Title
Dual Processes on Dual Dimensions: Associative and Propositionally-Mediated Discrimination and Peak Shift.

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Johnson, Toby D
McLaren, R.P.
Civile, Ciro
et al.

Publication Date
2021

Peer reviewed

Dual Processes on Dual Dimensions: Associative and Propositionally-Mediated Discrimination and Peak Shift.

Toby Johnson (T.Johnson@exeter.ac.uk)
Ciro Civile
Rossy McLaren
IPL McLaren (i.p.l.mclaren@exeter.ac.uk)

Psychology, University of Exeter, Exeter, UK

Abstract

Dual-process accounts posit that human learning can occur as a consequence of both associative and propositional processes. This can be contrasted with single process accounts that suggest learning is entirely propositional. In this paper, we offer evidence for both associative and propositional processes using a within-subjects two alternative forced choice discrimination paradigm. Stimuli that varied concurrently along two dimensions were created and each participant's awareness was directed toward one, facilitating rule induction (i.e., propositional processing) on that dimension. Performance on the other dimension was then used to assess associatively-based performance. We report results that are initially inconsistent with both single process and dual-process accounts of discrimination learning. However, we then show how an associative network, that represents stimuli integrally, can predict the performance shown by participants in the experiment, providing evidence for a dual-process account.

Keywords: Associative learning; Dual-Process; Single Process; Peak shift;

Introduction

There is an on-going debate concerning the mechanisms responsible for learning in humans. Some argue that it is accomplished by a single process that relies solely on a propositional system (Mitchell et al., 2009), akin to the explicit system 2 of the dual process theory of reasoning (Stanovich & Keith, 2000). Learning via this system requires awareness and effort, producing conscious knowledge about the relationship between events and/or stimuli. Others suggest it is best explained via two systems (dual-process account; McLaren et al., 1994, 2019). Dual-process theorists agree that learning can rely upon the propositional processes outlined above, but also stipulate the existence of associative processes that automatically respond to contingencies by forming links between event representations, without necessarily relying upon conscious knowledge or requiring any additional effort on the part of the subject. These processes are analogous to the implicit, system 1 of the dual process theory of reasoning (Stanovich & Keith, 2000).

One piece of evidence in support of associative processes comes from the peak shift effect, first described by Hanson (1957). Using stimuli that varied along a dimension, Hanson trained pigeons to discriminate between two lights with wavelengths of 550nm (S+) and 560nm (S-). During testing, responding was greatest to 540nm, declining towards 530nm. The peak of responding had shifted away from (beyond) S+ in the direction opposite to S-. One reason for believing that this result is indicative of associative processes is that it can be easily explained by instantiations of this type of processing in connectionist networks (e.g. McLaren & Mackintosh, 2002) using elemental representational assumptions.

Peak shift has also been demonstrated in humans, although responding often follows the monotonic gradient, indicative of propositional processes (e.g., Livesey & McLaren, 2009). That is, as the test stimulus moves toward the ends of the dimension, responding increases or plateaus. For example, if human participants are trained to categorize two similar stimuli, dark blue (S+) and light blue (S-), into two separate categories, they will respond most accurately to the darkest and lightest blue they are presented with. This is because they are likely to infer the rule "if darker blue respond category one; if lighter blue respond category two"; thus, the darkest and lightest of stimuli are the most easily differentiated and can be responded to the most accurately.

Wills and Mackintosh (1998) successfully demonstrated peak shift using a series of abstract icons to represent individual stimuli. Each stimulus was composed of a different number of different icons that roughly followed a Gaussian distribution. The authors hypothesized that the relationship between the icons and responding was difficult to verbalize, providing the conditions for associative processes, rather than propositional, to dominate learning. In line with this analysis, peak shift has been reported in humans using a variety of stimuli where the basis for discrimination is not easily articulated (e.g. McLaren & Mackintosh, 2002; Derenne, 2019). Similarly, it has been demonstrated using other behavioral manipulations, such as reducing training or the contingency between the stimulus and outcome (Jones & McLaren, 1999).

As indicated earlier, peak shift can be readily modelled by an associative system that represents stimuli as a set of elements. McLaren and Mackintosh (2002) provide one such model that utilizes an error correcting learning rule. In this model, stimuli on a dimension are represented by activation of overlapping units. Each stimulus is represented primarily by the activation of one unit corresponding to its position on the dimension, but also to a lesser extent by neighboring units. This pattern of activation follows a Gaussian gradient. Imagine stimulus S+, which is represented primarily by the activation of unit 6, but also by partial activation of units 5 and 7, and weak activation of units 4 and 8. If stimulus S- is represented in a similar fashion with primary activation of unit 5, an error correcting learning rule will ensure that the units that differentiate S+ and S- accrue the highest excitatory value, and that those that are shared between the stimuli don't. That is, for S+, units 7 and 8 will accrue the highest excitatory value. When tested along the dimension, the stimuli that are represented primarily by the activation of these units will show greater associative strength than S+, giving peak shift.

Recent research, however, has provided evidence to support the argument that the effect can be explained entirely by propositional processes (Lee, Hayes, & Lovibond, 2018). The authors suggest that as well as a 'linear' rule, resulting in a monotonic increase (or decrease) in responding toward the end of the dimension, some individuals may employ a 'similarity' rule, leading to peaked gradients around S+ and S-. Averaging of the response functions produced by these two different rules can lead to a peak shifted gradient. Using a fear conditioning paradigm, Lee et al. (2018) demonstrated that averaging across participants who self-reported using these two rules created the peak shift effect. Further research has provided support for this theory (e.g., Lovibond, Lee, & Hayes, 2019).

The present study aims to demonstrate both associatively driven and rule-based responding using a within-subjects experimental design. To achieve this, stimuli were constructed that varied along two dimensions, brightness, and hue. Participants were pre-trained using two stimuli belonging to mutually exclusive categories that varied more obviously on one dimension compared to the other. This was designed to promote awareness and rule induction to the more obvious dimension (aware), whilst still allowing associative strength to be generated toward the second dimension (unaware), as this dimension is nevertheless predictive of category membership. Participants then received further training with two more stimuli that varied to the same degree on both dimensions. This allowed further hypothesis testing, by the participant, of any rule that may have been induced regarding the aware dimension, whilst providing further opportunity for the other, unaware dimension, to garner more associative strength.

Using this design, a single process account would predict that, in the limiting case that participants are truly unaware that stimuli vary along this second dimension, they will show no learning toward it when tested, and thus will produce a flat generalization gradient that does not differ from chance. A dual-process account would predict that even if participants fail to notice the unaware dimension, associative processes will still apply, and associative strength will still accrue. When tested on stimuli that vary along this dimension, participants might well show peak shift. Both accounts predict rule-based responding to the aware dimension.

Method

Participants and Apparatus

200 individuals were recruited using online recruitment platforms, 99 from Prolific, and 101 from the University of Exeter. This was sufficient according to a power analysis based on a small to medium effect size for peak shift. Participants were required to have full color vision, not have any reading disorders, and be a native speaker of English. All participants reported normal color vision. The experiment was created and hosted using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). All participants completed the experiment using a desktop or laptop. However, no restrictions were set on browser type, screen size, or resolution.

Stimuli

All stimuli used measured 426 × 426 pixels and appeared individually in the center of a white screen. Two different types of stimuli were created and used, experimental stimuli and filler stimuli. The experimental stimuli comprised of colored squares which varied along the Hue, Saturation, and Value (HSV) scale (see Table 1 for their properties). Each stimulus was created by modifying two parameters of the HSV scale, hue (i.e., color) and value (i.e., brightness), to create stimuli that varied by equal increments along these two dimensions. 11 such stimulus values were created, but only seven were used (1, 3, 5, 6, 7, 9, and 11; see Table 1). In total, 49 experimental stimuli were created by combining all the dimension values (7 × 7). Saturation remained at 50 for all stimuli. The 49 filler stimuli were made of an array of abstract icons and have been used in previous discriminative studies (see Wills & Mackintosh, 1998). These were used to prevent trial by trial comparison of the experimental stimuli.

Table 1: Brightness (Value) and Color (Hue) properties of the stimulus values used.

	Stimulus Value						
Position	**1**	**3**	**5**	**6**	**7**	**9**	**11**
Value	75	65	55	50	45	35	25
Hue	118	134	150	158	166	182	198

Stimuli were denoted as varying along dimensions X (value; brightness) and Y (hue; color). Each stimulus will now be represented as a position, X,Y in that two-dimensional space.

Design and Procedure

Table 2 gives the experimental design. Participants were required to categorize visual stimuli using either the 'X' key, or the '.' (full stop) key. They were initially given feedback if they were incorrect and told that, from this feedback, they could learn the correct responses for the stimuli presented. Participants were not informed of the nature of the stimuli, only that the correct response depended entirely upon the appearance of the stimuli. Finally, there was a testing phase, in which no feedback was given, and a questionnaire at the end of the experiment.

Table 2: Experimental Design

	Condition	
	Brightness	**Color**
Pre-Training	3,5 vs 9,7	5,3 vs 7,9
Further Training	5,5 vs 7,7	
Test	Stimulus Space (1,1→ 11,11)	

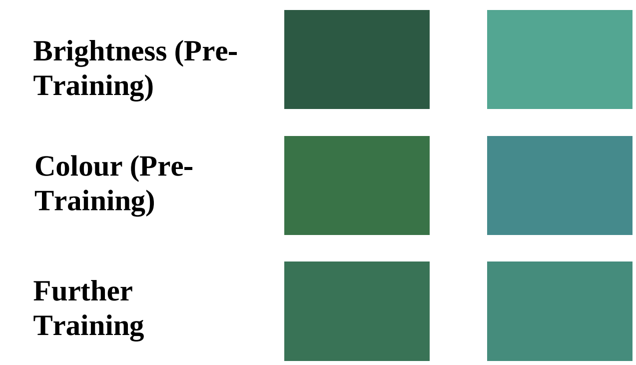

Figure 1: Stimuli used during pre-training for Brightness and Color groups, and further training for both.

Participants were divided into two groups, ‘Brightness’ and ‘Color’, referring to which dimension was designed to be the "aware" one. Training was split into two stages, pre-training and further training; the group manipulation affected only pre-training. All training consisted of alternating presentations of the experimental and filler stimuli. In all training stages, participants were required to categorise one stimulus as belonging to the left category (‘X’ key) and the other as belonging to the right category (‘.’ key). For each participant, each side of the dimension was associated with the same category. Feedback after incorrect responses consisted of the message ‘WRONG!’ appearing in the centre of the screen. This was displayed for 500ms.

Each training stage was eight blocks in length, with 12 trials per block. Each block contained three trials each of both experimental stimuli and three trials each of both filler stimuli. Trial order within each type of stimulus was random. Four blocks were presented sequentially, such that participants were required to categorise 48 trials, after which they were given a break before the next 48 trials. Trial length was four seconds.

During the testing stage participants were required to categorise all possible combinations of stimulus values without feedback. There were 49 of these, presented in a random order, interleaved with the filler stimuli. Testing consisted of 196 trials, presented as one block.

Finally, participants were required to complete a questionnaire that was adapted from Lee, Hayes, and Lovibond (2018). Among other things, this aimed at exploring whether participants noticed any relationships between the stimuli and the categories they belonged to.

The questionnaire started with a two-alternative forced-choice (2AFC) question asking whether they thought there was a relationship between the coloured rectangles and the categories they belonged to (yes or no). If they answered yes, they were asked to describe that relationship. The next 2AFC question asked participants whether they figured out the relationship in the first part (the training stages) or the second part (the testing stage) of the study. If participants answered no to the first 2AFC question they moved on to a 10AFC question which asked them to select the rule they think best applied to the coloured rectangles and the categories they belonged to. This included two linear rules with each dimension stated independently “GREEN (DARK) rectangles were associated with one category. BLUE (LIGHT) rectangles were associated with the other”, two linear rules with both dimensions combined “DARK GREEN (LIGHT GREEN) rectangles were associated with one category. LIGHT BLUE (DARK BLUE) rectangles were associated with the other”, two similarity rules with each dimension stated independently “Rectangles of a PARTICULAR GREEN (PARTICULAR DARKNESS) were associated with one category. Rectangles of a PARTICULAR BLUE (PARTICULAR LIGHTNESS) were associated with the other”, two similarity rules with both dimensions combined “Rectangles of a PARTICULAR LIGHT GREEN (PARTICULAR DARK GREEN) were associated with one category. Rectangles with a PARTICULAR DARK BLUE (PARTICULAR LIGHT BLUE) were associated with the other”. Finally, there was another option for “NO RELATIONSHIP”.

Data Analysis

Participants were excluded from the analysis if they had substantial missing data at test or reported a relational rule containing the "unaware" stimulus dimension. If participants did not describe a rule on the open ended question, then their choice from the 10AFC question was used. 96 participants were excluded based on this criteria. Had we included all participants, then the associative predictions would be different for those who reported a relational rule regarding the aware dimension, and those who reported one for the unaware dimension. This is because asymmetrical stimulus values were used during pre-training (e.g., 5,3 & 7,9) and thus some participants would have been trained with stimulus values 3, 5, 7, and 9 of the unaware dimension, whereas others would have only been trained with stimulus values 5 and 7 of the unaware dimension.

The 104 remaining participants reported a rule containing only the aware dimension, 50 from the color condition and 54 from the brightness condition. Their data were analysed by holding one dimension constant whilst varying the stimulus values of the other. The stimulus values were then collapsed and averaged to create ‘Training’ (stimulus values 5 and 7), ‘Near’ (3 and 9), and ‘Distant’ (1 and 11). The data were also grouped by dimension: unaware and aware. Finally, two data sets were created, ‘Stimulus 6’ and ‘Without 6’. Stimulus 6 data remained constant at stimulus value 6 on one dimension, whilst varying across the other. Without 6 data were the averages for Training, Near, and Distant at stimulus values 1, 3, 5, 7, 9, and 11 on the dimension held constant. Stimulus 6 is located between the two training values and conveys no information from the dimension held constant (Without 6 does), allowing a ‘pure’ look at the processes governing responding. For example, imagine a participant is unaware of that stimuli vary from light to dark. If we hold the color dimension (which would be the dimension they are aware of) constant at stimulus 6, a color which they cannot use to aid them with categorization, then responding to stimuli varying in brightness would be influenced solely by associative processes.

The response variable for all analyses was accuracy across the dimension being varied. The data were analysed initially using a repeated measures ANOVA with dimension and position as within-subjects factors, and condition (color and brightness) and rule use (linear and similarity) as between subjects factors. Independent, repeated measures ANOVA’s, with position as a within subjects factor, were then used to analyse the data from the aware and unaware dimensions separately. Trend analyses were conducted for position. Paired t-tests were then run for Distant, Near, and Training positions for each dimension to help interpret any effects. To control for the multiple-comparisons problem in the initial four factor ANOVA, a Bonferroni correction was used. With four factors and 15 F values, the alpha level of this analysis was adjusted to $p<.003$. The alpha level for the individual ANOVAs was $p<.05$. Degrees of freedom of the ANOVA were corrected using the Huyhn-Feldt method.

For the current study, we have not analysed the icon filler stimuli. Previously, discrimination training with this type of stimuli has resulted in peak shift (Wills & Mackintosh, 1998), however only when the arrangement of icons for each stimulus is randomised on a trial-by-trial basis, which was not the case here.

Results

Training

Training data were divided into two blocks per training phase, with blocks one and two from pre-training, and blocks three and four from further training. Accuracy increased from 73% in the first block, to 88.1% in block two, then reduced to 84.4% in the first block of the further training (block three), reflecting the harder discrimination required before rising to 89.4% at the end of training. A repeated measures ANOVA showed a significant effect of block, $F(2.3,234.6)=52.124$, $p<.001$, $\eta_p^2=.147$ confirming that learning had taken place. There was no significant main effect of condition, $F(1,102)=.678$, $p=.412$, $\eta_p^2=.004$ and no interaction between condition and training block, $F(2.3,234.6)=.944$, $p=.42$, $\eta_p^2=.003$. Mean accuracy was significantly above chance, $t(103)=26.78$, $p<.001$, $\eta_p^2=.206$.

Generalization Test

Stimulus 6: Figure 2 shows the test accuracy for the stimulus 6 data. Repeated measures ANOVA revealed a significant main effect of dimension, $F(1,100)=54.535$, $p<.001$, $\eta_p^2=.353$, as well as a significant dimension × position interaction, $F(2,200)=11.972$, $p<.001$, $\eta_p^2=.107$. There was no significant effect of: position, $F(2.0,198.0)=2.914$, $p=.057$, $\eta_p^2=.006$; condition, $F(1,100)=1.502$, $p=.223$, $\eta_p^2=.015$; or rule, $F(1,100)=1.347$, $p=.248$, $\eta_p^2=.013$. There were no other significant or near significant interactions

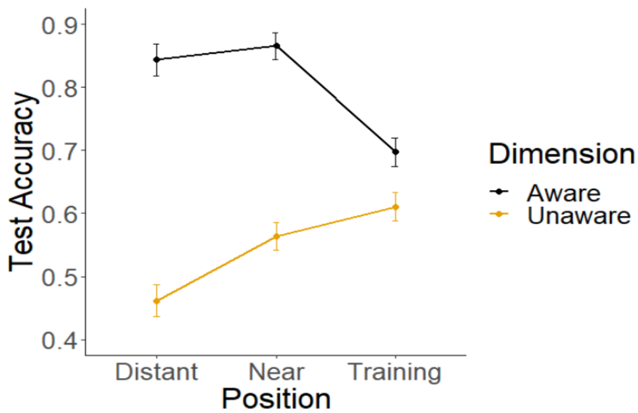

Figure 2: Test accuracy for the Stimulus 6 data.

Due to the non-significant main effect of condition and lack of any interaction between condition and dimension, participants from each condition were collapsed into one group for the individual analysis of each dimension.

For the aware dimension, there was a significant main effect of position, $F(1.7,185.8)=29.891$, $p<.001$, $\eta_p^2=.225$, as well as a significant linear $F(1,103) = 31.438$, $p<.001$, $\eta_p^2=.234$, and quadratic trend, $F(1,103) = 27.472$, $p<.001$, $\eta_p^2=.211$. Paired t-tests on Distant, Near, and Training showed a significant increase between Training (mean=.697, se=.023) and Near (mean=.865, se=.022), $t(103)=6.676$, $p<.001$, $\eta_p^2=.061$, but no significant difference between Near (mean=.865, se=.022) and Distant (mean=.844, se=.025), $t(103)=-1.135$, $p=.259$, $\eta_p^2=.011$.

For the unaware dimension, there was a significant effect of position $F(2,206) = 14.514$, $p<.001$, $\eta_p^2=.124$, and a significant linear trend $F(1,103) = 26.248$, $p<.001$, $\eta_p^2=.203$. The proportion of correct category responding was much lower in the unaware dimension and a one sample t-test revealed the average of the responses for the Training (mean=.611, se=.022), $t(103)=4.946$, $p<.001$, $\eta_p^2=.046$, and Near (mean=.563, se=.022), $t(103)=2.928$, $p=.004$, $\eta_p^2=.028$, positions significantly differed from chance, whereas the average of the Distant (mean=.462, se=.026) position did not, $t(103) = -1.508$, $p=.135$, $\eta_p^2=.014$. Paired t-tests revealed a non-significant decrease in accuracy from Training to Near, t(103)=1.656, p=.101, η_p^2=.016, and a significant decrease from Near to Distant, t(103)=3.750, p<.001, η_p^2=.035.

Without 6: Figure 3 shows the test accuracy for the without 6 data. Repeated measures ANOVA revealed a significant main effect of dimension F(1,100)=78.580, p<.001, η_p^2=.440, and position, F(1.9,185.1) = 30.835, p<.001, η_p^2=.236 as well as a significant position × dimension interaction, F(1.7,171.3) = 35.974, p<.001, η_p^2=.265. There was no significant main effect of: condition, F(1,100)=1.191, p=.278, η_p^2=.012; or rule, F(1,100)=1.038, p=.311, η_p^2=.01. There were no other significant interactions.

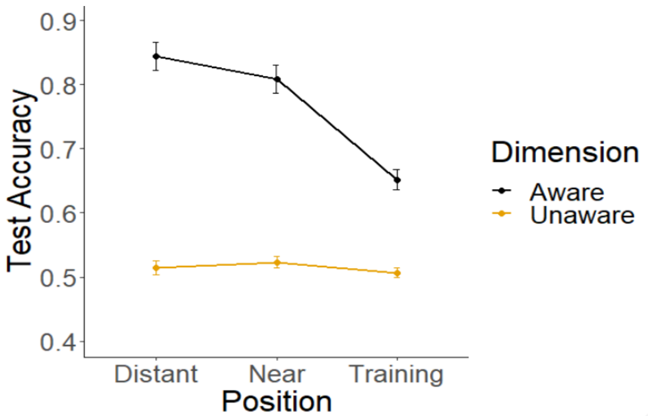

Figure 3: Test accuracy for the Without 6 data.

Given the non-significant main effect of condition and interaction between condition × dimension, participants from each condition were collapsed into one group for the individual analysis of each dimension.

For the aware dimension, there was a significant main effect of position, F(1.5,153.2)=137, p<.001, η_p^2=.571, and a significant linear F(1,103) = 166.62, p<.001, η_p^2=.618, and quadratic trend, F(1,103) = 59.069, p<.001, η_p^2=.364. A paired t-test on Distant, Near, and Training showed a significant increase between Training (mean=.651, se=.016) and Near (mean=.807, se=.021), t(103)=12.016, p<.001, η_p^2=.104, and between Near (mean=.807, se=.021) and Distant (mean=.842, se=.022), t(103)=4.306, p <.001, η_p^2=.04.

For the unaware dimension, there was no significant effect of position F(2,206) = 1.219, p=.298, η_p^2=.012. A one sample t-test revealed the average of the responses for the Near position (mean=.523, se=.009), t(103)=2.533, p=.013, η_p^2=.024, significantly differed from chance, whereas the means of the Training (mean=.507, se=.008), t(103)=.828, p=.409, η_p^2=.008, and Distant (mean=.514, se=.011), t(103)=1.248, p=.215, η_p^2=.012, positions did not. Paired t-tests revealed a non-significant numerical increase in accuracy from Training to Near t(103)=1.692, p=.094, η_p^2=.016, and non-significant decrease from Near to Distant, t(103)=.89, p=.376, η_p^2=.009 .

General Discussion

The results from the experiment were not entirely consistent with either dual or single-process theories. For the aware dimension, participants' gradient of accuracy was roughly monotonic, peaking and plateauing at the Near and Training positions for Stimulus 6, and monotonically increasing for the Without 6 data. This is consistent with previous literature (Lee et al., 2018; Livesey & McLaren, 2009). For the unaware dimension, a larger difference was seen between the Stimulus 6 and Without 6 data. For Stimulus 6 responses, accuracy monotonically decreased. For the Without 6 data, there was a numerical peak shift, with a marginally significant increase in accuracy between the Training and Near positions, and a non-significant decrease to the Distant position.

Aware dimension

Propositional Analysis: It is very likely that responding along the aware dimension was driven by rule use and, hence, propositional processes. This dimension was classified as such by looking at participants' answers to the question "describe the relationship between the [colored squares] and the categories they belonged to". Therefore, we can be certain that participants perceived the differences between the stimuli along this dimension, which would naturally lead to rule induction. Similarly, accuracy was consistently higher to this dimension compared to the unaware dimension, in line with previous research (e.g., Livesey & McLaren, 2009).

Contrary to the findings in Lee et al. (2018), we found no significant effect of the particular rule identified by the participant, or interaction between rule and dimension, indicating that participants' gradients of responding were not different depending on whether they employed a linear (n=89) or similarity rule (n=15). This may be due to the different demands of the task we employed. We used a two alternative forced choice (2AFC) categorization task, compared to Lee and colleagues who measured expectancy ratings for a hypothetical or real shock. In our 2AFC task, a similarity rule would result in participants making obviously incorrect decisions to certain stimuli. For instance, if a participant is presented with a blue that is obviously different to the blue they were trained with, it is still unlikely that they would report it as belonging to the green category; whereas a low expectancy of shock to any stimulus that is not S+ is plausible whatever side of the dimension the stimulus lies on.

Unaware Dimension

Propositional Analysis: The gradient of responding along the unaware dimension was dependent upon whether the stimulus conveyed category information from the aware dimension or not. That is, the gradient differed for the averages at the Stimulus 6 position and Without 6. However, simply looking at the accuracy of responding toward this dimension provides problems for a single-process account. If we are to assume that participants were not aware of this dimension, then the increased accuracy to the Training and Near positions for Stimulus 6, and the Near position for Without 6, are difficult to explain on a single process account that assumes learning cannot occur outside of awareness. Although it could be argued that participants were somewhat aware of this dimension and simply lacked the confidence to report this when asked, we then gave them ample opportunity to state that the stimuli varied in more than one way by providing them with a list of several pre-specified relationships. Furthermore, of the 96 participants that were excluded for stating the 'incorrect' rule, 28 reported a relational rule toward both dimensions, demonstrating that it was possible for participants to do this. Therefore, it is reasonable to conclude that they had learned about a dimension they were unaware of.

It is also possible that some of those included in the final analysis noticed both dimensions but based their decisions only, or mainly on one. This type of strategy might result in different responding to Stimulus 6 and Without 6 for the unaware dimension, as Stimulus 6 has no information from the aware dimension, and thus must be responded to solely on the basis of what has been learned about that dimension. With that in mind, participants responding to the Stimulus 6 stimuli could be argued to be consistent with a similarity rule, as accuracy was highest to the Training stimuli. However, as we have already mentioned, such a rule is, in some sense, irrational in a 2AFC categorization task, and if participants had noticed the difference between Training stimuli, it seems reasonable to suggest that they would have verbalized this difference and used it to aid in categorization. Such a rule would also, in itself, predict no differences between the Stimulus 6 and Without 6 data. In both, the Training stimulus values are still the most similar to the actual training stimuli and, thus we would have no reason to expect a difference in the gradients.

Associative Analysis: We initially suggested that associative processes would predict a peak-shifted gradient for both the Stimulus 6 and Without 6 data. This was not supported by the experimental data. However, the initial prediction assumed that the stimuli would be perceived as varying along two separable dimensions. Dimensions are separable if they are able to be perceived independently of one another, and therefore do not interact (Soto, Quintana, Pérez-Acosta, Ponce, & Vogel, 2015). However, it is not unreasonable to suggest that the two dimensions we used for our stimuli, brightness and hue, are not perceived independently and therefore are best represented as integral dimensions. Indeed, brightness and saturation are generally agreed to be integral dimensions, so it is likely that hue and brightness are also (e.g., Nosofsky, 1987; Soto et al., 2015).

To test the impact this would have on our associative predictions, we modified the connectionist network that was used to validate our initial associative predictions (that peak shift would occur to stimuli varying along the unaware dimension). This earlier network represented stimuli as varying along two separable dimensions, with 10 inputs units for each dimension. The new network represented the stimuli on an 11 × 11 matrix (representing the stimulus space). An activation function was used that peaked at the cell representing the stimulus (e.g., cell 5,5; 7,7; etc.) and varied in a Gaussian manner as a function of each neighboring cell's Euclidean distance from the peak. Figure 4 shows the networks output of the unaware dimension (black lines), compared against the experimental data from the unaware dimension (yellow lines). It appears that when an associative network is trained with stimuli that are arguably more representative of those which were used in the present experiment, the output does not significantly differ from the experimental data (all $p>.05$).

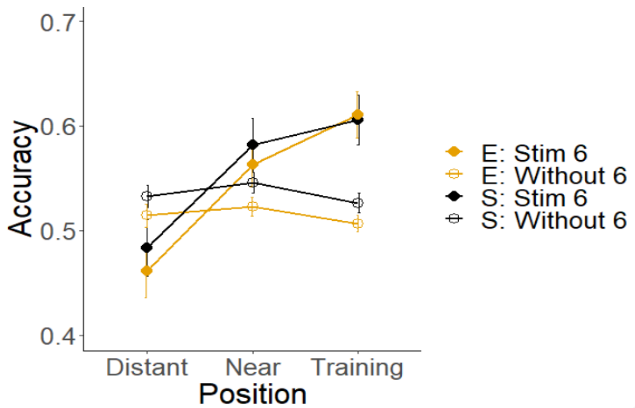

Figure 4: Graph showing simulation (S; black lines) of a connectionist network compared against experimental data (E; yellow lines). Output for the unaware dimension.

The model predicts a clear difference between the Stimulus 6 and Without 6 data. Similarly, the model predicts the monotonically decreasing accuracy for Stimulus 6 data, and numerically peak shifted accuracy for the Without 6 data. Looking more closely, the model also predicts the greater decline in accuracy from Near to Distant stimulus positions for the Stimulus 6 data, as well as the greater increase in accuracy from the Training to Near positions of the Without 6 data. That is, where the experimental data has shown a greater change in accuracy between stimulus positions, the model has also. However, it has predicted numerically higher accuracy for all stimulus positions except the Stimulus 6 Training position but note that we did not try to formally fit the model to the data. Equally, we are not suggesting that associative learning is in some way tied to integral stimulus representation in our model, dual-process accounts imply this can occur with separable dimensions as well; but we have shown that an associative network can still successfully model these results with only a slight modification to the way stimuli are represented.

We have already discussed whether a similarity rule similar to that envisaged by Lee et al. (2018) could explain responding to the Stimulus 6 stimuli and concluded that this type of rule does not seem a good fit in this type of experiment. A similarity rule might be reasonable however, if it was to involve what is, essentially, an associative process. For example, in Nosofsky's (1986) Generalized Context Model, categories are learned by tagging stored exemplars, against which new stimuli are compared using a similarity function. This rather different type of similarity "rule" could predict the function assumed by Lee et al. (2018) and would also overcome the limitation Lee et al.'s similarity rule faces with regards to rule induction in 2AFC categorization tasks. It may be that the verbal similarity rule reported by participants is their intuitive explanation of the associative processes just outlined. Of course, it could be argued that this is still a rule, as it certainly can have a propositional component. The participants make a judgement of similarity, and then use that as the basis for their decision. But that process is most easily captured by the kind of link-based computation that associative models excel at. And it also allows for the influence of the other category on this judgement, more plausibly explaining the somewhat peak-shifted gradients typically seen in people who classify themselves as employing this type of rule. It may be that, here, we have an instance of the associative in the service of the propositional that more accurately captures the true nature of cognition than either considered on their own.

To conclude, we have provided evidence for both associative and propositional processes using a within-subjects design. Participants responding was dependent upon whether they had inferred a relational rule to aid with categorization. When stimuli varied along a dimension for which they had inferred a rule, the gradient of responding was clearly indicative of propositional processes. For stimuli that varied along a dimension to which no rule had been inferred, the gradient of responding could be successfully modelled using an associative network that represented stimuli integrally, providing support for a dual-process account.

References

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

Derenne, A. (2019). Gradient Shift is Sensitive to the Width of the Generalization Test Range. *The Psychology Record, 69*, 401-414. https://doi.org/10.1007/s40732-019-00342-9

Hanson, H. M. (1957). Discrimination Training Effect on Stimulus Generalization Gradient for Spectrum Stimuli. *Science, 125,* 888-889. https://doi.org/10.1126/science.125.3253.888

Jones, F., & McLaren, I. P L. (1999). Rules and Associations. In *Proceedings of the Twenty-First Annual Conference of the Cognitive Science Society, NJ,* LEA.

Lee, J. C., Hayes, B. K., & Lovibond, P. F. (2018). Peak Shift and Rules in Human Generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*,*44*(12),1955-1970. https://doi.org/10.1037/xlm0000558

Livesey, E. J., & McLaren, I. P. L. (2009). Discrimination and Generalization Along a Simple Dimension: Peak Shift & Rule-Governed Responding. *Journal of Experimental Psychology: Animal Behavior Processes, 35*(4), 554-565. https://doi.org/10.1037/a0015524

Lovibond, P. F., Lee, J. C., & Hayes, B. K. (2019). Stimulus Discriminability and Induction as Independent Components of Generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(6), 1106-1120. https://doi.org/10.1037/xlm0000779

McLaren, I. P. L., & Mackintosh, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior, 30*(3), 177-200. https://doi.org/10.3758/BF03192828

McLaren, I.P.L**.**, Green, R.E.A. and Mackintosh, N.J. (1994). Animal learning and the implicit/explicit distinction. In N.C. Ellis (Ed.) *Implicit and explicit learning of languages*. Academic Press.

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

Mitchell, C. J, De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32,* 183-246. https://doi.org/10.1017/S0140525X09000855

Nosofsky, R. M. (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology: General, 115,* 39-57. https://doi.org/10.1037/0096-3445.115.1.39

Nosofsky, R. M. (1987). Attention and Learning Processes in the Identification and Categorization of Integral Stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87-108. https://doi.org/10.1037//0278-7393.13.1.87

Soto, F. A., Quintana, G. R., Pérez-Acosta, A. M., Ponce, F. P., & Vogel, E. H. (2015). Why are some dimensions integral? Testing two hypotheses through causal learning experiments. *Cognition, 143,* 163-177. https://doi.org/10.1016/j.cognition.2015.07.001

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23,* 645-726. https://doi.org/10.1017/S0140525X00003435

Wills, S., & Mackintosh, N. J. (1998). Peak Shift on an Artificial Dimension. *Quarterly Journal of Experimental Psychology*, *51B,* 1-32.