UNM05

The aim of this experiment was to develop a recognition memory test that was sensitive to detecting differences in the predictive validity of cues. All participants undertook the same training with the first phase of a “learned predictiveness” design (see Table 1). In this design, cues A and B are perfectly predictive of the outcomes that follow them, while cues X and Y are non-predictive. This learning phase was followed by two recognition memory tests. In both tests, two images were presented side-by-side, with one image being a cue presented in the training phase and the other image being a variation of one of the cues (hereafter, the “foil”). In Test 1, the foil had the same shape as the target but there was variation in its colours. In Test 2, the foil was a variation of one of the other cues. There were two different types of test trial in Test 2: congruent trials, in which the target and the foil were of the same predictiveness, and incongruent trials, in which the target and foil differed in their predictiveness. There were three between-subject groups, which differed in terms of the similarity that the foils had to the cues: High, Medium and Low similarity (see [Figure 1](#fig-foil_example)). The design of this can be seen in **?@tbl-expdesing**. Note that during the training stage cues A, B, X and Y were all presented equally frequently and equivalently recently to the recognition memory tests. Other things being equal, then, their memory traces should be equivalent. Thus, if recognition scores differ between these stimuli, it would indicate a difference in the level of processing that the cues underwent during training.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Design of the experiment   | Training | Test 1 | Test2 - congruent trials | Test2 - incongruent trials | | --- | --- | --- | --- | | AX-O1 | A vs *a* | A vs *b* | A vs *x* or *y* | | AY-O1 | B vs *b* | B vs *a* | B vs *x* or *y* | | BX-O2 | X vs *x* | X vs *y* | X vs *a* or *b* | | BY-O2 | Y vs *y* | Y vs *x* | Y vs *a* or *b* | |

## Methods

### Participants

90 participants were recruited through Prolific. The sample consisted of 35 women, 54 men and one non-binary person, with 24 different nationalities. The mean age was 31.5 for the 86 participants that reported their age, being (range 19 - 55. Pre-screening of participants in Prolific ensured that they had normal or corrected to normal vision, fluency in English language, and had not participated in previous studies from our lab. Participants were rewarded with £2.70 for their participation in the study. Participants were randomly allocated to each of the three groups, according to the foil-to-cue similarity during the memory tests (High, Medium and Low). Six participants were excluded, as they failed at least one of the comprehension checks before the memory tests, resulting on 27 in group High, 30 in group Medium, and 27 in group Low.

### Apparatus and stimuli

Participants were presented with a task built in PsychoPy [v. 2022.2.4, @peircePsychoPy2ExperimentsBehavior2019] and hosted in Pavlovia. The task was designed so it could only be run on a computer, but not on mobile devices. The screen background colour was grey (RGB: 128, 128, 128) and all stimuli and instructions were presented against this background. The four cues presented to each participant (A, B, X and Y) were randomly selected from a set of eight images, representing imaginary chemical compounds made of three red circles and three blue circles connected with black lines. Each cue was 945 x 945 pixels, automatically re-scaled to 0.4 x 0.4 of the window height. The outcomes (O1 and O2) were two images displaying a mutant creature, black with yellow details. Each was 332 x 664 pixels, automatically re-scaled to 0.16 x 0.2 of the window height.

Examples of the cue images are shown in [Figure 1](#fig-foil_example). The foils used in the tests were colour-modifications of the original cues: for the High similarity group, the colours of one red and one blue circle were switched (four remained unchanged); for the Medium similarity group, the colours of two red and two blue circles were switched (two remained unchanged); and for the Low similarity group, the colours of all circles were switched. All of the images used in the experiment are presented in Appendix I.

knitr::include\_graphics("foil\_example.png")

|  |
| --- |
| Figure 1: Example of the modifications in the original cue image to create the foils. |

### Design

The experiment used a mixed design, with cue-predictiveness (P and NP) manipulated within-subjects and the similarity between targets and foils manipulated between-subjects. The training phase consisted of four blocks, with each block consisting of 20 trials. As shown in [Table 1](#tbl-expdesign), there were four trial types (compound cues), with cues A and B predictive of outcome 1 and 2, respectively. Cues X and Y were paired equally often with outcomes 1 and 2 and were therefore non-predictive. Each compound cue was presented 5 times per block. The position of the cues and the outcomes (right-left), as well as the order of presentation of the trials, was fully randomized within each block.

Test 1 consisted of two presentations of each cue (8 trials in total). On each trial the target cue was presented with its corresponding foil. The left-right display of the target and the foil was counterbalanced, in such a way that each target appeared once on the left and once on the right.

Test 2 consisted of six presentations of each of the four cues (24 trials in total). Each cue was presented twice with each of the three foils shown in [Table 1](#tbl-expdesign). For example, cue A was presented with the foil corresponding to cue B, the foil corresponding to cue X, and the foil corresponding to cue Y.

### Procedure

Participants were presented with the study information and responded to a series of questions to give informed consent. If participants gave consent, they proceeded to the training phase instructions. These described the initial learning task, in which they would see two images of fictitious chemicals that would be mixed to produce a mutant creature. Participants were told that their task was to predict which mutant will result from each combination of chemicals. They were also instructed to use the feedback provided after their decision to make their future choices more accurate.

After reading the instructions, participants were presented with a comprehension check, in which the instructions were summarised, and they were asked to select the answer that best described what they had to do in the task. The experiment ended if participants failed this comprehension check twice. If participants passed the comprehension check, they proceeded to the training phase.

All the trials in the training phase started with a 0.5 second blank screen. After that, two cues were presented in the top part and the two outcomes in the bottom part. The coordinates (x/y PsychoPy height units) of the centre of the four stimuli images were: left cue, -0.3 x 0.2; right cue, 0.3 x 0.2 left outcome, -0.125 x -0.2; right outcome, 0.125 x -0.2. All stimuli were rescaled according to the height of the monitor. The participants had to select one of the outcomes by clicking on them, which was indicated by a yellow frame surrounding the selected outcome. Feedback was provided after 0.5 seconds: the correct outcome was surrounded by a green frame. If the correct outcome was selected by participants, the message “CORRECT!” was displayed in the centre of the screen in green (RGB: 0, 255, 0), otherwise the message “INCORRECT!” appeared in red (RGB: 255, 0, 0). After two seconds, the next trial started. If participants failed to select an outcome within 10 seconds of the trial starting, all images disappeared from the screen and the message “TIMEOUT - TOO SLOW” was presented in the centre of the screen in red, and the next trial started after one second. An example of a training trial can be seen in [Figure 2](#fig-trainexample).

knitr::include\_graphics("trainexample.png")

|  |
| --- |
| Figure 2: An example of a Training Trial. |

Once participants had completed the training phase, the instructions for Test 1 were displayed, telling participants that they were about to see two similar chemicals on each trial, and only one of them had appeared in the previous task. They had to select that chemical and then rate their confidence on a scale from 1 to 10 (see below). No feedback on response accuracy was provided. After the instructions, a comprehension check was included, presented only once and participants continued to the test, irrespective of the response given.

All test trials started with a 0.5 second blank screen. After that, the images of a target (a cue presented in the training phase), and a foil were presented in the top half of the screen. These images had the same size and position as in the training phase. Participants had to click on the image they thought they had seen in the previous phase, after which a rating scale was displayed for them to give a confidence rating. Above this rating scale, the question *How confident are you of your response?* was displayed. The rating scale had 10 points, with the labels *I am guessing* on the left end (point 1), and *I am certain* on the right end (point 10), and a red dot in the middle. Participants had to click on the rating scale to move the red dot to give their confidence rating. After this, a button with the word *CONTINUE* appeared. All responses in the test phase had no time limit and participants could advance to the next test trial at their own pace.

Once participants had completed Test 1, the instructions for Test 2 were displayed. They were very similar to Test 1 instructions, except it stated that the two chemicals would not be similar, but that it was still the case that only one of them had been presented in the first task. Another comprehension check was included before the start of Test 2. The test trials procedure is shown in [Figure 3](#fig-testexample).

|  |
| --- |
| Figure 3: An example of a Test Trial. |

## Results

In all statistical tests, we adopt a significance level of .05. Greenhouse–Geisser corrected degrees of freedom were used where Mauchly’s test indicated that the assumption of sphericity was violated.

Participants that failed either the comprehension check before Test 1 or Test 2 were excluded from these analyses. Six participants were excluded based on this criterion (three from group High, and three from group Low).

The dependent variable during training was the proportion of responses on a block in which participants selected the correct outcome. In [Figure 4](#fig-training) the mean accuracy for each similarity group is displayed across the four blocks of training. All groups showed a similar increase in responding across blocks, reaching an approximate accuracy of 0.85 in the final block of the experiment.

|  |
| --- |
| Figure 4: Accuracy during the training phase of Experiment. |

This was confirmed by a mixed model Analysis of Variance (ANOVA) with the within-subjects factor *block* (1-4), and the between-subjects factor *similarity* (High, Medium, and Low). This analysis found a significant main effect of the *block*, *F*(2.28, 184.81) = 54.69, *p* < .001, = .40, BF10 = 3.7x1023 ± 0.58%. There was no main effect of *similarity*, *F*(2, 81) = 0.23, *p* = .799, < .01, BF10 = 0.15 ± 3.19% nor an interaction effect, *F*(4.56, 184.81) = 1.45, *p* = .214, = .03, BF10 = 0.14 ± 9.87%. These results indicate that the training was equally effective for the three similarity groups, all of them increasing their performance as the phase progressed.

The mean accuracy for each group and type of cue (predictive vs non-predictive) on Test 1 is displayed in @fig\_acctest1. In group High similarity, the accuracy for non-predictive cues was higher than for predictive cues, whereas in groups with Medium and Low similarity, this relationship was inverted.

Coordinate system already present. Adding new coordinate system, which will  
replace the existing one.

|  |
| --- |
| Figure 5: Accuracy on Test 1. |

However, the mixed methods ANOVA did not find any significant effects, *similarity*: *F*(2, 81) = 2.03, *p* = .138, = .05, BF10 = 4.3x10-1 ± 0.69%, *predictiveness*: , BF10 = 0.18 ± 2.8% nor a *similarity x predictiveness* interaction effect, , BF10 = 0.46 ± 4.74%.

For the test data, we calculated a “memory score” for each trial by multiplying by 1 the confidence rating given by participants on the trials in which the target was selected (correct responses), and by 0 on the trials in which the foil was selected (incorrect responses). Mean memory scores for Test 1 are displayed in [Figure 6](#fig-test1). In groups Low and Medium similarity, memory score was higher for the predictive than the non-predictive cue, whereas this tendency was inverted in the High group. Also, groups Low and Medium showed higher memory score than group Low.

|  |
| --- |
| Figure 6: Memory scores during Test 1. |

These differences, however, were very small, and a mixed model ANOVA of the memory scores revealed none of the main effects or the interaction were significant: *similarity*: *F*(2, 81) = 1.29, *p* = .282, = .03, BF10 = 0.35 ± 1.05%; *predictiveness*: *F*(1, 81) = 1.17, *p* = .282, = .01, BF10 = 0.29 ± 1.02%; *similarity x predictiveness*: *F*(2, 81) = 0.85, *p* = .432, = .02, BF10 = 0.2 ± 2%. Of note, there was moderate evidence for a null-effect of predictiveness, suggesting that Test 1 failed to reveal any differences in cue-processing between predictive and non-predictive cues.

The mean accuracy for each group and type of cue (predictive vs non-predictive) on Test 2 is displayed in @fig\_acctest2. In group High similarity, the accuracy was higher for predictive than non-predictive cues, with higher accuracy for congruent (in which target and foil were either both predictive or both non-predictive) than incongruent cues (in which the target and foil differed in their predictiveness). In groups with Medium and Low similarity, accuracy was very similar for all types of cues, being generally lower for non-predictive than predictive cues, and lower for congruent than incongruent cues.

Coordinate system already present. Adding new coordinate system, which will  
replace the existing one.

|  |
| --- |
| Figure 7: Accuracy on Test 2. |

These data were analysed using a mixed ANOVA of memory scores with the factors of *predictiveness* (predictive vs. non-predictive target cue), *congruence* (target and foil same predictiveness or not), and *similarity* (High, Medium, Low), that did not find any significant effects or interactions, *similarity*: *F*(2, 81) = 0.07, *p* = .931, < .01, BF10 = 0.1 ± 0.45%; *predictiveness*: *F*(1, 81) = 0.79, *p* = .377, < .01, BF10 = 0.21 ± 2.11%; *congruence*: *F*(1, 81) = 0.40, *p* = .530, < .01, BF10 = 0.13 ± 0.78%; *similarity x predictiveness*: *F*(2, 81) = 0.24, *p* = .787, < .01, BF10 = 0.09 ± 8.07%; *predictiveness x congruence*: *F*(1, 81) = 0.07, *p* = .791, < .01, BF10 = 0.18 ± 6.69%; *similarity x congruence*: *F*(2, 81) = 1.83, *p* = .167, = .04, BF10 = 0.13 ± 4.59%; *similarity x predictiveness x congruence*: *F*(2, 81) = 0.04, *p* = .961, < .01, BF10 = 0.1 ± 6.76%.

[Figure 8](#fig-test2) shows the data from Test 2. Overall, mean memory scores were lower for the non-predictive than for the predictive cues. This difference was more pronounced in the High group than in the other groups. However, there were no apparent differences between the congruent trials and incongruent trials.

|  |
| --- |
| Figure 8: Memory scores during Test 2. |

These data were analysed using a mixed ANOVA of memory scores with the factors of *predictiveness* (predictive vs. non-predictive target cue), *congruence* (target and foil same predictiveness or not), and *similarity* (High, Medium, Low). This analysis found that the only significant differences were due to the main effect of *predictiveness*, *F*(1, 81) = 5.58, *p* = .021, = .06, BF10 = 11.56 ± 1.23%. No other significant effects were found, *similarity*: *F*(2, 81) = 0.41, *p* = .668, < .01, BF10 = 0.21 ± 3.22%; *congruence*: *F*(1, 81) = 0.29, *p* = .595, < .01, BF10 = 0.12 ± 0.85%; *similarity x predictiveness*: *F*(2, 81) = 0.97, *p* = .383, = .02, BF10 = 0.28 ± 6.39%; *predictiveness x congruence*: *F*(1, 81) = 0.42, *p* = .518, < .01, BF10 = 0.2 ± 3.45%; *similarity x congruence*: *F*(2, 81) = 2.61, *p* = .080, = .06, BF10 = 0.13 ± 7.01%; *similarity x predictiveness x congruence*: *F*(2, 81) = 0.03, *p* = .970, < .01, BF10 = 0.12 ± 17.51%. The results indicate that, in Test 2, participants showed better memory for predictive than non-predictive cues, but there were no differences due to the congruence of the test trial, nor the similarity of the foils to the training stimuli.

Despite there being no meaningful effects of similarity, and since the aim of this experiment was to determine the most sensitive way of testing participants’ memory, exploratory analyses were conducted on the data from each similarity group separately. For each, we conducted a within-subjects ANOVA with the factors *predictiveness* (predictive vs non-predictive) and *congruence* (congruent vs incongruent) for the accuracy and the memory scores. In group Low, there was no main effect of *predictiveness*, *F*(1, 26) = 0.46, *p* = .505, = .02, BF10 = 0.28 ± 1.71%, nor of *congruence*, *F*(1, 26) = 1.20, *p* = .283, = .04, BF10 = 0.23 ± 1.34% and no interaction, , BF10 = 0.27 ± 4.22%. For group Medium, there was also no effect of *predictiveness*: *F*(1, 29) = 0.74, *p* = .398, = .02, BF10 = 0.38 ± 1.88%, *congruence*: *F*(1, 29) = 2.80, *p* = .105, = .09, BF10 = 0.27 ± 0.83%, and no interaction *group x predictiveness*: , BF10 = 0.81 ± 67.27%). However, for group High there was a significant effect of *predictiveness*, *F*(1, 26) = 7.71, *p* = .010, = .23, BF10 = 72.36 ± 0.97%, but no effect of *congruence*, *F*(1, 26) = 1.87, *p* = .183, = .07, BF10 = 0.3 ± 0.89%, and no interaction, *F*(1, 26) = 0.52, *p* = .477, = .02, BF10 = 0.28 ± 7.44%. These results indicate that the only group that had a better memory for predictive stimuli over non-predictive stimuli was the High similarity group, irrespective of the congruence between foil and target in test.

## Discussion

This experiment aimed to find an effective method for examining memory cues which had been established as either predictive or non-predictive of an outcome during a previous learning task. Two different tests were used, both involving the presentation of a target (a previously trained cue) and a foil (a novel cue), which had the colours of some its elements swapped. In Test 1, the foil was the same shape as the target while in Test 2, the foil was derived from a different cue from the target. Three groups of participants received different levels of foil-to-cue similarity. Test 1 revealed no differences in memory due to predictiveness nor due to the foil-to-cue similarity. Test 2 showed better memory scores for predictive than non-predictive cues, and, although there was no significant effect of, nor any interaction with, the factor of the similarity, further analysis showed that only the High similarity group (in which the colours changed in two out of the six circles) showed an effect of cue-predictiveness on recognition memory.

To the best of our knowledge only one other study has investigated the relationship between the predictive validity of cues and recognition memory. Griffiths and Mitchell [-@griffithsSelectiveAttentionHuman2008] also used a recognition memory test as a measure of cue processing and in Experiment 4 of their study used a learned predictiveness design adapted from Le Pelley and McLaren [-@lepelleyLearnedAssociabilityAssociative2003], and similar in nature to the design of Experiment 1. However, Griffiths and Mitchell [-@griffithsSelectiveAttentionHuman2008] did not observe differences in the recognition performance for predictive and non-predictive cues. There are two main procedural differences between this experiment and Experiment 4 of Griffiths and Mitchell [-@griffithsSelectiveAttentionHuman2008]. First, their experimental task involved learning about predictive and non-predictive categories of stimuli, with each exemplar within each category being presented only once during training, whereas in our task, the same set of four stimuli was repeatedly presented during training. Second, their learned predictiveness task consisted of two phases: the first phase was similar to the training phase of the present study, but this was followed by a second phase in which all the categories were predictive of new outcomes. The memory test followed this second phase. The addition of this second phase is certainly the most likely reason for the differences in the findings: the established cue-processing bias that is established in phase 1 of the design is likely to be considerably attenuated by the second phase in which all cues were established as valid predictors of the outcomes.

Somewhat surprisingly there was no effect of foil-to-cue congruence in the data, with the predictiveness effect being observed both when the pairs of stimuli were congruent in their predictiveness (i.e., P-Con; NP-Con) and incongruent (i.e., P-Incon; NP-Incon). This is surprising, because there are seemingly different sources of memory that might contribute to these two trial types. Take for example the two trial types where the NP cue is the target stimulus. On NP-Con trials, the participant must rely on their memory of only NP cues to make a response, while on “NP-Incon” trials, the participant should be able to use their (presumably better) memory about P cues to detect the P foil. In principle, this should lead to better performance on the NP-Incon trials compared to NP-Con trials: this was clearly not the case. The critical factor in determining memory scores is whether the “P cue” was the target or the foil on the given trial: “P-Incon” trials are performed better than “NP-Incon” trials. It is possible that there is strong generalisation from the P cue to the presented P foil, and this is mistakenly identified as the target stimulus (especially so in the case of the High similarity condition). What is clear is that the only variable that affected memory scores was the predictiveness of the target stimulus, and so Test 2 represents a useful measure of recognition memory for stimuli in the task.