Supplementary Materials

Appendix A: Additional Rule Option

We also included a ‘step-function’ option designed to capture participants who derived a cut-off point and grouped all stimuli to one side as causing the outcome and all other stimuli as not causing the outcome in both experiments. This was motivated by the step-function gradients obtained in Wong and Lovibond (submitted). A very small number of participants (17/181) spontaneously reported these rules in the open-ended question. Due to a high degree of overlap between gradients obtained for the Step and Linear subgroups (see Appendix E and Appendix K) which was perhaps due to a high degree of overlap between the definitions of the rules themselves, we collapsed the Step and Linear subgroups into a single ‘Linear’ subgroup. The results in the paper are reported with these subgroups collapsed.

Appendix B: Rule Options in Experiment 1

If you had to choose, which of the following options do you think BEST describes the relationship between the circles and snack delivery? (Please read each rule carefully, and ask the experimenter if you are unclear on what any of the rules mean.)

a. As the circles became more BLUE, the likelihood of snack delivery INCREASED.

b. As the circles became more GREEN the likelihood of snack delivery INCREASED.

c. A circle of a particular intermediate colour (GREENISH-BLUE or BLUEISH-GREEN) led to snack delivery, and all other circles did not.

d. Circles that were BLUE or BLUER led to snack delivery, and circles that were GREEN or GREENER led to NO snack delivery.

e. Circles that were GREEN or GREENER led to snack delivery, and circles that were BLUE or BLUER led to NO snack delivery.

f. There was NO relationship between the circles and the likelihood of snack delivery.

 g. Other rule: (please specify)
Table 3.

Hue values used for stimuli along the dimension. Note that saturation and brightness were held constant at 100% and 75% respectively, and all stimuli were presented on a grey (RGB: 200 200 200) background. Hue values in the table are presented for counterbalancing condition 1 (green-blue). The hue values were reversed for counterbalancing condition 2 (blue-green).

<table>
<thead>
<tr>
<th>Exp</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>CS-</th>
<th>S5</th>
<th>CS+</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>S11</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.403</td>
<td>.4182</td>
<td>.4334</td>
<td>.4486</td>
<td>.479</td>
<td>.4942</td>
<td>.5094</td>
<td>.5246</td>
<td>.5398</td>
<td>.555</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.3955</td>
<td>.4122</td>
<td>.4289</td>
<td>.4456</td>
<td>.4623</td>
<td>.479</td>
<td>.4957</td>
<td>.5124</td>
<td>.5291</td>
<td>.5458</td>
<td>.5625</td>
</tr>
</tbody>
</table>

Appendix D: Rule Subgroups with small $n$ in Experiment 1

Figure 7. Generalization gradients for rule subgroups in Experiment 1 with small n. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).
Appendix E: Linear and Step Subgroups in Experiment 1

Figure 8. Generalization gradients for Linear and Step subgroups in Experiment 1. Linear and Step subgroups were combined for analysis due to a high degree of similarity between the respective gradients. Participants in the Differential group who derived an incorrect rule (inconsistent with their training contingencies) are plotted separately. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).

Appendix F: Linear Gradients by Direction in Experiment 1

Figure 9. Generalization gradients from the Linear and Step subgroups in Experiment 1, divided by direction of reported rule (greener/bluer stimuli led to the outcome). Note that the Step subgroup was combined with the Linear subgroup for the analyses. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).
Appendix G: Subgroup ns in Experiment 1

Table 4. Number of participants in each rule subgroup in Experiment 1. Linear and Step subgroups were combined due to a high degree of similarity between the generalization gradients.

<table>
<thead>
<tr>
<th>Group</th>
<th>CB</th>
<th>Lin B</th>
<th>Lin G</th>
<th>Step B</th>
<th>Step G</th>
<th>Sim</th>
<th>Other</th>
<th>No R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single</td>
<td>G-B</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>B-G</td>
<td>3</td>
<td>13</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Diff</td>
<td>G-B</td>
<td>11</td>
<td>0</td>
<td>6</td>
<td>2*</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>B-G</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

* Participants who derived a rule that was inconsistent with their training contingencies.

Appendix H: Individual Gradients in Differential-Similarity Subgroup in Experiment 1

Figure 10. Individual generalization gradients in the Differential-similarity subgroup in Experiment 1. Note that most display an area shift but no clear evidence of peak shift.
Appendix I: Full Procedure in Experiment 2

Shock workup. After attaching the shock electrodes, participants were told that the shock level would be calibrated to a level that they personally deemed to be “definitely uncomfortable, but not painful”. Participants self-administered shocks to themselves starting from the lowest level while the experimenter progressively increased the shock level. The shock level for each participant was set at the maximum level the participant could tolerate, and the experimenter made sure that the chosen level was sufficient to make the participants anxious. Participants were told to rest their hand with the electrodes attached on the desk in front of them and keep that hand as still as possible throughout the experiment. There were 4 phases in total. At the beginning of each phase, the screen froze for 30s and instructions came on screen telling the participant to wait for the experimenter. The experimenter went inside the cubicle at the beginning of each phase to deliver instructions verbally.

Phase 1 (Training – electrodes detached). Participants were told that in phases 1 and 2, they would be presented with symbols and that each symbol would either be followed by a shock or not followed by a shock. They were told that each symbol would appear for 10s, and they were to make an expectancy rating of shock during the stimulus presentation, and return the dial to the ‘Off’ position when there were no stimuli on screen. They were assured that shocks would only ever follow a symbol, and that when the screen was blank, no shocks would occur. They were also told that because of ethical reasons, the number of shocks administered to them was limited and thus, they would not be administered shocks in the first phase (first two blocks of training). Instead, they would be presented with a picture of shock (lightning bolt, accompanied by the word “SHOCK”), and were asked to imagine themselves getting shocked whenever they saw that picture, and to use the picture as feedback to enable them to predict when an actual shock would occur in phase 2. After checking for understanding of the instructions, the experimenter asked the participants to put on headphones with low-frequency white noise to help them mask external sounds, and left participants to read the written instructions (which were the same as the verbal instructions) on screen.
Appendix J: Rule Options in Experiment 2

If you had to choose, which of the following options do you think BEST describes the relationship between the circles and shock delivery? *(Please read each rule carefully, and ask the experimenter if you are unclear on what any of the rules mean.)*

a. As the circles became more BLUE, the likelihood of shock INCREASED.

b. As the circles became more GREEN the likelihood of shock INCREASED.

c. A circle of a particular intermediate colour (GREENISH-BLUE or BLUEISH-GREEN) led to shock, and all other circles did not.

d. Circles that were BLUE or BLUER led to shock delivery, and circles that were GREEN or GREENER led to NO shock.

e. Circles that were GREEN or GREENER led to shock, and circles that were BLUE or BLUER led to NO shock.

f. There was NO relationship between the circles and the likelihood of shock.

g. Other rule: (please specify)

Appendix K: Rule Subgroups with small $n$ in Experiment 2

*Figure 11. Generalization gradients in shock expectancy for rule subgroups in Experiment 2 with small $n$. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).*
Appendix L: Linear and Step Subgroups in Experiment 2

Figure 12. Generalization gradients in shock expectancy for Linear and Step subgroups in Experiment 2. Linear and Step subgroups were combined for analysis due to a high degree of similarity between the respective gradients. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).

Appendix M: Linear Gradients by Direction in Experiment 2

Figure 13. Generalization gradients from the Linear and Step subgroups in Experiment 1, divided by direction of reported rule (greener/bluer stimuli led to the outcome). Note that the Step subgroup was combined with the Linear subgroup for the analyses. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).
Appendix N: Skin Conductance Data

**Training**

Since shocks only occurred in the last block of training, only the skin conductance data for those four trials were analyzed. It appeared that both groups showed high skin conductance responses on the first CS+ trial, but that these responses suffered rapid habituation, with the Single Cue group showing a near non-significant decrease in ratings, $F(1,37) = 3.37, p = .075, \eta^2_p = .083$. In the Differential Group, skin conductance response was significantly higher to the CS+ than to the CS-, $F(1,32) = 23.33, p < .001, \eta^2_p = .422$, showing a differential conditioning effect. There was a significant CS x trial interaction, $F(3,96) = 3.41, p = .021, \eta^2_p = .096$, likely reflecting a convergence of ratings for the CS+ and CS- over the 4 trials due to the high initial skin conductance response to the first CS+ trial.

![Skin conductance data](image.png)

*Figure 14.* Skin conductance training data from Experiment 2. Shocks begin on presentation 9. There was a significant conditioning effect in the Differential group, with higher skin conductance levels to the CS+ than CS-. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).
Test

There were no significant main effects, trends, nor interactions, highest $F(1,69) = 1.123$, $p = .293$, $\eta^2_p = .016$.

Figure 15. Generalization gradients for the skin conductance data in Experiment 2. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).
Figure 16. Skin conductance gradients for each rule subgroup in Experiment 1. Note that the scale is different in e) since participants in this subgroup had very high SCL. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).
Appendix O: Individual Gradients in the Differential-Similarity Subgroup in Experiment 2

Figure 17. Individual generalization gradients in the Differential-Similarity subgroup in Experiment 2. Note that most display an area shift but no clear evidence of peak shift.
Appendix P: Analyses of Training Data by Rule Subgroup

Additional ANOVAs and trend contrasts were performed on the training data to verify whether acquisition differed between rule subgroups. We also averaged ratings for the CS+ and CS- for the final block of training (final 3 presentations) to check whether there were any subgroup differences at the end of training. Note that we only analyzed the Linear, Similarity, and No Relationship subgroups since there were only a small number of participants in the Other Rule subgroup in both experiments. For these additional analyses we report only the results of relevance to the rule subgroups.

Experiment 1

The training data for causal ratings to the CS+ in the Single Cue group were analyzed in a 3 (subgroup: Linear vs. Similarity vs. No Relationship) x (12) (training presentation) ANOVA (see Figure 18). There was no main effect of rule subgroup, $F(2,66) = 1.98, p = .146, \eta^2_p = .057$, and rule subgroup failed marginally to interact with the linear trend for presentation, $F(2,66) = 2.98, p = .057, \eta^2_p = .083$. Inspecting Figure 18, the acquisition curve for the CS+ in the Single Cue-Linear subgroup (Figure 18a) does appear slightly flatter compared to the Single Cue-Similarity subgroup (Figure 18b). In any case, there were no significant differences between rule subgroups in the Single Cue group in ratings to the CS+ in the final block of training, $F(2,66) = 1.67, p = .196, \eta^2_p = .048$.

The training data for the Differential group were analyzed in a 3 (subgroup: Linear x Similarity x No Relationship) x (2) (trial type: CS+ vs. CS-) x (12) (training presentation) ANOVA. There was no overall main effect of rule subgroup, $F < 1$, no interaction between rule subgroup and trial type, $F < 1$, and no three-way interaction between rule subgroup, trial and presentation, $F < 1$.
type, and the linear trend over presentations, \( F < 1 \). There was also no effect of subgroup on ratings to the CS+, \( F(2,43) = 1.97, p = .152, \eta_p^2 = .084 \), or CS−, \( F(2,43) = 2.26, p = .116, \eta_p^2 = .095 \), in the final block of training in the Differential group.

**Experiment 2**

The expectancy data in the training phase of Experiment 2 were analyzed in the same way as in Experiment 1 (see Figure 19). In the Single Cue group, there was a marginally non-significant overall effect of rule subgroup, \( F(2,33) = 3.20, p = .054, \eta_p^2 = .162 \). Inspection of Figure 19 suggests that this was due to the higher overall ratings given in the Similarity subgroup (b) than the other subgroups. In any case, there was no significant interaction between rule subgroup and the linear trend over CS+ presentations, \( F < 1 \), and no difference in ratings to the CS+ in the final block of training, \( F(2,33) = 2.32, p = .114, \eta_p^2 = .123 \).

![Figure 19](image)

*Figure 19.* Acquisition data (expectancy ratings) for the training phase in Experiment 2 for the a) Linear, b) Similarity, and c) No relationship subgroups. See Table 2 in the manuscript for subgroup ns. Within-subject error bars calculated according to the method proposed by Cousineau (2005) with the correction by Morey (2008).

In the Differential group, there was no overall main effect of rule subgroup, \( F(2,30) = 2.64, p = .088, \eta_p^2 = .150 \), and rule subgroup did not interact with trial type, \( F < 1 \). There was also no three-way interaction between trial type, rule subgroup, and the linear trend over presentations, \( F < 1 \). Similar to Experiment 1, there were no rule subgroup differences in ratings given to the CS+, \( F < 1 \), or CS−, \( F < 1 \), during the last block of training.