across time for natural stimuli, $r\,(219) = .23$, $p < .001$, but not artificial stimuli, $r\,(219) = .04$, $p = .578$. These attributes prompted the decision to subject the data to an Aligned Rank Transform, or ART, using the *ARTools* R package (Elkin et al., 2021). This involved transforming the *d*-scores into rank scores which were then aligned for each main effect and interaction term before being submitted to ANOVA (Elkin et al., 2021). Unlike conventional non-parametric approaches (e.g. Friedman, Welch, Kruskal-Wallis), ART can handle multiple factors and test for interactions while remaining robust to normality violations, unbalanced samples and autocorrelation artefacts (Wobbrock et al., 2011). Partial-eta-square (η^2p) estimates were computed from observed F-values and the degrees of freedom. Even though ART operates on rank-transformed data, η^2p remains a robust and standard metric for evaluating the magnitude of these effects, reflecting the

proportion of variance explained within the aligned-rank space rather than the raw observations (Wobbrock et al., 2011).

The ART-ANOVA produced a significant three-way interaction, F (3, 217) = 18.58, $p < .001$, η^2p = .204, between group, time of evaluation, and stimulus category. Significant two-way interactions were observed between time of evaluation and group affiliation, F (3, 217) = 7.72, $p < .001$, η^2p = .096, between stimulus category and group affiliation, F (3, 217) = 11.83, $p < .001$, η^2p = .141, and between time of evaluation and stimulus category, F (1, 217) = 78.36, $p < .001$, η^2p = .265. The pattern of significant interactions reveals that the specific relations assigned to groups explained the variance across stimulus categories and time points. Finally, main effects were significant for stimulus category, F (1, 217) = 4.03, $p = .046$, η^2p = .018, evaluation time, F (1, 217) = 39.47, $p < .001$, η^2p = .154, and group affiliation, F (3, 217) = 14.67, $p < .001$, η^2p = .169.

2.3. Post-hoc pairwise contrasts

To unpack the significant three-way interaction and directly test our hypotheses, we examined the specific pairwise contrasts of interest (with Tukey adjustments for multiple comparisons; for the full contrast matrix, see Supplementary Table S2). First, addressing our confirmatory baselines, evaluations of the natural relational poles remained highly stable across the experiment. There were no significant changes in the evaluation of natural words from Time 1 to Time 2 across the AFFECTIVE, VERTICAL, HORIZONTAL, or TEMPORAL groups (all $ps > .99$). This confirms that familiar relational poles maintained stable valence differentials, irrespective of the MTS contingencies.

In contrast, exposure to MTS contingencies significantly influenced AGC evaluations. Specifically, the evaluative differentiation of artificial categories increased significantly from pre-MTS to post-MTS for participants in the baseline AFFECTIVE group ($t = -8.124$, $p < .001$), confirming that direct emotional contingencies readily transfer to novel structural categories. A significant transfer effect was also observed for participants in the VERTICAL condition, with novel AGCs linked to "above" and "below" poles becoming significantly differentiated following the MTS task ($t = -6.08$, $p < .001$). However, the pre-to-post MTS evaluative shifts for AGCs in the TEMPORAL condition ($t = -3.125$, $p = .129$) and the HORIZONTAL condition ($t = -1.141$, $p > .99$) were non-significant. These findings indicate abstract grammatical structures can inherit the evaluative properties of vertical spatial axes just as effectively as explicit affective terms.

2.4. Awareness of evaluative strategies

All participants were asked 'To what extent are you aware of what makes this string likeable/dislikeable for you?' while a (natural/artificial) evaluated string remained displayed on the screen (Fig. 1, Panel B). Participants could respond with one of four options, which were scored as 1 (*I don't have any clue about what makes it likeable/dislikeable*), 2 (*I am more or less guessing what makes it likeable/dislikeable, but I could not describe what it is that makes it likeable/dislikeable*), 3 (*I think I know what makes it likeable/dislikeable*) and 4 (*I know what makes it likeable/dislikeable*). These scores were interpreted to respectively indicate minimal (1), partial (2), moderate (3) or maximum awareness (4) of an evaluative strategy. Frequencies of awareness responses for each combination of group, stimulus

Table 3
Outcomes following 1-sample tests (CH1, EH1) and Spearman's correlations (CH2, EH2) across four groups.

Tested Hypotheses	Implied Outcomes	Affective	Horizontal	Vertical	Temporal
CH1a: $D\text{-}NAT^{t1}$ > .1	Natural poles are differentiated before MTS	0.18 [0.16, 0.2]; $p < .001$; $g = 1.08$	0.19 [0.16, 0.21]; $p < .001$; $g = 0.97$	0.17 [0.16, 0.19]; $p < .001$; $g = 1.23$	0.16 [0.14, 0.18]; $p = 0.002$; $g = 0.86$
CH1b: $D\text{-}NAT^{t2}$ > .1	Natural poles are differentiated after MTS	0.14 [0.12, 0.15]; $p = 0.026$; $g = 0.62$	0.17 [0.15, 0.19]; $p < .001$; $g = 0.97$	0.17 [0.15, 0.19]; $p < .001$; $g = 1.06$	0.15 [0.13, 0.16]; $p = 0.002$; $g = 0.87$
EH1a: $D\text{-}AGC^{t1}$ > .1	AGC categories are differentiated before MTS.	0.1 [0.08, 0.11]; $p = 0.895$; $g = -0.04$	0.08 [0.06, 0.09]; $p = 0.148$; $g = -0.39$	0.09 [0.07, 0.1]; $p = 0.273$; $g = -0.3$	0.09 [0.08, 0.11]; $p = 0.545$; $g = -0.16$
EH1b: $D\text{-}AGC^{t2}$ > .1	AGC categories are differentiated after MTS.	0.34 [0.31, 0.36]; $p < .001$; $g = 2.43$	0.1 [0.08, 0.12]; $p = 0.994$; $g < 0.1$	0.23 [0.21, 0.25]; $p < .001$; $g = 1.75$	0.14 [0.12, 0.15]; $p = 0.013$; $g = 0.67$
CH2a: $\rho\text{-}NAT^{t1}$ > .1	Natural pole evaluations correlate with awareness before MTS.	ρ (463) = 0.47; $p < .001$	ρ (471) = 0.23; $p < .001$	ρ (431) = 0.26; $p < .001$	ρ (472) = 0.41; $p < .001$
CH2b: $\rho\text{-}NAT^{t2}$ > .1	Natural pole evaluations correlate with awareness after MTS.	ρ (464) = 0.55; $p < .001$	ρ (471) = 0.38; $p < .001$	ρ (430) = 0.28; $p < .001$	ρ (472) = 0.44; $p < .001$
EH2a: $\rho\text{-}AGC^{t1}$ > .1	AGC evaluations correlate with awareness before MTS.	ρ (464) = −0.14; $p = 0.002$	ρ (472) = −0.06; $p = 0.191$	ρ (432) = −0.27; $p < .001$	ρ (472) = −0.11; $p = 0.015$
EH2b: $\rho\text{-}AGC^{t2}$ > .1	AGC evaluations correlate with awareness after MTS.	ρ (464) = −0.08; $p = 0.074$	ρ (472) = −0.1; $p = 0.03$	ρ (432) = −0.29; $p < .001$	ρ (472) = −0.08; $p = 0.077$

Note. Tested hypotheses, and their implied outcomes, are listed across Columns 1 and 2 respectively. Remaining columns display the results of FDR-corrected one-sided one-sample tests (mean difference with 95% confidence intervals, alongside p-values and Hedge's g estimates - see Rows 1–4) and Spearman correlations (Rows 5–8). All hypotheses were tested against noise thresholds of.1 rather than conventional null estimates. Evaluative differences between natural poles were non-negligibly ($d > .1$) significant before and after MTS across all dimensions (Rows 1, 2). Evaluative differences between AGC poles after MTS reached significance for groups who linked AGCs with affective or vertical poles. AGCs linked with temporal poles were not significantly different at the Bonferroni-adjusted threshold ($p = .006$). AGCs linked with horizontal poles were non-significant (Row 4). Spearman correlations revealed strategy awareness was strongly and positively associated with natural word evaluations ($.55 > \rho > .23$, Rows 5, 6). Conversely, strategy awareness was weakly and negatively associated with AGC evaluations ($-.29 > \rho > -.06$, Rows 7, 8).

category and evaluation time are summarized in Table 3). Chi-square tests for linear trends across awareness categories indicated that participants generally reported higher strategy awareness following the MTS task compared to baseline (pre-MTS: $\chi^2(1) = 247.94$, $p < .001$; post-MTS: $\chi^2(1) = 580.42$, $p < .001$). Crucially, awareness levels differed substantively between stimulus categories. High awareness responses were significantly more likely during natural word evaluations, $\chi^2(1) = 1980.78$, $p < .001$, whereas AGC evaluations were consistently coupled with much lower awareness responses, $\chi^2(1) = 21.60$, $p < .001$.

To directly test our hypotheses, Spearman correlations between raw string evaluations and awareness levels were calculated at each time point (see Fig. 3, and Rows 9–16 in Table 2). Supporting our second confirmatory hypothesis (CH2), awareness responses were strongly and positively correlated with natural word evaluations, with all Spearman correlation coefficients (ρ) ranging between.23 and.55 (Columns 3 and 4 in Fig. 3). In contrast, and in support of our exploratory hypothesis EH2, AGC evaluations were weakly and negatively associated with awareness at all time points. Across AGCs, ρ estimates varied between −.29 and −.06, with only half reaching statistical significance (Columns 1 and 2 in Fig. 3). This divergence confirms that valence differentiation across natural poles strongly co-varied with explicit strategies, whereas valence generalization to novel AGCs occurred in the relative absence of such awareness.

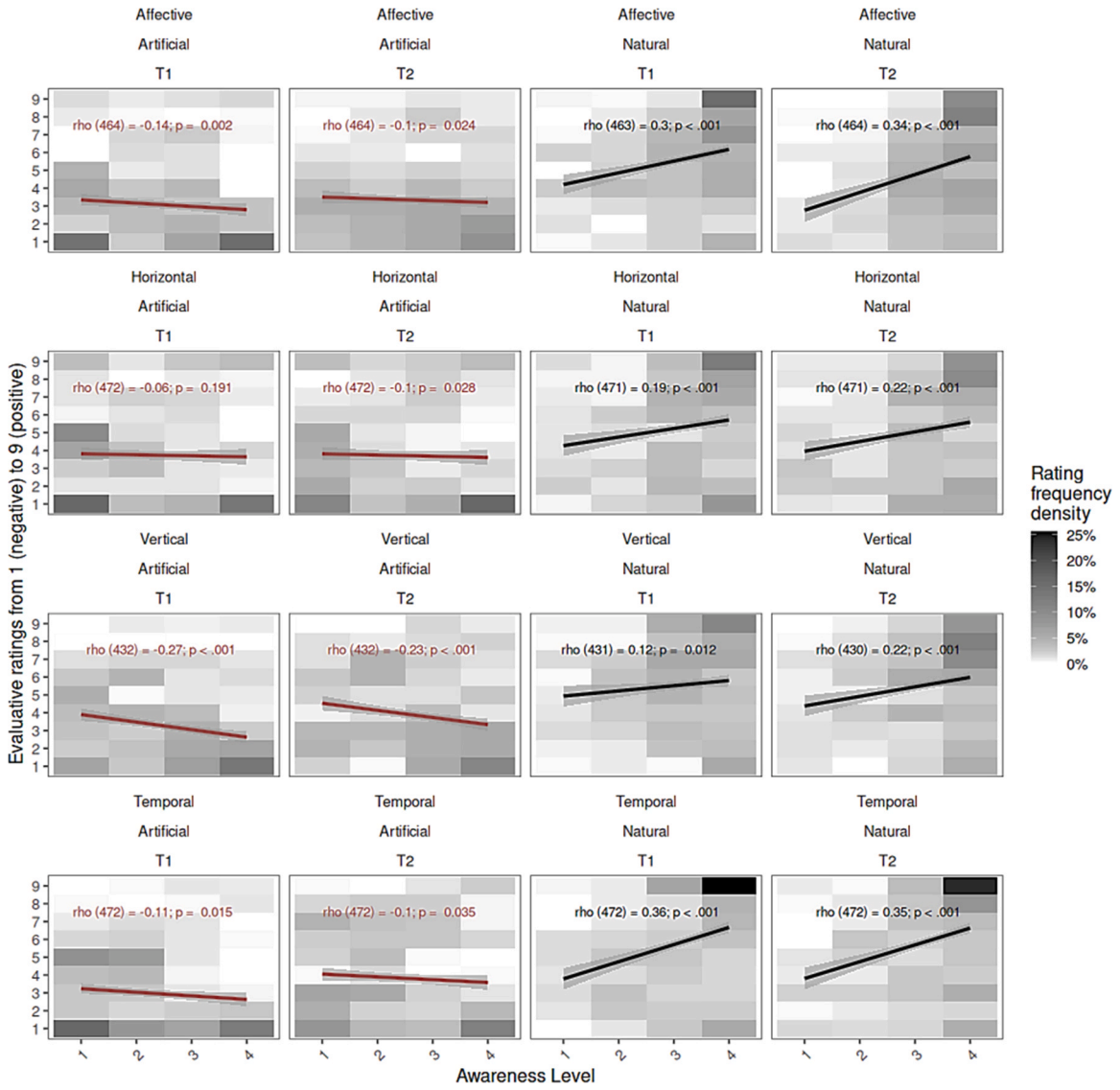

Fig. 3. Heat-map identifying frequencies of evaluative ratings (*y*-axis) at each awareness level (*x*-axis) across groups (rows), stimulus category and time points (columns). Spearman's correlation coefficients are provided with linear regression trends to convey the relationship between awareness levels and evaluative ratings.

2.5. Matching-to-sample performances

All participants completed a fixed set of 44 matching-to-sample (MTS) trials across three blocks. This consisted of a 16-trial pre-training block and a 20-trial training block, during which individual responses were followed by corrective feedback. These were followed by an 8-trial testing block, where responses did not produce differential feedback. Accurate and inaccurate responses were coded as 1 and 0, respectively, and then averaged across blocks to yield three point estimates per participant (see Fig. 4). During pre-training, participants matched natural words with their representative pole comparisons (Fig. 1, Panel E). Mean categorization accuracies during this phase were uniformly high (ranging between 81% and 100%) when natural poles had to be mapped to the respective comparisons, confirming basic task comprehension. Variability in mean accuracy increased substantially when novel, non-recurring AGC strings appeared as samples during the training (ranging from 25% to 85%) and testing (ranging from 8% to 83%) blocks, indicating that participants typically performed at or near chance levels when attempting to explicitly categorize the artificial strings with their linked poles. To test of the second exploratory hypothesis (EH2), we examined whether explicit MTS categorization performance correlated with AGC evaluative difference scores (*d*-scores) at either time point. Only non-significant correlation coefficients were observed before and after the MTS task ($-.04 < \rho < .04$; $p > .6$), demonstrating evaluation trends and AGC categorizations during MTS were statistically unrelated.

3. Discussion

The current study investigated whether natural and structurally novel 'artificial' terms denoting opposite relational poles become evaluatively differentiated. Four groups of participants contingently linked AGCs with vertical (above/below), horizontal (left/right), temporal (before/after), or affective (happy/unhappy) relational poles across matching-to-sample (MTS) trials. Before and after the MTS task, participants evaluated novel AGC strings and familiar natural poles, coupled with trial-by-trial strategy awareness checks. Following analysis, natural relational poles were observed to be consistently and significantly differentiated along valence, independent of MTS exposures. In contrast, AGC evaluative differences were directly contingent on the MTS task. Specifically, perceived

Fig. 4. Mean response accuracies (*y*-axes) across three trial blocks (*x*-axes) during the instrumental matching-to-sample task across four participant groups (panels). Individual points represent the mean accuracy with error for each participant. The dashed horizontal intercept indicates 50% response accuracies.

AGC categories were not differentially evaluated by any participant group prior to training. After MTS, AGCs linked to natural poles came to be evaluatively differentiated for the most part, with the exception of AGCs linked with horizontal and temporal poles (more on this in a moment). Analysis of strategy awareness checks and MTS performances revealed divergent outcomes between vertical poles and their linked AGCs. Evaluations of familiar poles were strongly and positively associated with high strategy awareness before and after MTS, as well as with high categorization accuracy during MTS, suggesting natural pole evaluations were explicitly mediated by consciously articulated semantic rules. Conversely, evaluations of AGCs were weakly and negatively associated with awareness of underlying evaluative strategies. Relatedly, categorization performances during the MTS task varied around chance levels when novel AGC strings appeared as samples. Taken together, these data suggest it is highly unlikely that explicit inferential processes mediated the AGC evaluations.

The observed effects suggest how the evaluative differentiation between explicitly affective poles ('happy' versus 'unhappy') parallels the effects observed across vertical poles, aligning with previous demonstrations of space-valence associations (Pohl & Miklashevsky, 2025; von Sobbe et al., 2019). The present study extends upon this literature by demonstrating, for the first time, how SVAs embodied across vertical spatial dimensions may *implicitly* generalize to abstracted grammar structures, transferring with a robustness similar to directly conditioned semantic valences (Amd, 2022; Jurchiș et al., 2020). By 'implicit,' we refer strictly to the lack of strategy awareness and chance-level MTS categorization accuracy observed when AGC strings were deployed as stimuli. Furthermore, because each tested AGC string was structurally novel, any evaluative differentiation between categories presupposes that the underlying grammar rules had been effectively abstracted rather than memorized at the item level (Reber, 1967).

While these findings strongly support the notion that binary relational poles are inherently valenced, the lack of significant transfer within the horizontal (left/right) and temporal (before/after) conditions warrants consideration. Unlike the vertical axis, which is highly robust and universally anchored in human cognition, horizontal and temporal SVAs are intrinsically more malleable. Previous literature has demonstrated that left/right evaluative mappings, which also serve as the spatial foundation for temporal "before" and "after" representations (de la Fuente et al., 2014; von Sobbe et al., 2019), can be subject to significant inter-individual variability. For instance, body-specific handedness (Casasanto, 2009; Kong, 2013), context-dependent motor consequences (Brouillet et al., 2015), and cumulative language experience (Tsaregorodtseva, Rinaldi & Marelli, 2025) can significantly impact horizontal SVAs.

It may be noted that the current sample comprised a demographic whose primary languages (English, iTaukei, and Hindi) follow a left-to-right script, implying variability stemming from differences in culturally assigned reading directions were not substantive (Román et al., 2015). Consequently, the failure of both the horizontal and temporal poles to implicitly transfer valence to the AGCs in the present study may reflect this established empirical noise.

Another concern is that participant handedness, a critical determinant of horizontal SVAs (Casasanto, 2009), was neither recorded nor controlled for. This could suggest that if left- and right-handed participants held opposing spatial biases, their evaluative shifts could have averaged out to zero at the group level, masking a true effect. However, our analytical approach mitigated this specific confound by computing absolute normalized difference scores across individual participants. Doing so isolated the *magnitude* of evaluative differentiation between poles, independent of their directional valences (Amd, 2024). Because this metric captures absolute differentiation, strong biases in *either* direction would yield larger difference scores, and would not cancel each other out during group aggregation. Therefore, the observed null effects in the horizontal condition reflect a genuine lack of evaluative differentiation magnitude rather than an artifact of opposing, unmeasured handedness biases. These outcomes suggest that while embodied valences can generalize to novel structures, the fundamental stability of the anchor dimension (e.g., a robust vertical axis versus comparatively malleable horizontal/temporal axes) corresponds to the magnitude of that transfer.

We conclude our discussion after addressing four additional limitations of the present design. First, the decision to require participants to explicitly evaluate strings on a continuous Likert scale cannot fully absolve the influence of conscious strategies on observed performances. If evaluations are conceptualized as explicitly specified relations, they remain variably susceptible to other propositions salient in awareness (Nudler et al., 2024). A future replication could incorporate valence-sensitive outcome measures entirely divorced from articulation, such as neurological responses to emotional stimuli during pre-verbal time windows (Amd & Baillet, 2019). Identifying whether, say, event-related potentials (ERPs) sensitive to stimulus valences are differentially elicited by novel AGC strings before and after MTS would provide direct physiological evidence of embodied SVAs activating prior to explicit articulation (Amd et al., 2013).

A second concern relates to the interpretation of chance-level performances during the AGC MTS trials. Critics might argue that chance-level categorization implies participants simply did not learn the task contingencies, and/or misunderstood the task requirements. However, this criticism conflates the acquisition of AGC structures with explicit contingency awareness, which are dissociable processes (Jurchiș et al., 2020). For instance, Amd (2022) showed how perceiving non-recurring AGC samples with category-congruent and category-incongruent comparisons eventually leads to increasing category-congruent mappings, even without corrective feedback (Experiment 3), implying grammar structures can be abstracted implicitly. Furthermore, the argument that participants failed to understand the task is contradicted by near-perfect MTS performances in the presence of natural poles. The fact that participants readily mapped natural poles to their correct visual comparisons confirms a thorough understanding of task parameters.

Third, the awareness checks deployed could be criticized for being too coarse-grained, potentially failing to differentiate between objective (performance-based) and subjective (report-based) awareness (Kiefer et al., 2023). One could argue that participants were aware of the task contingencies but did not consider them a valid justification for their AGC evaluations. Since we did not include any verbal measure of contingency awareness, we cannot obviate this possibility. However, a verbal measure of contingency awareness would be functionally redundant in the present context, as the MTS categorizations already served to capture these contingencies behaviorally. Moreover, if participants had explicitly derived that the AGCs represented opposite relational poles (e.g., that AGC1

meant 'above'), the Polarity Correspondence Principle dictates they should have differentiated them accordingly (Proctor & Cho, 2006). The fact that they differentiated natural words explicitly, but differentiated AGC strings while simultaneously reporting ignorance and failing at explicit categorization, strongly points toward an implicit process rather than a conscious appraisal of contingently trained rules.

Finally, given the unsupervised online nature of the study, one might question whether variability in participants' hardware (e.g., screen sizes, aspect ratios) or testing environments influenced the outcomes, particularly for a task relying on visual-spatial comparisons. However, this concern is largely mitigated by the study's design and screening protocols. The visual comparisons were rendered as scalable vector graphics (SVGs), ensuring that the relative spatial relations (e.g., 'above' versus 'below') remained visually consistent regardless of absolute screen dimensions. Because the task relied on the appraisal of semantic and *relative* poles rather than absolute visual angles, display variability is unlikely to have systematically confounded the evaluations. Additionally, the pre-training inclusion criterion, which required greater than 80% accuracy during the natural word MTS trials, confirmed that retained participants were sufficiently attentive and capable of cleanly discriminating the visual stimuli within their respective testing environments prior to encountering the novel AGC strings.

The present study corroborates earlier claims that opposite relational poles are inherently valenced (Amd, 2024; Pohl & Miklashevsky, 2025; Proctor & Cho, 2006). We expand upon the existing literature by providing novel evidence that these embodied polar valences along a vertical axis can implicitly generalize across abstracted structural categories, similar to directly conditioned emotional stimuli (Amd, 2022). These findings suggests the potential impact of embodied space-valence associations on symbolic behavior can operate at an abstracted, implicit processing level.

CRediT authorship contribution statement

Micah Amd: Writing – review & editing, Project administration, Conceptualization.

Author note

All procedures were approved by the University of the South Pacific Research Ethics Committee (REC). The study was funded by an internal research stipend (#FE082-FAL15-71502-001) to the author MA, who conceptualized, programmed, and conducted the study, analyzed the data, and prepared the manuscript. Materials, data, and the analysis script have been included in a supplementary file with the main manuscript. The author(s) have no conflicts of interest to disclose.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.lmot.2026.102316.

Data Availability

The data and analysis script have been included as supplemental files.

References

Amd, M. (2022). Valence generalization across nonrecurring structures. *Journal of Experimental Psychology: Animal Learning and Cognition, 48*(2), 105. https://doi.org/10.1037/xan0000317

Amd, M. (2023). Disentangling affect from self-esteem using subliminal conditioning. *Behavioural Processes, 213*, Article 104965. https://doi.org/10.1016/j.beproc.2023.104965

Amd, M. (2024). Relational cues are affectively differentiated. *The Psychological Record, 74*, 1–14. https://doi.org/10.1007/s40732-024-00600-5

Amd, M., & Baillet, S. (2019). Neurophysiological effects associated with subliminal conditioning of appetite motivations. *Frontiers in Psychology, 10*, 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

Brouillet, D., Milhau, A., & Brouillet, T. (2015). When 'good' is not always right: Effect of the consequences of motor action on valence-space associations. *Frontiers in Psychology, 6*, 237. https://doi.org/10.3389/fpsyg.2015.00237

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General, 138*(3), 351–367. https://doi.org/10.1037/a0015854

Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An aligned rank transform procedure for multifactor contrast tests. *En Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology* (pp. 754–768). Association for Computing Machinery. https://doi.org/10.1145/3472749.3474784

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Freddi, S., Brouillet, T., Cretenet, J., Heurley, L. P., & Dru, V. (2015). A continuous mapping between space and valence with left- and right-handers. *Psychonomic Bulletin & Review, 23*(3), 865–870. https://doi.org/10.3758/s13423-015-0950-0

de la Fuente, J., Santiago, J., Román, A., Dumitrache, C., & Casasanto, D. (2014). When you think about it, your past is in front of you: How culture shapes spatial conceptions of time. *Psychological Science, 25*(9), 1682–1690. https://doi.org/10.1177/0956797614534695

Gromer, D. (2023). APA: Format outputs of statistical tests according to APA guidelines. *R Package Version 0 3 4*. https://CRAN.R-project.org/package=apa.

Hester, J., & Bryan, J. (2024). Glue: Interpreted string literals. *R Package Version 1 8 0*. https://CRAN.R-project.org/package=glue.

Janczyk, M., Tucholski, K., Kaup, B., & Ulrich, R. (2025). Mental association of time and valence revealed with a novel chronometric approach: The positive-future effect. *Memory and Cognition, 53*(3). https://doi.org/10.3758/s13421-025-01715-y
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. https://doi.org/10.1037/xge0000734
Kassambara, A. (2023). rstatix: Pipe-friendly framework for basic statistical tests. *R Package Version 0 7 2*. https://CRAN.R-project.org/package=rstatix.
Kiefer, M., Frühauf, V., & Kammer, T. (2023). Subjective and objective measures of visual awareness converge. *Plos One, 18*(10), Article e0292438. https://doi.org/10.1371/journal.pone.0292438
Koch, S., Schubert, T., & Blankenberger, S. (2023). The spatial representation of loudness in a timbre discrimination task. *I-Perception, 14*(6). https://doi.org/10.1177/20416695231213213
Kong, F. (2013). Space–valence associations depend on handedness: Evidence from a bimanual output task. *Psychological Research, 77*(6), 773–779. https://doi.org/10.1007/s00426-012-0471-7
Lenth, R. (2024). emmeans: Estimated marginal means, aka least-squares means. *R Package Version 1 10 4*. https://CRAN.R-project.org/package=emmeans.
Mahr, T. (2020). rprime: Functions for working with 'eprime' text files. *R Package Version 0 1 2*. https://CRAN.R-project.org/package=rprime.
Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15*(4), 243–247. https://doi.org/10.1111/j.0956-7976.2004.00659.x
Nudler, Y., Zvi, M., Levy, G., & Bar-Anan, Y. (2024). Experimental evidence that demand characteristics do not play a dominant role in the evaluative conditioning effect, 19485506241252306 *Social Psychological and Personality Science*. https://doi.org/10.1177/194855062412523.
Pohl, J., & Miklashevsky, A. (2025). Vertical and horizontal space-valence associations: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 170*, Article 106054. https://doi.org/10.1016/j.neubiorev.2025.106054
Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin, 132*(3), 416–442. https://doi.org/10.1037/0033-2909.132.3.416
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X
Román, A., Elgendi, A., Tenenbaum, H. R., & Santiago, J. (2015). Reading direction causes spatial biases in mental model construction in language understanding. *Scientific Reports, 5*, Article 18248. https://doi.org/10.1038/srep18248
von Sobbe, L., Scheifele, E., Maienborn, C., & Ulrich, R. (2019). The space–time congruency effect: A meta-analysis. *Cognitive Science, 43*(1), Article e12709. https://doi.org/10.1111/cogs.12709
Staats, W. W. (1996). *Behavior and personality: Psychological behaviorism*. Springer Publishing Company.
Tsaregorodtseva, O., Rinaldi, L., and Marelli, M. Systematic Spatial Association Effects Can Arise from Language Experience. *Available at SSRN 5658577*. https://doi.org/10.2139/ssrn.5658577.
Verdonschot, R. G., van Steenbergen, H., & Spapé, M. M. (2019). *The E-Primer: An introduction to creating psychological experiments in E-Prime* (2.ª ed.). Leiden University Press. https://doi.org/10.1017/9789400603363
Wickham, H. (2023). Forcats: Tools for working with categorical variables (Factors). *R Package Version 1 0 0*. https://CRAN.R-project.org/package=forcats.
Wickham, H., Vaughan, D., & Girlich, M. (2023). *purrr: Functional programming tools* (Versión 1.0.2). *Comprehensive R Archive Network (CRAN)*. https://CRAN.R-project.org/package=purrr.
Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only Anova procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 143–146. https://doi.org/10.1145/1978942.1978963
Xie, Y., Dervieux, C., & Riederer, E. (2024). *knitr: A general-purpose package for dynamic report generation in R* (Versión 1.45). *Comprehensive R Archive Network (CRAN)*. https://CRAN.R-project.org/package=knitr.
Zhu, H. (2024). kableExtra: Construct complex table with 'kable' and Pipe Syntax. *R Package Version 1 4 0*. https://CRAN.R-project.org/package=kableExtra.