Sub-sampling of cues in associative learning

Omar D. Perez[1, 3], Edgar H. Vogel[4], Sanjay Narasiwodeyar[2] & Fabian A. Soto[2]

[1] Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California

[2] Department of Psychology, Florida International University, Florida, USA

[3] Nuffield College CESS-Santiago, Faculty of Business and Economics, University of Santiago, Santiago, Chile

[4] Faculty of Psychology, University of Talca, Talca, Chile

Author Note

Correspondence concerning this article should be addressed to Omar D. Perez, 1200 E. California Blvd., Pasadena, California, CA91125. E-mail: odperez@caltech.edu

Abstract

Theories of learning distinguish between elemental and configural stimulus processing depending on whether stimuli are processed independently or as whole configurations. Evidence for elemental processing comes from findings of summation in animals where a compound of two dissimilar stimuli is deemed to be more predictive than each stimulus alone, whereas configural processing is supported by experiments employing similar stimuli that fail to find summation. However, in humans the summation effect is robust and impervious to similarity manipulations. In three experiments in human causal learning, we show that summation can be obliterated when partially reinforced cues are added to the summands in training and test (AXY+, BXY+, CXY-, DXY-; test with ABX and ABY). This lack of summation only holds when the partially reinforced cues are similar to the reinforced cues (Experiment 1) and seems to depend on participants sampling only the most salient cue in each trial (Experiments 2a and 2b). We propose a limited-sampling model that captures the present results and others from the animal literature where complex visual stimuli are employed and which have been usually interpreted as evidence in support of configural processing.

Keywords: generalization, Rescorla-Wagner, configural, summation, elemental

Word count: 10,106

Sub-sampling of cues in associative learning

Introduction

What do organisms do when confronted with a situation where two cues that provide independent information about the same amount of reward are present? How much reward do they predict? In animal studies, a typical procedure to study this type of problem is the summation design, where two cues A and B are separately paired with the same amount of reinforcement during a first stage of training and responding to a compound of the two stimuli, AB, is assessed in a final test. In animals, a summation effect is obtained when when there is more responding to AB than to each of A or B alone (Aydin and Pearce, 1997; Kehoe *et al.*, 1994). In humans, an analogous effect is obtained in causal learning tasks if the AB compound is deemed more predictive of an outcome than each of A or B alone (Pérez *et al.*, 2018; Soto *et al.*, 2009; Thein *et al.*, 2008).

Two types of computational models in the Pavlovian conditioning literature have been pitted against each other using this summation design. In these models, the internal representation of stimuli is taken to be either elemental (meaning that an agent represents and processes each stimulus independently) or configural (if compounds of stimuli are represented and processed as whole configurations that enter into a single association with the outcome). The cardinal example of elemental models is the influential Rescorla-Wagner model (1972), which assumes that subjects represent A and B independently, and that the presence of the compound AB simply makes subjects sum their individual predictions. What the Rescorla-Wagner model proposes is that the conjoint presentation of stimuli A and B will produce a total prediction given by $v_{total} = v_{AB} = v_A + v_B$. Using this rule, the model can directly explain summation with the presentation of a compound of cues that have been previously trained with the same outcome value, and makes efficient use of the available evidence under the assumption that the two cues are independent causes of such outcome (Pérez *et al.*, 2018).

Configural models, by contrast, take a sharply different approach. Under this view, subjects process and associate whole configurations with the outcomes that follow. Total responding to a compound AB depends on the similarity between the training configurations A and B, and the testing configuration AB. In the basic formulation of configural theory proposed by Pearce (1987; 1994), responding to the compound AB in a simple summation design will depend on the similarity between the training configurations A and B, and the testing configuration AB. More formally, the total predictive value of AB is given by $v_{AB} = E_{AB} + \sum_j e_{j,AB}$, where $e_{j,AB}$ is the predictive value of configuration j that generalizes to configuration AB (including stimuli A and B). Under this rule, the outcome prediction for AB when both A and B have converged to the same predictive value during training is equal to $v_{AB} = 1.0$, which is simply the average of the prediction from each of its components A and B, anticipating an absence of summation.

Empirically, studies in animals have shown that summation is modulated by a number of factors, among which the similarity between the stimuli forming the compound appears critical. Indeed, summation is usually stronger when the components A and B are from different sensory modalities (e.g., auditory and visual) (Kehoe *et al.*, 1994; Redhead, 2007; Thein *et al.*, 2008) but not with stimuli from the same modality (e.g., two visual stimuli) (Aydin and Pearce, 1995; Aydin and Pearce, 1997). Apart from modality, spatial and temporal contiguity can also impact summation (Glautier, 2002; Glautier *et al.*, 2010). These results have led elemental and configural models to incorporate similarity as a critical factor determining the level of summation obtained (Soto *et al.*, 2014b; Thorwart *et al.*, 2012; Wagner, 2008). Under these flexible models, dissimilar stimuli (such as those coming from different modalities) will prompt elemental processing and a level of summation closer to the one predicted by elemental models. Configural processing, on the other hand, will be prompted by similar stimuli (such as those coming from the same modality) and summation should be weaker and closer to the average of the predictive value of each of the components alone (Harris, 2006; McLaren and Mackintosh, 2002;

Melchers *et al.*, 2008; Pérez *et al.*, 2018; Soto *et al.*, 2014a; Thorwart *et al.*, 2012; Wagner, 2008). More recent models predict that higher spatial or temporal contiguity should also encourage configural processing and reduce the summation effect (Soto *et al.*, 2014b; see also Thorwart *et al.*, 2012).

In spite of being a widely-accepted factor in modulating summation, evidence of similarity affecting summation in humans is scarce. In a recent series of studies, we tested this prediction and found no evidence supporting this hypothesis across several manipulations of similarity between A and B, including color, shape and spatial contiguity (Pérez *et al.*, 2018). Across six studies, we consistently obtained strong summation, with participants disregarding similarity and spatial contiguity when making their predictions. In that work, we concluded that the properties of the visual stimuli where such that participants could easily identify them on the screen, assume them as independent causes of the outcome and, sum the individual predictions when presented with the compound (Pérez *et al.*, 2018). Perhaps an additional condition to obtain an effect of similarity on summation is that the components A and B must be difficult to identify.

In the present work, we designed a task that made difficult for participants to identify the component stimuli A and B, with the goal to evaluate whether such a task would lead to an effect of similarity on summation. In this new design, training and testing compounds always consisted of three cues (target cues A and/or B, which predicted the outcome, plus additional non-target cues, which did not predict the outcome) joined at a central point (see Figure 2). Cues could be identified by the spatial configuration of their eight dot components, and the compound of three cues was rotated around its central point across trials, so that the spatial position and direction of a cue did not provide information about its identity.. In Experiment 1, similarity among cues within a compound was manipulated by changing their color. We found that similarity did modulate summation in our task but, against the expectations of an associative analysis, the key variable was

similarity between target and non-target cues, rather than similarity between the two target cues A and B. In Experiments 2a and 2b we tested a potential mechanism for this lack of summation: limited sampling and processing of stimuli within compounds. Our data suggest that under circumstances that hinder component identification (which could be present in previous animal experiments where summation was not obtained) subjects may not process all elements in a configuration, but rather sample a subset of the stimulus configuration and respond according to prior experience with that subset only. We show in a formal model how such limited sampling process might be formulated under similar principles to those of classic prediction-error models of Pavlovian conditioning, and how it captures results that are usually interpreted as evidence of configural processing of stimuli.

Results

Experiment 1

In the three experiments reported in this paper participants were presented with a task in which they were asked to play the role of an allergist who had to predict the levels of allergy caused by different drugs in a hypothetical patient, Mr. X (Pérez *et al.*, 2018; Soto *et al.*, 2009). During training, a configuration comprising one of the target and the two non-target cues was presented at the center of the screen (see Figure 1). Participants were required to give an assessment of the level of allergy that each drug would cause in a fictitious patient, Mr. X, on a scale of 0 to 35. After clicking with the mouse on the rating scale to make their predictions, participants had to confirm them by clicking on a button below the rating scale. Cues were presented 30 times each, in randomized order. Trials were self-paced.

Before the training phase, participants were presented with the following instructions:

In this experiment, we will ask you to imagine that you are an allergist; that is, a medical specialist whose work is to discover the causes of allergic reactions in

people. A new patient, Mr. X, visits you asking for your help. Mr. X has allergic reactions to some drugs, but not to others. In an attempt to discover which drugs cause allergic reactions to Mr. X, you apply the drug on Mr. X's skin and observe the magnitude of the allergic reaction. On each trial, the computer will show you a shape representing the drug that has been applied to Mr. X's skin. You will be asked to predict the magnitude of the allergic reaction that Mr. X will have to the drug. Enter your prediction using the rating scale that will be shown at the center of the screen. This scale ranges from 0 to 35 points, where 0 points means that the drug produces no allergic reaction and 35 points means that that the allergic reaction is of maximum intensity. You can adjust your prediction as many times as you want and take as long as you like to enter the prediction. Once you feel satisfied with your prediction, you must confirm it by either pressing the "Enter" key in your keyboard or clicking on the grey button below the rating scale. You will receive feedback about the actual magnitude of the allergic reaction. At first, you will need to guess the correct answer, but your predictions should get more accurate as the experiment progresses.

During the training stage, and after entering their prediction and confirming it, participants were presented with a feedback screen with the message "Correct!" if participants' estimation was correct, or the message "Incorrect!" if the estimation was incorrect (see Figure 1). After entering their prediction, they were presented with the outcome message "Allergic reaction: 10 points in the allergy scale" for cues that predicted allergy or "No allergic reaction: 0 points in the allergy scale" for cues that did not predict allergy. To encourage learning and attention to the task, each incorrect assessment was followed by a 5-second message (i.e., a time-out), while a 1-second message was employed for correct assessments.

The testing stage was similar, with two key differences. First, participants were presented not only with the same stimuli they had seen during training (AXY, BXY, CXY, DXY), but also, to test the summation effect, with stimuli including A and B in the same configuration (ABX or ABY). Second, no feedback was provided during the test stage. Each stimulus was presented two times during the test stage.

Participants were presented with the following instructions before the test phase:

> *Now you will have to make a final evaluation of the drugs that produce allergic reactions in Mr. X. The computer will show you single drugs or combinations of drugs and, as before, you must enter your prediction using a scale that ranges from 0 to 35 points, where 0 points means that the drug produces no allergic reaction and 35 points means that the drug produces an allergic reaction of maximum intensity. Use intermediate values to indicate different degrees of allergic reaction between no allergic reaction and maximum allergic reaction. To confirm your choice, press "Enter" in your keyboard or click the gray button that will be shown below the rating scale. You can change your decision as many times as you want before confirming it.*

In Experiment 1, we hypothesized that the strong and robust summation found in previous human studies (Pérez *et al.*, 2018) resulted from the use of cues that could be easily parsed and recognized by participants. Thanks to properties like their reliable spatial position and the absence of non-target cues, a simple counting strategy brings about reliable and strong summation independently of similarity. Therefore, a task where stimuli are part of a configuration consisting of a fixed number of components should impair participants' deployment of such strategy.

In this new design, in all trials participants were presented with a compound of three cues that was either reinforced ("10 points of allergy out of a total of 35") or not ("0 points

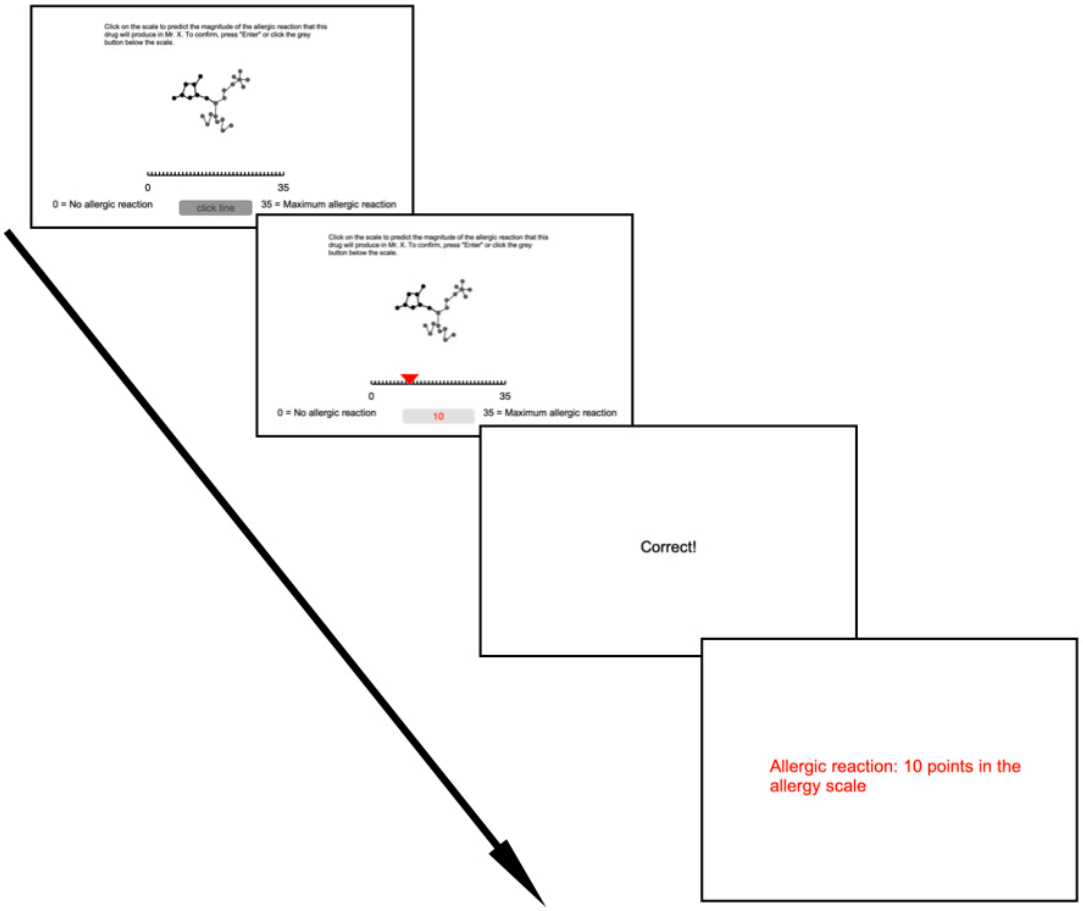


Figure 1. **Trial design for all experiments.** In each training trial, participants observed one target cue (A,B, C or D) and two non-target cues (X and Y) in the same configuration (AXY, BXY, CXY, DXY) and were asked to rate the level of allergy they thought would be produced by the stimulus on a scale from 0 to 35. After an inter-stimulus interval, feedback was presented ("Correct!" if the prediction was correct, "Incorrect!" if the prediction was incorrect), followed by the message "Allergic reaction: 10 points in the allergy scale" for cues that predicted allergy or "No allergic reaction: 0 points in the allergy scale" for cues that did not predicted allergy. During the test phase, the cues were presented in the same way as in training, but now the compound AB, together with one non-target cue (X or Y) was presented in some trials (i.e., ABX, ABY). The predictions were assessed in the same way as in training but no feedback was given to participants. Each cue was presented two times.

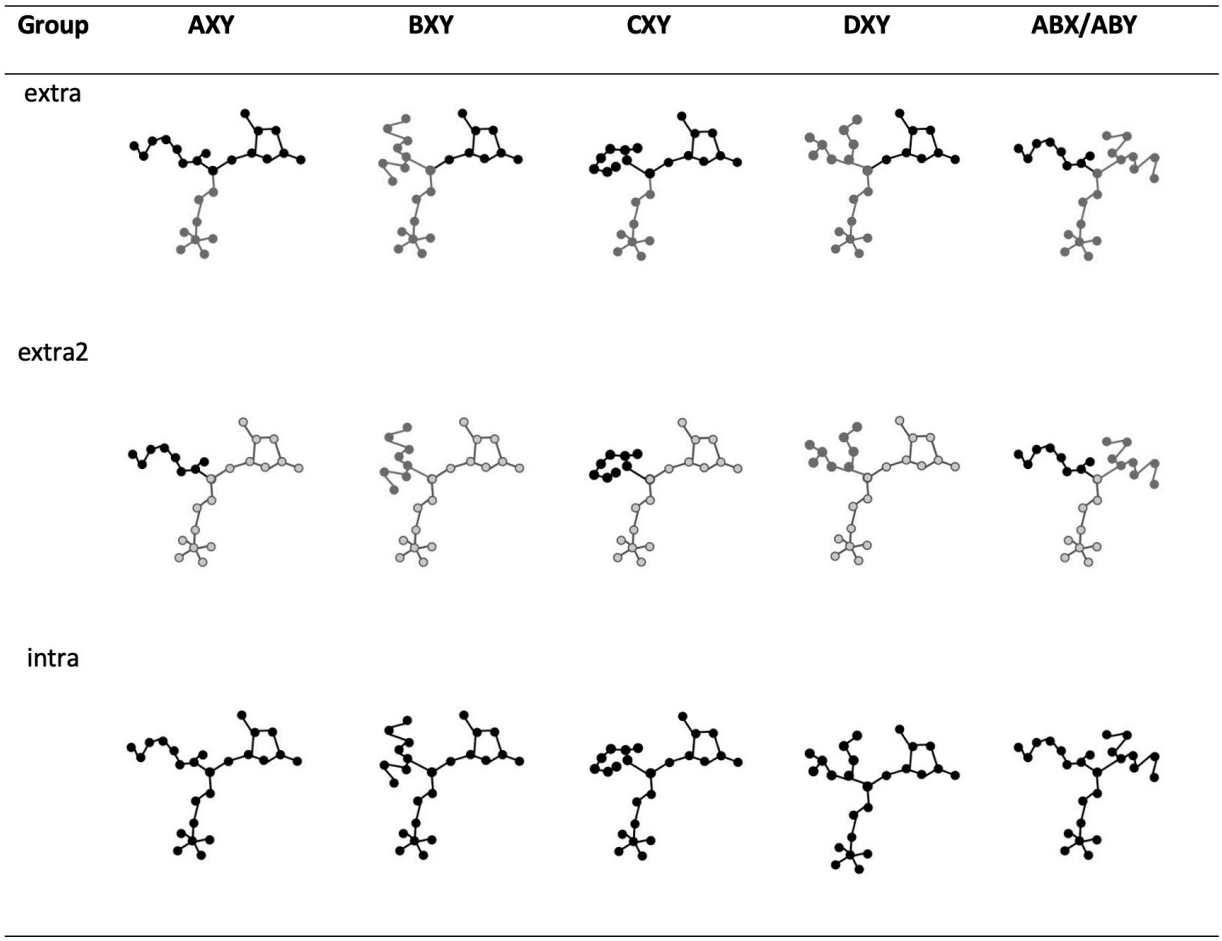


Figure 2. **Stimuli used in the experiments reported in this paper.** Each stimulus was formed by a central point which branched out into three different components. Target cues A and B were accompanied by two other non-target cues, X and Y. The compound AB was accompanied by one of the non-target stimuli X or Y. All stimuli were presented in three different planar orientations, so that each component was equally likely to appear in one of three positions. Cues A, B, C and D were the target cues. Cues X and Y were contextual or non-target cues that accompanied target cues in each configuration. Stimuli in group *intra* varied only in shape. Stimuli in groups *extra* and *extra2* varied in both shape and color.

of allergy out of a total of 35”). Each trial included one target cue (i.e., a cue predictive of the outcome; A, B, C or D) and two non-target cues (i.e., cues not predictive of the outcome; X and Y). The positions of each of the cues varied across trials. As shown in Figure 2, all compounds (AXY, BXY, CXY, DXY, ABX/ABY) consisted of the three cues joined at a central solid circle, with single cues consisting of 8 solid circles each. Therefore, each compound consisted of 25 circles connected by black lines. In addition, we rotated the compound around its central circle across trials, so that each cue was presented in three possible spatial positions and directions. This ensured that the number of stimuli on each configuration was the same for both training and test phases; the spatial position and direction of the target components did not facilitate their identification.

To test for summation, during the test stage in some trials both A and B were presented in the same configuration, but with only one of the non-target cues (either X or Y, counterbalanced). Cues C and D were included to check which participants correctly learned the contingencies between the cues and their respective outcomes. Since our aim was to investigate generalization of learning, only those participants who correctly identified the contingencies during training (i.e., showed learning) were included in the final analysis. We employed the same criteria as in our previous studies on summation (Pérez *et al.*, 2018). Because the cues were designed to be as abstract as possible so as to avoid any prior preference for any of them, the assignment of cues and shapes was kept fixed in all groups.

The critical manipulation concerned the similarity between cues within a configuration. We did this in three different groups. Group *intra* included cues that varied only in shape (i.e., intra-dimensional differences, see Figure 2). For the two *extra* groups, cues varied both in shape and in color (i.e., extra-dimensional differences). In group *extra*, A and B had different color (black vs gray), but the colors were shared with the non-target cues X and Y (one of them black and the other gray, see top-left stimulus in Figure 2). This means that although A and B can be easily distinguished, they cannot be easily

distinguished from the non-target cues X and Y. Finally, in group *extra2*, A and B had different color (black vs. gray), and they also differed in color with the non-target cues X and Y (light gray with black contours). As a consequence, the similarity between target and non-target cues was lowest in the *intra* group and highest in the *extra2* group.

Figure 3 shows our expectations for this experiment, based on the prediction from associative learning theory of an effect of similarity on summation. If similarity between cues A and B is the critical variable driving summation, we should expect higher summation in groups *extra* and *extra2*, where A and B are dissimilar, than in group *intra* (Figure 3; left panel). If, in addition, similarity between the targets A and B and non-target cues X and Y influences summation, we should expect higher summation in group *extra2*, where those cues are dissimilar, than in group *extra* (Figure 3; right panel.)

The results are shown in Figure 4, where we present the mean causal ratings to the trained (AXY and BXY) and novel (mean of ABX and ABY) compounds in the three groups of Experiment 1. The figure indicates a summation effect in the form of higher ratings to the novel compounds than to trained compounds in group *extra2*, but no evident summation in groups *extra* and *intra*. A 3 (group) X 3 (compound) mixed design ANOVA confirmed the reliability of these observations in showing a significant group x compound interaction ($F(2, 55.5) = 4.00, p = .02, \eta_p^2 = 0.13, 90\%CI\ [.04, .17]$). Subsequent simple effects of compound in each group revealed that the mean ratings to ABX/ABY reliably differed from ratings to the trained AXY and to BXY in group *extra2* ($t(26) = 4.82, p < .001, D = 0.77, 95\%CI\ [2.81, 6.91]$ and $t(26) = 4.81, p < .001, D = 0.77, 95\%CI\ [2.83, 6.79]$, respectively) but did not in groups *extra* ($t(12) = 0.77, p = .46, BF_{01} = 2.79$ in both cases) and *intra* ($t(17) = 1.16, p = .26, BF_{01} = 2.30$, and $t(17) = 1.17, p = .26, BF_{01} = 2.28$, respectively). The main effects of compound and group were also reliable ($F(1, 55.05) = 14.97, p < .001, \eta_p^2 = 0.21, 90\%CI\ [.10, .31]$ and

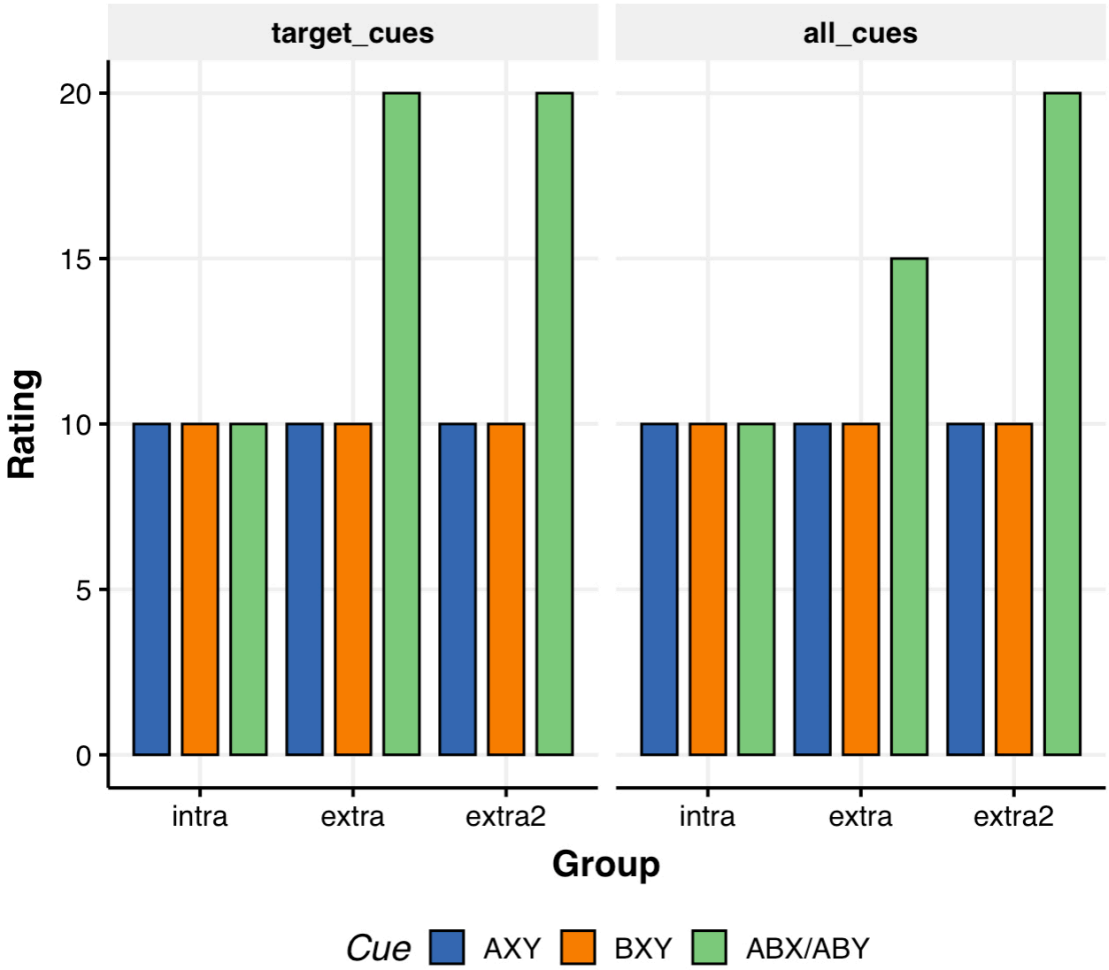


Figure 3. **Expectations for Experiment 1, based on the prediction from associative learning theory of an effect of similarity on summation. Left panel.** Expected results when similarity of only the target cues A and B influences summation. **Right panel.** Expected results when similarity of all cues (i.e., non-target cues as well as target cues) influences summation.

$F(2, 55) = 3.80, p = .03, \eta_p^2 = 0.12, 90\%CI$ [.01, .25], respectively), but they are mainly accounted by the summation effect of group *extra2* described above.

The summation effects observed in this experiment can be further illustrated through a summation index computed as the mean rating for the novel compound (ABX and ABY) minus the mean causal ratings for the trained compounds (AXY and BXY). The outcome of this computation is summarized in Table 1, which indicates that the summation effect was not significantly different from zero in group *extra* ($t(12) = 0.80, p = .46, BF_{01} = 2.80$)

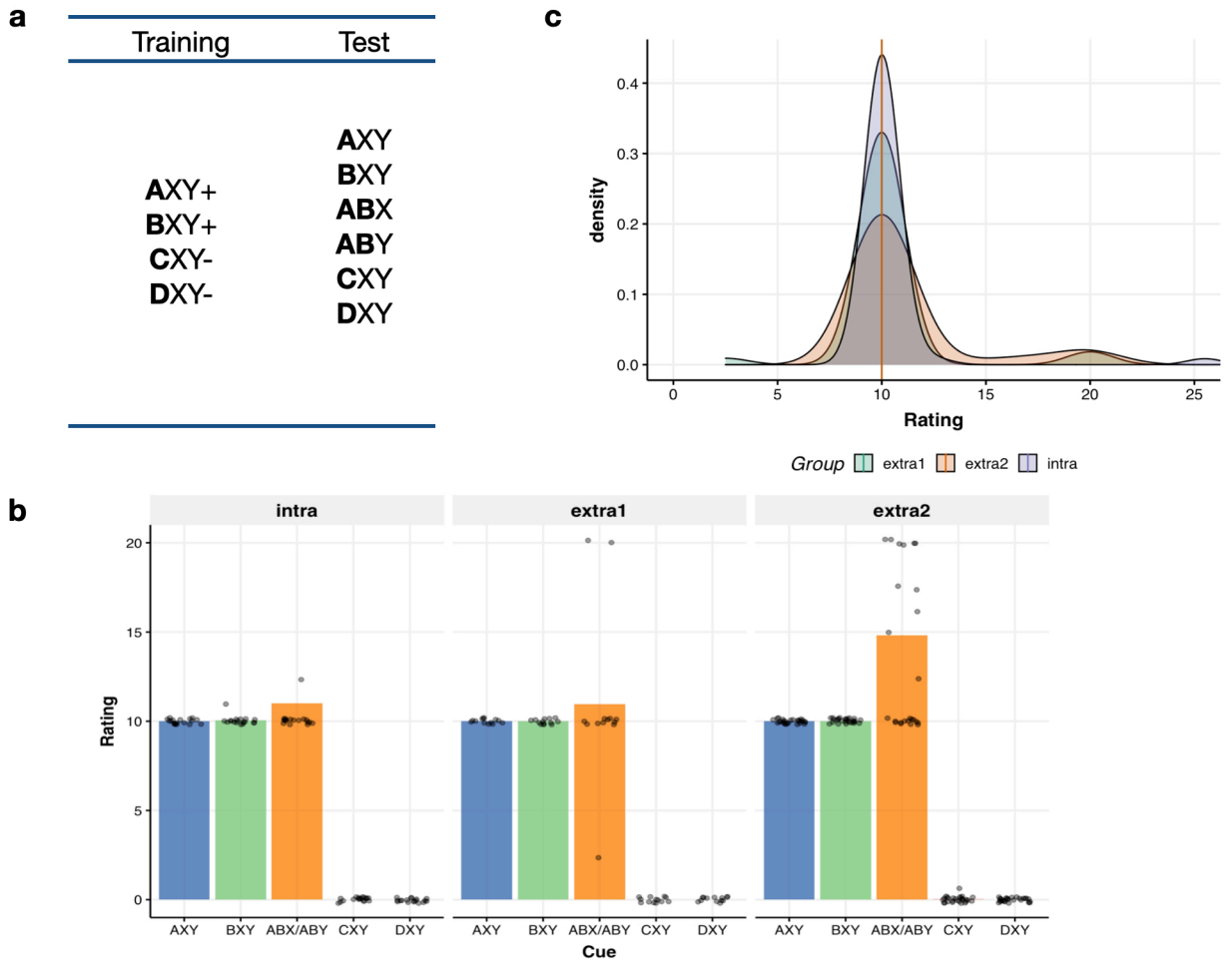


Figure 4. **Design and results of Experiment 1**. **(a)** Letters denote cues, represented by different chemical shapes that can cause different levels of allergy to a fictitious patient. Each cue is represented by a capital letter. Cues A, B, C and D were the target cues (shown in bold), whereas X and Y were non-target cues. In Experiment 1, all cues were followed (+) or not followed (-) by the same level of allergy (10 points of allergy out of a total of 35). **(b)** Average ratings given to each cue during the test. Jittered points represent individual ratings. **(c)** Density plots of ratings given to the ABX/ABY cue during the test in the three different groups.

and group *intra* ($t(17) = 1.16, p = .26, BF_{01} = 2.30$) while a significant summation effect was found for group *extra2* ($t(26) = 4.02, p < .001, BF_{01} = 0.01$). The effect of similarity between target and non-target cues on summation was evident in a significant difference between groups *extra2* and *extra* ($t(31.77) = 2.22, p = .03, D = 0.67, 95\%\, CI[-0, 01, 1.35]$), and between *extra2* and *intra* ($t(42.50) = 4.54, p = .01, D = 0.71, 95\% CI\ [0.09, 1.32]$). There was no significant difference in summation between groups *extra* and *intra* ($t(22.55) = -0.03, p = .98, D = -0.01, 95\% CI\ [-0.75, 0.74]$)

Group	N	Summation
extra	13	1.57 [-0.42 3.56]
extra2	27	4.13 [2.33 5.93]
intra	18	1.58 [0.27 2.89]

Table 1

Mean and 95% confidence intervals for the summation effect for all groups of Experiment 1. Confidence intervals were calculated following the method proposed by Morey (2008).

To our knowledge, this is the first demonstration of humans being sensitive to manipulations of similarity in a predictive learning task. Firstly, we obtained evidence of participants showing no summation in group *intra*, a result that is relatively uncommon in the human literature, previously obtained only by (Glautier *et al.*, 2010). This was achieved by using a design which made it more difficult for participants to distinguish the cues being presented, discouraging a counting strategy.

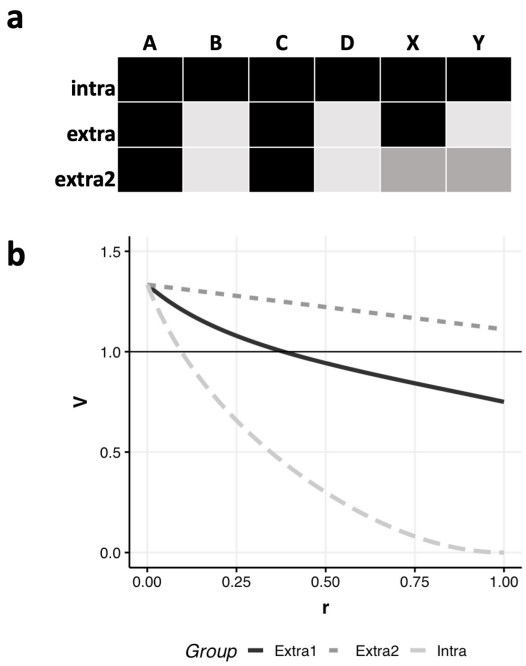
What is notable about this result is that the main driver of the summation effect was the similarity between the target (A and B) and non-target (X and Y) cues, and not the similarity between the two targets A and B, which is the variable assumed in associative learning models to be the main driver of summation. As can be seen in Figure 2, the two target cues A and B were as dissimilar in group *extra* as in group *extra2*. If the effect of cue

similarity acted through the mechanisms proposed by associative models, then both of these groups should have shown stronger summation than that shown in group *intra*.

A Formal Associative Analysis. To confirm our intuitions through a more formal associative analysis, we performed simulations with the Replaced Elements Model (REM) proposed by Wagner and colleagues (see Brandon *et al.*, 2000; Vogel *et al.*, 2017; Wagner, 2008 for details). What makes REM particularly attractive is that the model can use representations that vary in a continuum from fully elemental to fully configural, through changes in a replacement parameter r ($0 < r < 1$). This allows the model to accommodate results that are consistent with either configural or elemental approaches, and to capture the effect of cue similarity by making r dependent on it. [1]. Figure 5c shows in matrix form the assumptions that we made in terms of the replacement parameter in each group. First, in order to give the model a chance to produce summation, we made the very conservative assumption of $r = 0$ for any pair of cues of different color. With this assumption the groups would differ only in the number of cues having the same color, in which case r was varied across a broad parametric range from 0 (no replacement) to 1 (full replacement) in steps of 0.02.

Figure 5c presents the predictions of REM for the three groups reported in Experiment 1. In general, the model predicts a weaker summation effect under this common-cue design compared to a simple summation design. As expected, summation decreases monotonically for all groups as the number of elements replaced increases. Furthermore, negative summation is predicted by REM for values of $r >= 0.1$ in group *intra*, whereas positive summation is predicted for the group *extra* across the same range of

[1] Simulations with additional associative models, such as Harris' (2006) attentional model, are difficult because, to capture our design, the theory should be flexible enough to allow for different levels of similarity between the three different cues in a compound. Such flexibility so far has only been formalized and implemented in REM. Still, the explanation for the effect of similarity advanced by Harris' model is that when cues A and B are similar, they share common cues that foster inhibitory connections between them (i.e., an "Espinet" effect). According to this mechanism, the presence of explicit common cues in our design should lead to absence of summation in all groups, given that asymptotic learning is achieved.



c

intra	B	C	D	X	Y
A	r	r	r	r	r
B		r	r	r	r
C			r	r	r
D				r	r
X					r

extra	B	C	D	X	Y
A	0	r	0	r	0
B		0	r	0	r
C			0	r	0
D				0	r
X					0

extra2	B	C	D	X	Y
A	0	r	0	0	0
B		0	r	0	0
C			0	r	0
D				0	0
X					r

Figure 5. **Simulations of the REM model for Experiment 1. (a)** Schematic representation of the cues that have the same or different color in the three groups of Experiment 1. Notice that the black, white and grays colors were used for illustrative purposes only. **(b)** Predicted asymptotic associative strength for Experiment 1. The simulations were conducted with the ALTSim simulator (Thorwart *et al.*, 2009) using the learning rule suggested by Schulteis et al. (2008). The parameters employed were $\alpha = 0.1, \beta^{+} = 0.2, \beta^{-} = 0.04$. **(c)** r values for different pairs of cues in the three groups. It was assumed that the level of replacement (r) of different-color cues was zero.

r values. When $r = 0.1$, the model predicts no summation in group *intra* but positive summation in both *extra* groups. Thus, at no point in its parametric space does REM predict the pattern of results of Experiment 1, with positive summation for group *extra2* and zero summation for groups *intra* and *extra*. Indeed, when *extra* approaches zero summation, it does so in the range of negative summation for *intra*.

Potential Visual Mechanisms. If associative learning mechanisms have trouble explaining these results, the question remains as to what other non-associative mechanisms might be at play. There are at least two hypotheses that come to mind when one considers the visual perception literature. The first one is a perceptual grouping hypothesis, depicted in the left panel of Figure 6. The cues in our design varied in shape only and were all

composed of eight circles connected by lines. Given that the three different shapes in a compound were connected with one another via the central circle, perceptual grouping of cues of the same color may result in them being perceived and encoded as a single configuration (Palmer, 1999). This is, therefore, akin to a perceptual version of configural theory. In Figure 6, one can see that for group *intra* this would result in encoding of a single three-cue group during training, and encoding of a novel but similar three-cue group during testing. For this group, one would expect a rating for the testing compound ≤ 10, with the actual value depending on how much learning about the training group one assumes generalizes to the testing group.

The predictions are more complex for groups *extra* and *extra2*. In group *extra*, during training people should parse the compound into two groups: one composed of a target cue and a non-target cue, being relevant for outcome prediction (solid red ellipse in Figure 6), and one composed of a single non-target cue, being irrelevant for outcome prediction (dotted red circle in Figure 6). The testing compound would be parsed again into two groups: one being a trained relevant group of two cues, associated with 10 degrees of allergy (solid red ellipse in Figure 6), and one being a novel group composed of a single target cue (green circle in Figure 6). For this group, one would expect a rating for the testing compound > 10, with the actual value depending on how much learning about the two-cue training group one assumes generalizes to the single-cue testing group. Finally, in group *extra2*, during training people should parse the compound into two groups: one composed of the target cue, being relevant for outcome prediction (solid red circle in Figure 6), and one composed of the two non-target cues, being irrelevant for outcome prediction (dotted red ellipse in Figure 6). The testing compound would be parsed into three groups of single cues, two of them being trained groups associated with ten degrees of allergy. For this group, one would expect a rating for the testing compound ≈ 20. In sum, the predictions of this perceptual grouping hypothesis are that ratings to the ABX/ABY testing compounds should be: $extra2 \approx 20 > extra > extra2 \leq 10$. Thus, the perceptual

grouping hypothesis is more than superficially similar to configural theory, as it makes the same rank predictions as all versions of REM implementing some configural processing (i.e., $r > 0$; see Figure 5).

The second hypothesis inspired by the visual perception literature represents a shift farther away from traditional associative learning theory. Depicted in the right panel of Figure 6, this hypothesis proposes that the difficulty to individuate cues in our design led participants to perform a visual search for the target cues in the display (Wolfe, 2018). Because during training there was always a single target cue in a compound, all groups should learn to sample that cue during training (albeit with varied difficulty, depending on similarity between target and non-target cues). More importantly, groups would differ on their sampling strategy during testing. In visual search, the similarity between target and non-target cues is a key factor determining the strategy shown by participants to sample information from a display (Duncan and Humphreys, 1989), with similar cues leading to serial processing and dissimilar cues leading to parallel processing. Similarity between target and non-target cues should foster serial processing in groups *extra2* and *intra*. In addition, participants might show a tendency to stop their search and report the rating associated with the first target found, a phenomenon known as satisfaction of search in the visual search literature (Wolfe, 2018). Such a strategy reduces cognitive and timing costs compared to exhaustive search. The result is that participants in groups *extra2* and *intra* would mostly sample a single cue from the testing compounds ABX/ABY and report its associated rating of ten. On the other hand, dissimilarity of all cues in group *extra* would result in parallel processing of the two target cues, and reporting of an allergy rating higher than ten (i.e., a summation effect).

Thus, it is possible that a serial search process of overt attention makes participants sample either one (groups *intra* and *extra*) or all (group *extra2*) the stimuli in the compound to assess their predictions during the test with ABX/ABY. A direct implication

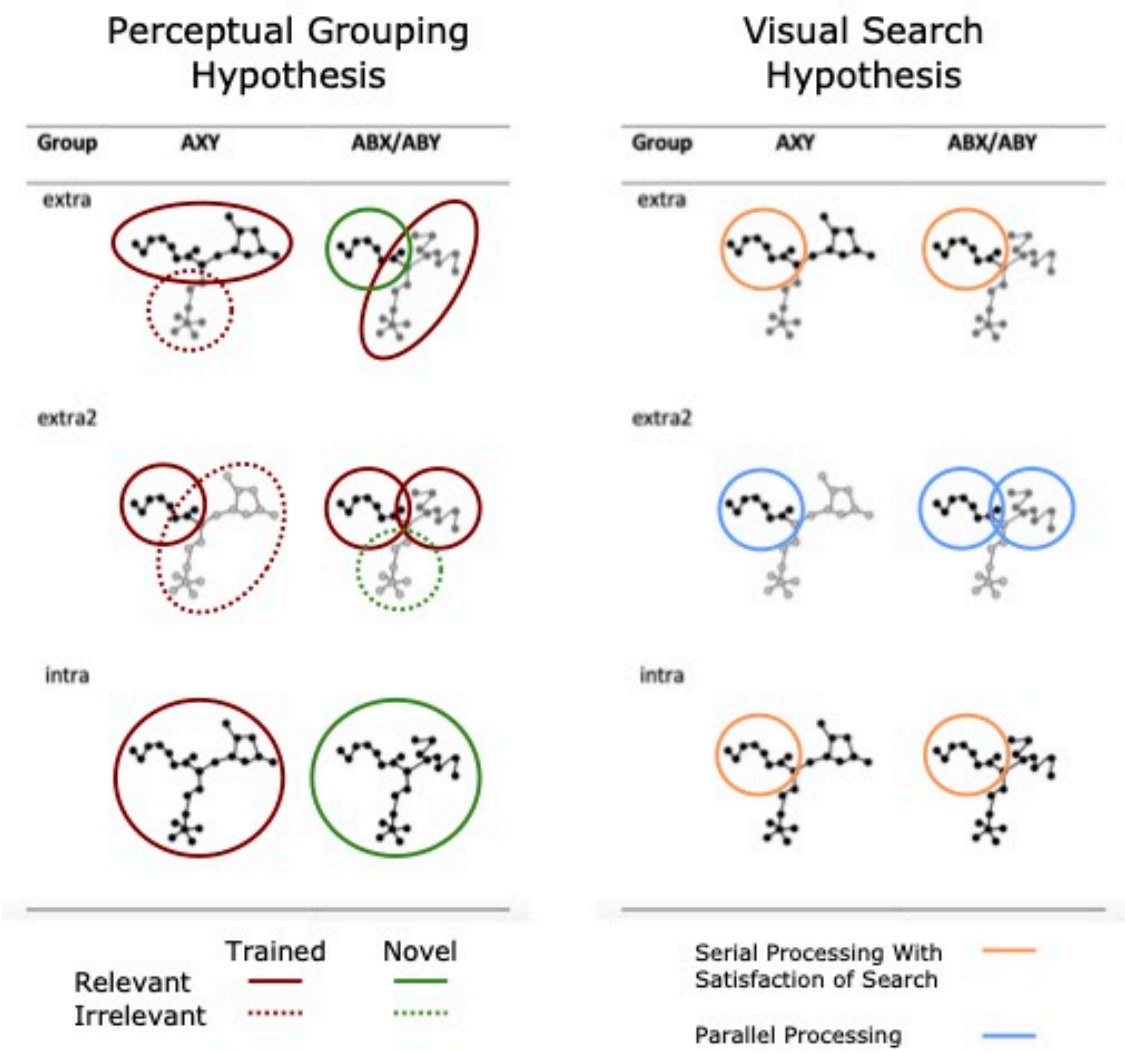


Figure 6. **Comparison of Perceptual Grouping and Visual Search hypotheses for Experiment 1 (see main text for details).**

of this idea is that the scores for this novel ABX/ABY compound in the different groups should follow a bimodal distribution with peaks in 10 and 20, reflecting the fact that participants either find one or both of the target cues to make predictions. This could be appreciated in the distribution of points in Figure 4b, but also in Figure 4c, which shows a density plot of score for the compound ABX/ABY in each group. The most noticeable feature of this plot is that the difference in summation between the groups is given by the proportion of participants showing either zero summation (10 points of allergy) or full summation (20 points of allergy) in the different groups. As similarity decreases from *intra*

to *extra2*, the proportion of participants processing all cues and consequently showing full (null) summation increases (decreases), with virtually no participants scoring in between these two values—save some in group *extra2*. The bimodality of scores was confirmed by Hardigan's Dip Test: $D = 0.10, p < .001$ and a mixture model which yielded a generative model with two Gaussians as the best explanation of the distribution of scores (Gaussian 1: $\mu_1 = 9.75, \sigma = 2.65$; Gaussian 2: $\mu_2 = 22.10, \sigma = 2.65$).

In sum, our data suggest that the difference in summation between groups in these experiments may be reflecting a search process where participants subsample one of the cues or the two target cues during the course of the test phase. The following two experiments were designed to further test the hypothesis of a sub-sampling process as the best explanation for the lack of summation in groups *intra* and *extra*, where cues are more difficult to be parsed and discriminated during the test stage.

Experiments 2a and 2b

To further test the sub-sampling hypothesis, in Experiments 2a and 2b we used the same stimuli as in group *intra* of Experiment 1 and divided participants in two groups. Group *intra* was a replication of group *intra* of Experiment 1 in which the outcome associated to both AXY and BXY was 10 points of allergy. As in Experiment 1, we expected the majority of participants to rate the compound as producing 10 points of allergy.

Critically, to test if participants are searching and sampling only one of the cues to make their predictions—and explaining why summation is not observed in this group—, group *intra2* involved two different outcome values. In Experiment 2a, we assigned 10 points of allergy for cue AXY and 8 points of allergy for cue BXY. Critically, the assignment of 10 and 8 points of allergy was reversed in Experiment 2b. To the extent that our stimulus manipulation prompts a search process and limited sampling and processing

of a single cue, the majority of participants in group *intra2* should rate the compound ABX/ABY as producing either 8 or 10 points of allergy, indicating that they have responded in accord with the value of only one of the two components.

The predictions of this sub-sampling hypothesis differ from those of contemporary associative theories in two respects. First, with their usual parametric assumptions, associative models predict lower response to ABX/ABY than to both AXY and BYX in both groups (*intra* and *intra1*) (see Figure 5). By contrast, a sub-sampling process anticipates a mean of ABX and ABY equal to the mean of AXY and BXY in both groups. Second, and perhaps most importantly, according to a sub-sampling hypothesis driven by visual search, provided that ratings to AXY differ from BXY in group *intra2*, it is expected that participants show a tendency to rate ABX/ABY equally to one of the reinforced cues A or B (either 8 or 10 points of allergy). On the other hand, associative theories (as well as the perceptual grouping hypothesis presented earlier) make a variety of predictions about the ratings to the compound ABX/ABY. In some cases, the prediction is that such ratings should be equal to *either* the higher *or* the lower of the two trained ratings (i.e., 8 or 10), but *both* ratings are never predicted for the same stimuli.

Given that the goal of Experiments 2a and 2b was to test whether participants rate the ABX/ABY compound similarly to either AXY or BXY, we focus our analysis on the distribution of scores and the non-parametric Friedman's test. Figure 7b presents the barplot and individual scores shown in jittered points for the trained (AXY and BXY) and novel (ABX and ABY) cues in the two groups of Experiment 2a. As expected, most participants rated AXY and BXY in accord with their trained outcomes (10 for AXY in both groups, and 8 and 10 for BXY in groups *intra* and *intra1*, respectively). In group *intra*, median scores were virtually identical for cues AXY, BXY and ABX/ABY with very little variation around 10 points of allergy which led to no reliable differences in the three distributions ($X^2(2) = 0.75, p = .16$). Differences were apparent in group *intra2*

a

Training	Test
AXY —> 10	
BXY —> 10 (*intra*)	**A**XY
BXY —> 8 (*intra2*)	**B**XY
CXY —> 0	**AB**X
DXY —> 0	**AB**Y
	CXY
	DXY

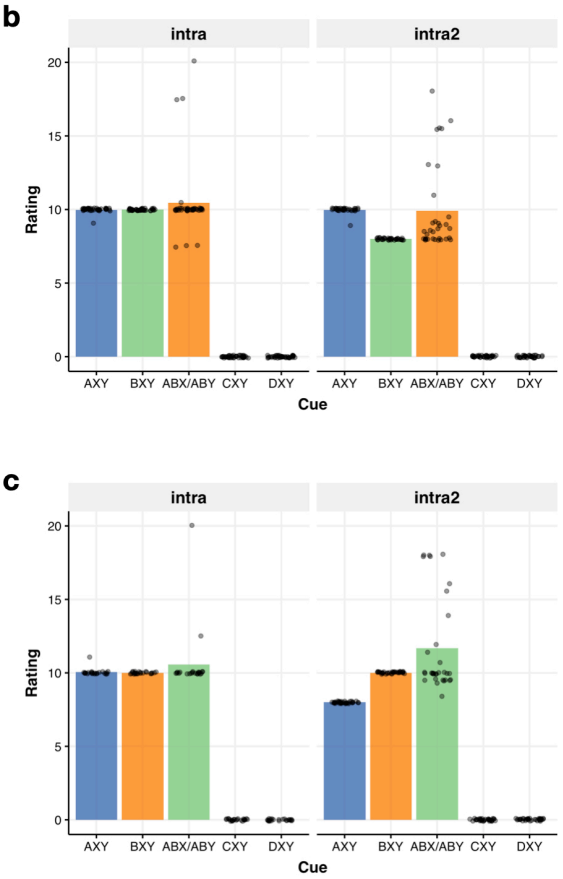


Figure 7. **Design and results of Experiments 2a and 2b**. **(a)** In Experiments 2a and 2b, cues were followed by different levels of allergy, which are represented by the numbers shown next to each of them. The only difference between Experiments 2a and 2b was the assignment of different levels of allergy outcome to cues AXY and BXY in group *intra2* (8 or 10 points of allergy, counterbalanced across the two experiments). **(b)** Barplots and individual ratings (shown in jittered points) given to each cue during the test in Experiment 2a. **(c)** Barplots and individual ratings (shown in jittered points) given to each cue during the test in Experiment 2b.

($X^2(2) = 44.69, p < .001$), where the scores to ABX/ABY (med=8.5) were closer to BXY (med=8) than to AXY (med=10). Dunn-Bonferroni post-hoc tests indicated that the distribution of ABX/ABY differed significantly from both AXY ($p = .01$) and BXY ($p =< .01$) after Bonferroni adjustments. These results are consistent with a sub-sampling process where the great majority of participants rate the predictions of ABX/ABY based on the predictive value of just of its constituent elements, rather than a combination of the two.

The results of Experiment 2b are summarized in the bottom plot of Figure 7c. As can be appreciated, the outcome of this experiment is basically the mirror image of Experiment 2a: In the case of group *intra*, we replicated for the third time the absence of any reliable difference among the cues ($X^2(2) = 2.46, p = .30$) which varied in a very small range around 10 points of allergy. In contrast, in group *intra2* the Friedman's test indicates reliable differences among the cues ($X^2(2) = 55.00, p < .001$), where the ratings to ABX/ABY (med=10) were again very similar to BXY (med=10) and different from AXY (med=8) ($p = 1.00$ and $p < .001$). Thus, ratings to the novel compounds ABX and ABY seem to be determined by the predictive value of B. This change in the distribution of ratings from Experiment 2a to Experiment 2b is not anticipated by any theory of associative learning (or by the perceptual grouping hypothesis) but is consistent with a visual search approach where participants sample only one of the two target cues to make their predictions of the compound under conditions of highly-complex visual stimuli.

A model of limited cue sampling

Taken together, the results reported above provide evidence for the hypothesis that sub-sampling of cues can bring about weak to no summation in humans. Our data suggest that when subjects cannot easily identify the cues forming a compound, they may engage in a search process which stops when one of the cues forming the compound is found.

It is worth noting that, in the animal conditioning literature, the majority of data where summation is not obtained come from auto-shaping experiments in pigeons using visual stimuli (Aydin and Pearce, 1995; Aydin and Pearce, 1997). In an auto-shaping procedure, pigeons receive pairings of visual stimuli and food, which results in the animals approaching the stimuli and pecking at them. Two features of autoshaping designs may encourage a visual search strategy. First, the proximity of the pigeon to the screen in which stimuli are displayed limits its ability to sample the large and complex stimuli usually displayed in compound generalization experiments. Second, just before the final ballistic movement that produces a peck, the peck's target is centered on the area dorsalis of the pigeon's retina, which is considered a "second fovea" due to its high density of cells. That is, pigeons specifically sample visual information from the pecked area, and they tend to peck on a circumscribed area of the screen displaying visual features predictive of reinforcement (Soto *et al.*, 2012).

Therefore, just as in the present experiments, it is possible that pigeons sample and process only the visual features of the stimuli that they happened to be pecking during a particular trial. At this point, they may continue to sample information from the target while relevant information from the rest of the display is never sampled (analogous to satisfaction of search), leading to a sub-optimal sampling of useful information in the stimulus. Under this process, summation should not be observed.

The purpose of this section is to show how these animal results, which have traditionally been conceived of as being a consequence of configural processing [2], may also be explained by sub-sampling of stimuli within a complex visual configuration. We simulate a formal model based on this idea.

[2] In what follows, we will refer to both REM with $r = .5$ and Pearce's model as "configural", recognizing that we are referring to the predictions made by these models and not the type of stimulus representation assumed in their formulations.

Learning rule

The model we propose learns according to reinforcement prediction-error (Bush and Mosteller, 1951). The predictive value or *associative strength* of stimulus i in trial n, v_i^n, is updated in accord with Equation 1

$$v_i^{n+1} = v_i^n + \alpha_i \beta(\lambda^n - v_i^n) \tag{1}$$

This algorithm assumes that the change in the predictive value of stimulus i is determined by the difference between the observed outcome and the current outcome expected from that stimulus.

Stimulus perception

The agent perceives cues in isolation and process them independently, updating their predictive value in accord with Equation 1. However, when a compound of cues is presented, the agent perceives visual features of the compound which are absent when the cues are presented in isolation. For example, as long as cues are presented in close proximity during compound trials (i.e., either overlapping or close to one another), it is possible that the subject will sample visual information in some areas of the display that are unique to compound trials, such as line intersections, corners, etc. This makes the compound trial different to a single cue trial by the addition of those features in the configuration. We assume that the agent treats those added features the same as any other cue [3].

[3] A similar idea is proposed in Wagner and Rescorla (1972) and REM. However, in contrast to these models, we do not rely on internal representations; rather, we simply recognize that objective properties of stimuli are added when compounds of cues are presented in close spatial proximity.

Sampling process

Finally, we assume that the probability of stimulus i being processed in any given trial is given by a *softmax* function incorporating salience and predictive value of each stimulus (Sanderson and Bannerman, 2011; Sutton and Barto, 1998):

$$p_i = \frac{e^{(\eta \alpha_i v_i)}}{\sum_j e^{(\eta \alpha_j v_j)}} \tag{2}$$

, where η is a decisiveness or *temperature* parameter that determines the extent to which the agent is biased to sample cues with low salience or predictive value ($j = [1, 2, ..., i, ...k]$) [4]. Finally, in each trial the agent samples from the stimulus array $S = [S_1, S_2, ..., S_k]$ for a single stimulus to process. We assume that the sampled cue S_i follows a categorical distribution:

$$Prob[S = S_i|\mathbf{p}] = Categorical(\mathbf{p}) \tag{3}$$

where $\mathbf{p} = [p_1, p_2, ..., p_k]$ is the vector of probabilities given by Equation 2 and k is the number of stimuli presented in a given trial during training and testing.

To illustrate how this model operates, see the schematic representation of Figure 8. In this example, we assume that the compound AB is presented. In this case, the configuration produces an additional unique cue X which is also perceived by the agent for this particular combination of cues. Once the agent samples cue A (shown in red), its value is updated according to the difference between the current value or prediction (v_i) and the

[4] Formally, the model should include the absolute value of v, since cues with high inhibitory strength (negative v) should command more attention than other cues in a given array (Parkhurst *et al.*, 2002). However, such implementation is only relevant in a model that allows for negative associative values for cues, which is not the focus of our model.

outcome observed (λ) in that trial.

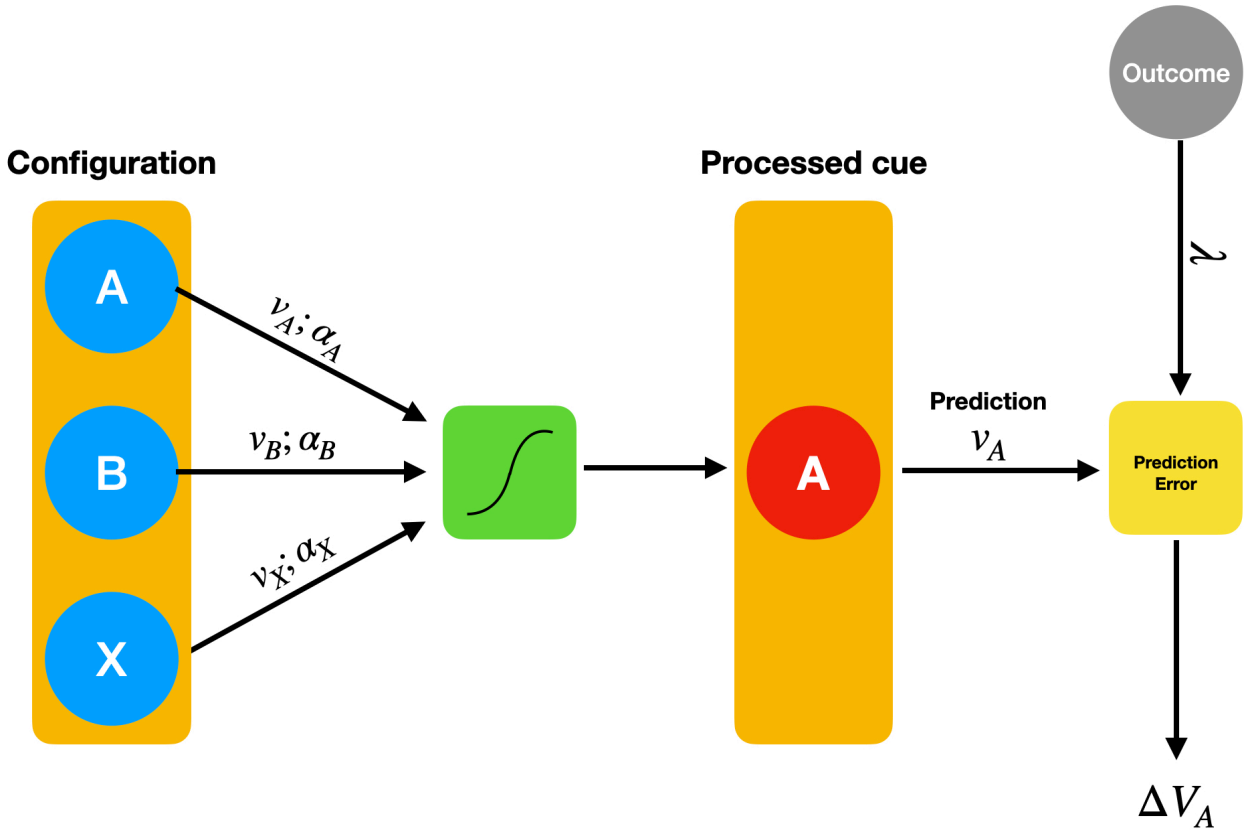


Figure 8. **A sub-sampling model.** On each trial where the agent is presented with a compound of cues that predict an outcome, the agent searches and samples one of the cues according to their current value (v) and salience (α). In this example, the compound AB is presented and an additional cue X is perceived for the compound. For illustrative purposes, we assume that cue A has been sampled in this trial (colored in red). Once the agent samples the cue, it updates its value according to a prediction error rule. Only the value of the sampled cue (A, in this example) is updated for the following trial.

For the following simulations, we assume that the salience of the unique cues is equal to the salience of the target cues; that is, $\alpha_{\text{target}} = \alpha_{\text{nonTarget}}$, and that the order of presentation of different trial types is random [5].

[5] The former assumption—that the salience of the cues presented and the unique cues represented by the subject in compound trials are equal—is the assumption generally held by Pearce (1994; 2002) in his modelling of configural theory.

To allow the agent to sample cues with low predictive value we set the temperature parameter η to 30 and the initial predictive value of each cue, both target and unique cues, v_i, $i \in [1, ..., k]$, to 0.05. Unless otherwise noted, the value of λ was set to 1 for reinforced trials and to zero for non-reinforced trials. Finally, we assume that an additional cue for each one of the possible combinations of cues is added when a compound is presented [6]. We ran 80 simulations for each experimental design. The values shown in the figures are the average values across all simulations.

In this basic formulation of a sub-sampling model, we do not attempt to provide a definitive solution to the elemental-configural controversy, but rather to highlight some cases in which a sub-sampling process driven by visual search might be the best explanation of the data. Therefore, to facilitate the exposition we focused on the simple summation task in which many studies have found absence of summation in animals rather than the more complex summation design with common cues employed in this paper.

In addition, note that the model presented here is an attempt to capture only the *serial processing with satisfaction of search* strategy presented in Figure 6 to explain results from groups *extra* and *intra*. The *parallel processing* strategy assumed to happen in group extra2 is not captured by the model. We assume that traditional elemental models, such as the Rescorla-Wagner model (Rescorla and Wagner, 1972), can explain behavior under conditions that foster parallel processing (i.e., clearly identifiable component cues).

[6] For example, if the compound ABC is presented, we assume that there is an additional unique cue represented for each possible pair—AB, AC, BC—and the compound of three cues—ABC. The compound ABC is thus represented as ABCXYZV, where X, Y, Z, and V are the additional unique cues.

Results

Summation

We start by replicating the conditions of a simple summation design in pigeon autoshaping. To this end, we assume for simplicity that both components predict the same outcome value and have equal saliencies ($\alpha_A = \alpha_B = .4$). The results of this simulation are shown in the right panel of Figure 9a. As shown in the figure, the sub-sampling model correctly predicts the failure to find a summation effect in these experiments. Under the sub-sampling model, the system randomly samples the unique cue that is only perceived in the testing phase (ABX, where X is the unique cue), which tends to bring down responding to AB (ABX) compared to the elements A and B. The left hand panel of Figure 9a shows a similar design reported by Rescorla and Coldwell (1995) using pigeon autoshaping, where the authors failed to find a summation effect.

In a second set of simulations, we investigated the predictions of a sub-sampling approach when cue B predicts a lower outcome value than A, trying to match the conditions of Experiment 2a. To account for the fact that B seemed to be more salient and drive responding in that experiment, we set the value of α_A to .4 and the value of α_B to .5. A value of .4 was also set to the unique cue X ($\alpha_X = .4$). To account for different outcome values predicted by each cue, we set $\lambda_A = 1 > \lambda_B = .95$. The model correctly predicts the pattern of results of Experiment 2a, in that responding to A, the most salient cue, is higher than to B (see Figure 9b). The model also replicates our finding that responding to AB would be closer to B than to A. Lastly, we tried to match the conditions of Experiment 2b by reversing the roles of A and B, so that B predicts a higher outcome value than A ($\lambda_A = .95, \lambda_B = 1$). Again, the sub-sampling model correctly captures the pattern of behavioral results in this experiment, in that responding to AB should be closer to the outcome predicted by B, and higher than in Experiment 2a (see Figure 9c).

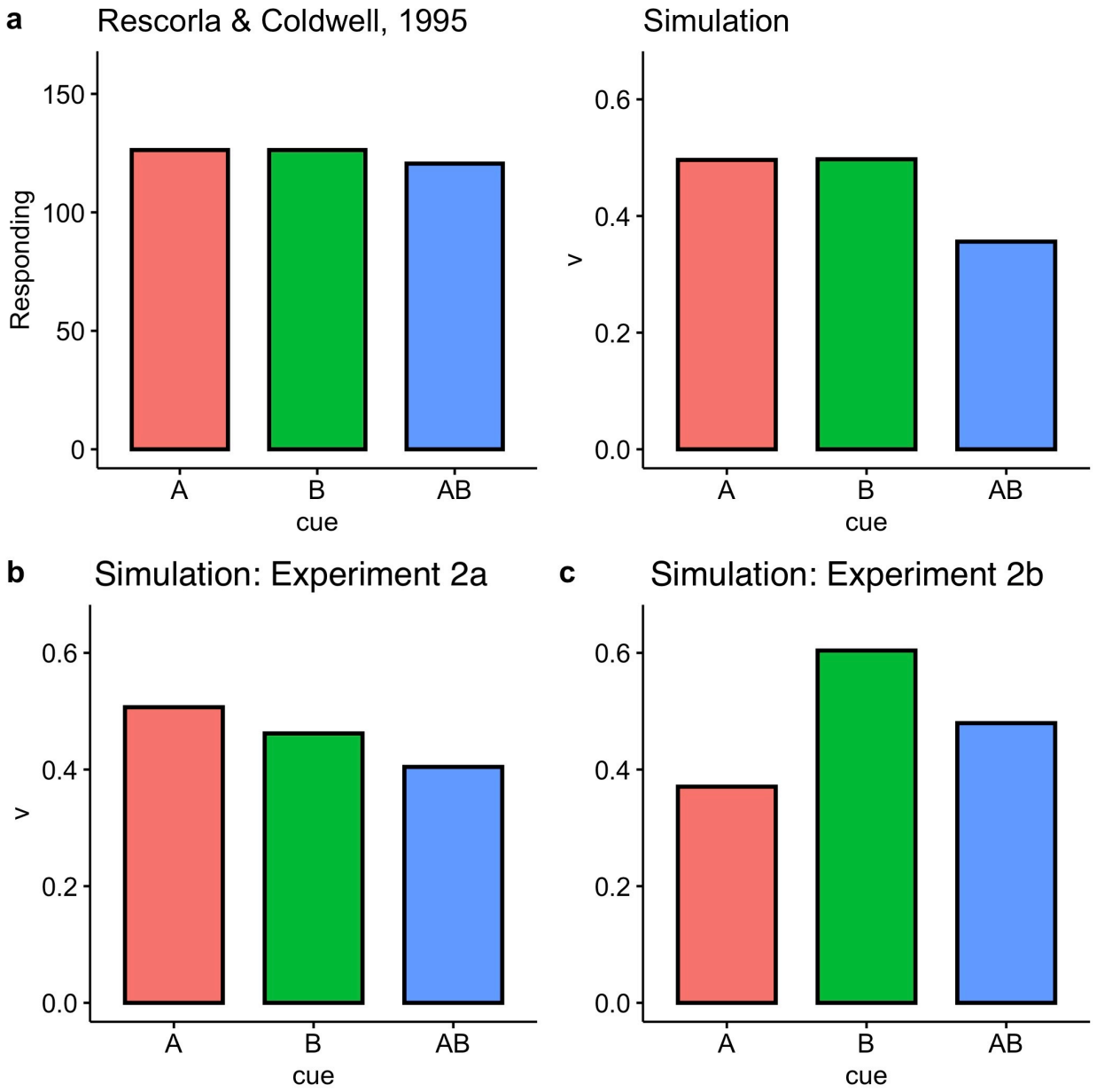


Figure 9. (a) Simulations of the sub-sampling model for different summation designs. The left panel shows the results obtained by Rescorla and Coldwell (1995) in pigeon autoshaping. The right panel shows the simulations of the model assuming that the saliencies of A and B are equal. (b) Simulations of a sub-sampling model for a summation experiment where the salience of B is assumed to be higher than that of A, but B predicts a lower outcome value. (c) Simulations of a sub-sampling model for a summation experiment where the salience of B was higher A but the outcome value predicted by B is higher than the value predicted by A.

Differential summation

In another experiment in pigeon auto-shaping, Pearce and colleagues (1997) found that responding at test for the compound of three cues, ABC, was weaker when the three cues were separately paired with a reinforcer (A+, B+, C+) than when the cues were paired with the same reinforcer, but in compounds (AB+, AC+, BC+) (see Figure 10, left panel). Pearce's configural model can readily account for these results under the assumption that the similarity between the compounds comprising two cues (AB, AC, BC) and the tested compound ABC is higher than that between each A, B and C and the compound ABC.

Figure 10 (right panel) depicts the simulations of this design in the sub-sampling model, which also correctly captures these data. In the case of single-cue training, at the end of training ABC is composed of four unique cues with very low predictive values (.05 each, by assumption), whereas the same ABC compound is composed of the same four unique cues, some of which have been experienced by the agent during training and have therefore acquired higher predictive values than in the case of single-cue training. During the presentation of ABC at test, the agent sometimes sample these higher-valued unique cues. As a consequence, the agent tends to respond more to the compound ABC after compound training than after single-cue training, in agreement with the results observed by Pearce et al. (1997). In other words, our model proposes that it is not the similarity between two-stimuli compounds like AB and the test compound ABC what increases summation in this design, but rather the availability of additional cues perceived by the subject which can be sampled during both training and testing.

Reversing a conditioned inhibitor

Pearce and Wilson ran an experiment which included a feature-negative design of the form A+, AB- in a first phase, and reinforced presentations of B alone in a second phase (B+; completing a negative patterning design across phases). In the learning literature, the

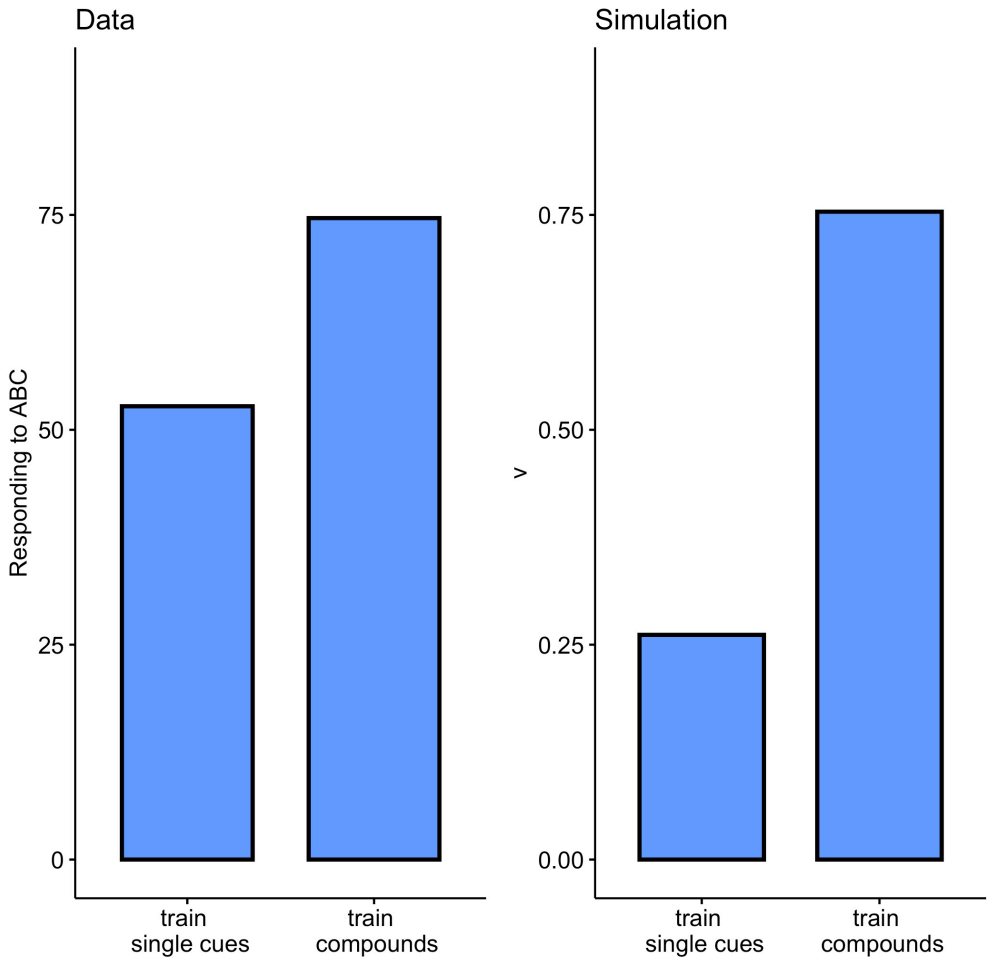


Figure 10. **Simulations of a sub-sampling model for a differential summation design.** Left panel. Results obtained by Pearce and colleagues (1997) in pigeon autoshaping. The bars show the responding to the compound ABC after training with the single cues A, B and C or after training with the compounds AB, BC, and AC. Right panel. Simulations of the sub-sampling model for the same design.

first phase turns cue B to what is referred to as a *conditioned inhibitor*, meaning that B signals the absence of an otherwise present reinforcer. In contrast to an elemental theory which would predict B to recover its predictive value so that responding to AB should be higher than A or B alone at the end of the second phase, Pearce and Wilson observed instead lower responding to AB than A or B alone (see Figure 11, left panel). This result is also anticipated by a sub-sampling approach (see Figure 11, right panel). During the first stage, A acquires more value than both B and the unique cue represented for AB, call it X. During the second stage, B acquires value independently of the sub-sampling process. During the final test with AB, however, the agent still samples the unique cue X, whose value has not been modified during the second phase, staying at a low level. Responding to AB is therefore lower than to either A or B.

Discussion

This paper reports the first empirical evidence of similarity affecting summation in humans. In Experiment 1, when stimulus properties were designed to make it difficult for subjects to parse them and deploy counting strategies, the difference in similarity between target and non-target cues within a compound was reflected in different ratings for the compound ABX/ABY during the test. In addition, although previous studies had demonstrated that summation could disappear when participants assumptions are affected by prompting rational rules concerning the independence of cues in a compound (Pérez *et al.*, 2018), in Experiment 1 we observed null summation in the absence of such manipulations. We replicated this absence of summation across three different experiments.

By simulation, we found that these data cannot be easily accommodated by current associative learning models. Indeed, for the type of complex summation design employed in our experiments, flexible models such as REM predict negative summation for a wide range of values in its parametric space. These predictions are not consistent with the results obtained in Experiment 1.

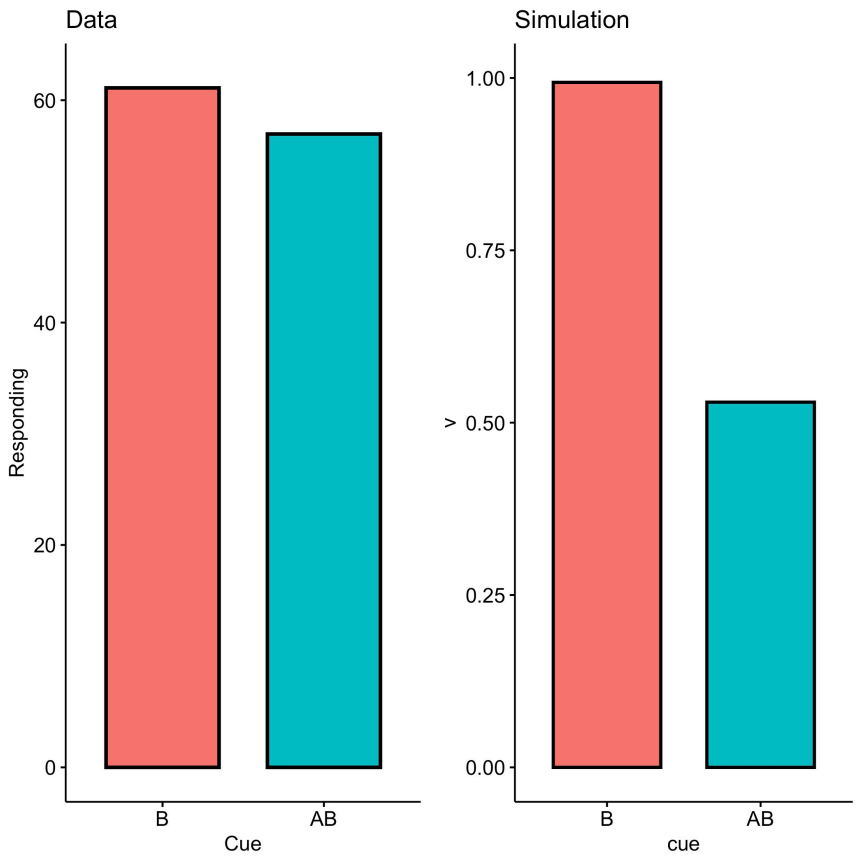


Figure 11. **Simulations of a sub-sampling model for Experiment 3 of Pearce and Wilson (1991).** The experiment comprises a first phase where A is reinforced and the compound AB is not (A+, AB-). In the second phase B is reinforced in isolation. The left panel shows the original data obtained by these authors. Simulations of the sub-sampling model are shown in the right panel.

We hypothesized an alternative view for these results. We thought the data in groups *intra* and *extra* could be more readily interpreted as coming from a visual earch process where participants sample one of the stimulus in the compound and learn only about that sampled stimulus in any given trial. This sub-sampling phenomenon should be a natural consequence of a design with compounds that are not easily parsed into different components. In Experiments 2a and 2b we obtained evidence for this hypothesis by showing that the majority of participants rated a compound ABX/ABY equal to the value of one of the cues in the compound. Again, associative models cannot explain these results as they predict lower responding to ABX/ABY in groups *intra2* than in group *intra*—in

the former group the asympotic values for AXY and BXY are 8 and 10, respectively, whereas these values are both 10 for group *intra*. Moreover, when we swapped the outcome values predicted by AXY and BXY between Experiments 2a and 2b, participants followed the value of one of the stimuli, BXY, probably due to it having higher salience. Such rating change is not predicted by current associative learning models.

We presented a computational implementation of this sub-sampling hypothesis. Through simulation we showed how the model is able to capture not only the pattern of data presented here, but also a wider range of previous phenomena that are usually attributed to configural processing of stimuli in the animal literature. In contrast to these models, however, we obtained these results by taking a view that relies on objective properties of stimuli—some of which are absent when stimuli are presented in isolation—rather than on internal representations, as it is usually assumed in associative learning models, which makes this approach more parsimonious than current theories based on internal stimulus representations.

The sub-sampling model outlined in this paper does not contain a mechanism for sampling multiple cues in a compound and is therefore not meant to capture summation data. We have assumed that traditional elemental models, such as the Rescorla-Wagner model (Rescorla and Wagner, 1972), can explain behavior under conditions that foster parallel processing during visual search (i.e., clearly identifiable component cues). A simple extension could be formulated to formalize a mechanism that chooses between sub-sampling and full sampling strategies. As a first approximation, the mechanism could be implemented in an all-or-none fashion. For example, the agent could arbitrate between strategies following a rule such as

$$M = \begin{cases} m_{sub} & \text{if } d > D \\ m_{full} & \text{if } d \leq D. \end{cases} \tag{4}$$

where d is a parameter that determines how difficult is to parse components—which should be a direct function of the similarity between them—and D is a threshold that varies across individuals and which will determine whether sub-sampling or full sampling is deployed.

Regardless of the possible extensions of a sub-sampling approach, our work provides empirical and computational evidence suggesting that under conditions in which components in a compound are difficult to parse and identify, participants may engage in an inefficient search process by which they only sample a subset of the compound. This sub-sampling approach can explain the empirical evidence presented in this paper better than current models of learning, and a simple formalization of its principles can also capture previous data from the animal literature usually interpreted as supporting configural processing. We believe this interpretation is testable, important, and should be considered as a plausible alternative hypothesis in the literature on learning and generalization.

Materials and Methods

Participants were tested in desktop Windows (c) computers running Psychopy (version 1.75 for Experiment 1; version 1.82.4 for Experiments 2 and 3; Peirce, 2007). Responses were recorded from standard PC keyboards.

Statistical analysis

Statistical analyses were performed using RStudio and SPSS 27.0. For all the pre-planned comparisons we calculated a Welsh t-test and included Cohen's D, along with

a 95% confidence interval on this estimate, as a measure of effect size. When reporting interactions between factors, we computed η_p^2. Following Steiger (2004), we report a 90% confidence interval on this estimate. The reliability of the results was contrasted against the usual criterion of $\alpha = .05$. When a null effect was expected, a Bayes Factor in favor of the null (BF_{01}) is reported.

Experiment 1

Participants. Eighty-six undergraduate students from Florida International University participated in Experiment 1. Participants did not have previous experience with the experimental procedure and were tested simultaneously and in the same room. The Institutional Review Board of Florida International University approved all the studies in this paper (IRB-15-0460). Written informed consent was obtained from all participants. They were given course credit for their participation.

Participants were randomly assigned to one of three groups: *intra* ($n_{intra} = 27$), *extra* ($n_{extra} = 28$) and *extra2* ($n_{extra2} = 31$). In all of the studies in this paper, we determined our exclusion criteria before data collection by following the criteria employed in Pérez *et al.*, 2018. Participants that did not give on average a rating between 7 and 13 points of allergy to cues that predicted allergy and between 0 and 3 points of allergy to cues that did not predict allergy, were left out from the analysis. The final number of participants in Experiment 1 was $n_{intra} = 18$, $n_{extra} = 13$, $n_{extra2} = 27$.

Procedure

Groups differed in the similarity between cues in the display (see Figure 2). Each stimulus was created from three different cues that "branched out" from a central point. Among these branches, only one of them represented the target cue associated with either allergy or no allergy during training (A, B, C or D). The other two branches were non-target cues that could not predict the presence or absence of allergy (X and Y).

During the test, the compound AB was comprised by two target branches together with an additional non-target cue (ABX or ABY). In group *intra*, all these "branches" were of the same color (black), but differed in shape. In group *extra*, A and B differed in color (grey and black), but they shared color with the non-target cues (X and Y, one grey and one black). In group *extra2*, the target cues were the same as in group *extra*, but now the non-target cues had a distinctive color as well. In all groups, A and B, which predicted allergy, shared color with cues C and D, which predicted no allergy. Thus, all participants, regardless of group, had to attend to shape; color was irrelevant to solve the discrimination.

Experiment 2a

Participants

75 undergraduate students from Florida International University were randomly assigned to one of two groups ($n_{intra} = 40, n_{intra2} = 35$) and were compensated with course credit for their participation. The exclusion criteria were the same as those of Experiment 1, except that the admittance interval for the mean rating to A and B in group intra2 was set to [5-11]. The final number of participants per group was $n_{intra} = 39$, $n_{intra2} = 33$.

Procedure

The procedure was the same as described for group *intra* of Experiment 1, with only one exception: In group *intra2*, stimulus BXY was associated with 8 points of allergy during training (see Figure 7a).

Experiment 2b

Participants

80 undergraduate students from Florida International University were randomly assigned to one of two groups ($n_{intra} = 42, n_{intra2} = 38$) and were tested under the same

conditions of Experiment 2a. The final number of participants per group was $n_{intra} = 24$, $n_{intra2} = 33$.

Procedure

The procedure was the same as in Experiment 2, except that the outcome values for A and B were interchanged in group *intra2*. In this experiment, the outcome assigned to BXY was 10 while a value of 8 was assigned to AXY.

Acknowledgements

We thank Jaron Colas and Tony Dickinson for their valuable comments on a previous version of this manuscript. S.D.G.

Open Practices Statement

The data and materials for all experiments are available at https://osf.io/xqnpk/

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Aydin, A. and Pearce, J. M. (1995). Summation in autoshaping with short-and long-duration stimuli. *The Quarterly Journal of Experimental Psychology*, **48**: 215–234.

Aydin, A. and Pearce, J. M. (1997). Some determinants of response summation. *Animal Learning & Behavior*, **25**: 108–121.

Brandon, S. E., Vogel, E. H., and Wagner, A. (2000). A componential view of configural cues in generalization and discrimination in Pavlovian conditioning. *Behavioural brain research*, **110**: 67–72.

Bush, R. R. and Mosteller, F (1951). A model for stimulus generalization and discrimination. *Psychological review*, **58**: 413–423.

Duncan, J. and Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, **96**: 433–458.

Glautier, S (2002). Spatial separation of target and competitor cues enhances blocking of human causality judgements. *Q J Exp Psychol B*, **55**: 121–135.

Glautier, S, Redhead, E., Thorwart, A., and Lachnit, H. (2010). Reduced summation with common features in causal judgments. *Experimental psychology*, **57**: 252–259.

Harris, J. a. (2006). Elemental representations of stimuli in associative learning. *Psychological review*, **113**: 584–605.

Kehoe, E. J., Horne, A. J., Horne, P. S., and Macrae, M. (1994). Summation and configuration between and within sensory modalities in classical conditioning of the rabbit. *Animal Learning {&} Behavior*, **22**: 19–26.

McLaren, I. P. L. and Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal learning & behavior*, **30**: 177–200.

Melchers, K. G., Shanks, D. R., and Lachnit, H. (2008). Stimulus coding in human associative learning: flexible representations of parts and wholes. *Behavioural processes*, **77**: 413.

Morey, R. D. *et al.* (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, **4**: 61–64.

Palmer, S. E. (1999). *Vision science: Photons to phenomenology*. MIT press.

Parkhurst, D., Law, K., and Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research*, **42**: 107–123.

Pearce, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological review*, **94**: 61–73.

Pearce, J. M. (1994). Similarity and discrimination: a selective review and a connectionist model. *Psychological review*, **101**: 587–607.

Pearce, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal learning & behavior*, **30**: 73–95.

Pearce, J. M., Aydin, A., and Redhead, E. S. (1997). Configural analysis of summation in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**: 84.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, **162**: 8–13.

Pérez, O. D., San Martín, R., and Soto, F. A. (2018). Exploring the Effect of Stimulus Similarity on the Summation Effect in Causal Learning. *Experimental Psychology*, **65**: 183–200.

Redhead, E. (2007). Multimodal discrimination learning in humans: Evidence for configural theory. *Quarterly Journal of Experimental Psychology*, **60**: 1477–1495.

Rescorla, R. and Coldwell, S. E. (1995). Summation in autoshaping. *Animal Learning & Behavior*, **23**: 314–326.

Rescorla, R. and Wagner, A. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning II: Current research and theory*, **2**: 64–99.

Sanderson, D. J. and Bannerman, D. M. (2011). Competitive short-term and long-term memory processes in spatial habituation. *Journal of experimental psychology. Animal behavior processes*, **37**: 189–199.

Schultheis, H., Thorwart, A., and Lachnit, H. (2008). Rapid-REM: A MATLAB simulator of the replaced-elements model. *Behavior Research Methods*, **40**: 435–441.

Soto, F. A., Gershman, S., and Niv, Y (2014a). Toward a General Theory of Generalization: Rational Principles of Dimensional Generalization Can Explain Compound Generalization Phenomena. *Psychological review*: 1–79.

Soto, F. A., Vogel, E. H., Castillo, R. D., and Wagner, A. (2009). Generality of the summation effect in human causal learning. *Quarterly journal of experimental psychology (2006)*, **62**: 877–889.

Soto, F. A., Gershman, S. J., and Niv, Y. (2014b). Explaining compound generalization in associative and causal learning through rational principles of dimensional generalization. *Psychological review*, **121**: 526.

Soto, F. A., Siow, J. Y., and Wasserman, E. A. (2012). View-invariance learning in object recognition by pigeons depends on error-driven associative learning processes. *Vision Research*, **62**: 148–161.

Steiger, J. H. (2004). Beyond the F test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological methods*, **9**: 164–182.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. Vol. 1. 1. MIT press Cambridge.

Thein, T., Westbrook, R. F., and Harris, J. A. (2008). How the associative strengths of stimuli combine in compound: Summation and overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes*, **34**: 155–166.

Thorwart, A., Livesey, E. J., and Harris, J. A. (2012). Normalization between stimulus elements in a model of Pavlovian conditioning: Showjumping on an elemental horse. *Learning & behavior*, **40**: 334–346.

Thorwart, A., Schultheis, H., König, S., and Lachnit, H. (2009). ALTSim: A MATLAB simulator for current associative learning theories. *Behavior Research Methods*, **41**: 29–34.

Vogel, E. H., Ponce, F. P., and Wagner, A. R. (2017). A theoretical analysis of transfer of occasion setting: SOP with replaced elements. *Behavioural processes*, **137**: 19–32.

Wagner, A. (2008). Evolution of an elemental theory of Pavlovian conditioning. *Learning & behavior*, **36**: 253–65.

Wagner, A. and Rescorla, R. (1972). Inhibition in Pavlovian conditioning: Application of a theory. *Inhibition and learning*: 301–336.

Wolfe, J. M. (2018). Visual Search. en. *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*. Ed. by J. T. Serences. Vol. 2. Hoboken, NJ: John Wiley & Sons: 1–55.