

Observational learning strategies impact the neural correlates of declarative memory formation

Corresponding Author: Dr Avi Mendelsohn

This manuscript has been previously submitted at another journal. This document only contains information relating to versions considered at Communications Biology.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors present an interesting experiment that combines a simple reinforcement learning task interleaved with an incidental memory task that involves own and observed decisions with different reward contingencies. The design is quite complex as there are multiple within- and between-participant factors involved that are defined based on the computational modeling results. Overall, the task has some potential to unravel a detailed relationship between RL and memory formation, but there are several flaws in the modeling and analysis results that prevent a publication at this time. However, most of the criticism can probably be addressed by a thorough revision of the paper and the underlying analyses.

Computational modeling

The authors fitted two simple RL models (with either a single or separate learning rates for exp. and obs. trials) to the choice task and used maximum likelihood estimation (MLE) as implemented in the “fmin” functions in MATLAB. While the table in the Supplement is lacking the information about the estimated parameters for the single-LR model, the data presented there already reveals a common problem when using MLE with the fmin-function in MATLAB: learning rates and softmax temperatures are often estimated at (or very close) to the range boundaries of the parameters indicating problems in the optimization procedures. Learning rates close to zeros would suggest that these participants did not learn at all, whereas temperatures near zeros would indicate that participants were only making random choices inconsistent with learning. These are effects due to a flawed estimation procedure, not a problem with the data itself. In the data presented in the supplement only about half of the participants had temperatures below 95 or learning rates above 0.02.

These problems could have been detected in a posterior predictive check, an essential component of model validation in a computational modeling study (see e.g. Wilson & Collins, E-life, 2019, <https://elifesciences.org/articles/49547>), which the authors did not conduct. Essentially, this involves simulating new data with the fitted parameters and comparing them to the experimental data on a model-free performance indices (e.g. accuracy). One solution to this problem is a hierarchical estimation procedure (as implemented in MCMC sampling technique in Stan) or in the CBM toolbox for MATLAB (paper: <https://doi.org/10.1371/journal.pcbi.1007043>, manual: <https://payampiray.github.io/cbm>), which uses variational techniques for parameter estimation. This estimation procedure regulates individual parameter estimates, thus avoid the typical “at-boundary estimates” of fmincon. Interestingly, the toolbox will also provide “model responsibilities” that indicate with model is the best for each participant, thus avoiding the individual model comparison via BIC and protected exceedance probabilities that indicate which model is best in explaining the data in the entire sample overall. Because the modeling results are essential for the paper, all subsequent analyses are likely to be biased, if not flawed and should be taken with skepticism at this point.

Definition of participant groups

The definition of the between-participant factor Group (social learning vs. social ignoring) was made based on the model comparison between the single-LR vs. double-LR model. Because the model estimation in its current state is flawed, the analytical decision about the group should not be based on it. In fact, I would find it much more convincing, if the participant

groups were defined on the (model-free) performance in congruent and incongruent trials (e.g. split at 50% in Fig. 2 A) and then fit the models to these groups of participants. Overall, it is quite striking that a substantial number of participants cannot ignore the incongruent contingencies in observed trials, when making their own choices in incongruent trials.

Model-free behavioral analyses

There is one puzzling results in the behavioral data (Figs 2 and 3) that remains unmentioned and unexplained: the social learning group commits much more errors on own choices in incongruent trials (Fig 2D), but they have a much higher memory confidence in observed, incongruent trials. Why is that and what does that say about the interaction between value-based decision-making and declarative memory? I would welcome an interpretation and some conceptual thoughts about this finding.

fMRI analyses

The contrast of experience vs observed trials (thresholded leniently at $p < 0.001$ uncorr.), which serves as a functional localizer is not completely orthogonal to condition-specific ROI analysis. This ROI analysis pools over all voxels in the ROI thus avoiding the multiple comparisons problem. A “normal” mass-univariate fMRI analysis would have not yielded any significant findings in these analyses. However, the functional localizer essentially serves as a mask, which precludes the analysis of any results that show the opposite effect (exp. > obs. trials). Thus, we don’t really get to see, which brain regions are activated, when own decisions are made (compared to observe decisions).

Minor issues and typos

page 6, bottom: at this point, it is not clear that the participants did not know about the subsequent memory test. Please include this information here

page 7, Recognition Memory Test: was there an equal number of old and new items?

page 9: description of the fMRI preprocessing is very sparse.

Figure 1: please use different choice cues for the obs trials structure in panel A

Figure 2 and 3: The different shades of red are virtually indiscriminable, please choose a different color scheme for the 4 conditions.

Legend of Fig 2: (C-D) Please stick with the group labels “social learning” and “social ignoring”

Figure 4: Some of the bars for “social learning exp” in run 2 have the wrong shade of blue.

page 22, bottom: ... tendency to follow the choices of others even of the outcome ... “of” should be change to “if”

page 22, bottom: “Generalized” should be changed to “General”

page 23: “metallization” should be changed to “mentalization”

page 23: “This may indicate that the mental representation of others’ decision process was taken into account, at least to some extent, not only when observing others decide, but also during one’s own decision-making process.” This reverse inference (attributing cognitive function only because a certain brain area is activated) and should be removed.

Reviewer #2

(Remarks to the Author)

This study examined behavior and brain activation associated with different reinforcement learning styles. More specifically, the authors tested the hypothesis that subjects would differ regarding how much they are inclined to incorporate information from ‘socially’ observed decision making trials in addition to self-performed trials. I think this is a timely and interesting research question, and the experimental design is suitable (with one exception, see specific comments below). Behaviorally, this study partly replicated previously published results from the same authors suggesting that a ‘social ignoring’ subgroup primarily uses information gained from self-performed (or ‘experienced’) trials whereas another ‘social learning’ group uses information gained from both trial types. The brain imaging results, however, are not convincing in its current form. The main reason is that large parts of the statistical analysis remain unclear and insufficient thereby compromising the reliability of the reported results.

Major points:

In general, I was quite disappointed by the lack of detail regarding many aspects of the fMRI analysis. This includes pretty much all aspects of the analysis: data acquisition, preprocessing (see ‘other comments’), modeling, and statistical evaluation.

1. To start with, I could not find information as to whether the model regressors were time-locked to the stimulus pair, the outcome, or the response? This of course is a critical piece of information when it comes to reinforcement learning where it is

important to know if we are looking at stimulus-related effects (evolution of Q values) or outcome-related effects (evolution of prediction error).

2. The statistical analysis comprised two steps. First a whole-brain contrast was computed to identify voxels that were more strongly activated for observed trials than for experienced trials. Second, based on this contrast 4 ROIs were identified, which were then further scrutinized regarding more specific effects involving group, congruency, and run (as a proxy for learning progress).

As far as I can tell, the whole brain contrast was not corrected for multiple comparisons. Instead, an uncorrected threshold of $p < .001$ was applied. This is certainly not acceptable. Given the rather large sample size ($N = 41$), it should be possible to reveal a reliable effect based on a proper correction method (voxel level or cluster level FWE-corrected). Hence, the basis of the subsequent ROI analysis is already flawed. Also, a complete report of the whole brain analysis results is missing (e.g. a table with all clusters and statistics). Finally, no information whatsoever is given regarding the reverse contrast (i.e., stronger activity for experienced vs. observed trials)

For the ROI analysis, again a complete report of the results is missing. Information regarding the main effects of group, congruency, and run is missing. Also, and even more importantly, no information regarding any of the interaction effects is presented. Of particular importance here is the interaction group $\times$ run. If this interaction is not significant the selective report of group differences in each run is not valid (a difference in significance is not a significant difference). The main point the authors want to make is that there is an increasingly stronger activation from run 1 to run 2 specifically in the social learning group in observed trials. However, if there is no significant interaction between group and run in observed trials (actually, between group, run, and trial type), this claim cannot be supported statistically.

3. Detailed information on how the ROI analyses were conducted is missing. Specifically, which voxels were included in each ROI at which threshold? Mean across voxel values or only peak voxel?

4. I was surprised that the authors did not use estimated Q-values (and estimated prediction error for that matter) as trial-wise parametric regressors during the fMRI analysis? Wouldn't this be a much more powerful way of tracking learning-related changes in brain activation than comparing run 1 and run 2 as done in the current paper version?

5. Although I generally liked the experimental design, I am worried that the timing differences between observed and experienced trials (choice time 2 sec vs. < 4 sec) precludes a clean comparison of the two conditions in terms of brain activity. I might be an option to include choice time as an addition covariate.

Other comments

1. Is the equation for computing the reward prediction error correct (page 7)?
 $1 - R - Q$ instead of $1 - (R - Q)$?

2. The social learning group includes two sub-types
a) Assigning same weight to observed trials
b) Assigning more weight to observed than experienced trials

How many subjects are in (a) vs. (b). Are the results affected by including only (a) subjects (the group the was exclusively identified in the previous publication)?

3. Imaging Methods; data acquisition: functional data were collected with $TR=1000$ and 64 slices. I am wondering how that (extremely high temporal AND spatial resolution) can be possible without any acceleration factor?

4. Imaging Methods; Preprocessing:
- Why only the first volume discarded (instead of 2 or even 3 which seems more standard to me)?
- What was the resulting spatial resolution after normalization?

5. Fig. 3: what do the error bars represent? In Fig. 3.A) it looks like quite a consistent difference between experienced and observed trials. However, the statistical analysis reported in the main text, this effect is far from being significant.

6. Same in Fig 4: what do the error bars represent?

Reviewer #3

(Remarks to the Author)

In this manuscript, the authors examine whether learning and memory processes vary with the learning domain (experiential versus observational) and contingencies (whether observed action-outcome associations are congruent with experienced ones), as well as individual differences in these processes based on whether participants engage in observational learning processes (social learning group) or not (social ignoring group). The authors find several differences in behavior and brain activation patterns between the two groups. Specifically, while the social learning group performed worse especially on incongruent trials (because they fail to learn that contingencies are reversed in those trials, therefore learning from observation lead them to perform below chance), they exhibit stronger subjective memory confidence, stronger overall brain activity, and possibly stronger encoding of confidence rating during observed trials.

The task design is interesting and the findings somewhat novel; however, I have several concerns, in particular about the robustness and validity of the computational modelling analysis and of the fMRI analyses and statistics, which question the interpretation of most findings and need to be addressed. The rationale for many of the analyses is also difficult to follow, probably due to the fact that no clear hypotheses are formulated.

Major concerns:

- 1) My first major concern relates to the computational modelling analyses. Many robustness checks are missing and the parameter estimates provided in Table S1 suggests that there may be several issues with the models or model-fitting procedure, which question their validity.
 - 1a) If the reward R took values of 1 and 0, the Q values should be initialized to 0.5. Otherwise, with Q -values initialized at 0, this means that a no-reward outcome ($R=0$) is fully expected (i.e. the reward prediction error $R-Q$ will be equal to 0).
 - 1b) Because there is an obvious learning difficulty difference between congruent and incongruent trials and the clear difference in the learning curves in Fig 2C-D, it seems key that the model should include different learning rates for the two trial types.
 - 1c) The 'fmin' function in Matlab usually does not handle boundaries very well and parameter estimates often get 'stuck' at extreme values (e.g. 0 or 1 for learning rate) – this problem appears to be quite significant when looking at parameter values in Table S1, with several participants having both learning rates very close to 0, or softmax beta closed to 100 (which I assume was used as an upper bound). I would recommend using more sophisticated methods to handle this issue, such as transforming parameters after fitting instead of imposing bounds, or using the cbm toolbox in Matlab (the tutorial explains how to transform parameters).
 - 1d) Parameter recovery analyses should be performed to ensure that parameter values are meaningful and can be accurately recovered. Similarly, because multiple models are being tested, model recovery analyses are key to ensure different models are not confused (i.e. data generated by one model is best fit by that model).
 - 1e) I understand why the authors might want to ensure that the likelihood of the chosen model is greater than chance; however, how about the case where both models are higher than chance but with BICs very close to each other? In that case, using BIC to assign a participant to a group appears somewhat arbitrary.
 - 1f) It appears that the authors are equating the learning rate with a strategy weight, i.e. higher learning rate for observed compared to experienced trials means greater weight to observed trials. I do not believe this is correct. The learning rate quantifies the speed of learning (i.e. how quickly new outcome information is used to update beliefs) and in some circumstances a lower learning rate can be more optimal for learning.
 - 1g) Posterior predictive checks are needed to show that the behavioral patterns shown in Figure 2A-D can be appropriately predicted by model-generated data.
- 2) My second major concern pertains to the lack of robustness, as well as unclear information, related to the fMRI analyses.
 - 2a) First, it is unclear what first-level GLMs were defined. Please explain in more details exactly how many GLMs were run, what regressors were included in each GLM, time-locked to what onset, including duration and parametric modulators. The methods section is currently too vague to ensure that those analyses could be reproduced by the reader.
 - 2b) Perhaps more concerning, the fMRI findings appear not to be corrected for multiple comparisons.
 - 2c) For the analyses shown in Figure 4, a simple contrast between observed and experienced trials does not provide much insight into the mechanisms of learning. It would be more interesting to study the signals associated with specific learning computations, such as prediction errors or Q -values across trials.
 - 2d) In the fMRI GLM testing for memory effects, how were the differential reaction times from Figure 3D accounted for? Could the activations be driven by the shorter exposure time to the picture in the remembered versus non-remembered trials?
 - 2e) Could the effects shown in Table 2 simply be driven by the possibility that brain activity is in general stronger during observed than experienced trials, regardless of the region? A more specific result would be to show that the memory effect (i.e. the remembered vs non-remembered contrast) differs between trial types.
 - 2f) For both analyses, why were run 1 and run 2 analyzed separately? In order to make a claim that the effects are specific to run 2 (in Table 1), the authors should test for a group*run interaction – visually inspecting the plots in Figure 4B, signal just appears to be generally lower in the social ignoring group, which again is difficult to interpret since this mean signal does not represent a well-controlled contrast or meaningful computation.
 - 2g) Figure 6 – it's unclear whether this analysis is ROI-based or whole brain. There are also no stats provided to ensure that there is a significant trial type by group interaction.
- 3) A third concern relates to the fact that most of the authors' claims rely on the differences between the two groups – beyond the concerns described above about the somewhat arbitrary nature of defining the groups, the concerns that parameter values are unreliable, and some missing statistics in the fMRI analyses not properly assessing the group differences/interaction with other factors, I wonder if the analyses testing for such group differences are appropriately powered? With only 20 and 21 participants in each group, power is likely to be low.
- 4) Generally, I found that the analyses that were conducted lacked rationale. Perhaps this is due to the authors' sole hypothesis mentioned in the introduction being that "observing the decisions of others would activate brain networks associated with social cognition". This hypothesis is very general, does not justify the experimental design nor the analyses that were conducted, and leaves the reader wondering what to expect in the results and whether the observed results are consistent with hypotheses or not. More specific hypotheses need to be put forth about the expected findings, so that the rationale for the proposed analyses is justified in line with the question and hypotheses, and findings that do not support those hypotheses need to be discussed appropriately.
- 5) An example related to the point above pertains to the memory data analyses. Based on the authors' previous findings, the

hypothesis should be that participants who prioritize their own experiences would display better memory for events associated with high reward values specifically; however, in their current analyses, the authors did not consider reward value (i.e. whether a trial was rewarded or not). If I understand their previous result correctly, memory performance should be higher for rewarded versus unrewarded trials and this effect might be stronger or only present in those individuals that learn from experience rather than observation (i.e. social ignoring group). Why wasn't the effect of reward outcome tested? Related to this, could this potential effect of reward underlie the confusing finding of shorter exposure time being associated with better memory performance? It could be that the expectation/anticipation of the reward makes participants be both faster and remember the picture better.

6) My final major concern is that the authors do not appear to have made their data and code available online (or at least for the reviewers, with an explicit plan to share openly upon publication).

Other comments, in order of text:

7) p.3: "the neural correlates underlying inferential social learning, i.e., updating our beliefs by gathering information from social evidence, remain largely unknown" – this is a bit misleading as there is a lot of literature on this topic (e.g. Hampton et al 2008 PNAS, Yoshida et al 2010 Journal of Neuroscience, Hill et al 2017 Nature Neuroscience, Collette et al 2017 eLife, Charpentier et al 2020 Neuron, Diaconescu et al 2020 eLife, Zhang & Glascher 2020 Science Advances, Konovalov et al 2021 Neuron, to just cite a few), so it would help to be a bit more specific in identifying what the gap in the literature really is that this study aims to fill.

8) p.4: "observing the decisions of others would activate brain networks associated with social cognition" – on p.3 the authors seem to imply from previous literature that it is the mental representation of others' intentions that primarily activates 'social cognition networks' – simply observing others' decisions is not the same process as actively representing their intentions, so it would be great if the authors could disentangle those two processes.

9) Was the bonus payment contingent on rewards earned or was it always 80 NIS regardless of performance? If the latter, did participants believe that it would be contingent on rewards earned when entering the study?

10) Could you clarify the following details about the task design?

10a) How were the 50 experienced and 50 observed trials split? Were there two blocks or were trials interleaved? If there were two blocks, was block order counterbalanced across participants (and if so were there any order effects)? If they were interleaved, were they always alternating (one trial of each type) or randomly interleaved?

10b) How many congruent versus incongruent trials were shown?

10c) What was the rationale behind simulating different accuracy rates for the second player's action between the congruent and incongruent trials? I understand that accuracy ought to be higher on congruent trials because of the 80:20 contingencies, but it doesn't seem very believable that the performance of another player would drop to chance when contingencies become 70:30 (what was used for incongruent trials).

10d) On a similar note, how did the authors determine the 70% and 50% of trials where the second player chooses accurately – were those trials randomly assigned or did they follow some sort of learning rule (since the second player is also supposed to be learning). Generally, it is a bit of a concern in this task design that the second player's simulated choices may not appear very realistic, and I wonder if participants may have picked up on that.

11) The authors appear to introduce the rationale for their study with previous literature on inferring others' intentions. However, nothing in the model is learning the other player's intentions... Instead, the model is simply trying to learn from the action-outcome associations regardless of whether it's the participant or the other player's turn. A social inference model in this task should be able to learn that the other player may have different goals and intentions from the participants (i.e. on incongruent trials, learning that the other player's goal may be opposite the participant's and as a result use this information in a productive manner – e.g. when the other player's chooses the blue fractal, I should choose the orange one because our goals are different). Overall, this comment also relates to points 4), 7) and 8) above, whereby the analyses do not appear well justified by the introduction and rationale – I would recommend either toning down the relevance to social inference processes in the introduction, or actually testing whether social inference is taking place with the appropriate models and analyses.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors adequately responded to all my question and concerns and the re-analysis of the data has resulted in a greatly improved manuscript. I now recommend the publication in Communications Biology.

Reviewer #2

(Remarks to the Author)

My concerns have been addressed.No further comments from my side!

Reviewer #3

(Remarks to the Author)

This manuscript has greatly improved, and I appreciate the new insights into memory-value interactions and overall clearer flow. However, some of my concerns were not totally addressed or some of the new analyses have raised new concerns. I have added more details below (the numbers/letters refer to my original points for clarity) and believe most of those should be addressed with minimal clarification or analyses.

As a side note, I will add that it was often quite hard to track where the authors made changes to the manuscript, and to avoid putting this additional burden on reviewers, I would encourage them in future manuscript revisions to always copy-paste in the rebuttals all the changes that were made to the manuscript text (as well as detailed page and line numbers), so the reviewers can directly see how changes to the manuscript addresses each of their comment.

1ab) The new best-fitting model involves separate learning rates for experienced and observational trials. However, no behavioral difference between the two conditions is reported (unless I missed them) – in what way is behavior (model-free learning performance) different between Exp and Obs and is that tendency captured by the difference in learning rates across participants? Also, the difference in learning rates (as seen in Fig 2E) appears to be primarily driven by the social ignoring group. Would the same model win in both groups? Because participants are split into two groups early on according to their behavioral performance, it would make more sense to fit the models separately to each group to assess not only parameter value differences but overall learning strategy differences between groups.

1d) The posterior predictive checks are a great addition, thank you for also adding model recovery. It would also be great to show parameter recovery, i.e. the correlation between simulated parameter values and recovered parameter values for the two alphas and softmax beta. It's okay to only perform recovery for the best-fitting model, and show it in the supplementary materials to not perturb the flow of the paper.

2a) There are still some missing details about the second model, specifically what onset was used the fractal phase. In the first model it is mentioned that the fractal phase onset was actually the onset of the trial type presentation – is that the same for the second model? If so that does not seem appropriate to track value estimation, which should only start at the time the fractals appear. Please clarify. Event durations for the second model are also missing.

p.11, l.333-336: "Specifically, we tested for regions with stronger BOLD signals during experienced trials using the contrast experienced > observed trials, and for preferential activity during observation using the inverse contrast observed > experienced." What GLM was used for this analysis (first or second?), and what onset/phase (trial type/fractal, picture or feedback)?

p.11, l.337: is the FWE correction whole-brain FWE? If so please add 'whole-brain' for clarification.

p.11, l.339, l.343 and l.347: Please clarify which GLM was used in each of these analyses.

Figure 4 and Table 1 mention time-locking to the presentation of fractal pairs. Are those images obtained with the first or second GLM? If first, the methods for that GLM state that the onset of the fractal phase is in fact the onset of the trial type, so this needs to be clarified. It would be stronger to show activation time-locked to the fractal presentation (than to the trial type presentation) to ensure that differences are driven by the condition and not by visual differences between the text presented on the screen.

2c) If I understand the second model correctly, value regressors were only included at the time of the picture (and not at the time of the fractal presentation). In order to examine learning signals, expected value is likely to be computed during fractal presentation and prediction error should be the main signal at feedback. While it's interesting to observe value signal during the picture presentation, it is unclear why the authors would expect value to only be represented during that event.

The authors state in their response that "brain regions that were selectively activated during the presentation of images that were subsequently correctly recollected were modulated by Q-value, mirroring our behavioral findings". First, it is unclear whether the value signals shown in Fig 5C for Obs are significant, since the only difference that's reported is the p-value of the main effect of condition. The authors should show that there is significant value representation 'on its own' and not solely as a difference from Exp. Full statistics and effect sizes should be reported for this analysis, not just a P-value in the figure legend. The so-called mirroring of the behavioral finding does not seem to be accurate, since the authors themselves describe in their discussion that "while the behavioral effect of trial value on memory formation was restricted to the 'Social ignoring', group, the accompanying neural activation during observed trials was apparent in both groups" (l.715), thus suggesting more of a discrepancy than a mirroring.

2d) The authors state in the rebuttal that "using the new group assignment, we did not find significant differences between remembered and non-remembered trials across groups and conditions.". Figure 3C legend still reports a significant difference in RT between experienced and observed trials. Please clarify.

2e) The authors state in the rebuttal that "As suggested by the reviewer, we examined the activation to image presentation separately for each group in remembered vs non-remembered trials and showed that in observed trials the regions revealed by this contrast are modulated by trial value (see above)." While that analysis is interesting, it is not the analysis I suggested, which would be to model an interaction between subsequent memory (remembered > non-remembered) and trial type (observed > experienced), or at the very least visualize the 'subsequent memory' map separately for observed and experienced trials.

4) I have re-read the introduction in detail and still do not see where the hypotheses are specified. I think that l.112-129 should not be in the introduction as those lines are summarizing the results, and would be better placed in a 'conclusion'

section at the end of the discussion. Instead, the end of the introduction should detail the hypotheses and the rationale behind them, similar to what the authors did in the rebuttal, although I wouldn't necessarily frame the hypotheses in terms of social learning and social ignoring groups since those groups were not pre-determined and instead defined based on the data.

5) The following analysis might help better relate the value-memory effects from behavior to brain. Instead of the binning procedure the authors used, a more statistically valid approach for assessing the behavioral effect of value on memory would be to perform a trial-by-trial mixed effect logistic regression, where the dependent variable on each trial is whether the picture will be remembered (coded as 1) or not (coded as 0) and the predictors are cue expected value, condition, group and their interaction (with at least a random intercept and maybe some random slopes if the covariance structure allows it). This framework would allow better parsing of variance in the data.

6) Details plans for sharing the data and code should be mentioned and links made available in the manuscript. Code and materials should be well organized prior to submission, not upon acceptance, in order to minimize the risk of errors in the analysis pipeline and to ensure reproducibility.

Minor comments and typos:

- The name 'pairwise' for the model with separate learning rate for incongruent vs congruent pairs is confusing as in the logic of 'oneAlpha', 'twoAlphas', 'fourAlphas' and 'pairwise' it would suggest more than four alphas. It would make more sense to rename the 2 twoAlpha models something like 'twoAlphas(exp/obs)' and 'twoAlphas(cong/incong)' or something along those lines.
- Figure 2A x-axis should say 'Incongruent' (not 'lcongruent')
- Figure S2A title should say 'Experienced' (not 'Experinced')
- Figure 3B legend states that "The 'Social Ignoring' group (blue shades) exhibited significantly lower confidence ratings (shown in z-scores) during the recognition test in comparison to the 'Social learning' group (red shades)". This is not evident at all from the graph, which appears to show no differences between groups or conditions. Differences in RT in panel 3C are also not evident. In panels 2E, 3A, 3B, and 3C, it would be great if the authors could highlight significant effects of interest for the reader.
- Figure 3B legend, l.528, in the phrase 'measured as proportion of correct responses to presented images – hit-rates', is the '–' a minus? If not I would rephrase as 'hit rate, measured as proportion of correct responses to presented images'.
- In Figure 3D-G, is each dot a participant? According to the methods, it seems like each dot is a bin of high-reward value. This should be made clear in the Figure legend, though probably the mixed-effects logistic regression approach suggested above (point 5) would be a better way to analyze value effects on memory.
- Figure 5C is missing a legend for the blue/red colors

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)

Thank you to the authors for addressing the concerns. The only remaining minor concern is about the new analysis in response to my point 5. Specifically, the authors first state that they performed "a trial-by-trial linear mixed-effects regression, with subsequent memory success as the dependent variable (0/1), and Q-value, trial type (experienced/observed), group (learning/ignoring) as predictors, and a random intercept for subject", then they state that they actually performed separate regressions for each group. From a statistical standpoint, it would be better to run a single model with group as a predictor, then only run a separate regressions for each group if an interaction with group is observed in the first model.

Minor point: please cite the new Fig S2 (parameter recovery) somewhere in the main text.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

The authors present an interesting experiment that combines a simple reinforcement learning task interleaved with an incidental memory task that involves own and observed decisions with different reward contingencies. The design is quite complex as there are multiple within- and between-participant factors involved that are defined based on the computational modeling results. Overall, the task has some potential to unravel a detailed relationship between RL and memory formation, but there are several flaws in the modeling and analysis results that prevent a publication at this time. However, most of the criticism can probably be addressed by a thorough revision of the paper and the underlying analyses.

Computational modeling

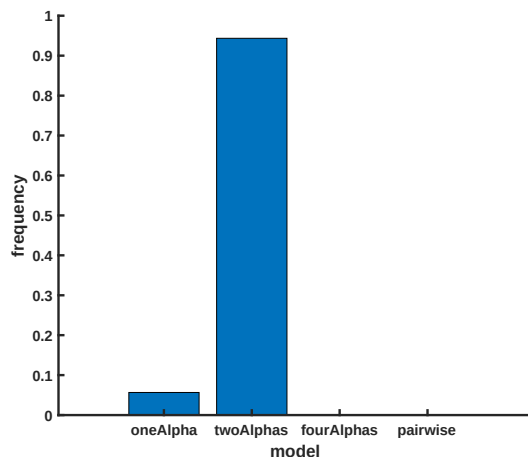
The authors fitted two simple RL models (with either a single or separate learning rates for exp. and obs. trials) to the choice task and used maximum likelihood estimation (MLE) as implemented in the “fmin” functions in MATLAB. While the table in the Supplement is lacking the information about the estimated parameters for the single-LR model, the data presented there already reveals a common problem when using MLE with the fmin-function in MATLAB: learning rates and softmax temperatures are often estimated at (or very close) to the range boundaries of the parameters indicating problems in the optimization procedures. Learning rates close to zeros would suggest that these participants did not learn at all, whereas temperatures near zeros would indicate that participants were only making random choices inconsistent with learning. These are effects due to a flawed estimation procedure, not a problem with the data itself. In the data presented in the supplement only about half of the participants had temperatures below 95 or learning rates above 0.02.

These problems could have been detected in a posterior predictive check, an essential component of model validation in a computational modeling study (see e.g. Wilson & Collins, E-life, 2019, <https://elifesciences.org/articles/49547>), which the authors did not conduct. Essentially, this involves simulating new data with the fitted parameters and comparing them to the experimental data on a model-free performance indices (e.g. accuracy). One solution to this problem is a hierarchical estimation procedure (as implemented in MCMC sampling technique in Stan) or in the CBM toolbox for MATLAB (paper: <https://doi.org/10.1371/journal.pcbi.1007043>, manual: <https://payampiray.github.io/cbm>), which uses variational techniques for parameter estimation. This estimation procedure regulates individual parameter estimates, thus avoid the typical “at-boundary estimates” of fmincon. Interestingly, the toolbox will also provide “model responsibilities” that indicate which model is the best for each participant, thus avoiding the individual model comparison via BIC and protected exceedance probabilities that indicate which model is best in explaining the data in the entire sample overall. Because the modeling results are essential for the paper, all subsequent analyses are likely to be biased, if not flawed and should be taken with skepticism at this point.

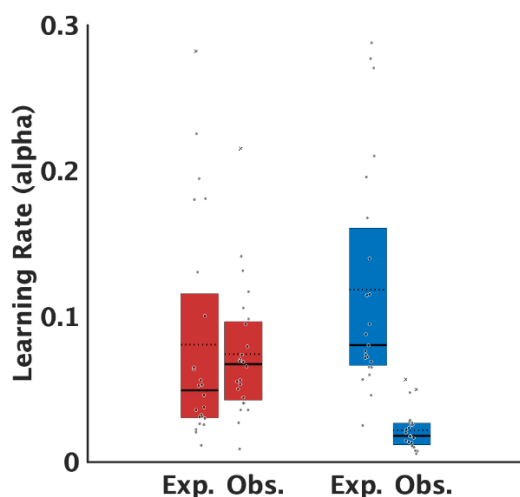
We thank the reviewer for this important comment (a similar comment was raised by reviewer 3). As suggested, we modelled the behavioral data anew, with hierarchical Bayesian inference

using the CBM toolbox, and compared among four alternative models consisting a single learning-rate for all trial types, separate learning rates for experienced/observed trials, separate learning rates for congruent/incongruent pairs across other trial types, or four learning rates for experienced/observed by congruent/incongruent trials.

At the population level, we found overwhelming support for the model with separate learning rates for experienced and observed trials. The mean model frequency across the group was as follows: one-alpha model = 0.07; two-alfas model for experienced and observed trials = 0.93; two-alfas model for congruent/incongruent pairs = 0; four-alfas model = 0. Based on this result, we applied the model with separate learning rates for experienced and observed trials to all participants.

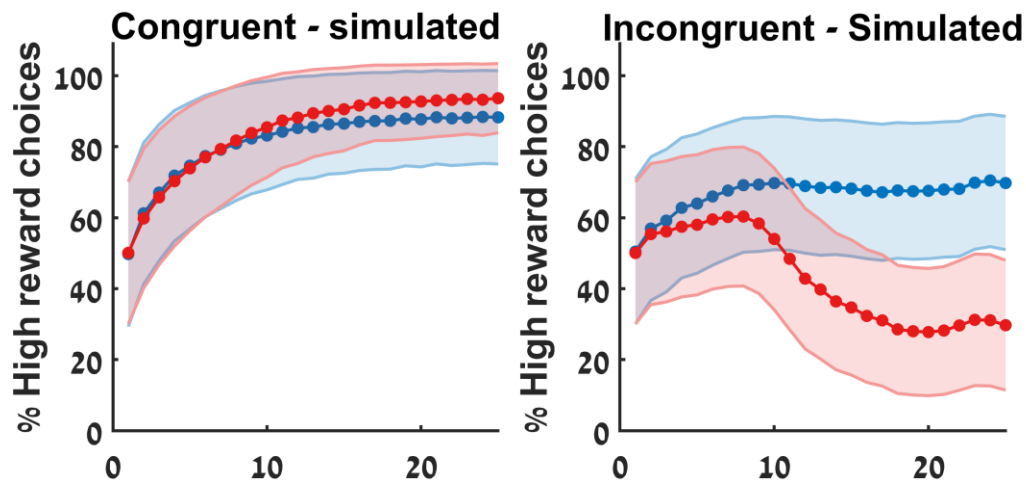


After separating our participants into one of two groups: the ‘Social learning’ or ‘Social ignoring’ group according to task performance (see below), we found that the two groups differed dramatically in their learning rates of the observed trials (see image below, which is now presented as Figure 2E in the revised manuscript):



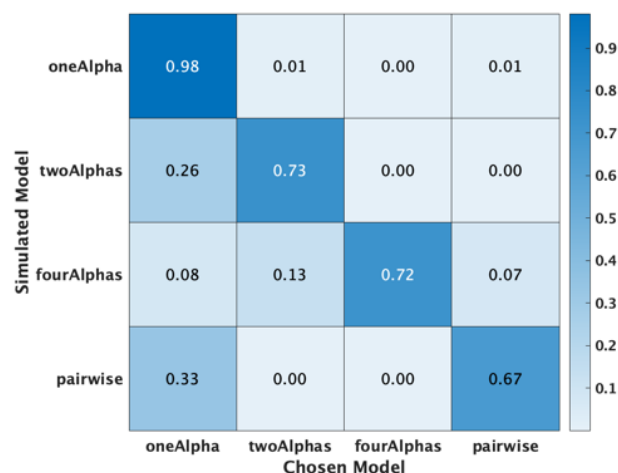
Learning rates are similar for experienced and observed trials in the ‘Social learning’ group (red), but significantly different in the ‘Social ignoring’ group, where learning rates for observed trials are relatively low.

To validate our parameter estimation procedure, we conducted a posterior predictive check. For each participant, we simulated choice behavior using the model-derived parameters, repeated 1000 times, and grouped the data according to the group assignment, aiming to reproduce the learning curves originally shown in Figure 2C,D. The resulting simulated curves are now presented in Figure 2G,H.



Posterior predictive curves shown here represent the learning curves for the participants assigned to the ‘Social learning’ group (blue) and ‘Social ignoring’ group (red), as well as the 95% variance of simulated data, using the parameters obtained for each participant in the CBM analysis. These data are shown separately for congruent trials (left) and incongruent trials (right).

To further validate our modeling procedure, we conducted 100 simulated experiments, each including 60 datasets (representing 60 participants). Choices were simulated based on the fitted parameters of the four alternative models: a single learning rate for all trials, separate learning rates for experienced and observed trials, four separate learning rates (experienced/observed by congruent/incongruent), and another assigning distinct learning rates to each stimuli pair (congruent/incongruent across trial type). The resulting confusion matrix—depicting model recovery performance—is shown below and in Figure 2F of the revised manuscript.

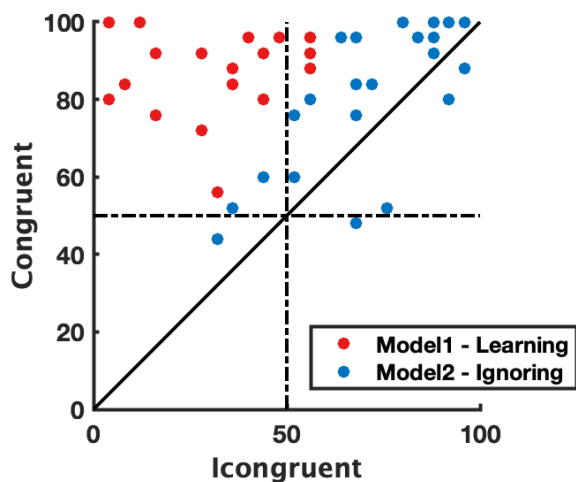


Definition of participant groups

The definition of the between-participant factor Group (social learning vs. social ignoring) was made based on the model comparison between the single-LR vs. double-LR model. Because the model estimation in its current state is flawed, the analytical decision about the group should not be based on it. In fact, I would find it much more convincing, if the participant groups were defined on the (model-free) performance in congruent and incongruent trials (e.g. split at 50% in Fig. 2 A) and then fit the models to these groups of participants. Overall, it is quite striking that a substantial number of participants cannot ignore the incongruent contingencies in observed trials, when making their own choices in incongruent trials.

We agree that a model-free division of the groups according to behavioral profiles and then assessing model fit is a more straightforward approach, and have implemented this suggestion. We divided the data based on participants' performance in incongruent trials, using their scores in congruent trials as baseline. For each participant, we calculated the ratio of their mean scores in incongruent trials and in congruent trials. Group assignment was made using a cutoff of the median of these ratios, resulting in two balanced groups, where participants with scores above the median were assigned to the 'Social learning' group, and those below the median to the 'Social ignoring' group.

As shown in the image below, this division effectively distinguishes between participants whose scores in both conditions are relatively similar (presented in blue, clustering around the diagonal) and those whose scores deviate from the diagonal (shown in red). This is depicted in the revised manuscript in Figure 2A.



Regarding the reviewer's remark about the surprisingly large proportion of subject who fail to ignore the incongruent contingencies, we note that, by design, participants are unaware that the incongruent trials differ from congruent. Moreover, the contingencies in the incongruent condition are more ambiguous and therefore harder to learn.

Model-free behavioral analyses

There is one puzzling result in the behavioral data (Figs 2 and 3) that remains unmentioned and unexplained: the social learning group commits much more errors on own choices in incongruent trials (Fig 2D), but they have a much higher memory confidence in observed, incongruent trials. Why is that and what does that say about the interaction between value-based decision-making and declarative memory? I would welcome an interpretation and some conceptual thoughts about this finding.

After reassigning participants into groups based on their behavior in the RL task, as detailed in the response to the above comment, this is no longer the case. Both groups show comparable confidence levels on average across all conditions.

fMRI analyses

The contrast of experience vs observed trials (thresholded leniently at $p < 0.001$ uncorr.), which serves as a functional localizer is not completely orthogonal to condition-specific ROI analysis. This ROI analysis pools over all voxels in the ROI thus avoiding the multiple comparisons problem. A "normal" mass-univariate fMRI analysis would have not yielded any significant findings in these analyses. However, the functional localizer essentially serves as a mask, which precludes the analysis of any results that show the opposite effect (exp. > obs. trials). Thus, we don't really get to see, which brain regions are activated, when own decisions are made (compared to observe decisions).

Thank you for this comment. We have taken a number of measures to address this concern. First, we strengthened the statistical thresholds, and now show statistical maps that are corrected using FWE at $P < 0.05$, with a cluster size threshold of $k > 20$.

Moreover, we added the results of the experienced > observed contrast during the time of presentation of the fractal pairs. As anticipated, this contrast yields robust activation in motor networks (Figure 4), as well as in anterior cingulate cortex, anterior insula, and ventral striatum. These findings are detailed and discussed throughout the manuscript. Finally, we decided to omit the ROI analysis shown in the previous manuscript version, as the groups and model approach substantially changed during the revision procedure.

Minor issues and typos

page 6, bottom: at this point, it is not clear that the participants did not know about the subsequent memory test. Please include this information here.

We added the following sentence in p.7 of the revised manuscript:
'Participants were not informed that they would undergo a memory test for the images presented during this stage.'

page 7, Recognition Memory Test: was there an equal number of old and new items?
Yes, this is now mentioned in p. 7 of the revised manuscript, under 'Recognition memory test'.

page 9: description of the fMRI preprocessing is very sparse.

The fMRI processing steps are now elaborated in detail, in p. 10 of the revised manuscript.

Figure 1: please use different choice cues for the obs trials structure in panel A

The choice to show similar stimuli pairs in experienced and observed panels was intentional, since we wish to emphasize the fact that the two stimuli pairs were presented in both conditions.

Figure 2 and 3: The different shades of red are virtually indistinguishable, please choose a different color scheme for the 4 conditions.

We changed the color scheme and hope that the new color scheme is now satisfactory.

Legend of Fig 2: (C-D) Please stick with the group labels "social learning" and "social ignoring"

Corrected

Figure 4: Some of the bars for "social learning exp" in run 2 have the wrong shade of blue.

Figure 4 has now been changed.

page 22, bottom: ... tendency to follow the choices of others even if the outcome ... "of" should be changed to "if"

This has been modified.

page 22, bottom: “Generalized” should be changed to “General”

Done

page 23: “metallization” should be changed to “mentalization”

Changed

page 23: “This may indicate that the mental representation of others’ decision process was taken into account, at least to some extent, not only when observing others decide, but also during one’s own decision-making process.” This reverse inference (attributing cognitive function only because a certain brain area is activated) and should be removed.

This was removed.

Reviewer #2 (Remarks to the Author):

This study examined behavior and brain activation associated with different reinforcement learning styles. More specifically, the authors tested the hypothesis that subjects would differ regarding how much they are inclined to incorporate information from socially observed decision making trials in addition to self-performed trials. I think this is a timely and interesting research question, and the experimental design is suitable (with one exception, see specific comments below). Behaviorally, this study partly replicated previously published results from the same authors suggesting that a 'social ignoring' subgroup primarily uses information gained from self-performed (or 'experienced') trials whereas another 'social learning' group uses information gained from both trial types. The brain imaging results, however, are not convincing in its current form. The main reason is that large parts of the statistical analysis remain unclear and insufficient thereby compromising the reliability of the reported results.

Major points:

In general, I was quite disappointed by the lack of detail regarding many aspects of the fMRI analysis. This includes pretty much all aspects of the analysis: data acquisition, preprocessing (see 'other comments'), modeling, and statistical evaluation.

1. To start with, I could not find information as to whether the model regressors were time-locked to the stimulus pair, the outcome, or the response? This of course is a critical piece of information when it comes to reinforcement learning where it is important to know if we are looking at stimulus-related effects (evolution of Q values) or outcome-related effects (evolution of prediction error).

We regret omitting important details of the analysis. These have been added to the revised manuscript in the Methods and Results sections.

Figure 4 presents the statistical maps from contrasts comparing experienced and observed trials, time-locked to the presentation of the pairs of choice cues.

Figure 5 displays the results of contrasts using regressors time-locked to the presentation of the images that were later tested for recognition. Panel A shows regions that were activated during the presentation of images that were subsequently recognized, compared to non-recognized images. Panels B and D display significant activation associated with a regressor designed to track the parametric modulation of the maximum available Q value. This information is now elaborated in the main text (pp. 11-12, pp. 20-22) and in the figure legend.

2. The statistical analysis comprised two steps. First a whole-brain contrast was computed to identify voxels that were more strongly activated for observed trials than for experienced trials. Second, based on this contrast 4 ROIs were identified, which were then further scrutinized regarding more specific effects involving group, congruency, and run (as a proxy for learning progress).

As far as I can tell, the whole brain contrast was not corrected for multiple comparisons. Instead, an uncorrected threshold of $p < .001$ was applied. This is certainly not acceptable. Given the rather large sample size ($N = 41$), it should be possible to reveal a reliable effect based on a proper correction method (voxel level or cluster level FWE-corrected). Hence, the basis of the subsequent ROI analysis is already flawed. Also, a complete report of the whole brain analysis

results is missing (e.g. a table with all clusters and statistics). Finally, no information whatsoever is given regarding the reverse contrast (i.e., stronger activity for experienced vs. observed trials).

We agree with the reviewer's comment and have re-analyzed the data using the appropriate statistical thresholds. In light of revisions made throughout the manuscript, we removed the ROI analysis from Figure 4 and now focus on statistical maps of the Experienced > Observed and Observed > Experienced contrasts, time-locked to the presentation of the choice cues. These statistical maps were thresholded using family-wise error (FWE) correction at $P < 0.05$, applied at both voxel and cluster levels.

Additionally, the locations and statistics of the activated clusters corresponding to all figure panels are now provided in the accompanying tables.

For the ROI analysis, again a complete report of the results is missing. Information regarding the main effects of group, congruency, and run is missing. Also, and even more importantly, no information regarding any of the interaction effects is presented. Of particular importance here is the interaction group x run. If this interaction is not significant the selective report of group differences in each run is not valid (a difference in significance is not a significant difference). The main point the authors want to make is that there is an increasingly stronger activation from run 1 to run 2 specifically in the social learning group in observed trials. However, if there is no significant interaction between group and run in observed trials (actually, between group, run, and trial type), this claim cannot be supported statistically.

In light of broad revisions made to the manuscript, we have removed the ROI analysis related to the Observed > Experienced trials from the results.

3. Detailed information on how the ROI analyses were conducted is missing. Specifically, which voxels were included in each ROI at which threshold? Mean across voxel values or only peak voxel?

Figure 5C shows ROI analysis of voxel clusters that were activated by the contrast of subsequently remembered vs. non-remembered images. The procedure is now detailed in pp. 11-12 of the revised manuscript:

“To explore subsequent memory effects, we contrasted between subsequently remembered and non-remembered trials during the picture presentation phase. Second-level statistical maps were generated by contrasting the parametric modulators against baseline, using a statistical threshold of $P < 0.001$, FWE corrected for cluster size at $P < 0.05$.”

“We also conducted an ROI analysis based on the activations identified in the subsequent-memory contrast (subsequently remembered > non-remembered trials during picture presentation). For each region, we extracted parameter estimates associated with the high-value parametric regressors. Specifically, for each participant, the mean parameter estimate was calculated across all voxels within the cluster defined by the second-level analysis. We then performed a mixed-effects ANOVA on the parameter estimates to assess potential effects of group and condition (experienced vs. observed).”

4. I was surprised that the authors did not use estimated Q-values (and estimated prediction error for that matter) as trial-wise parametric regressors during the fMRI analysis? Wouldn't this be a

much more powerful way of tracking learning-related changes in brain activation than comparing run 1 and run 2 as done in the current paper version?

As suggested by the reviewer, we conducted a parametric modulation analysis, using trial-by-trial Q-values, as derived from the Rescorla-Wagner model for each participant, separately for Experienced and Observed trials, time-locked to image presentation. The results of this analysis are shown in Figure 5 and Table 2 of the revised manuscript. This analysis revealed significant activation in vmPFC in both experienced and observed trials. In observed trials, significant activation correlated to trial-by-trial Q-values was observed in the bilateral extrastriate cortex, right fusiform gyrus and posterior cingulate cortex (Table 2). Importantly, in observed trials, brain regions that were selectively activated during the presentation of images that were subsequently correctly recollected were modulated by Q-value, mirroring our behavioral findings.

5. Although I generally liked the experimental design, I am worried that the timing differences between observed and experienced trials (choice time 2 sec vs. < 4 sec) precludes a clean comparison of the two conditions in terms of brain activity. I might be an option to include choice time as an addition covariate.

The mean choice time for observed trials were set a-priori at 1970ms, ranging from 1550 to 2350ms. The mean duration for the choices made by participants (experienced trials) were 1235.5 ± 313 ms. As now described in p. 11, the regressors time-locked to the choice stimuli presentation phase were defined from the onset of the trial and extended until the end of the choice display. Accordingly, the mean modeled durations across participants were 3735.5 ± 313.5 ms for experienced trials and 4470ms for observed trials. Notably, these regressors were modeled individually for each participant based on the actual duration of each trial, ensuring that the modeled durations accurately captured the specific timing of the phase of interest for each and every event.

Other comments

1. Is the equation for computing the reward prediction error correct (page 7)?
 $1 - R - Q$ instead of $1 - (R - Q)$?

δ_t^{uc} and Q_t^{uc} both relate to the unchosen option, whereas R_t represent the actual received reward. Thus, the reward that would have been granted to the unchosen option is $1 - R_t$ and the resulting prediction error $1 - R_t - Q_t$.

2. The social learning group includes two sub-types

- a) Assigning same weight to observed trials
- b) Assigning more weight to observed than experienced trials

How many subjects are in (a) vs. (b). Are the results affected by including only (a) subjects (the group that was exclusively identified in the previous publication)?

The group assignment has changed during the course of the revisions made to the manuscript as suggested by Reviewer 1. In the revised manuscript, participants are assigned in a model-free manner, based on the ratio of their choices during incongruent and congruent trials. The cutoff

value was set as the median of these ratios, creating two equally-sized groups. Learning rates were then assessed for each participant using a model that included separate parameters for experienced and observed trials. As shown in Figure 2E, the participants assigned to the ‘Social ignoring’ group had significantly lower learning rates in observed vs. experienced trials, whereas the learning rates in the ‘Social learning’ group were similar.

3. Imaging Methods; data acquisition: functional data were collected with TR=1000 and 64 slices. I am wondering how that (extremely high temporal AND spatial resolution) can be possible without any acceleration factor?

Indeed, a multi-band acceleration factor = 6 was used during scanning. This has been added to the Methods section.

4. Imaging Methods; Preprocessing:

- Why only the first volume discarded (instead of 2 or even 3 which seems more standard to me)?

Thank you for raising this point. While it's true that discarding 2–3 initial volumes is common practice, especially in the absence of active shimming, in our case advanced shimming procedures were performed prior to each run, ensuring that the magnetic field was well-stabilized before acquisition began.

To further deal with potential confounds, we additionally modeled the first event of each run as an event of no interest, so any small remaining fluctuations wouldn't affect the results. Based on this, we determined that removing only the first volume was sufficient for our analysis. We added this information to the revised manuscript (p. 11, para. 1).

- What was the resulting spatial resolution after normalization?

After normalization, the images were resampled to a spatial resolution of $2 \times 2 \times 2 \text{ mm}^3$, in line with the standard MNI template.

5. Fig. 3: what do the error bars represent? In Fig. 3.A) it looks like quite a consistent difference between experienced and observed trials. However, the statistical analysis reported in the main text, this effect is far from being significant.

Since the group division has changed following the revisions, the graphs in Figure 3 have changed as well. Despite some apparent differences among groups, there were no statistical differences in memory performance.

6. Same in Fig 4: what do the error bars represent?

The error bars in the bar graphs throughout the manuscript represent standard error of the mean.

Reviewer #3 (Remarks to the Author):

In this manuscript, the authors examine whether learning and memory processes vary with the learning domain (experiential versus observational) and contingencies (whether observed action-outcome associations are congruent with experienced ones), as well as individual differences in these processes based on whether participants engage in observational learning processes (social learning group) or not (social ignoring group). The authors find several differences in behavior and brain activation patterns between the two groups. Specifically, while the social learning group performed worse especially on incongruent trials (because they fail to learn that contingencies are reversed in those trials, therefore learning from observation lead them to perform below chance), they exhibit stronger subjective memory confidence, stronger overall brain activity, and possibly stronger encoding of confidence rating during observed trials.

The task design is interesting and the findings somewhat novel; however, I have several concerns, in particular about the robustness and validity of the computational modelling analysis and of the fMRI analyses and statistics, which question the interpretation of most findings and need to be addressed. The rationale for many of the analyses is also difficult to follow, probably due to the fact that no clear hypotheses are formulated.

Major concerns:

1) My first major concern relates to the computational modelling analyses. Many robustness checks are missing and the parameter estimates provided in Table S1 suggests that there may be several issues with the models or model-fitting procedure, which question their validity.

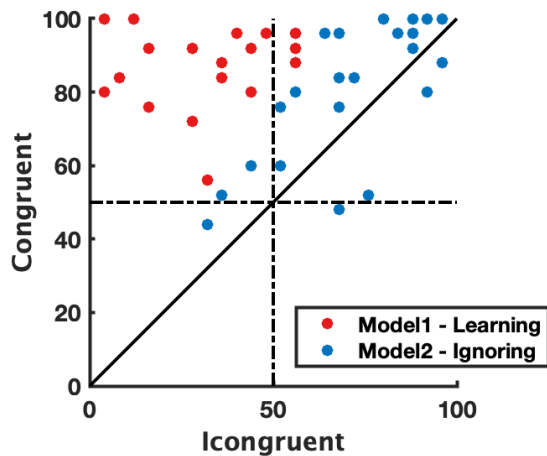
1a) If the reward R took values of 1 and 0, the Q values should be initialized to 0.5. Otherwise, with Q -values initialized at 0, this means that a no-reward outcome ($R=0$) is fully expected (i.e. the reward prediction error $R-Q$ will be equal to 0).

1b) Because there is an obvious learning difficulty difference between congruent and incongruent trials and the clear difference in the learning curves in Fig 2C-D, it seems key that the model should include different learning rates for the two trial types.

Following the reviewer's comments, as well as similar ones raised by Reviewer 1, we changed our modeling procedure. We now perform hierarchical Bayesian inference using the CBM toolbox for Matlab (Piray et al., 2019) to model the participants' behavior, using initial values of $Q=0.5$.

We also changed the analysis approach, such that we initially divide the participants into two groups according to their behavioral performance on the reinforcement learning task.

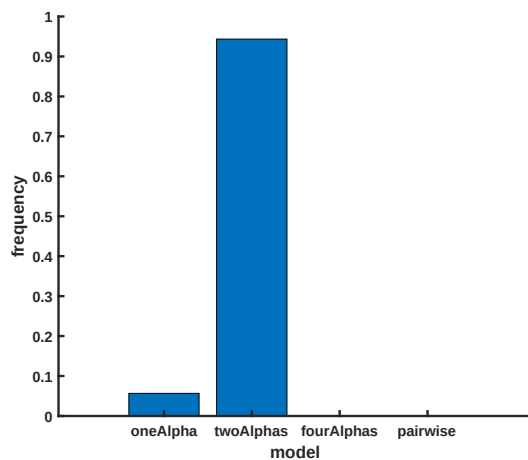
We divided the data according to their performance in incongruent trials, using their scores in congruent trials as baseline (using the median of the ratio between the incongruent and congruent scores as the cutoff value). As shown in the image below, this division effectively separates between participants whose scores on both conditions are relatively similar (in blue - around the diagonal), from those whose scores deviate from the diagonal (in red).



In light of the reviewer’s comment 1b, we compared among four alternative models consisting a single learning-rate for all trial types, separate learning rates for experienced/observed trials across pair type (congruent/incongruent), separate learning rates for congruent/incongruent pairs across experienced/observed trials, and four learning rates for experienced/observed by congruent/incongruent trials.

At the population level, we found overwhelming support for the model with separate learning rates for experienced and observed trials. The mean model frequency across the group was as follows: one-alpha model = 0.07; two-alphas model for experienced and observed trials = 0.93; two-alphas model for congruent/incongruent pairs = 0; four-alphas model = 0. Based on this result, we applied the model with separate learning rates for experienced and observed trials to all participants.

This issue is now detailed in p. 15 of the revised manuscript.



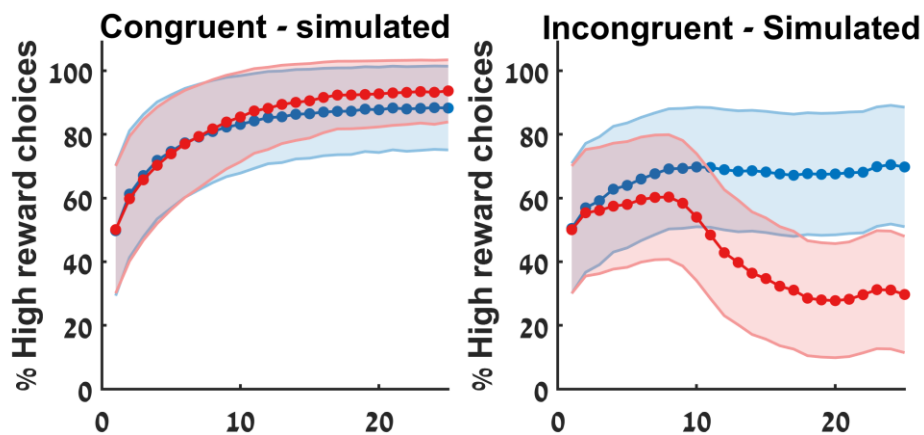
1c) The ‘fmin’ function in Matlab usually does not handle boundaries very well and parameter estimates often get ‘stuck’ at extreme values (e.g. 0 or 1 for learning rate) – this problem appears

to be quite significant when looking at parameter values in Table S1, with several participants having both learning rates very close to 0, or softmax beta closed to 100 (which I assume was used as an upper bound). I would recommend using more sophisticated methods to handle this issue, such as transforming parameters after fitting instead of imposing bounds, or using the cbm toolbox in Matlab (the tutorial explains how to transform parameters).

Thank you for this suggestion, which we have now implemented. Following other comments, we now divide the groups according to their behavior, and independently use the CBM toolbox to model each subject's behavior as detailed in the answers to the above comment. As noted above, the model with two learning rates for experienced and observed trials provided a significantly better fit than the other modeled alternatives. Therefore, all subsequent analyses — including the extraction of individual alpha values, trial-by-trial expected values (Q), and prediction errors — were based on the two-rate model for each participant.

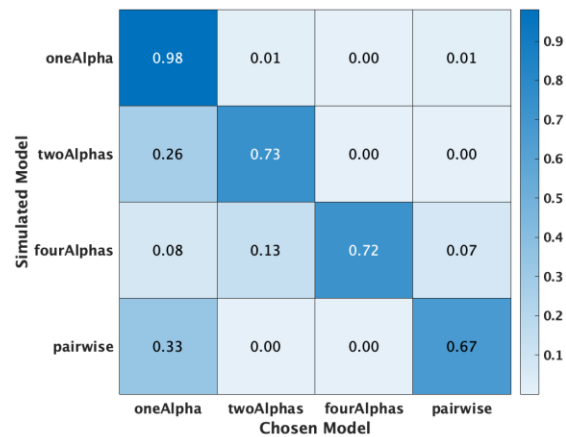
1d) Parameter recovery analyses should be performed to ensure that parameter values are meaningful and can be accurately recovered. Similarly, because multiple models are being tested, model recovery analyses are key to ensure different models are not confused (i.e. data generated by one model is best fit by that model).

Thank you for this comment. Figure 2G, H (see image below) of the revised manuscript now show the posterior-predictive learning curves estimated by simulations that use the same parameters as the ones computed by the model for our original participants. Specifically, for each participant, we simulated choice behavior using the model-derived parameters, repeated 1000 times, and grouped the data according to the group assignment, aiming to reproduce the actual learning curves shown in Figure 2C, D.



To further validate our modeling procedure, we conducted 100 simulated experiments, each including 60 datasets (participants). Choices were simulated based on the fitted parameters of our primary model, as well as two alternative models: one incorporating four separate learning rates (experienced/observed by congruent/incongruent), and another assigning distinct learning rates to each stimuli pair (congruent/incongruent across trial type). The resulting confusion matrix—

depicting model recovery performance—is shown below and in Figure 2F of the revised manuscript.



1e) I understand why the authors might want to ensure that the likelihood of the chosen model is greater than chance; however, how about the case where both models are higher than chance but with BICs very close to each other? In that case, using BIC to assign a participant to a group appears somewhat arbitrary.

Based on the comments above, we have modified the analysis such that participants are assigned to different groups according to their behavioral profiles in the RL task, and subsequently modelled using the CBM toolbox.

1f) It appears that the authors are equating the learning rate with a strategy weight, i.e. higher learning rate for observed compared to experienced trials means greater weight to observed trials. I do not believe this is correct. The learning rate quantifies the speed of learning (i.e. how quickly new outcome information is used to update beliefs) and in some circumstances a lower learning rate can be more optimal for learning.

We agree with this comment. As described above, we now use a different approach for group separation, relying on behavior during incongruent and congruent trials of the RL task.

1g) Posterior predictive checks are needed to show that the behavioral patterns shown in Figure 2A-D can be appropriately predicted by model-generated data.

We now show posterior predictive curves in Figure 2G, H.

2) My second major concern pertains to the lack of robustness, as well as unclear information, related to the fMRI analyses.

2a) First, it is unclear what first-level GLMs were defined. Please explain in more details exactly how many GLMs were run, what regressors were included in each GLM, time-locked to what onset, including duration and parametric modulators. The methods section is currently too vague to ensure that those analyses could be reproduced by the reader.

We regret omitting important details of the analysis. These have been added to the revised manuscript in the Methods and Results sections.

Figure 4 presents the statistical maps from contrasts comparing experienced and observed trials, time-locked to the presentation of the pairs of choice cues.

Figure 5 displays the results of contrasts using regressors time-locked to the presentation of the images that were later tested for recognition. Panel A shows regions that were activated during the presentation of images that were subsequently recognized, compared to non-recognized images. Panels B and D display significant activation associated with a regressor designed to track the parametric modulation of the maximum available Q value. This information is now clearly stated in the main text (pp. 11-12, pp. 20-22) and in the figure legend.

2b) Perhaps more concerning, the fMRI findings appear not to be corrected for multiple comparisons.

We thank the reviewer for raising this important issue. All statistical maps presented in the revised manuscript are corrected for multiple comparisons. Figure 4 shows the comparison between experienced and observed conditions, time-locked to stimulus-pair presentation, using a family-wise error (FWE) correction at $P < 0.05$ with a minimum cluster size of 20 consecutive voxels. Figure 5 presents contrasts for conditions time-locked to image presentation onset, using a threshold of $P < 0.001$ and corrected for cluster size with FWE at $P < 0.05$ ($k > 40$ voxels).

2c) For the analyses shown in Figure 4, a simple contrast between observed and experienced trials does not provide much insight into the mechanisms of learning. It would be more interesting to study the signals associated with specific learning computations, such as prediction errors or Q-values across trials.

We agree with this comment, in addition to the contrasts comparing experienced vs. observed trials, we now show the results of a parametric modulation analysis, using Q-values derived from the chosen two-alpha model to show activations related to the evolution of expected values during the time of image presentation, separately for Experienced and Observed trials. This analysis revealed significant activation in vmPFC in both experienced and observed trials. In observed trials, significant activation correlated to trial-by-trial Q-values was observed in the bilateral extrastriate cortex, right fusiform gyrus and posterior cingulate cortex (Table 2). Importantly, in observed trials, brain regions that were selectively activated during the presentation of images that were subsequently correctly recollected were modulated by Q-value, mirroring our behavioral findings. This is now reported in Figure 5 and Table 2 of the revised manuscript.

2d) In the fMRI GLM testing for memory effects, how were the differential reaction times from Figure 3D accounted for? Could the activations be driven by the shorter exposure time to the picture in the remembered versus non-remembered trials?

The regressors in this contrast were defined according to trial-by-trial reaction times, such that each event started upon image presentation and ended at button press. Moreover, using the new group assignment, we did not find significant differences between remembered and non-remembered trials across groups and conditions.

2e) Could the effects shown in Table 2 simply be driven by the possibility that brain activity is in general stronger during observed than experienced trials, regardless of the region? A more specific result would be to show that the memory effect (i.e. the remembered vs non-remembered contrast) differs between trial types.

We now show contrasts of observed > experienced trials as well as experienced > observed trials time-locked to stimuli-pair presentation, showing robust differential activation for each contrast. As suggested by the reviewer, we examined the activation to image presentation separately for each group in remembered vs non-remembered trials and showed that in observed trials the regions revealed by this contrast are modulated by trial value (see above).

2f) For both analyses, why were run 1 and run 2 analyzed separately? In order to make a claim that the effects are specific to run 2 (in Table 1), the authors should test for a group*run interaction – visually inspecting the plots in Figure 4B, signal just appears to be generally lower in the social ignoring group, which again is difficult to interpret since this mean signal does not represent a well-controlled contrast or meaningful computation.

We have removed the separation into two runs in the revised manuscript, and now show statistical maps and ROI analysis using models of concatenated datasets across the two sessions.

2g) Figure 6 – it's unclear whether this analysis is ROI-based or whole brain. There are also no stats provided to ensure that there is a significant trial type by group interaction.

As part of the revisions made to the manuscript, we have removed the analysis and results from the study.

3) A third concern relates to the fact that most of the authors' claims rely on the differences between the two groups – beyond the concerns described above about the somewhat arbitrary nature of defining the groups, the concerns that parameter values are unreliable, and some missing statistics in the fMRI analyses not properly assessing the group differences/interaction with other factors, I wonder if the analyses testing for such group differences are appropriately powered? With only 20 and 21 participants in each group, power is likely to be low.

We acknowledge the reviewer's concerns regarding the sample size. We should first mention that, as indicated above, group definition was changed to rely on RL behavior, the parameter values are now based on a more sophisticated modeling procedure using the CBM toolbox, and the fMRI statistics were amended. Moreover, the 'Social learning' and 'Social ignoring' groups now consist of 23 and 24 participants, respectively.

Since most of the group comparisons are performed on the behavioral data, we carried out power analysis on several of the behavioral results, and report effect size and observed power.

The main between-group effect concerned the difference in correct choices during incongruent trials between the 'Social Learning' and 'Social Ignoring' groups. Using G*Power, we calculated the effect size (Cohen's d) with the formula $d = (\text{Mean}_{\text{group1}} - \text{Mean}_{\text{group2}}) / \text{pooled SD}$, and conducted a two-tailed power analysis. The resulting Cohen's d was 1.488, yielding an observed power of 0.998.

We also conducted a power analysis on the difference in learning rates for observed trials, which may be less dependent on the behavioral scores initially used to divide the groups, and therefore less likely to be biased. Cohen's d was 1.237, resulting in an observed power of 0.985. This is now reported in p. 15 of the revised manuscript.

4) Generally, I found that the analyses that were conducted lacked rationale. Perhaps this is due to the authors' sole hypothesis mentioned in the introduction being that "observing the decisions of others would activate brain networks associated with social cognition". This hypothesis is very general, does not justify the experimental design nor the analyses that were conducted, and leaves the reader wondering what to expect in the results and whether the observed results are consistent with hypotheses or not. More specific hypotheses need to be put forth about the expected findings, so that the rationale for the proposed analyses is justified in line with the question and hypotheses, and findings that do not support those hypotheses need to be discussed appropriately.

We appreciate this comment, and have adjusted hypotheses according to the focus of the revised manuscript. Specifically,

Our study aimed to explore the formation of memories during trial-and-error learning, and examine the effect of the different components of RL on the formation of incidental memory. Based on previous results by us and others, we deemed the observational learning paradigm particularly suited for this goal, since in such situations expectation of reward can be dissociated from decision making, and in the case of subjects who prioritize their own experience, also from the update of values for learning. Thus, the current study aimed to "examine the neural mechanisms involved in forming memories for information encoded during observational and active reinforcement learning, considering the propensity of different individuals to adopt diverse learning strategies".

We hypothesized that the relationship between reinforcement learning (RL) behavior and memory would be most pronounced during the observation of others' choices, as this condition dissociates the learning process from direct behavioral engagement. Within this framework, we predicted that the 'Social ignoring' group would show positive relationships between RL-based updating and memory formation during observation trials, as the reduced pressure to update expected value would free cognitive resources for acquiring new information.

We hope that this rationale is now better explained in the introduction and throughout the manuscript.

5) An example related to the point above pertains to the memory data analyses. Based on the authors' previous findings, the hypothesis should be that participants who prioritize their own experiences would display better memory for events associated with high reward values specifically; however, in their current analyses, the authors did not consider reward value (i.e. whether a trial was rewarded or not). If I understand their previous result correctly, memory performance should be higher for rewarded versus unrewarded trials and this effect might be stronger or only present in those individuals that learn from experience rather than observation (i.e. social ignoring group). Why wasn't the effect of reward outcome tested? Related to this, could this potential effect of reward underlie the confusing finding of shorter exposure time being associated with better memory performance? It could be that the expectation/anticipation of the reward makes participants be both faster and remember the picture better.

In our previous publication, participants who prioritize their own experiences displayed better memory for events associated with higher expectation of reward, as reflected in the trial value. This result was reproduced in the current study: As part of the revisions made to the manuscript, we have added analyses to test the interaction between trial-by-trial estimated expected values and memory performance. Indeed, we found a positive correlation between expected values and subsequent memory performance, specifically in the ‘Social ignoring’ group, and during observed trials (see Figure 3). This is also corroborated in the fMRI findings, showing parametric modulation of activation in vmPFC and extrastriate regions in this group corresponding to the evolution of expected value throughout the task. There was no effect of outcome (rewarded/non-rewarded) in either study.

6) My final major concern is that the authors do not appear to have made their data and code available online (or at least for the reviewers, with an explicit plan to share openly upon publication).

We are more than willing to share our data and code on a public server. Upon acceptance, we will organize the materials and upload them to an open access server.

Other comments, in order of text:

7) p.3: “the neural correlates underlying inferential social learning, i.e., updating our beliefs by gathering information from social evidence, remain largely unknown” – this is a bit misleading as there is a lot of literature on this topic (e.g. Hampton et al 2008 PNAS, Yoshida et al 2010 Journal of Neuroscience, Hill et al 2017 Nature Neuroscience, Collette et al 2017 eLife, Charpentier et al 2020 Neuron, Diaconescu et al 2020 eLife, Zhang & Glascher 2020 Science Advances, Konovalov et al 2021 Neuron, to just cite a few), so it would help to be a bit more specific in identifying what the gap in the literature really is that this study aims to fill.

We have now removed the statement regarding inferential social learning, as this is not included as a study aim. We now emphasize that the main goals of the current study are to examine the neural mechanisms involved in forming memories for information encoded during observational and active reinforcement learning, considering the propensity of individuals to adopt diverse learning strategies.

8) p.4: “observing the decisions of others would activate brain networks associated with social cognition” – on p.3 the authors seem to imply from previous literature that it is the mental representation of others’ intentions that primarily activates ‘social cognition networks’ – simply observing others’ decisions is not the same process as actively representing their intentions, so it would be great if the authors could disentangle those two processes.

The reviewer is correct in pointing out that observing others’ decisions is a distinct process from mental representation of others’ intentions. The current task was not designed to examine the latter. We have revised the introduction and the text throughout to ensure that we do not imply that participants can infer others’ intentions by observing their choices. However, we do mention that observational learning engages neural networks that support, among other things, making social inferences about the goals and intentions of others (p. 3, para. 4).

9) Was the bonus payment contingent on rewards earned or was it always 80 NIS regardless of performance? If the latter, did participants believe that it would be contingent on rewards earned when entering the study?

Participants believed they would get paid according to their performance, on top of the base payment of 120 NIS for their participation, where in practice all participants received the same payment of 200NIS.

10) Could you clarify the following details about the task design?

10a) How were the 50 experienced and 50 observed trials split? Were there two blocks or were trials interleaved? If there were two blocks, was block order counterbalanced across participants (and if so were there any order effects)? If they were interleaved, were they always alternating (one trial of each type) or randomly interleaved?

The order of experienced and observed trials was drawn randomly across trials. This is now mentioned in p. 6, para. 4 of the revised manuscript.

10b) How many congruent versus incongruent trials were shown?

The task included 40 congruent trials and 60 incongruent trials (now added to p. 6).

10c) What was the rationale behind simulating different accuracy rates for the second player's action between the congruent and incongruent trials? I understand that accuracy ought to be higher on congruent trials because of the 80:20 contingencies, but it doesn't seem very believable that the performance of another player would drop to chance when contingencies become 70:30 (what was used for incongruent trials).

Our rationale for this choice was to allow the participants to collect enough data points of each combination of choice X outcome in the observed trials to allow them, theoretically, to notice the reversed reward contingency. Furthermore, as figure 2A shows (presented also in response to comment 1a above), a substantial proportion of participants performed at chance or below chance level on incongruent trials, and therefore would not have had trouble believing chance performance by their partner.

10d) On a similar note, how did the authors determine the 70% and 50% of trials where the second player chooses accurately – were those trials randomly assigned or did they follow some sort of learning rule (since the second player is also supposed to be learning). Generally, it is a bit of a concern in this task design that the second player's simulated choices may not appear very realistic, and I wonder if participants may have picked up on that.

The percentages of correct answers for the second player were predetermined and identical for each participant. We believe that the probabilistic design, combined with the abovementioned probabilistic structure of the others' choices, created a high level of uncertainty, preventing participants from successfully tracking the dynamics of choices within 50 trials. In interviews performed with participants during previous pilot experiments (reported in Yifrah et al., 2021), they did not report noticing any suspicious behavior during the observed trials.

11) The authors appear to introduce the rationale for their study with previous literature on inferring others' intentions. However, nothing in the model is learning the other player's intentions... Instead, the model is simply trying to learn from the action-outcome associations regardless of whether it's the participant or the other player's turn. A social inference model in this task should be able to learn that the other player may have different goals and intentions from the participants (i.e. on incongruent trials, learning that the other player's goal may be opposite the participant's and as a result use this information in a productive manner – e.g. when the other player's chooses the blue fractal, I should choose the orange one because our goals are different). Overall, this comment also relates to points 4), 7) and 8) above, whereby the analyses do not appear well justified by the introduction and rationale – I would recommend either toning down the relevance to social inference processes in the introduction, or actually testing whether social inference is taking place with the appropriate models and analyses.

The reviewer is correct in stating that the current task is not designed to examine how individuals represent or learn the intentions of others. We in fact do not claim that participants learn the intentions of others, but that individuals can “learn about the value of stimuli not only through their own behavior, but also by observing the behaviors of others...” (p. 3). We have revised the introduction and the text throughout to ensure that we do not imply that participants can infer others' intentions by observing their choices. However, we do mention that observational learning engages neural networks that support, among other things, making social inferences about the goals and intentions of others (p. 3, para. 4).

Point-by-point answers to comments raised by reviewer 3:

We thank the reviewer for their helpful comments and suggestions. We have carefully studied the comments, and have implemented the requested analyses, revised the text accordingly, and amended the figures. We believe these changes have improved the manuscript significantly.

This manuscript has greatly improved, and I appreciate the new insights into memory-value interactions and overall clearer flow. However, some of my concerns were not totally addressed or some of the new analyses have raised new concerns. I have added more details below (the numbers/letters refer to my original points for clarity) and believe most of those should be addressed with minimal clarification or analyses.

As a side note, I will add that it was often quite hard to track where the authors made changes to the manuscript, and to avoid putting this additional burden on reviewers, I would encourage them in future manuscript revisions to always copy-paste in the rebuttals all the changes that were made to the manuscript text (as well as detailed page and line numbers), so the reviewers can directly see how changes to the manuscript addresses each of their comment.

We appreciate this comment, and have implemented the suggestion in the response letter.

1ab) The new best-fitting model involves separate learning rates for experienced and observational trials. However, no behavioral difference between the two conditions is reported (unless I missed them) – in what way is behavior (model-free learning performance) different between Exp and Obs and is that tendency captured by the difference in learning rates across participants?

Participants' behavior is naturally captured only in experienced trials. Observed trials play a role in learning, and thus have a latent role in the choices made by the participants in experienced trials. This can readily be seen in their choices on trials of the incongruent pair, where the behavior of the two groups differ (by definition, this is how the groups were determined, following reviewers' requests).

Also, the difference in learning rates (as seen in Fig 2E) appears to be primarily driven by the social ignoring group. Would the same model win in both groups? Because participants are split into two groups early on according to their behavioral performance, it would make more sense to fit the models separately to each group to assess not only parameter value differences but overall learning strategy differences between groups.

In accordance with the reviewer's comment, we compared the four models for each group separately. In both groups, the Exp/Obs two-alpha model showed the best fit to the participants' behavior. In the Social learning group, the Exp/Obs model was the winning model for 22 out of the 24 participants, with the remaining 2 participants preferring the one-Alpha model. In the Social ignoring group, all participants were assigned with the Exp/Obs two-alpha model.

This analysis is now added in the revised manuscript, p. 15, l.437:

“Since the group assignment was based on behavioral performance, we also inspected the model frequency for each group separately to test for differences in model assignment at the population level. Similar to the pattern observed in the entire sample, the two-alphas model (involving

separate learning rates for experienced and observed trials) was preferred in both groups (Learning group: one-alpha model = 0.084, two-alfas (exp/obs) = 0.913, and near-zero frequencies for the other models; Ignoring group: one-alpha model = 0.0130, two-alfas (exp/obs) = 0.984, with again near-zero frequencies for the other models).”

1d) The posterior predictive checks are a great addition, thank you for also adding model recovery. It would also be great to show parameter recovery, i.e. the correlation between simulated parameter values and recovered parameter values for the two alphas and softmax beta. It’s okay to only perform recovery for the best-fitting model, and show it in the supplementary materials to not perturb the flow of the paper.

As the reviewer suggested, we now show scatter plots depicting parameter recovery for the two_alfas Exp/Obs model, see Figure S3:

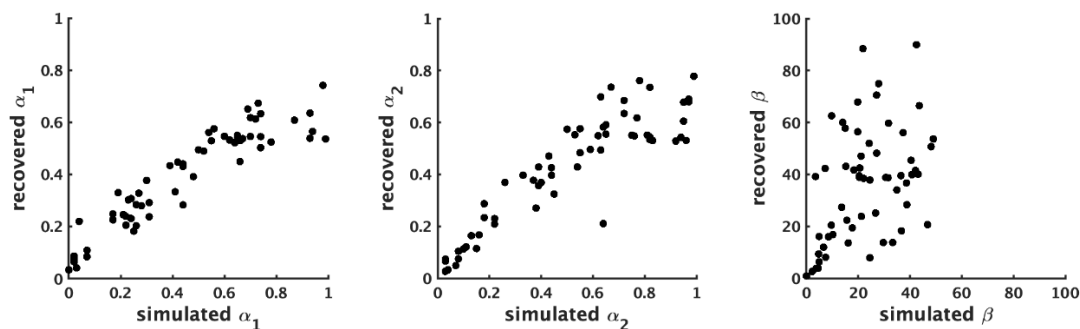


Figure S2. Parameter recovery for the twoAlphas Exp/Obs model. Simulated values, drawn from the range of actual parameters, vs. recovered parameters values, shown for (A) learning-rate 1 (experienced trials, $r = 0.925$), (B) learning-rate 2 (observed trials, $r = 0.896$), and (C) inverse temperature ($r = 0.440$).

2a) There are still some missing details about the second model, specifically what onset was used [for] the fractal phase. In the first model it is mentioned that the fractal phase onset was actually the onset of the trial type presentation – is that the same for the second model? If so that does not seem appropriate to track value estimation, which should only start at the time the fractals appear. Please clarify. Event durations for the second model are also missing.

We added the missing details to the Methods section, p. 11:

“The second GLM tested for neural activations associated with trial-by-trial expected values, as calculated from the Rescorla-Wagner model applied for each participant (see above). This GLM included separate regressors for experienced and observed trials during the stimulus-pair presentation phase, picture presentation (without division into subsequently remembered and non-remembered items), and feedback phase. As for the first model, the stimulus presentation phase onset was aligned with the time of trial type presentation and the durations were self-paced. The trial durations for the picture phase were self-paced as well, determined by button presses to advance the trial. Trial durations for the feedback phase were 1 second.”

The onsets of the stimulus presentation condition were aligned with the trial-type presentation, as this one-second phase could not be satisfactorily separated from the stimulus presentation phase itself. Since the presentation of the stimuli always succeeded the trial-type instructions, the BOLD responses were bound to convolve and overlap, rendering their separation into two conditions problematic. Considering the short duration of the trial-type display and the slow BOLD response, it is common to unite such events into one condition.

p.11, l.333-336: “Specifically, we tested for regions with stronger BOLD signals during experienced trials using the contrast experienced > observed trials, and for preferential activity during observation using the inverse contrast observed > experienced.” What GLM was used for this analysis (first or second?), and what onset/phase (trial type/fractal, picture or feedback)?

This contrast was conducted by using the first GLM, comparing activation during stimulus pair presentation. This is now stated in the revised manuscript:

pp. 11-12:

“To examine differences in neural activation during experienced choices versus the observation of others' choices, we performed two group-level contrasts using the first GLM. Specifically, we tested for regions with stronger BOLD signals during stimulus pair presentation in experienced > observed trials, and for preferential activity during stimulus pair presentation using the inverse contrast - observed > experienced. Statistical significance was determined using a threshold corrected for multiple comparisons, with a height threshold of $P < 0.05$ (family-wise error (FWE)-corrected) and a cluster size threshold of $k > 20$ voxels at $P < 0.05$ FWE-corrected.”

p.11, l.337: is the FWE correction whole-brain FWE? If so please add ‘whole-brain’ for clarification.

Yes, this is now mentioned in the text:

p. 12:

“Statistical significance was determined using a threshold corrected for multiple comparisons, with a whole-brain height threshold of $P < 0.05$ (family-wise error (FWE)-corrected) and a cluster size threshold of $k > 20$ voxels at $P < 0.05$ FWE-corrected.”

p.11, l.339, l.343 and l.347: Please clarify which GLM was used in each of these analyses.

This is now clarified in the text:

p.12:

“Using the same model (GLM 2), activations corresponding to parametric modulation of high-reward stimuli during picture presentation were examined by contrasting each of the parametric regressors (separately for experienced and observed trials) against baseline, using a statistical threshold of $P < 0.001$, FWE corrected for cluster size at $P < 0.05$.

We also conducted an ROI analysis based on the activations identified in the subsequent-memory contrast (GLM 2 - subsequently remembered > non-remembered trials during picture

presentation). For each region, we extracted parameter estimates associated with the high-value parametric regressors. Specifically, for each participant, the mean parameter estimate was calculated across all voxels within the cluster defined by the second-level analysis. We then performed a mixed-effects ANOVA on the parameter estimates to assess potential effects of group and condition (experienced vs. observed).”

Figure 4 and Table 1 mention time-locking to the presentation of fractal pairs. Are those images obtained with the first or second GLM? If first, the methods for that GLM state that the onset of the fractal phase is in fact the onset of the trial type, so this needs to be clarified. It would be stronger to show activation time-locked to the fractal presentation (than to the trial type presentation) to ensure that differences are driven by the condition and not by visual differences between the text presented on the screen.

Figure 4 pertains to the results of the analysis performed by using GLM 1. The onsets are the trial-type, and the events extend until the end of the fractal phase (see also response to comment above). This detail is now clarified in the text, Figure 4 legend, and Table 1 legend.

2c) If I understand the second model correctly, value regressors were only included at the time of the picture (and not at the time of the fractal presentation). In order to examine learning signals, expected value is likely to be computed during fractal presentation and prediction error should be the main signal at feedback. While it’s interesting to observe value signal during the picture presentation, it is unclear why the authors would expect value to only be represented during that event.

We agree with the reviewer that expected value is typically computed at the time of cue (fractal) presentation to guide decision-making, and that prediction errors are generated following feedback presentation. However, we were specifically interested in the factors that may influence memory encoding of the picture stimulus, which occurred after fractal presentation and prior to the feedback. At the time of picture presentation, participants anticipate the upcoming outcome, and this expectation is based on the value of the preceding cue. To better reflect this rationale, we have revised the text to clarify that the value regressor was designed to examine the anticipatory influence on encoding and subsequent memory performance:

P. 21 l.585:

“To explore subsequent memory-related activation patterns, we focused on the time of the picture presentation to capture effects on memory encoding, and conducted a GLM analysis contrasting trials in which the presented images were subsequently remembered against non-remembered image trials.”

P. 23 l.615-620:

“At the time of picture encoding, prior to the feedback, subjects are engaged in reward expectation, according to the expected value. To examine whether the expected value of cue-related outcomes interacts with neural activation during memory formation, we further tested for neural activation during image presentation associated with trial-by-trial parametric modulation of expected value, using expected value estimates derived from individual model parameters.”

The authors state in their response that “brain regions that were selectively activated during the presentation of images that were subsequently correctly recollected were modulated by Q-value, mirroring our behavioral findings”.

First, it is unclear whether the value signals shown in Fig 5C for Obs are significant, since the only difference that’s reported is the p-value of the main effect of condition. The authors should show that there is significant value representation ‘on its own’ and not solely as a difference from Exp. Full statistics and effect sizes should be reported for this analysis, not just a P-value in the figure legend.

The so-called mirroring of the behavioral finding does not seem to be accurate, since the authors themselves describe in their discussion that “while the behavioral effect of trial value on memory formation was restricted to the ‘Social ignoring’ group, the accompanying neural activation during observed trials was apparent in both groups” (1.715), thus suggesting more of a discrepancy than a mirroring.

The reviewer is correct in that the sentence we wrote in the rebuttal letter was not accurate. As the reviewer indicates, whereas the behavioral effect of expected value on subsequent memory was emphasized in observed trials within the Social ignoring group, the fMRI results show that value modulated subsequent memory in both groups.

As suggested by the reviewer, we now report full statistics for each ROI, including mean beta values, test statistics, and effect sizes from one-sample t-tests conducted separately for each condition (Exp/Obs) across participants. While the parameter estimates in the Experienced condition are not significantly different from zero in any of the ROIs, all ROIs show significant value-related activation in the Observed condition. Since we did not detect interaction effects in the original ROI ANOVA tests, we did not carry out single-group comparisons.

p. 23-24, starting from line 626:

“Four regions of interest (ROIs) that were detected in the subsequent memory contrast were subjected to further analysis, by extracting from each ROI the mean beta value of the parametric modulation regressors, separately for each group and for experienced and observed trials (Fig. 5C). All ROIs showed a main effect for trial type, such that activation was stronger during observed vs. experienced trials: left middle occipital gyrus - $F(1,45)=6.956$, $P<0.01$; left lateral occipital cortex - $F(1,45)=8.582$, $P<0.005$; left superior occipital gyrus - $F(1,45)=4.779$, $P<0.05$; right superior occipital gyrus - $F(1,45)=3.229$, $P=0.079$. No between-group or group x trial type interactions were found in this analysis. Within these four ROIs, we tested whether the parameter estimates (beta values of the parametric regressor for high-reward values) differed from baseline, separately for experienced and observed trials. One-sample t-tests yielded significant effects for observed trials across the entire sample in all four ROIs: left middle occipital gyrus - $t(46)=2.895$, $P<0.01$, Cohen’s $d=.366$; left lateral occipital cortex - $t(46)=3.995$, $P<0.0001$, $d=.298$; left superior occipital gyrus - $t(46)=2.503$, $P<0.05$, $d=.270$; right superior occipital gyrus - $t(46)=3.656$, $P<0.001$, $d=.326$. Except for the left superior occipital gyrus, the effects remained significant after Bonferroni correction for multiple comparisons. In contrast, the parameter estimates did not differ from baseline in experienced trials: left middle occipital gyrus -

$t(46)=-.185$, $P=.854$, $d=.362$; left lateral occipital cortex - $t(46)=-.483$, $P=.631$, $d=.344$; left superior occipital gyrus - $t(46)=-.422$, $P=.675$, $d=.285$; right superior occipital gyrus - $t(46)=1.438$, $P=.157$, $d=.295$.”

2d) The authors state in the rebuttal that “using the new group assignment, we did not find significant differences between remembered and non-remembered trials across groups and conditions”. Figure 3C legend still reports a significant difference in RT between experienced and observed trials. Please clarify.

We are grateful to the reviewer for noticing this, we accidentally neglected to update the legend to Figure 3. The figure legend now reads:

“Figure 3. Memory performance of picture recognition. (A) Memory strength (d') was comparable on average across groups, conditions, and pairs. (B) No significant differences were found in confidence levels among groups, pair type or trial type. (C) Mean reaction times for button presses to reveal the trial outcome upon picture presentation in each trial. RT's did not differ between groups, yet a main effect was found for trial type indicating longer RTs for observed vs. experienced trials across groups and memory success ($p < 0.001$).”

Nevertheless, as indicated in the rebuttal letter, we did not find significant differences between remembered and non-remembered trials across groups and conditions, only differences in RTs between experienced vs. observed trials across the other conditions.

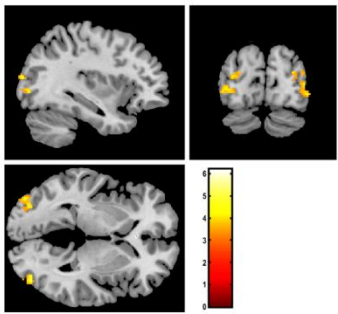
2e) The authors state in the rebuttal that “As suggested by the reviewer, we examined the activation to image presentation separately for each group in remembered vs non-remembered trials and showed that in observed trials the regions revealed by this contrast are modulated by trial value (see above).” While that analysis is interesting, it is not the analysis I suggested, which would be to model an interaction between subsequent memory (remembered > non-remembered) and trial type (observed > experienced), or at the very least visualize the ‘subsequent memory’ map separately for observed and experienced trials.

Thank you for this clarification. In the original comments (2e), the reviewer raised the “possibility that brain activity is in general stronger during observed than experienced trials, regardless of the region”.

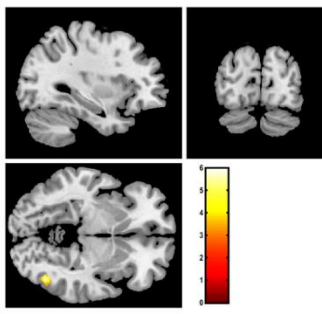
Running separate subsequent memory contrasts for experienced and observed trials showed broader activation profiles for the experienced trials, and rather limited activations in observed trials ($P<0.001$, uncorrected). Thus, the observation that the parametric modulation of max value was more pronounced during observed vs. experienced trials does not seem to stem from overall heightened activation for remembered vs, non-remembered events during observed trials.

Subsequent memory effect (remembered > non-remembered)

Experienced trials



Observed trials



Statistics: *p*-values adjusted for search volume

set-level		cluster-level					peak-level					mm mm mm							
p	c	$P_{FWE-cor}$	$q_{FDR-cor}$	k_E	p_{uncor}	$P_{FWE-cor}$	$q_{FDR-cor}$	T	(Z_p)	p_{uncor}									
0.000	3	0.000	0.000	270	0.000	0.018	0.038	6.16	5.23	0.000	42	-80	-2						
							0.074	0.064	5.75	4.97	0.000	34	-92	18					
							0.980	0.495	4.57	4.13	0.000	40	-88	10					
							0.862	0.447	4.82	4.32	0.000	-24	-92	30					
							0.970	0.495	4.61	4.16	0.000	-32	-86	16					
0.000		112	0.000			0.000	1.000	0.495	4.21	3.85	0.000	-24	-86	18					
												0.994	0.495	4.47	4.05	0.000	-34	-82	0

Statistics: *p*-values adjusted for search volume

set-level	<i>p</i>	<i>c</i>	cluster-level				peak-level						mm mm mm		
			<i>P</i> _{FWE-cor}	<i>q</i> _{FDR-cor}	<i>k</i> _E	<i>p</i> _{uncor}	<i>P</i> _{FWE-cor}	<i>q</i> _{FDR-cor}	<i>T</i>	(<i>Z</i> _{<i>p</i>})	<i>p</i> _{uncor}				
			0.002	0.001	72	0.000	0.037	0.044	5.95	5.10	0.000	46	-66	-6	
							1.000	0.909	3.60	3.36	0.000	40	-66	0	

4) I have re-read the introduction in detail and still do not see where the hypotheses are specified. I think that 1.112-129 should not be in the introduction as those lines are summarizing the results, and would be better placed in a ‘conclusion’ section at the end of the discussion. Instead, the end of the introduction should detail the hypotheses and the rationale behind them, similar to what the authors did in the rebuttal, although I wouldn’t necessarily frame the hypotheses in terms of social learning and social ignoring groups since those groups were not pre-determined and instead defined based on the data.

We appreciate this comment, and have added some text detailing the hypotheses and rationale of the current study.

We removed the results summary from the end of the introduction section, as they are stated also in the beginning of the discussion.

p. 5, 1.106-128:

“The current study aimed to examine the neural mechanisms involved in forming memories for information encoded during observational and active reinforcement learning, considering the propensity of different individuals to adopt distinct learning strategies. Participants underwent fMRI scans while engaging in a reinforcement learning task that encompassed both active decision-making trials (experienced trials) and observed trials, the latter featuring a counterpart with cue-outcome contingencies that were either congruent or incongruent with their own experiences. Each trial additionally incorporated a distinct pictorial item, subsequently presented in a recognition test conducted outside the scanner.

Participants were assigned to one of two distinct groups, labeled as ‘Social learning’ and ‘Social ignoring’ groups, according to their performance in the reinforcement learning task, reflecting the extent to which they learned from observed trials. Applying a reinforcement learning model to the

subjects' choices confirmed that subjects of the 'Social learning' group displayed similar learning rates for experienced and observed trials, whereas those of the 'Social ignoring' group had significantly higher updating rates for their own actions.

Our main aim was to examine the neural underpinnings of RL-memory formation interactions, focusing on the time of cue choice and of the ensuing picture presentation. We hypothesized that the relationship between RL behavior and memory would be most pronounced during trials that involve the observation of others' choices, as this condition dissociates the learning process from direct behavioral engagement. On the neural level, we hypothesized that brain regions associated with social cognition would activate during the observation of others' choosing among presented cue options, and that the fMRI signal of regions associated with subsequent memory of pictorial information would be mediated by trial-by-trial updating of cue value memory encoding."

5) The following analysis might help better relate the value-memory effects from behavior to brain. Instead of the binning procedure the authors used, a more statistically valid approach for assessing the behavioral effect of value on memory would be to perform a trial-by-trial mixed effect logistic regression, where the dependent variable on each trial is whether the picture will be remembered (coded as 1) or not (coded as 0) and the predictors are cue expected value, condition, group and their interaction (with at least a random intercept and maybe some random slopes if the covariance structure allows it). This framework would allow better parsing of variance in the data.

We thank the reviewer for this suggestion. As suggested, we performed a trial-by-trial linear mixed-effects regression, with subsequent memory success as the dependent variable (0/1), and Q-value, trial type (experienced/observed), group (learning/ignoring) as predictors, and a random intercept for subject. This approach allowed us to account for within-subject variability and directly assess how Q-values predict memory across conditions and groups. The results showed a significant interaction effect in the Social ignoring group, indicating that trial-by-trial q-values related differently to memory performance for observed vs. experienced trials. This finding coincides with our previously reported effects.

To assess whether decision value (Q-value) influenced subsequent memory differently for experienced versus observed trials, we conducted separate generalized linear mixed-effects models (GLME) for the 'Social learning' and 'Social ignoring' groups. Each model predicted trial-by-trial memory success (0/1) as a function of Q-values, trial type (experienced vs. observed), and their interaction, while including random intercepts for subjects to account for individual variability. In the 'Social learning' group, the interaction between Q-values and trial type was not statistically significant ($\beta = 0.053$, $SE = 0.086$, $t=0.622$, $P=0.533$), indicating that the relationship between value and memory did not differ across trial types. In contrast, the Social ignoring group showed a significant Q-value $\times$ trial type interaction ($\beta = 0.192$, $SE=0.092$, $t=2.074$, $P=0.038$), suggesting that the influence of value on memory is differently modulated during observed trials compared to experienced ones in this group.

This analysis is described in the Methods section, p. 9, starting from l.270:

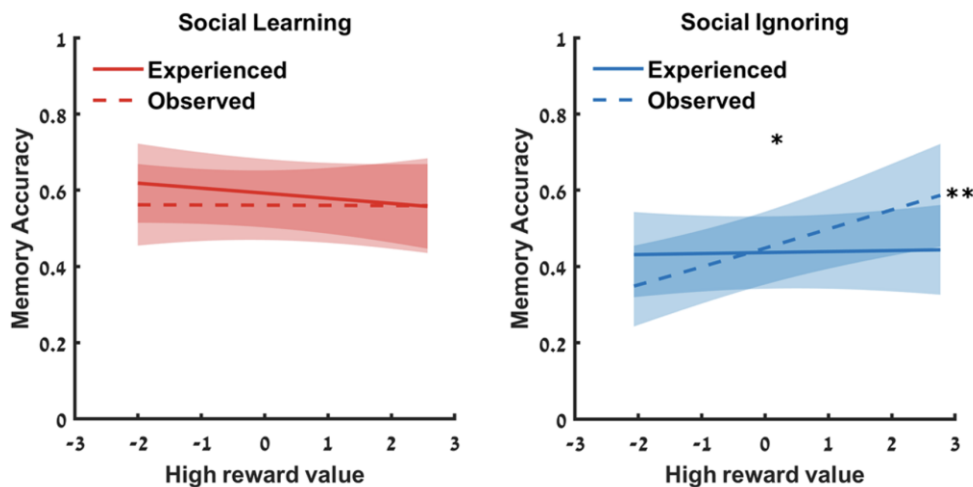
"To examine the relationship between high reward value and memory performance, we conducted generalized linear mixed-effects models (GLME) separately for the Social learning and Social ignoring groups. In each model, trial-by-trial memory success (coded as 0 = forgotten,

1 = remembered) was predicted by the normalized high reward value, trial type (experienced vs. observed), and their interaction. Subject-specific random intercepts were included to account for individual variability. GLME models were fit using a binomial distribution with a logit link function. The Models were implemented in MATLAB R2023b using the ‘fitglme’ function.”

The results are detailed in p. 18, starting from l.525:

“In attempt to examine the possible relationship between trial-by-trial model values and corresponding subsequent memory performance in each of the groups, we conducted generalized linear mixed-effects models (GLME) for the Social Learning and Social Ignoring groups. Each model predicted trial-by-trial memory success (0/1) as a function of normalized high-reward values in each trial, trial type (experienced vs. observed), and their interaction, while including random intercepts for subjects to account for individual variability. In the Social Learning group, the interaction between high reward values and trial type was not statistically significant ($\beta = 0.053$, $SE = 0.086$, $t=0.622$, $P=0.533$), indicating that the relationship between value and memory did not differ reliably across trial types. In contrast, the Social Ignoring group showed a significant cue value by trial type interaction ($\beta = 0.192$, $SE=0.092$, $t=2.074$, $P=0.038$), suggesting that the influence of value on memory is differently modulated during observed trials compared to experienced ones in this group (Fig. 3D). Simple effects analyses revealed that reward value significantly predicted memory performance only in the observed condition of the Social ignoring group ($P=0.0053$), while no significant effects were found in the other conditions ($P_s > 0.1$).”

Figure 3D:



Legend to Figure 3D:

“Predicted memory accuracy as a function of reward value, plotted separately for each group. The lines depict logistic regression fits for the effects of reward value on subsequent memory performance within each condition; solid and dashed lines denote experienced and observed trials, respectively. Shaded areas represent ± 1 standard error of the estimates. A significant interaction effect was found in the Social ignoring group for condition by memory (denoted by the asterisk, $*P < 0.05$), and the slope of the regression line for the observed condition was significant ($**P < 0.01$).”

6) Details plans for sharing the data and code should be mentioned and links made available in the manuscript. Code and materials should be well organized prior to submission, not upon acceptance, in order to minimize the risk of errors in the analysis pipeline and to ensure reproducibility.

We fully agree with the importance making the codes and materials available.

To that end, we have created an OSF project that contains the complete behavioral data, Matlab scripts for analyzing the behavioral data (including choice modeling and participant-level parameter estimation), and the computations and code used for creating the graphs and figures presented in the manuscript.

The repository also contains second-level fMRI contrast maps, which can be inspected in SPM12, as well as mat files containing the coordinates of the regions-of-interest (ROIs) reported in the paper.

Two documentation files -readme_behavior.txt and readme_fMRI.txt - are provided as well to guide users through the contents of the scripts and data.

The link to view the files uploaded to the repository is:

https://osf.io/8rcus/?view_only=9445a72998654a61b77b6a11752169b9

We will make the repository public upon acceptance of the manuscript.

We are currently examining the feasibility of sharing the raw fMRI data of all participants as well, in terms of the ethics approval and online storage.

Minor comments and typos:

- The name 'pairwise' for the model with separate learning rate for incongruent vs congruent pairs is confusing as in the logic of 'oneAlpha', 'twoAlphas', 'fourAlphas' and 'pairwise' it would suggest more than four alphas. It would make more sense to rename the 2 twoAlpha models something like 'twoAlphas(exp/obs)' and 'twoAlphas(cong/incong)' or something along those lines.

This suggestion was incorporated in the text and in Figure 2F.

- Figure 2A x-axis should say 'Incongruent' (not 'Icongruent')

- Figure S2A title should say 'Experienced' (not 'Experinced')

Thank you, these typos were fixed.

- Figure 3B legend states that "The 'Social Ignoring' group (blue shades) exhibited significantly lower confidence ratings (shown in z-scores) during the recognition test in comparison to the 'Social learning' group (red shades)". This is not evident at all from the graph, which appears to show no differences between groups or conditions.

The reviewer is correct. As mentioned in the above comment, the legend to Figure 3 was mistakenly left from the original manuscript version. This has been corrected now, stating that:

(B) No significant differences were found in confidence levels among groups, pair type or trial type.

Differences in RT in panel 3C are also not evident.

There actually are significant effects here, as now indicated in the legend:

“Mean reaction times for button presses to reveal the trial outcome upon picture presentation in each trial. RT’s did not differ between groups, yet a main effect was found for trial type indicating longer RTs for observed vs. experienced trials across groups and memory success ($p < 0.001$)”.

In panels 2E, 3A, 3B, and 3C, it would be great if the authors could highlight significant effects of interest for the reader.

Figures 3A and 3B don’t show significant effects.

Asterisks were added to Figures 2E and 3C.

- Figure 3B legend, 1.528, in the phrase ‘measured as proportion of correct responses to presented images – hit-rates’, is the ‘–’ a minus? If not I would rephrase as ‘hit rate, measured as proportion of correct responses to presented images’.

This is no longer relevant.

- In Figure 3D-G, is each dot a participant? According to the methods, it seems like each dot is a bin of high-reward value. This should be made clear in the Figure legend, though probably the mixed-effects logistic regression approach suggested above (point 5) would be a better way to analyze value effects on memory.

This is no longer relevant.

- Figure 5C is missing a legend for the blue/red colors

Corrected.