Confidence in approach-avoidance conflict

Oleg Solopchuk* **(o.solopchuk@uke.de)**
Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf
20246 Hamburg, Germany
Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics
72076 Tübingen, Germany

Leonard Asan* **(l.asan@uke.de)**
Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf
20246 Hamburg, Germany

Jan Gläscher (glaescher@uke.de)
Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf
20246 Hamburg, Germany

Christian Büchel (buechel@uke.de)
Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf
20246 Hamburg, Germany

Peter Dayan (dayan@tue.mpg.de)
Department of Computational Neuroscience, Max Planck Institute for Biological Cybernetics
72076 Tübingen, Germany
Department of Computer Science, University of Tübingen
72074 Tübingen, Germany

*Equal contribution

Abstract

Choices that balance several attributes involve mixed feelings. However, how this ambivalence is reflected in confidence judgments is unclear. We tested how people judge separate confidence dimensions corresponding to different choice attributes in the face of a conflict between approach and avoidance. We found confidence judgments to be partially dissociated, with the information about aversive consequences leaking into the confidence of the appetitive dimension of a choice. We also found that confidence decreases the perception of pain and increases momentary happiness, suggesting the need to refine accounts of affective judgments.

Keywords: metacognition; valence; pain; happiness

Introduction

Our decisions often involve multiple attributes that compete. However, how separately accessible these attributes are to metacognitive evaluation is not clear. For example, the exploration-exploitation dilemma involves a trade-off between short- and long-term reward, but despite a strong tendency to explore, people seem unable to express confidence judgments that dissociate these different timescales (Solopchuk & Dayan, 2025).

Here, we tested the multifacetedness of metacognition in a more explicit paradigm, in which each choice was associated with a monetary reward and a painful heat stimulus. We examined if people were generally able to dissociate their confidence judgments concerning these choice attributes, as well as whether information regarding one choice attribute systematically leaked into judgments about the other. Given the recent finding that metacognitive and affective judgments are strongly correlated in a perceptual discrimination task (Voodla, Uusberg, & Desender, 2025), we also tested how confidence about each task attribute affects the perception of pain and momentary happiness.

Methods

Forty eight participants performed a two-armed bandit task in which each bandit delivered both (thermal) pain and reward points ranging from 0 to 100 (Figure 1). Pain levels were calibrated to account for individual pain sensitivity, while reward points were converted to a bonus at the end of the experiment. All amounts were sampled from Gaussian distributions with a fixed standard deviation of 8 levels/points. The mean values for pain and reward were either low (30), medium (50), or high (70). We tested all combinations of pain and reward means as well as bandit sides, resulting in 81 unique offers that were randomized, and presented once each. We analyzed the sensitivity of choices to the difference of observed reward and pain averages with a logistic regression predicting the probability of choosing the left bandit. We used linear regression to analyze the sensitivity of reward and pain confidence judgments, as well as the pain intensity and happiness ratings to other task factors. We performed t-tests on the coefficients, with degrees of freedom = 47 in all tests.

Results

We found that people's choices were sensitive to the differences in average reward points (t=11.69, p<1e-3) and pain levels (t=-6.63, p<1e-3, Figure 2, top left), although reward had a much stronger influence on choices than pain (t-test on the difference of the regression coefficients, t=11.69, p<1e-3). Participants' confidence in their choice being better in terms of reward varied with the difference in average reward points between the chosen and the other bandit (factor 'reward difference', t=15.59, p<1e-3, Figure 2, top right) and total reward across both bandits (t=9.92, p<1e-3), in agreement with previous findings (De Martino, Fleming, Garrett, & Dolan, 2013). Similarly, confidence regarding pain varied with the difference in average pain levels between the chosen and the other bandit (factor 'pain difference', t=-11.77, p<1e-3) and the total pain (t=-6.30, p<1e-3). Unexpectedly, we found that 'reward' confidence increased with pain difference (t=3.59, p<1e-3) and decreased with pain sum (t=-5.26, <1e-3), even after controlling for pain confidence in the same regression. There

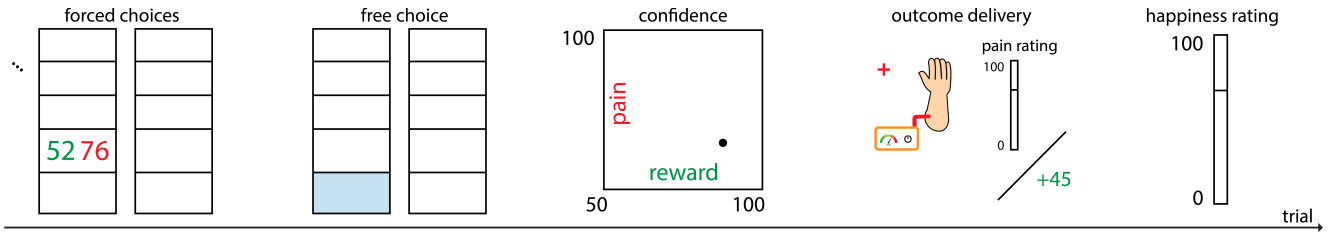

Figure 1: Task description. In each trial, the first four choices were forced, providing participants with 2 samples from each of the two bandits. Upon choosing the highlighted bandit, the pain level and reward points drawn from the means associated with the chosen bandit were displayed for one second (as shown). The fifth choice was free, and was followed by a two-dimensional confidence rating regarding the correctness of the choice in terms of maximizing rewards and minimizing pain. The assignment of pain and reward to the grid axes was counterbalanced across participants. Following the confidence rating, participants received the noxious heat stimulus and reward points sampled from the means of the chosen bandit. Pain intensity ratings were collected immediately after heat application, and each trial concluded with participants rating their happiness on a 0–100 scale.

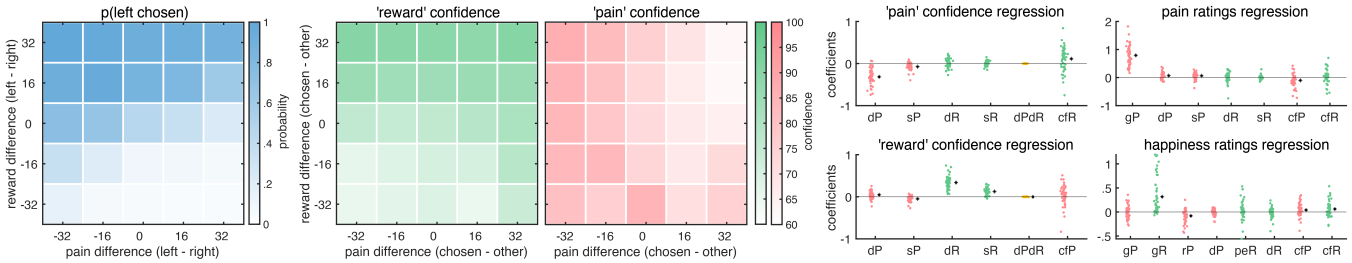

Figure 2: Left: probability of choosing the left bandit as a function of binned reward and pain difference between the left- and the righthand bandit. Middle: confidence about the choice being correct in terms of reward/pain as a function of reward and pain difference between the chosen and the other option. Right: regression coefficients predicting 'pain' and 'reward' confidence, pain ratings, and happiness ratings. Abbreviations: dP/sP - difference/sum of pain, dR/sR - difference/sum of reward, cfR/cfP - confidence in having made a correct decision in terms of maximizing reward/minimizing pain, gP - delivered pain, gR - obtained reward points, rP - pain rating, peR - reward prediction error. Predictors significant on the group level are marked with asterisks.

was also a small but consistent interaction effect of pain and reward difference on 'reward' confidence (average coefficient = -0.002, t=-3.18, p=0.003), suggesting that higher pain difference weakens the impact of reward difference on 'reward' confidence. There was no effect of either reward difference or total reward on 'pain' confidence. Conversely, 'reward' confidence was a significant predictor when included in the 'pain' confidence regression (t=2.70, p=0.01) but not the other way around.

We found that the strongest predictor of the pain rating was the true delivered pain intensity (t=15.12, p<1e-3). Pain perception also depended on the pain difference (t=4.18, p<1e-3) and the pain sum (t=3.86, p<1e-3), both increasing the perceived pain. Finally, pain was also perceived as less intense if participants were more confident that they made a better choice in terms of pain (t=-2.98, p=0.005). We found no effect of reward difference, reward sum, or reward confidence in predicting pain intensity judgments.

Finally, happiness ratings were most strongly and positively correlated with the number of earned reward points on the current trial (t=5.78, p<1e-3). Happiness also decreased when pain was perceived as more intense (t=-4.1, p<1e-3), and increased both with choice confidence regarding reward (t=2.72, p=0.009) and choice confidence regarding pain (t=2.08, p=0.043).

Discussion

We found that people's choices were accompanied by a multidimensional sense of confidence - confidence about each attribute depended most strongly on its objective determinants, namely the differences and sums of the average bandit outcomes. Surprisingly, information about pain leaked into the confidence about reward in dissociable ways - higher pain difference between the chosen and the other option increased 'reward' confidence, while the sum of average pain points across bandits decreased 'reward' confidence. The former suggests either a 'no pain no gain' rationalization bias, by which people increase their confidence in reward when more pain is chosen, or that people tend to choose the more painful option when they are more confident that the choice is also better in terms of reward. The latter points towards a broader inter-domain generalization of aversive contextual confidence modulation. Future work should test (e.g. delayed) non-painful monetary aversive consequences, as well as relating metacognitive bias, sensitivity, and efficiency pertaining to different choice attributes with pain and happiness ratings.

We also found that pain intensity ratings, apart from the objective intensity, depended on the same objective parameters as the 'pain' confidence, as well as on the 'pain' confidence itself. An analogous stronger dependence of happiness ratings on the confidence about reward hints towards a more complex relationship between affective ratings than simple correlations, and calls for further refinements of models of affective states (Rutledge, Skandali, Dayan, & Dolan, 2014).

Acknowledgments

We are grateful to Mira Rufeger for the help in collecting the data. This work was funded by the Max Planck Society, the Alexander von Humboldt Foundation, DFG (German Research Foundation) CRC 1528 Cognition of Interaction, DFG 5389 Dynamic Belief Updating, and DFG Projektnummer 449640848 grants. PD is a member of the Machine Learning Cluster of Excellence, EXC number 2064/1–Project number 39072764 and of the Else Kröner Medical Scientist Kolleg "ClinbrAIn: Artificial Intelligence for Clinical Brain Research".

References

De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110.

Rutledge, R. B., Skandali, N., Dayan, P., & Dolan, R. J. (2014). A computational and neural model of momentary subjective well-being. *Proceedings of the National Academy of Sciences*, *111*(33), 12252–12257.

Solopchuk, O., & Dayan, P. (2025). Multifaceted confidence in exploratory choice. *PLOS ONE*, *20*(1), e0304923.

Voodla, A., Uusberg, A., & Desender, K. (2025). Metacognitive confidence and affect – two sides of the same coin? *Cognition and Emotion*, *0*(0), 1–18.