Online supplementary material: A memory-based judgment account of expectancy-liking dissociations in evaluative conditioning

Frederik Aust[1], Julia Haaf[2], & Christoph Stahl[1]

[1] University of Cologne, Germany
[2] University of Missouri, MO

Reanalysis 1

Here we report an exploratory set of analyses of our Experiment 3 with different exclusion criteria. The analyses replicate those reported in the main text but we included participants who made no more than one incorrect response in the intermittent US identification task—some participants reported accidentally clicking the wrong button—but we excluded the three participants who responded with the scale mid-point in CS pleasantness ratings to all CSs.

Participants. For Experiment 3, we recruited 273 new participants. We excluded 17 participants who performed the category recognition task at chance level, that is, participants who respond correctly to 25% or less of all category recognition questions, 25 participants who gave more than one incorrect response to the identification task, three participants who invariably responded with the scale mid-point in CS pleasantness ratings to all CSs, and one participant who aborted the experiment. Thus, we stopped collecting data after 229 valid participants. Participants mean age was 23.39 years ($SD = 6.23$), 164 were female, and 32 studied psychology or media psychology. 8 participants reported vision impairments; five were red-green color blind, one had astigmatism and another had a blind eye. 82 participants reported to have prior knowledge about the CS pictures.

On average, participants took 49.62 minutes ($SD = 12.71$) to complete the study.

Results

US expectancy. End-of-study US expectancy ratings were consistent with the intermittent ratings that we observed in Experiment 2. As in the previous experiments, we analyzed expectancies of the correct US but show a difference score between expectancy of positive and negative US in Figure 1. As predicted, we found strong evidence that the changes in expectancy of the correct US category across experimental contexts differed between acquisition and extinction learning schedules, $BF_{10} = 3.98 \times 10^{31}$. We observed

Correspondence concerning this article should be addressed to Frederik Aust, Department Psychology, Herbert-Lewin-Str. 2, 50931 Cologne, Germany. E-mail: frederik.aust@uni-koeln.de

this pattern irrespective of US valence ($BF_{01} = 451.78$) and of whether US expectancy was assessed before or after CS pleasantness, $BF_{01} = 13.58$. We, thus, analyzed all data and averaged across positively and negatively paired CS.

As predicted, expectancy for the correct US category increased from the first to the second experimental context in the acquisition learning schedule ($BF_{10} = 19.42$, one-tailed) but decreased in the extinction learning schedule, $BF_{10} = 1.64 \times 10^4$, one-tailed. We found strong evidence that participants expected USs despite the previous extinction procedure, $BF_{10} = 3.54 \times 10^8$, one-tailed. Comparisons of US expectancy between participants whom we asked to take into account both contexts versus only the second context provided only weak evidence for a difference in both acquisition ($BF_{10} = 3.78$, one-tailed) and extinction schedules, $BF_{10} = 3.47$ (one-tailed). There was no conclusive evidence to suggest that there was any other effect of our manipulations, $BF_{01} \geq 1.66$. To conclude, participants' end-of-study US expectancies corresponded to CS-US contingencies and the intermittent momentary ratings observed in Experiment 2 when we referenced and reinstated the learning contexts.

Because we found no conclusive evidence for or against integrative judgments in the preregistered between-participant comparisons of the second and the new contexts we additionally compared the differences between acquisition and extinction schedules in each context. In the first experimental context participants expressed markedly higher US expectancies in the extinction than in the acquisition schedule, $BF_{10} = 1.33 \times 10^7$ (one-sided). This pattern was reversed in the second experimental context. Participants expressed markedly higher US expectancies in the acquisition than in the extinction schedule, $BF_{10} = 9.91 \times 10^3$ (one-sided). Critically, we found some—albeit weak—evidence indicating that US expectancies across both experimental contexts did not differ between acquisition and extinction learning schedules, $BF_{01} = 4.53$. These additional analyses indicate that, like the EC effect, US expectancy appeared to be resistant to extinction when we referenced both learning contexts. Hence, we successfully elicited integrative US expectancy judgments.

CS pleasantness. We were able to replicate our findings from Experiment 2 without repeated assessment of CS pleasantness. Referring to and reinstating specific experimental contexts had the predicted effect on the EC effect dependent on the learning schedule, $BF_{10} = 2.73 \times 10^3$, Figure 1. We observed this pattern irrespective of whether CS pleasantness was assessed before or after US expectancy ($BF_{01} = 7.80$) and, thus, analyzed all data. When participants rated CS pleasantness in the new context at the end of the experiment, we found evidence for an EC effect in the extinction conditions, $BF_{10} = 11.85$ (one-tailed). Moreover, we found evidence that this EC effect in the extinction schedule was comparable to the EC effect in the acquisition schedule, $BF_{01} = 7.06$ (one-tailed). When we compared participants CS pleasantness ratings for the first and second context, we observed, both, the predicted increase in the EC effect in the acquisition, $BF_{10} = 52.24$ (one-tailed), and the predicted decrease in the extinction schedule, $BF_{10} = 20.84$ (one-tailed). In the extinction schedule, we found some evidence indicating that the EC effect was not reduced when participants rated CS pleasantness in the context of CS-alone trials compared to the new context, $BF_{01} = 6.24$ (one-tailed). Participants´ ratings in the second context provided some evidence, however, that our learning procedure did not extinguish the EC effect completely, $BF_{10} = 72.51$. In this experiment, we did find evidence indicating that the EC effect was larger in the context of CS-US pairing trails than in the new context, $BF_{10} = 79.97$. Similarly, in the acquisition

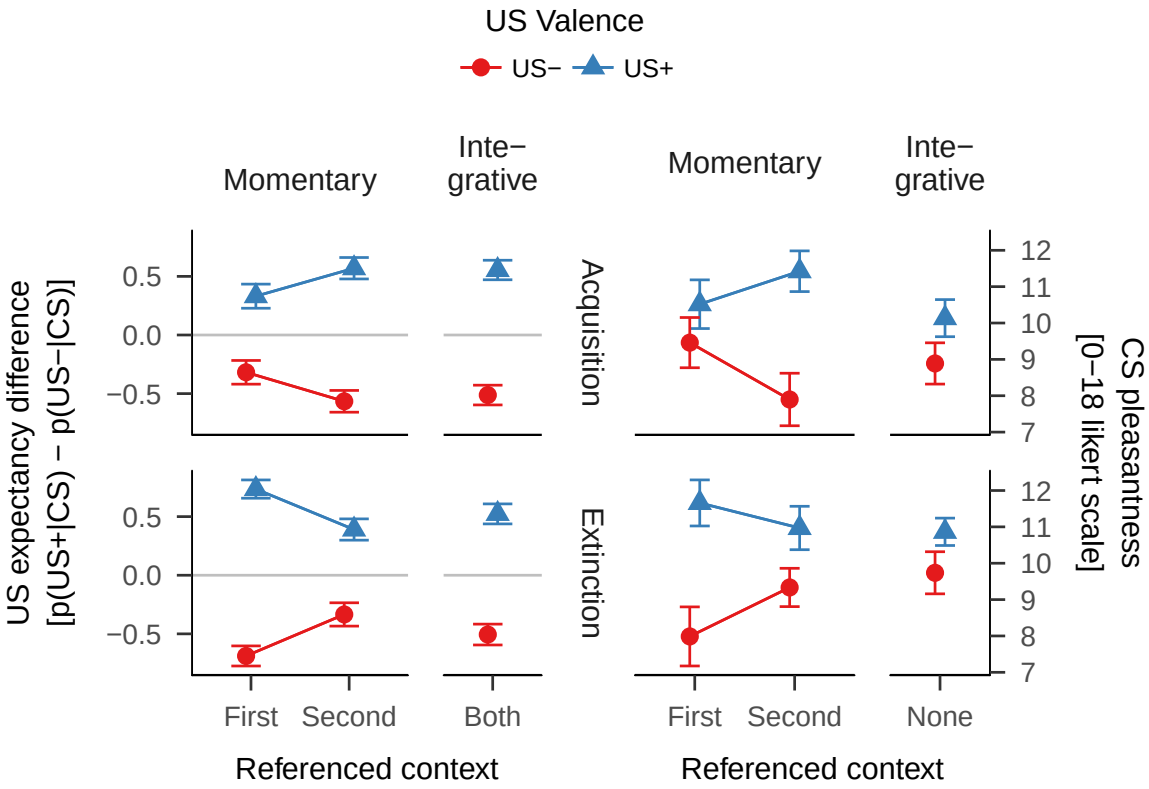

Figure 1. US expectancy and perceived CS pleasantness at the end of Experiment 3. The left plot shows observed differences in mean US expectancy for acquisition (top) and extinction (bottom) learning schedules. Expectancy for positive and negative USs is indicated by positive and negative values, respectively. The right plot shows observed mean CS pleasantness ratings for each learning schedule. Error bars represent 95% within-subject confidence intervals.

schedule, the EC effect was larger in the second context, in which CS were paired with US, than in the new context, $BF_{10} = 19.29$. The comparison between the EC effect in the first context (in which CSs were presented alone) and the new context, however, was inconclusive, $BF_{01} = 1.70$ (one-tailed). We found no noteworthy evidence for any other effects of our manipulations, $BF_{10} \leq 1.06$. In sum, we found comparable EC effects in the acquisition and extinction procedures when participants rated CS pleasantness in a new context at the end of the learning procedure. We did, however, also observe the predicted extinction effects on nondefault momentary CS pleasantness judgments: the EC effect was larger in the context of CS-US pairing trials than in the context of CS-alone trials.

In the first experimental context, participants exhibited a larger EC effect in the extinction than in the acquisition schedule, $BF_{10} = 84.68$ (one-sided). In the second experimental context, this pattern reversed: Participants exhibited a larger EC effect in the acquisition than in the extinction schedule, $BF_{10} = 10.54$ (one-sided). In this between-participant design, the data were uninformative as to whether participants' prior knowledge about CSs affected these findings, $BF_{01} = 1.31$.

CS-US pairing memory. Again, US category recognition was quite accurate. As predicted, we found that US category memory varied with referenced context and learning schedule, $BF_{10} = 5.16 \times 10^{12}$. Unlike in Experiment 2, we found evidence indicating that the recognition advantage for US absence was dependent on the learning schedule, $BF_{10} = 348.86$. Participants best remembered that a US was absent in the acquisition learning schedule ($M = .87$, SD = .28); however, memory for US absence in the extinction learning schedule ($M = .77$, SD = .35) was comparable in magnitude to the memory for the correct category when a CS had been paired with a US ($M = .76$, SD = .32, and $M = .73$, SD = .33 for acquisition and extinction, respectively). We found no evidence on whether collecting US expectancy or CS pleasantness first affected memory for CS-US pairings, $BF_{01} = 1.26$; we no noteworthy evidence for any other effects of our experimental manipulations, all $BF_{01} \geq 1.26$.

US identity recognition, too, was quite accurate in both acquisition ($M = .85$, SD = .29) and extinction learning schedule ($M = .84$, SD = .29). We found some evidence suggesting that memory for negative USs ($M = .86$, SD = .27) was better than for positive USs ($M = .81$, SD = .31) but there was no conclusive evidence indicating that any other experimental manipulation affected US identity recognition, all $BF_{01} \leq 0.85$.

As in the previous experiments, end-of-study pleasantness ratings of US categories indicated that participants remembered the valence of the US categories, $BF_{10} = 1.63 \times 10^{165}$. Without any exemplars available, participants rated the animal category as more pleasant than the object category, $BF_{10} = 1.40 \times 10^{35}$, and human category as less pleasant than object category, $BF_{10} = 2.10 \times 10^{88}$. Thus, recognition memory for US categories may be indicative or participants US valence memory.

As in the previous experiments, memory for CS-US pairings was too accurate to test whether the observed changes in CS pleasantness across contexts in the two learning schedules was contingent on memory for CS-US pairs.

Reanalysis 2

Here we report a second exploratory set of analyses of our Experiment 3 with different exclusion criteria. The analyses replicate those reported in the main text but we included participants who respond correctly to 25% or less of all category recognition questions.

Participants. For Experiment 3, we recruited 273 new participants. We excluded 57 participants who gave more than one incorrect response to the identification task and one participant who aborted the experiment. Thus, we stopped collecting data after 215 valid participants. Participants mean age was 23.72 years ($SD = 6.56$), 154 were female, and 32 studied psychology or media psychology. 7 participants reported vision impairments; five were red-green color blind, one had astigmatism and another had a blind eye. 79 participants reported to have prior knowledge about the CS pictures.

On average, participants took 49.62 minutes ($SD = 12.71$) to complete the study.

Results

US expectancy. End-of-study US expectancy ratings were consistent with the intermittent ratings that we observed in Experiment 2. As in the previous experiments, we analyzed expectancies of the correct US but show a difference score between expectancy of positive and negative US in Figure **??**. As predicted, we found strong evidence that

the changes in expectancy of the correct US category across experimental contexts differed between acquisition and extinction learning schedules, $BF_{10} = 2.03 \times 10^{31}$. We observed this pattern irrespective of US valence ($BF_{01} = 230.24$) and of whether US expectancy was assessed before or after CS pleasantness, $BF_{01} = 32.85$. We, thus, analyzed all data and averaged across positively and negatively paired CS.

As predicted, expectancy for the correct US category increased from the first to the second experimental context in the acquisition learning schedule ($BF_{10} = 10.40$, one-tailed) but decreased in the extinction learning schedule, $BF_{10} = 3.89 \times 10^{3}$, one-tailed. We found strong evidence that participants expected USs despite the previous extinction procedure, $BF_{10} = 3.49 \times 10^{5}$, one-tailed. Comparisons of US expectancy between participants whom we asked to take into account both contexts versus only the second context provided only weak evidence for a difference in both acquisition ($BF_{10} = 1.79$, one-tailed) and extinction schedules, $BF_{10} = 1.65$ (one-tailed). There was no conclusive evidence to suggest that there was any other effect of our manipulations, $BF_{01} \geq 1.61$. To conclude, participants' end-of-study US expectancies corresponded to CS-US contingencies and the intermittent momentary ratings observed in Experiment 2 when we referenced and reinstated the learning contexts.

Because we found no conclusive evidence for or against integrative judgments in the preregistered between-participant comparisons of the second and the new contexts we additionally compared the differences between acquisition and extinction schedules in each context. In the first experimental context participants expressed markedly higher US expectancies in the extinction than in the acquisition schedule, $BF_{10} = 8.68 \times 10^{6}$ (one-sided). This pattern was reversed in the second experimental context. Participants expressed markedly higher US expectancies in the acquisition than in the extinction schedule, $BF_{10} = 4.72 \times 10^{3}$ (one-sided). Critically, we found some—albeit weak—evidence indicating that US expectancies across both experimental contexts did not differ between acquisition and extinction learning schedules, $BF_{01} = 5.35$. These additional analyses indicate that, like the EC effect, US expectancy appeared to be resistant to extinction when we referenced both learning contexts. Hence, we successfully elicited integrative US expectancy judgments.

CS pleasantness. We were able to replicate our findings from Experiment 2 without repeated assessment of CS pleasantness. Referring to and reinstating specific experimental contexts had the predicted effect on the EC effect dependent on the learning schedule, $BF_{10} = 587.34$, Figure **??**. We observed this pattern irrespective of whether CS pleasantness was assessed before or after US expectancy ($BF_{01} = 8.19$) and, thus, analyzed all data. When participants rated CS pleasantness in the new context at the end of the experiment, we found evidence for an EC effect in the extinction conditions, $BF_{10} = 11.60$ (one-tailed). Moreover, we found evidence that this EC effect in the extinction schedule was comparable to the EC effect in the acquisition schedule, $BF_{01} = 7.03$ (one-tailed). When we compared participants CS pleasantness ratings for the first and second context, we observed, both, the predicted increase in the EC effect in the acquisition, $BF_{10} = 33.83$ (one-tailed), and the predicted decrease in the extinction schedule, $BF_{10} = 11.87$ (one-tailed). In the extinction schedule, we found some evidence indicating that the EC effect was not reduced when participants rated CS pleasantness in the context of CS-alone trials compared to the new context, $BF_{01} = 3.75$ (one-tailed). Participants´ ratings in the second context provided some evidence, however, that our learning procedure did not extinguish the EC effect completely, $BF_{10} = 7.13$. In

this experiment, we did find evidence indicating that the EC effect was larger in the context of CS-US pairing trails than in the new context, $BF_{10} = 10.48$. Similarly, in the acquisition schedule, the EC effect was larger in the second context, in which CS were paired with US, than in the new context, $BF_{10} = 15.56$. The comparison between the EC effect in the first context (in which CSs were presented alone) and the new context, however, was inconclusive, $BF_{01} = 1.80$ (one-tailed). We found no noteworthy evidence for any other effects of our manipulations, $BF_{10} \leq 2.35$. In sum, we found comparable EC effects in the acquisition and extinction procedures when participants rated CS pleasantness in a new context at the end of the learning procedure. We did, however, also observe the predicted extinction effects on nondefault momentary CS pleasantness judgments: the EC effect was larger in the context of CS-US pairing trials than in the context of CS-alone trials.

In the first experimental context, participants exhibited a larger EC effect in the extinction than in the acquisition schedule, $BF_{10} = 9.62$ (one-sided). In the second experimental context, this pattern reversed: Participants exhibited a larger EC effect in the acquisition than in the extinction schedule, $BF_{10} = 45.05$ (one-sided). In this between-participant design, the data were uninformative as to whether participants' prior knowledge about CSs affected these findings, $BF_{10} = 1.69$.

CS-US pairing memory. Again, US category recognition was quite accurate. As predicted, we found that US category memory varied with referenced context and learning schedule, $BF_{10} = 1.34 \times 10^9$. Unlike in Experiment 2, we found evidence indicating that the recognition advantage for US absence was dependent on the learning schedule, $BF_{10} = 247.78$. Participants best remembered that a US was absent in the acquisition learning schedule ($M = .84$, SD = .32); however, memory for US absence in the extinction learning schedule ($M = .74$, SD = .37) was comparable in magnitude to the memory for the correct category when a CS had been paired with a US ($M = .73$, SD = .33, and $M = .71$, SD = .34 for acquisition and extinction, respectively). We found no evidence on whether collecting US expectancy or CS pleasantness first affected memory for CS-US pairings, $BF_{01} = 11.34$; we no noteworthy evidence for any other effects of our experimental manipulations, all $BF_{01} \geq 5.75$.

US identity recognition, too, was quite accurate in both acquisition ($M = .81$, SD = .32) and extinction learning schedule ($M = .81$, SD = .31). We found some evidence suggesting that memory for negative USs ($M = .83$, SD = .30) was better than for positive USs ($M = .79$, SD = .32) but there was no conclusive evidence indicating that any other experimental manipulation affected US identity recognition, all $BF_{01} \leq 1.33$.

As in the previous experiments, end-of-study pleasantness ratings of US categories indicated that participants remembered the valence of the US categories, $BF_{10} = 6.08 \times 10^{160}$. Without any exemplars available, participants rated the animal category as more pleasant than the object category, $BF_{10} = 2.35 \times 10^{38}$, and human category as less pleasant than object category, $BF_{10} = 2.00 \times 10^{83}$. Thus, recognition memory for US categories is indicative or participants US valence memory.

As in the previous experiments, memory for CS-US pairings was too accurate to test whether the observed changes in CS pleasantness across contexts in the two learning schedules was contingent on memory for CS-US pairs.

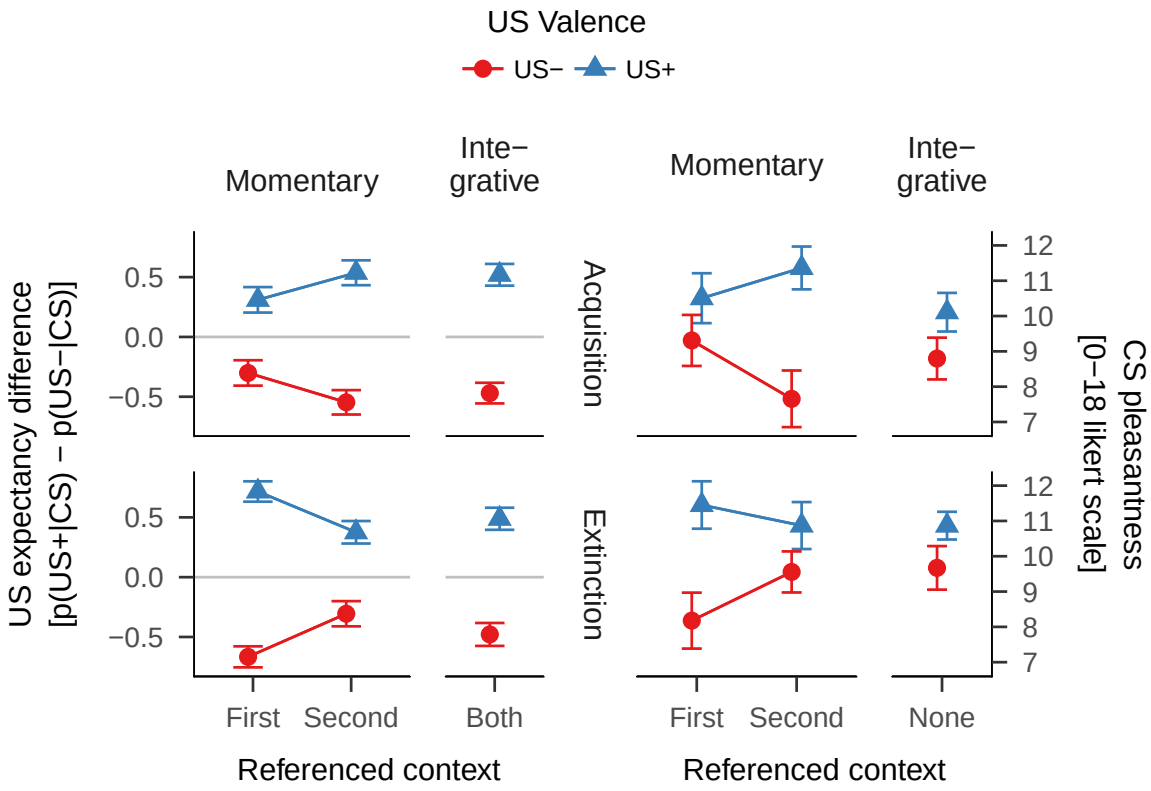

Figure 2. US expectancy and perceived CS pleasantness at the end of Experiment 3. The left plot shows observed differences in mean US expectancy for acquisition (top) and extinction (bottom) learning schedules. Expectancy for positive and negative USs is indicated by positive and negative values, respectively. The right plot shows observed mean CS pleasantness ratings for each learning schedule. Error bars represent 95% within-subject confidence intervals.