

Experience and advice consequences shape information sharing strategies

Corresponding Author: Dr Hernan Anllo

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Version 0:

Decision Letter:

**** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors ****

Dear Dr Anllo,

Thank you for your patience during the peer-review process. Your manuscript titled "Experience and advice consequences shape information sharing strategies" has now been seen by 3 reviewers, and I include their comments at the end of this message. They find your work of interest but raised some important points. We are interested in the possibility of publishing your study in Communications Psychology, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response to the reviewers. Please highlight all changes in the manuscript text file.

Editorially, we emphasize the need to address the referees' methodological requests, including the requested robustness checks, in addition to their presentational concerns. As for the interpretational issue highlighted by Reviewer #3, please bear in mind that statements about the absence of an effect in Communications Psychology require positive evidence for the null, derived from Bayesian statistics or equivalence tests.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

I am attaching an Editorial Requests Table that details critical reporting requirements for the revised manuscript. Please attend to each item and ensure your manuscript is fully compliant. We are requesting that your manuscript aligns with these requirements as this facilitates the evaluation of your manuscript, reducing delays in re-review and potential future acceptance. If your revised manuscript is not aligned with these requests on major issues, such as those concerning statistics, it may be returned to you for further revisions without re-review. Additional information can be found in our style and formatting guide <https://www.nature.com/documents/commspsychol-style-formatting-guide-accept.pdf> Communications Psychology formatting guide.

Please use the following link to submit your

- revised manuscript,
- point-by-point response to the referees' comments,
- cover letter (as a separate document),
- the Editorial Policy Checklist (see below),
- the Reporting Summary (see below), and
- the completed Editorial Request Table (attached):

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to receive your revised paper within 8 weeks; please let us know if you aren't able to submit it within this time so

that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, provided it still presents a significant contribution to the literature at that stage.

We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Neil Garrett

Neil Garrett, PhD
Editorial Board Member
Communications Psychology
orcid.org/0000-0003-1440-472X

REVIEWER EXPERTISE:

Reviewer #1: Social decision making, reinforcement learning

Reviewer #2: Social learning, computational modelling

Reviewer #3: Social decision making, computational modelling

REVIEWER REPORTS:

Reviewer #1 (Remarks to the Author):

The present study investigated a case in which participants decided whether or not to share their choices as advice to others. The authors found that a significant portion of the participants shared their choices as advice even when they did not learn anything about the environment. Moreover, the tendency to share advice was modulated by their expertise (i.e., the progress of the learning), task difficulty, individual differences in social anxiety, and other factors.

I believe this study addressed an important issue with a pre-registered experimental design and data analysis plan. While a growing body of literature has investigated advisees' behavior, few studies have assessed advisors' behavior. The present study, therefore, can provide significant insights into social psychology.

My comments below are mostly about computational modeling and interpretation of the data.

I would like to request that the authors examine to what extent the model fitting procedure captures the true $p(\text{Choice})$ and PE given the current experimental setting and RL model, by using simulation. That is, I would first simulate an RL agent in the experimental task. Notably, in the simulated data, I am able to know the TRUE $p(\text{Choice})$ and PE. I would then fit the RL model to the simulated data and obtain the estimated learning rate and inverse temperature. Finally, I would compare the trial-by-trial TRUE $p(\text{Choice})$ and PE with those estimated from the fitted learning rate and inverse temperature.

It is worth testing other models for a robustness check. I would suggest RL models with asymmetric learning rates, with perseveration, and with both asymmetric learning rates and perseveration. I do not insist that these are alternative better models, but it is useful to evaluate the robustness of the main findings in this study.

Having said that, I would like to note that the current description of computational modeling is unclear.

- Why is the learning rate denoted as α in the main text and ϕ in the supplementary information (SI)? Why is the inverse temperature depicted by β in the main text and λ in the SI?

- For MCMC, what prior was used? How did they evaluate the convergence of the sampling processes?

- I believe the authors obtained the joint posterior distribution (learning rate x inverse temperature). How did they obtain the point estimates of the learning rate and inverse temperature from the 2D posterior distribution?

- What was WAIC computed for? The authors did not conduct any model comparison. I find no reason to compute WAIC in this study.

- As far as I understand, the RL model is fitted to each participant's data separately. What are the random effects?

It would be better if the authors discussed potential interpretations of the observed effect of trial number. I guess that not only expertise but also many other factors changed with time.

I would like to know the strength of the association between advising behavior and SPIN (beyond statistical significance). I would like to ask the authors to report the effect size.

It would be great if the authors explained why they included the previous PE in the second regression model. I am especially wondering why PE was included rather than the outcome (i.e., reward).

Reviewer #2 (Remarks to the Author):

The study titled "Experience and advice consequences shape information sharing strategies" by Hernán Anlló and colleagues aimed to explore how individuals decide to give advice within a learning context, particularly focusing on the roles of accumulated knowledge, social anxiety, and prosocial tendencies, as well as the impact of different consequences (monetary costs, reputational stakes) on advising behavior. Using a reinforcement learning framework, the researchers found that individuals were more likely to share advice as they gained experience and moved from exploratory to exploitative decision-making. Additionally, the introduction of consequences significantly influenced advice-giving: monetary costs reduced it, while reputational stakes increased it. Socially anxious individuals were cautious about giving advice during uncertain, early stages but more likely to advise once confident in their choices. Prosocial individuals were generally more inclined to give advice, especially under reputational conditions. The study concluded that advice-giving is a complex behavior influenced by a combination of personal experience, psychological traits, and external incentives.

The paper is well written and addresses a significant gap in the literature. I have a few concerns regarding the modeling framework and the interpretation of the results.

MAJOR Criticisms:

1. Reputation feedback: The absence of real-time feedback on the adviser's reputation means participants were not influenced by immediate consequences of their advice on a trial-by-trial basis. This could have impacted their decision-making process, as they did not experience the direct and immediate repercussions of their advice on their reputation. Real-time feedback could potentially lead to different advising behaviors, as participants would continuously adjust their strategies based on the immediate feedback received after each trial. This may be acknowledged in the Discussion of the results.

2. The impact of uncertainty on decision-making: While the RL model in the study effectively captures the overall dynamics of learning and decision-making under uncertainty, it does not explicitly differentiate between informational/epistemic and irreducible uncertainty. Participants' decisions to give advice might be influenced differently by their ability to reduce uncertainty through learning (informational uncertainty) versus their strategies for handling inherent randomness (irreducible uncertainty). I was wondering if the authors considered comparing this RL model to Bayesian models that explicitly differentiate the roles of uncertainty in decisions to offer advice.

3. Over-Advising Complications: A significant portion of participants engaged in over-advising (giving advice indiscriminately), which complicates the interpretation of the data. How did the authors debrief participants, and to what extent was over-advising driven by specific task instructions?

4. Parameter recovery and test-retest reliability: Two parameters were central to this model: (i) Learning Rate: This parameter reflects how quickly participants adjusted their expectations based on new information, and (ii) Inverse Temperature: This parameter indicates the consistency of participants' choices, acting as a measure of decision-making stochasticity. Were these parameters well recovered from simulations (i.e., if you simulated behavior based on the empirical estimates of these parameters, did you recover the same values? Typically, intraclass coefficients are computed to evaluate this). Since parameter recovery is a lower bound on test-retest reliability, have the authors also explored whether both the key behavioral variables as well as the parameter estimates were reliable when testing the participants a second, third time, etc.? This would also clarify whether the cluster of over-advisers is a reliable, stable effect as well.

MINOR Criticisms:

1. It would be helpful to use exact p values when reporting the results.
2. It is unclear whether the study was adequately powered for the mixed-effects linear and logistic regression models. Perhaps this could be clarified.

Reviewer #3 (Remarks to the Author):

(First of all, I apologize for the delayed response due to teaching commitments and conference travels.)

Anlló and colleagues examined the question of whether information sharing (i.e., advice) is shaped by reinforcement learning processes (Study1) and how such relationship is modulated by advice consequences (Study2; ie, whether sharing information is costly, or is related to participants' reputation/responsibility). Across both studies, individual differences associated with social anxiety and prosociality are also examined. With appropriate statistical analyses and computational modelling, the authors reported that participants were indeed willing to share information on average, and the degree to share information is linked to knowledge of the task (as reflected in better choice accuracy, and computationally derived latent variables). Advice consequences modulated information sharing, and both social anxiety and prosociality played some role in modulating the extent of information sharing.

The manuscript is overall well-written and easy to follow, and the conclusions are supported by the data analysis approach and the results. Some aspects, however, could be unpacked, improved or clarified. I list my detailed comments below and hopefully they will help revise the paper.

(ps, It would be helpful if the authors could include page numbers and line numbers in their manuscript in the future, which will make reviewers' lives a lot easier. Thank you.)

(1) Despite a generally good introduction, it lacks detail regarding social anxiety, prosociality, and advice consequences. Indeed, all aspects are the main focuses of the study and are preregistered. But among other potentially interesting traits, why did the current study focus on social anxiety and prosociality? What are the justifications? Some brief background literature would be helpful. Similarly, the choice of advice consequences (costly, regulation, responsibility) among other possible consequences are not motivated/justified by literature.

(2) In Study 1, and also "no cost" condition of Study 2, how exactly were participants instructed regarding their advice sharing? Were all participants well-informed with their bonus was unrelated with their decision of broadcasting? It is vital, if participants linked the results of advice broadcasted with incentives, than it was merely a non-social reward learning game. This might also explain the "over-advising" pattern observed in the task. It would be helpful to clarify this.

(3) A few questions regarding the RL modeling:

(3a) How was the 'outcome' coded in the model? Is it 0/1 for no-win/win, or is it the actually amount (e.g., 6) used? This is important and relevant for the initialization of Q values. I checked the code on OSF, and it seems that Qs are initialized to be 0.5. This is fine if 0/1 were used for outcomes, but this might not accurately reflect the task if actual amounts were used.

(3b) Related to "Our models included within-subjects effects related to the task, [...], and effects that varied between participants [...]". I looked at the Stan model code on Github, but I cannot find how this covariates were included. Also, it would be very helpful to report the parameter results (learning rate, temperature) related to the task design and individual differences. Especially I am curious to see if there is any relationship between the inverse temperature parameter and SPIN and PTM, as the authors were indeed trying to interpret some of the results in the vein of exploration v. exploitation. (see also my comment #4).

(3c) In SI, the description of computational modeling seems to be grabbed directly from elsewhere. I have no problem with this practice per se. But as this study did not include any alternative models, also as the authors clarified, it is not necessary to include any description of WAIC or other model evaluation related information.

(3d) It might be helpful to consider some work on RL in social decision-making, e.g., Olsson et al., 2020; Zhang et al. 2020; Lockwood & Klein-Flügge, 2020.

(4) Results around SPIN traits are very interesting, but some aspects are somewhat confusing.

(4a) I am not sure why PE is also included in the GLM. In the formula, before Section 2, PE is included to test H1.3; but judging from H1.3, I am not sure what role is PE playing here. Regardless, now that PE is tested, I wonder why there is no interaction term with SPIN/PTM?

(4b) The interaction between SPIN and P(choice) seems to be one of the most interesting results in this work. But I do not fully agree with the interpretation of "[...] probability to give advice increased with SPIN (fear) score for exploitative trials (i.e., later trials)". Indeed, judging from the data, individuals with high SPIN tend to give more advice when P(choice) is high; but it is necessarily linked to exploitative trials (later trials)? This interpretation relied heavily on the assumption that in later trials P(choice) is high – this might be true, but not obvious. I wonder if trial number can be included in this GLM? Or alternatively, can there be a way to better define/differentiate explorative and exploitative trials? This will make the interpretation stronger.

(4c) Relatedly, the PE results on advice-giving are somewhat counterintuitive. The data says when PE_{t-1} is high, participants are more likely to share advice. But does this symbolize an explorative advising strategy: supposedly, in earlier trials, participants hadn't learnt well, so they had higher PEs, and then in later trials, PEs were reduced. So, is it inconsistent with the finding that individuals tend to advise more in exploitative trials? As I commented above, it might be interesting to see the interaction between PE and SPIN, pending PE should indeed be included. And as a side note, Q and PE are negatively correlated by nature in RL models, so I also wonder if there may be a collinearity problem.

(5) For an interpretation of Study 3, "In the responsibility consequence condition, we observed that advice and performance results were very similar to the No cost condition, which may contribute to the idea that responsibility may be implicitly driving advising even when not explicitly mentioned." This is true in some aspects, however, it is not very convincing in another aspect, e.g., the ratio of over-advising, and the asymmetry between the relationship between PTM and advice calibration. Thus, it is reluctant for readers to be convinced that they are "very similar", because if advising is largely driven by responsibility, then there would not be so many over-advisors. Besides, based on this similarity to point out the idea that responsibility driving advising is not reasonable. Also, relevant statistical results showing non-significant differences between these two conditions were not provided.

Minor:

- Introduction: "[...] will often rely more on advice than even their own experiential information". This might not be always true,

as a large body of literature has shown that advisees tend to weigh their own beliefs over external advice (i.e., the well-known 'egocentric bias', Morin et al., 2021; Rader et al., 2017; Yaniv et al., 2000).

- The notion of the softmax inverse temperature is not consistent between the main text (called beta) and the SI (called lambda).

- Fig1B, how did the "reputation" task work? Why the adviser is rated by 3 people? This is not explicit in the description.

References:

- Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour—a reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, 16(8), 761-771.
- Morin, O., Jacquet, P. O., Vaesen, K., & Acerbi, A. (2021). Social information use and social information waste. *Philosophical Transactions of the Royal Society B*, 376(1828), 20200052.
- Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, 21(4), 197-212.
- Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass*, 11(8), e12329.
- Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, 83(2), 260-281.
- Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, 15(6), 695-707.

EDITORIAL POLICIES

We ask that you ensure your manuscript complies with our editorial policies and reporting requirements.

To that end, we require revised manuscripts to be accompanied by two completed items: a reporting summary that collects information on study design and procedure, and an editorial policy checklist that verifies compliance with all required editorial policies.

- <https://www.nature.com/documents/nr-reporting-summary.zip>>Nature Research Reporting Summary
- <https://www.nature.com/documents/nr-editorial-policy-checklist.pdf>>Editorial Policy Checklist

All points on the policy checklist must be addressed. Your revised manuscript can only be sent back to the referees if these checklists are completed and uploaded with the revision.

Notes: If you have submitted a Stage 1 Registered Report, Review, Primer, Comment, or Perspective you do not need to submit these forms. If you have already submitted these forms, you may disregard this request.

*** TRANSPARENT PEER REVIEW:** Communications Psychology uses a transparent peer review system. This means that we publish the editorial decision letters including Reviewers' comments to the authors and the author rebuttal letters online as a supplementary peer review file. However, on author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If your manuscript has been previously reviewed at another journal, those Reviewers' comments would not form part of the published peer review file.

**** Visit Nature Research's author and referees' website at** <http://www.nature.com/authors>**>www.nature.com/authors for information about policies, services and author benefits****

Communications Psychology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the

instructions in the link below. Please also inform all co-authors that they can add their ORCID to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the [Platform Support Helpdesk](http://platformsupport.nature.com/).

Version 1:

Decision Letter:

**** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors ****

Dear Dr Anllo,

Thank you for your patience during the peer-review process. Your manuscript titled "Experience and advice consequences shape information sharing strategies" has now been seen again by 3 reviewers, and I include their comments at the end of this message.

You will see that the reviewers find your revisions have mostly addressed their concerns. Editorially, we ask that you address the remaining reviewer concerns, in particular those of Reviewer 1 who has asked if you could include some further clarifications in the manuscript around the modelling by providing some formal description (ideally with equations) with regard to: The priors used, Noise variables and The Hierarchical Model.

We are very interested in the possibility of publishing your study in Communications Psychology, but would like to consider your responses to these final concerns before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response to the reviewers. Please highlight all changes in the manuscript text file.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your

- revised manuscript,
- point-by-point response to the referees' comments,
- cover letter (as a separate document):

Link Redacted

**** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first ****

We hope to receive your revised paper within 8 weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, provided it still presents a significant contribution to the literature at that stage.

We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Neil Garrett

REVIEWER EXPERTISE:

Reviewer #1: Social decision making, reinforcement learning

Reviewer #2: Social learning, computational modelling

Reviewer #3: Social decision making, computational modelling

REVIEWER REPORTS:

Reviewer #1 (Remarks to the Author):

I appreciate the authors' efforts to revise the manuscript in response to my concerns. However, the description of the computational modeling remains unclear, and certain parts are even contradictory. I would like to request that the authors include the mathematical descriptions (i.e., equations) to clarify the following points.

> We used non-informative priors, where the learning rate was taken as the logit of a parameter α_{logit} with a prior of Gaussian ..., and the beta was an exponential of a parameter β_{exp} with a prior of Gaussian...

Logit or exponential transformations of the Gaussian are not "non-informative." Please double-check the definition of the non-informative prior.

> For individual-level noise variables, we used a Gaussian prior with a mean of 0 and a sigma of 1.

It is unclear what the "individual-level noise variables" refer to, as I could not locate relevant information in the manuscript.

The authors stated that a hierarchical modeling approach was used, but the structure remains difficult to understand. For instance, are the parameters, α_{logit} and β_{exp} , individual-level or group-level parameters? How are the individual-level parameters assumed to be drawn from their population-level parameters? Where and how is the Cholesky decomposition applied within the hierarchical model?

There are many ambiguities in the current description of the computational modeling. Once again, I request that the authors provide the mathematical descriptions (i.e., equations) of the hierarchical model (e.g., <https://www.princeton.edu/~ndaw/d10.pdf> and <https://osf.io/preprints/psyarxiv/vj2wp>).

Reviewer #2 (Remarks to the Author):

The authors have adequately addressed all of my comments.

Reviewer #3 (Remarks to the Author):

The authors have done a very thorough job at addressing my previous comments. They appear to explain the literature much clearer, make the results stronger, and support the conclusion and interpretation better. I have some additional questions for clarification.

(1) re: R3.4. Thanks for clarifying the details of initial Q values, as motivated by dividing reward magnitudes by 10 to range between 0-1. It'd be helpful to include this information in the main text (e.g., page 6).

(2) re: R3.10. "[...], it wouldn't be so surprising in our view that participants who are positively surprised with unexpected rewards would share more advice, and those who are penalized with lesser rewards decide not to share." This made me wonder if it may make sense to test a learning model with double learning rates (for positive/negative PE)? And generally why in the first place did the authors decide only to test one model without comparing it to other alternatives? I am aware that including the double learning rates model deviates from the pre-reg, but it would be helpful to justify in the main text why only one model is tested.

(3) I appreciate the new results presented in SI, but most supplementary Tables and Figures are not yet referenced in the main text.

EDITORIAL POLICIES

We ask that you ensure your manuscript complies with our editorial policies and reporting requirements.

To that end, we require revised manuscripts to be accompanied by two completed items: a reporting summary that collects information on study design and procedure, and an editorial policy checklist that verifies compliance with all required editorial policies.

- <https://www.nature.com/documents/nr-reporting-summary.zip>>Nature Research Reporting Summary
- <https://www.nature.com/documents/nr-editorial-policy-checklist.pdf>>Editorial Policy Checklist

All points on the policy checklist must be addressed. Your revised manuscript can only be sent back to the referees if these checklists are completed and uploaded with the revision.

Notes: If you have submitted a Stage 1 Registered Report, Review, Primer, Comment, or Perspective you do not need to submit these forms. If you have already submitted these forms, you may disregard this request.

* **TRANSPARENT PEER REVIEW:** Communications Psychology uses a transparent peer review system. This means that we publish the editorial decision letters including Reviewers' comments to the authors and the author rebuttal letters online as a supplementary peer review file. However, on author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If your manuscript has been previously reviewed at another journal, those Reviewers' comments would not form part of the published peer review file.

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Psychology is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on 'Modify my Springer Nature account' and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.

<https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research>

For more information please visit <http://www.springernature.com/orcid>

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Version 2:

Decision Letter:

** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your coauthors **

Dear Dr Anllo,

Your manuscript titled "Experience and advice consequences shape information sharing strategies" has now been seen by our reviewers, whose comments appear below. In light of their advice I am delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Psychology.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers and a list of editorial requests. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at <https://www.nature.com/documents/commsj-file-checklist.pdf>.

OPEN ACCESS:

Communications Psychology is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit <https://www.nature.com/commspsychol/open-access>

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

* TRANSPARENT PEER REVIEW: Communications Psychology uses a transparent peer review system. On author request, confidential information and data can be removed from the published reviewer reports and rebuttal letters prior to publication. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons.

Please use the following link to submit the above items:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Jennifer Bellingtier

Jennifer Bellingtier, PhD
Senior Editor
Communications Psychology

Neil Garrett, PhD
Editorial Board Member
Communications Psychology
orcid.org/0000-0003-1440-472X

REVIEWERS' EXPERTISE:

Reviewer #1 Social decision making, reinforcement learning

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have adequately addressed all the concerns I raised.

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

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/>

Please find below our point-by-point responses (**in black**, in-text modifications **in red**) to all of the Editor's and the Reviewers' remarks (copied **in blue**).

Best regards,

Hernán Anlló, Gil Salamander, Stefano Palminteri, Nichola Raihani, Uri Hertz

Reviewer #1 (Remarks to the Author):

R1.1. The present study investigated a case in which participants decided whether or not to share their choices as advice to others. The authors found that a significant portion of the participants shared their choices as advice even when they did not learn anything about the environment. Moreover, the tendency to share advice was modulated by their expertise (i.e., the progress of the learning), task difficulty, individual differences in social anxiety, and other factors. I believe this study addressed an important issue with a pre-registered experimental design and data analysis plan. While a growing body of literature has investigated advisees' behavior, few studies have assessed advisors' behavior. The present study, therefore, can provide significant insights into social psychology. My comments below are mostly about computational modeling and interpretation of the data.

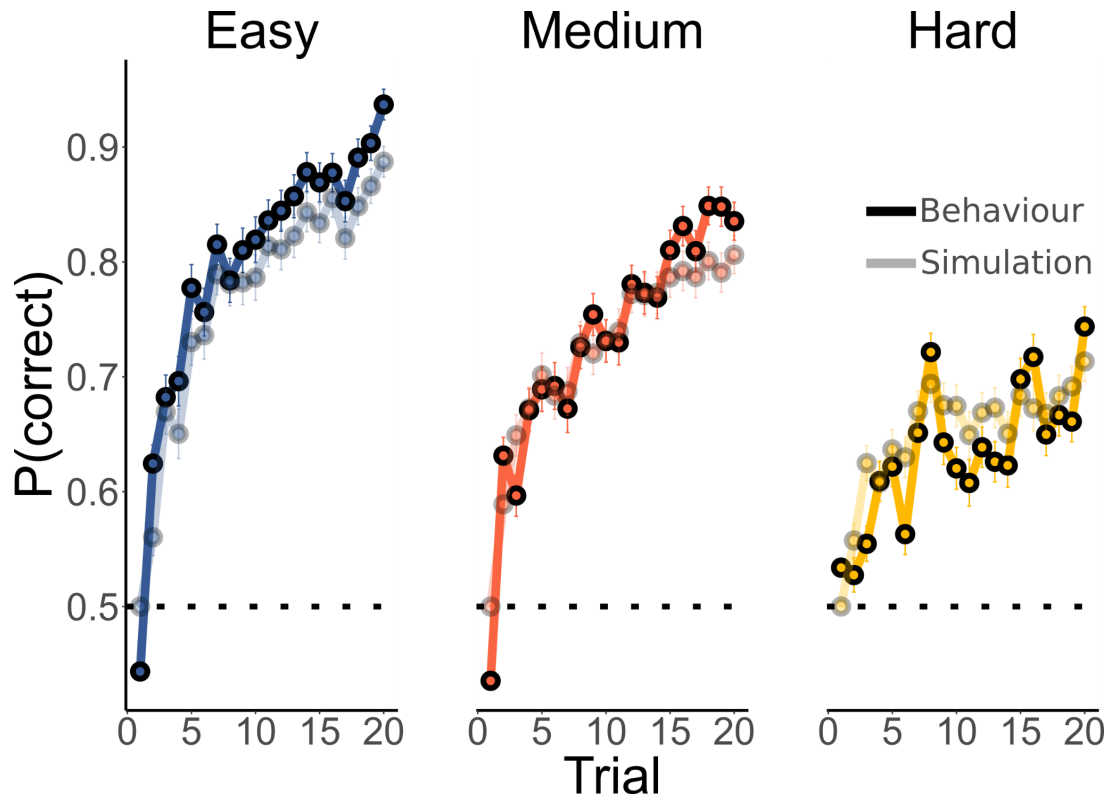
This summary is accurate. We thank the Reviewer for the interest they show in our study. We hope to have addressed all queries to their satisfaction below.

R1.2. I would like to request that the authors examine to what extent the model fitting procedure captures the true $p(\text{Choice})$ and PE given the current experimental setting and RL model, by using simulation. That is, I would first simulate an RL agent in the experimental task. Notably, in the simulated data, I am able to know the TRUE $p(\text{Choice})$ and PE. I would then fit the RL model to the simulated data and obtain the estimated learning rate and inverse temperature. Finally, I would compare the trial-by-trial TRUE $p(\text{Choice})$ and PE with those estimated from the fitted learning rate and inverse temperature.

We appreciate the reviewer's concern and are happy to comply with this request, and evaluate how adequate the fitting procedure is in regard to our sample through simulation. We carried two simulations. The first simulates the model using the parameters recovered from participants and then formally compares choices between the model and the participants (Bavard et al, 2021; Anlló et al, 2024). The second follows the suggested procedure of simulation and parameter recovery (Palminteri et al, 2017; Wilson et al, 2019). See detailed descriptions below, and the revised text in red.

1. We implemented a tried and tested approach (Bavard et al, 2021; Anlló et al, 2024), that consists of fitting our model to each individual participant's behaviour, recover the value of each individual's parameter, and then use said model and parameters to generate new simulated behaviour at the individual level. Crucially, as requested by the reviewer, this provides us with simulations that are comparable to specific participant behaviour. The figure below shows the result of this computation, as the mean performance per

block difficulty on a trial basis (i.e. correct choice ratio, where correct choice is the choice that maximizes value), for both participants and simulations:



Visual inspection of the results shows a simulated performance that closely mirrors actual behaviour. To confirm this similarity, we computed the Bayes Factor for an effect of a difference between real and simulated behaviour. We did so by using the Bayesian Information Criterion approximation to the Bayes Factor approach (Wagenmakers, 2007), fitting two identical regressuib models that differed only on a “Data Type” categorical factor (levels: real behaviour vs. simulation). Overall, results greatly favored the model without the “data type” factor ($P(\text{correct}) \sim \text{Difficulty}$), regardless of whether we considered it as a main effect ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type}$; $\text{BF} = 0.0024$), an interaction ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type}:\text{Difficulty}$; $\text{BF} = \text{BF} < 0.0001$), or both ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type} + \text{Data Type}:\text{Difficulty}$; $\text{BF} < 0.0001$). Overall, these results confirmed that for the purpose of estimating choice probability, which was the key element of interest here given its role as a proxy for the accumulation of knowledge of each option, our model provided reliable estimations.

In accordance with this novel analysis, we have modified our manuscript as follows:

(Page 11 - Supplementary materials)

To evaluate how adequate the fitting procedure was in regard to our sample, we implemented a tried and tested approach^{7,8}, that consists of fitting our model to each individual participant's behaviour, recover the value of each individual's parameter, and

then use said model and parameters to generate new simulated behaviour at the individual level. Crucially, this produces simulations that are comparable to specific participant behaviour. Results of this computation are shown in Figure S8.

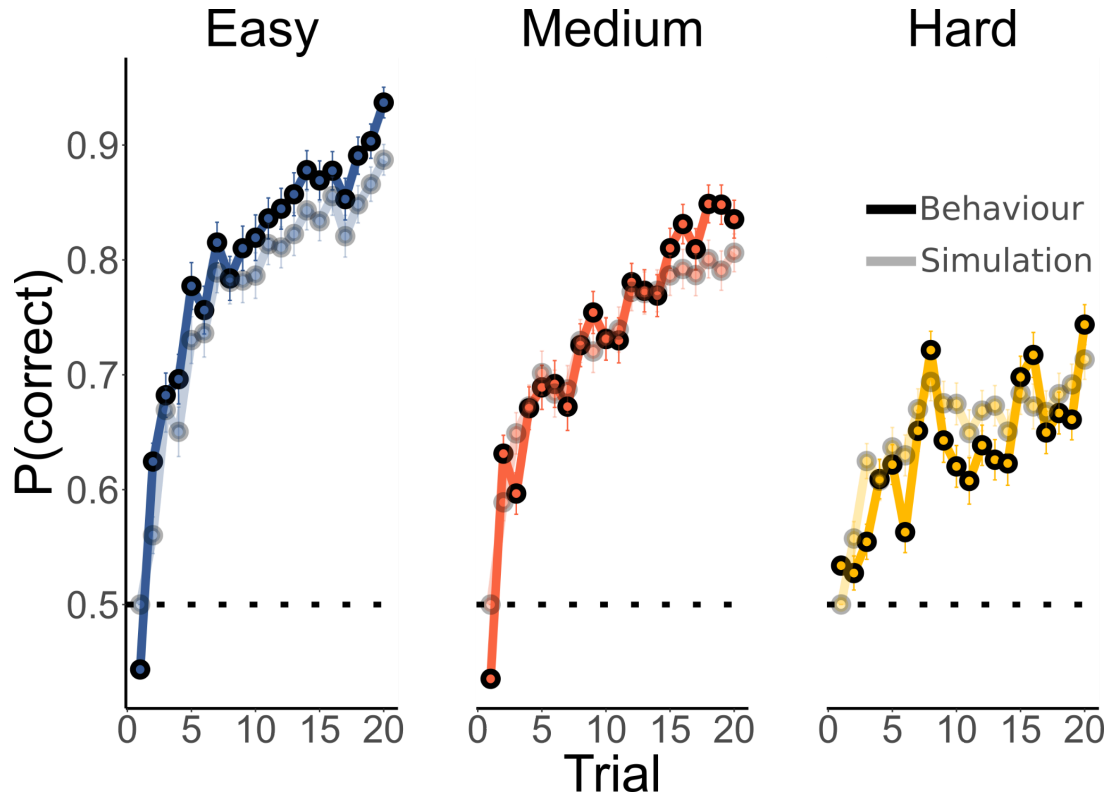


Figure S8. Simulations vs. participant behavior. Mean performance per block difficulty on a trial basis, for both participants and simulations. (Bars are standard error of the mean; $P(\text{correct})$ is correct choice ratio, where correct choice is the choice that maximizes value; $N = 373$, simulated agents $N = 373$).

Visual inspection of the results shows a simulated performance that closely mirrors actual behaviour. To confirm this similarity, we computed the Bayes Factor for an effect of a difference between real and simulated behaviour. We did so by using the Bayesian Information Criterion approximation to the Bayes Factor approach⁹, fitting two identical regressuib models that differed only on a “Data Type” categorical factor (levels: real behaviour vs. simulation). Overall, results greatly favored the model without the “data type” factor ($P(\text{correct}) \sim \text{Difficulty}$), regardless of whether we considered it as a main effect ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type}$; $\text{BF} = 0.0024$), an interaction ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type}:\text{Difficulty}$; $\text{BF} = \text{BF} < 0.0001$), or both ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type} + \text{Data Type}:\text{Difficulty}$; $\text{BF} < 0.0001$).

2. We followed this robustness check by implementing the “simulation plus parameter recovery” approach (Palminteri et al, 2017; Wilson et al, 2019). This technique requires, on the one hand, that we use our model to simulate the actual observed behaviour

(which we have done with success in the query above). On the other hand, it requires (1) that we generate simulated agents that solve our task with predefined random assortments of parameter values (and not just parameters coming from observed behaviour), (2) that we fit our model to these simulated agents' behaviour, and observe whether the parameters obtained match the original parameters used in the first place (i.e. if said parameters are "recovered" successfully). This approach is ideal due to its stringency: it pushes on the flexibility of the model by asking it to fit combinations of parameters that may lay outside the scope of human performance. Here, robustness is determined by the quality of recovery (i.e. a statistically significant correlation where R at least > 0.5). We have included this analysis in our manuscript as follows:

(Page 12 - Supplementary materials)

We finished our robustness tests by completing the implementation of the "simulation plus parameter recovery" approach^{10,11}. Following the successful use of our model to simulate participant behaviour above, we generated artificial agents with predefined random assortments of parameter values. Then, we presented these agents with our default task exactly as it was presented to human participants (i.e. three blocks of increasing difficulty, 20 trials per block), and refitted our model on their behaviour to evaluate if parameters could be successfully recovered. The model of our task relied fundamentally on two parameters (i.e. learning rate [α] and inverse choice temperature [β]). Thus, we picked 9 random values for each of these parameters, and conceived one simulated agent for every possible parameter combination (for a total of $N = 81$). As shown in Figure S9.A, trial-by-trial simulated performance was coherent with task design (i.e. learning and decreasing performance with increased difficulty were observed). Figure S9.B shows that the correlation between the randomly-generated parameters and the recovered parameters was acceptable both for learning rates (Pearson's product-moment correlation $t = 85$, $p < 0.001$, $R = 0.772$ (95% confidence interval: 0.760, 0.783) and inverse temperatures ($t = 92$, $p < 0.001$, $R = 0.800$ (0.788, 0.808)).

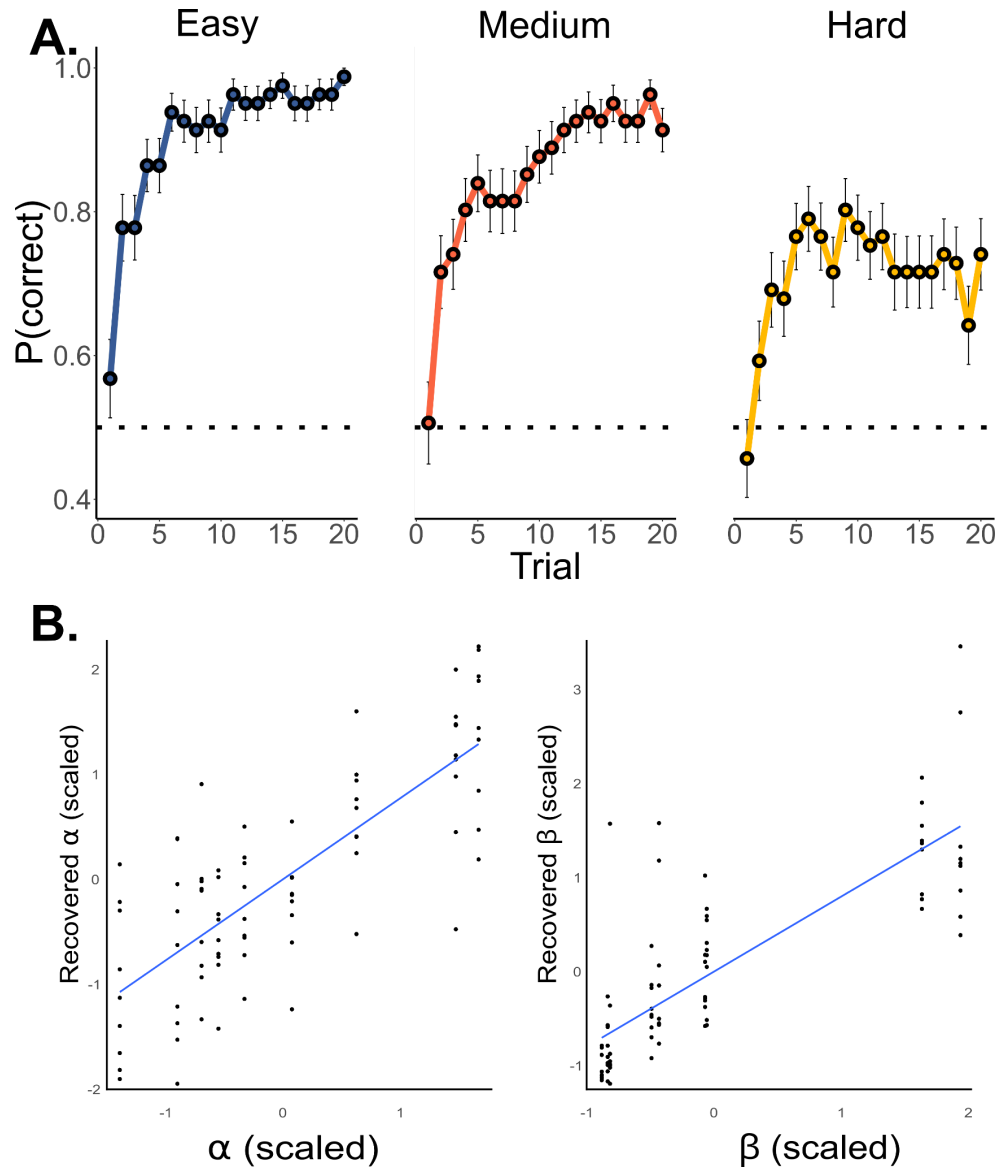


Figure S9. Parameter recovery. **A.** Per trial performance of simulated agents. **B.** Correlations of randomly generated and refitted learning rates and inverse temperatures (scaled). (*Bars are standard deviation; $P(\text{correct})$ is correct choice ratio, where correct choice is the choice that maximizes value; simulated agents $N = 81$*)

Following the recovery of the parameters, we conducted an Intraclass Correlation Coefficient analysis (ICC), for both learning rate and inverse temperature, and observed significant moderate-high to high reliability¹³. Results of this procedure are outlined in Table S8 below.

ICC Learning rate (α)							
	ICC	F	DF1	DF2	CI (lower)	CI (higher)	P
Single raters (absolute)	0.7	5.8	80	81	0.58	0.80	<.001
Single random raters	0.7	5.8	80	80	0.58	0.80	<.001

Single fixed raters	0.7	5.8	80	80	0.57	0.80	<.001
Average raters (absolute)	0.83	5.8	80	81	0.73	0.89	<.001
Average random raters	0.83	5.8	80	80	0.73	0.89	<.001
Average fixed raters	0.83	5.8	80	80	0.73	0.89	<.001
ICC Inverse temperature (°)							
Single raters (absolute)	0.65	4.7	80	81	0.51	0.76	<.001
Single random raters	0.65	4.9	80	80	0.51	0.76	<.001
Single fixed raters	0.66	4.9	80	80	0.52	0.77	<.001
Average raters (absolute)	0.79	4.7	80	81	0.67	0.86	<.001
Average random raters	0.79	4.9	80	80	0.67	0.86	<.001
Average fixed raters	0.80	4.9	80	80	0.68	0.87	<.001

Table S8. Intraclass correlation coefficient analysis for recovered vs randomly-generated parameters. *Single raters absolute: single raters using an absolute agreement definition. Each participant is rated by a different, randomly selected rater. Single random raters: single raters under the assumption that raters are randomly selected. Reflects consistency of ratings, not just agreement. Single fixed raters: raters are fixed, consistency is measured within this fixed set of raters. Average raters (absolute): the average of multiple raters using an absolute agreement definition. Average random raters: average of multiple raters assuming that raters are randomly selected. Average fixed raters: average of multiple raters with the assumption that the set of raters is fixed. ICC: intraclass correlation coefficient value. DF1: degrees of freedom for numerator. DF2: degrees of freedom for denominator. CI: 95% confidence interval bounds. P: p-value.*

Overall, these results confirmed that for the purpose of estimating choice probability, which was the key element of interest here given its role as a proxy for the accumulation of knowledge on each option, our model provided reliable estimations.

References for this answer

1. Bavard S., Rustichini A., Palminteri S. Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Sci. Adv.* 7, eabe0340 (2021). DOI:10.1126/sciadv.abe0340
2. Anlló, H., Bavard, S., Benmarrakchi, F. et al. Comparing experience- and description-based economic preferences across 11 countries. *Nat Hum Behav* 8, 1554–1567 (2024). DOI:10.1038/s41562-024-01894-9
3. Wagenmakers, E. J. A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779-804 (2007).
4. Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in cognitive sciences*, 21(6), 425–433. <https://doi.org/10.1016/j.tics.2017.03.011>
5. Wilson R.C., Collins A.G.E. (2019) Ten simple rules for the computational modeling of behavioral data *eLife* 8:e49547, <https://doi.org/10.7554/eLife.49547>

R1.3. It is worth testing other models for a robustness check. I would suggest RL models with asymmetric learning rates, with perseveration, and with both asymmetric learning rates and perseveration. I do not insist that these are alternative better models, but it is useful to evaluate the robustness of the main findings in this study.

We completely agree with the Reviewer that further robustness checks were an important addition to be made, and added the parameter recovery and simulation in the query above, which are the gold standard of model robustness checks. However, when it comes to fitting novel models that lie outside of our preregistered hypotheses, we ask the Reviewer to reconsider. The main focus of our paper is not an evaluation of the specific learning process in this task, which is well established, but to use said process to uncover how learning shapes advice-giving. Further, we have no cognitive hypothesis related to how or why multiple learning rates would or should affect the process of advice-giving. Thus, we believe that testing multiple learning models may be distracting from the main findings, and beyond the scope of this paper. However, it is indeed a great suggestion for future work, and incidentally we are currently examining this dimension in a different project, where learning about different social information sources is examined.

R1.4. Having said that, I would like to note that the current description of computational modeling is unclear.

- Why is the learning rate denoted as alpha in the main text and phi in the supplementary information (SI)? Why is the inverse temperature depicted by beta in the main text and lambda in the SI?

We thank the Reviewer very much for this very important remark. It appears that there has been an error in the typesetting of the document, which has now been corrected.

(Page 11 - Supplementary materials)

COMPUTATIONAL MODELING

We operationalized our Reinforcement Learning (RL) models through STAN⁵, which facilitates Bayesian statistical inference because it allows for the specification of complex statistical models together with their corresponding posterior inferences. Our RL model quantified how individuals updated the perceived value (Q-value) of different options based on choice outcomes (rewards). Thus, it was structured around two critical parameters: the learning rate (α) and the inverse temperature (β).

R1.5. - For MCMC, what prior was used? How did they evaluate the convergence of the sampling processes?

We used non-informative priors, where learning rate was taken as the logit of a parameter `alpha_logit` with prior of gaussian with mean 0 and sigma 2, and the beta was an exponential of a parameter `beta_exp` with prior of gaussian with mean 0 and sigma 2. For individual-level noise variables we used a gaussian prior with mean 0 and sigma of 1.

We checked the R-hat for any high values that could indicate problems with convergence and found that none exceeded 1.01.

We have now included these details and the convergence table in the supplementary materials ("Supplementary Materials - Model Parameters" section).

R1.6. - I believe the authors obtained the joint posterior distribution (learning rate x inverse temperature). How did they obtain the point estimates of the learning rate and inverse temperature from the 2D posterior distribution?

We used the mean of the posterior as the point estimate of the individual parameters. We now clarified this in the text:

(Page 15 - Supplementary materials)

To account for individual-level random effects for both learning rate and inverse temperature we implemented a Cholesky decomposition⁷ to efficiently parameterize the covariance matrix of random effects. We used non-informative priors, where learning rate was taken as the logit of a parameter `alpha_logit` with prior of gaussian with mean 0 and sigma 2, and the beta was an exponential of a parameter `beta_exp` with prior of gaussian with mean 0 and sigma 2. For individual-level noise variables we used a gaussian prior with mean 0 and sigma of 1. We checked the Rhat for any high values that could indicate problems with convergence and found that none exceeded 1.01. Full convergence and parameters are available in the CSV files accompanying the analysis code (<https://github.com/hrl-team/AdvisingRL>). For evaluation of individual level parameters we used the group level and random effects recovered from the model fitting procedure, and used the mean of the posterior as point estimate for further analyses.

R1.7. - What was WAIC computed for? The authors did not conduct any model comparison. I find no reason to compute WAIC in this study.

The Reviewer is correct, we agree that as a standalone measure the WAIC is not very informative. It has now been removed.

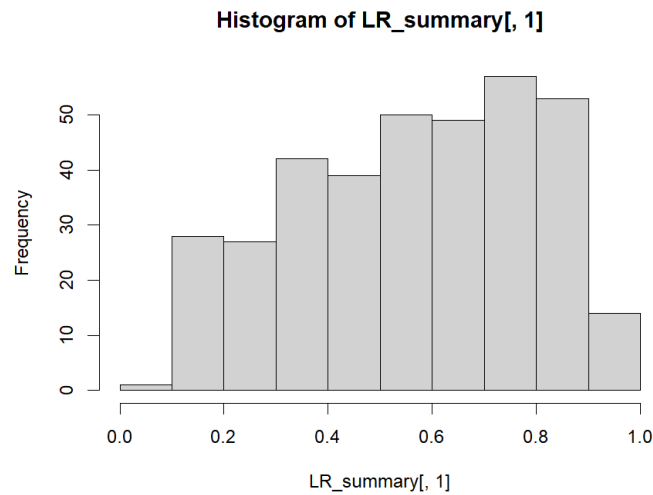
R1.8. - As far as I understand, the RL model is fitted to each participant's data separately. What are the random effects?

We now added all parameters recovered from the model to the supplementary materials. These include group level and random effects, their posterior, R-hat and mean value. Below are the histograms of the individual level parameters, indicating the spread of learning rates and beta

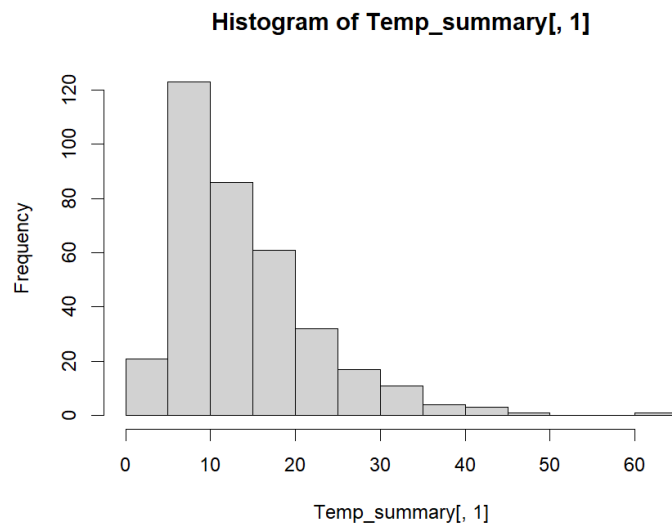
parameters across participants which were used to uncover individual level $p(\text{choice})$ and PE. Individual level parameters (full table available at <https://github.com/hrl-team/AdvisingRL>):

Experiment 1

Learning Rates - mean $\pm$ std 0.56 ± 0.23

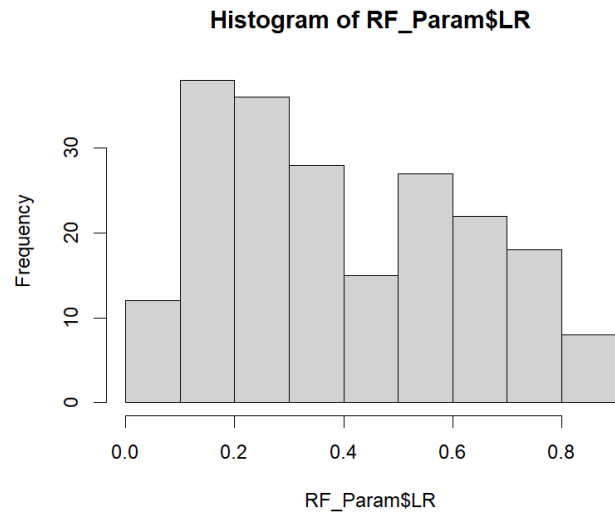


Inverse Temperature - mean $\pm$ std 13.96 ± 8.36

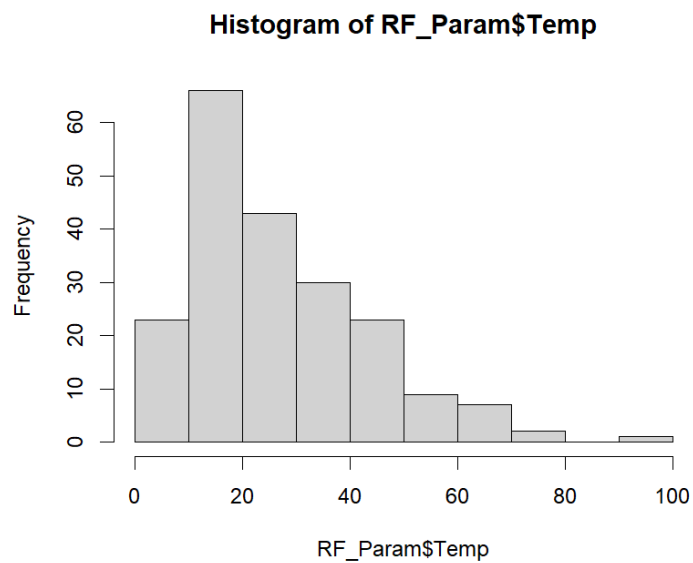


Experiment 2

Learning Rates - mean $\pm$ std 0.39 ± 0.23



Inverse temperature - mean $\pm$ std 26.46 $\pm$ 16.14



R1.9. It would be better if the authors discussed potential interpretations of the observed effect of trial number. I guess that not only expertise but also many other factors changed with time.

This is a very fair remark. Our general discussion now includes a commentary on possible concurrent/additional effects incarnated by trial progression.

(Page 15 - Main manuscript)

4. General discussion

[...]

It is important to note that the present work was constrained by certain limitations. [...]. Trial progression could affect our results in unexpected ways. For example, changes related to boredom or impatience in the face of how repetitive the task is⁶⁴, fluctuations in attention allocation⁶⁵, fluctuations in confidence⁶⁶, and even changes in evidence accumulation thresholds^{66,67} could affect participants' decisions. Future work should target these elements, possibly by including race models, drift diffusion models and subjective metacognitive trial-by-trial checks.

References for this answer

Agrawal, M., Mattar, M. G., Cohen, J. D., & Daw, N. D. The temporal dynamics of opportunity costs: A normative account of cognitive fatigue and boredom. *Psychological Review*, 129(3), 564–585. (2022) <https://doi.org/10.1037/rev0000309>

Zilker, V. Attentional dynamics of evidence accumulation explain why more numerate people make better decisions under risk. *Sci Rep* 14, 18788 (2024). <https://doi.org/10.1038/s41598-024-68969-5>

Lebreton M, Bacily K, Palminteri S, Engelmann JB. Contextual influence on confidence judgments in human reinforcement learning. *PLoS Comput Biol* 15(4): e1006973 (2019). <https://doi.org/10.1371/journal.pcbi.1006973>

Held, L.K., Vermeylen, L., Dignath, D. et al. Reinforcement learning of adaptive control strategies. *Commun Psychol* 2, 8 (2024). <https://doi.org/10.1038/s44271-024-00055-y>

R1.10. I would like to know the strength of the association between advising behavior and SPIN (beyond statistical significance). I would like to ask the authors to report the effect size.

This is a fair request. Please do consider however that pinpointing a standardized effect size within the framework of generalized linear models can be challenging, because of how participant-level variance is partitioned. Namely, to produce a standardized effect size, the effect on the response scale needs to be normalized by a measure of variability (e.g., in Cohen's D, standard deviation). And it is not really clear which estimate to take in the case of a mixed model, as there are usually several (see Rights & Sterba, 2019 for an expanded discussion on this). Accepted alternatives to classic standardized effect sizes in this context (to be interpreted with caution) include reporting model coefficients (<https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html>). For these coefficients to be more interpretable, we here present them in a standardized form by exponentiating them (i.e. $e^{\text{coefficient}}$). This transformation produces an approximation to the odds ratio (OR), namely the multiplicative change in the odds that the predicted outcome will happen. For example, the coefficient for the interaction between SPIN and p(choice) is 0.243(95% CI = 0.063, 0.4239),

which means $OR = 1.28(1.065, 1.530)$. This implies that there is a 28% increase in the odds that individuals will give advice for every unit of joint change in SPIN and $p(\text{choice})$.

We now address this issue explicitly in the manuscript, explain our rationale, and make the requested additions for all relevant effects where this method is applicable:

(Page 6 - Main manuscript)

(e) Analyses

[...]

P-values for effects associated with predictors were calculated through analysis of deviance (Type 2 Wald χ^2 tests). All analyses were conducted using the R language⁴¹, with the `dyplr` package⁴² for data management, the `lme4`⁴³ and `car`³⁷ packages for regressions, the `emmeans` package⁴⁴ for post-hoc contrasts, and the `ggplot2` package⁴⁵ for figure production. **Given the limitations imposed by the use of mixed-effect models that stem from how variance is partitioned⁴⁶, where applicable effect sizes were reported by measure of how much each predictor generated a change in the predicted variable. This was done by reporting regression coefficients, and transforming said coefficients through exponentiation to obtain an approximate measure of the Odds Ratios (OR)⁴⁷ for each predictor, as a standardized form of effect size. It is also worth noting that we report per-predictor chi-square values as well, as a measure of how much variance each predictor explains within their model.**

(Page 7 - Main manuscript)

[...] lower probability to be an over-adviser at higher levels of social anxiety (coefficient = -0.203 (95% CI = -0.418, 0.010), **OR = 0.82 (0.66, 0.99)**; Likelihood ratio Chi-square $\chi^2(1) = 3.49$, $p = 0.062$

higher probability of over-advising among more altruistic individuals (coefficient = 0.165 (-0.048, 0.385), **OR = 1.18 (0.95, 1.47)**; $\chi^2(1) = 2.30$, $p = 0.130$

(Page 8 - Main manuscript)

[...] Trial number main effect: **coefficient = 0.20 (95% CI = 0.165, 0.226), OR = 1.22 (1.18, 1.25)**; $\chi^2(1) = 109$, $p < 0.001$

[...] PTM (altruism) scores marginally predicting higher levels of advice (**coefficient = 0.292(0.005, 0.585), OR = 1.34(1.01, 1.79)**, $\chi^2(2) = 4$, $p = 0.046$

[...] The likelihood of broadcasting advice per trial was significantly predicted by the estimated choice probability ($P(\text{choice})$ main effect: coefficient = 2.99(2.82, 3.16), **OR = 19.89(16.78, 23.57)**, $\chi^2(1) = 1206$, $p < 0.001$).

[...] interaction between social anxiety and P(choice) that aligned with our prediction (see H1.4), as probability to give advice increased with SPIN (fear) score for exploitative trials (coefficient 0.243(0.063, 0.4239), OR = 1.28(1.065, 1.530), $\chi^2(1) = 7$, $p = 0.008$; see Figure 2B).

[...] Higher positive prediction errors in the preceding trial also predicted higher probability of advice-giving (PE main effect: coefficient = 0.54(0.31, 0.79), OR = 1.72(1.36, 2.20), $\chi^2(1) = 18$, $p < 0.001$)

(Page 13 - Main manuscript)

[...] choice probability (P(choice) main effect: coefficient = 2.1 (1.6, 2.5), OR = 8.17(4.95, 12.18), $\chi^2(1) = 666$, $P < 0.001$).

[...] P(choice) x SPIN interaction: coefficient = 0.255(-0.20, 0.788), OR = 1.29(0.82, 2.20), $\chi^2(1) = 27$, $p < 0.001$ [...] (P(choice) x PTM interaction: coefficient = -0.921(-1.425,-0.417), OR = 0.40(0.24, 0.66), $\chi^2(3) = 4$, $p = 0.036$; Fig. 5.C).

Reference for this answer

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, 24(3), 309–338. <https://doi.org/10.1037/met0000184>

Nakagawa, S. & Schielzeth, H. (2013). *Methods Ecol. Evol.* 4, 133–142. DOI: 10.1111/j.2041-210x.2012.00261.x

R1.11. It would be great if the authors explained why they included the previous PE in the second regression model. I am especially wondering why PE was included rather than the outcome (i.e., reward).

In our pre-registration we hypothesized that recent prediction error, i.e., how negatively or positively surprised the participants were with the outcome of their choice, would affect their likelihood to share their current choice as advice, in addition to the learned likelihood of receiving reward p(choice). This is based on previous literature showing how prediction errors affect reward-related wellbeing (Rutledge et al, 2014) and subjective estimation of one's experience (Keren et al, 2021). We did not include the recent reward as a predictor as p(choice) represents a time weighted average of the reward, and p(choice) was our main predictor of interest in these analyses. Our analyses have shown that while p(choice), representing previous rewards, affected the decision to share advice, PE had comparatively little predictive power when it came to sharing advice (as seen by how little variance was explained by this predictor in our models). We now added this motivation to the main text:

We fitted the model using uninformative priors, and used the mean of the posterior as the point estimate of the individual parameters. We then used these parameters to estimate trial-by-trial probability of choice $p(\text{choice})$ and prediction error. Probability of choice represents the accumulated knowledge about the relative rewards of the two options, and is expected to be high when choosing an option with much higher expected reward from the other option (see eq. 3). It therefore captures the knowledge of the participants about expected rewards. We therefore used this value, and its interaction with individual traits, to predict advice-giving behaviour. In addition, we hypothesized that recent prediction error, i.e., how negatively or positively surprised the participants were with the outcome of their choice, would affect their likelihood to share their current choice as advice. This last hypothesis was based on previous literature showing how prediction errors affect reward-related wellbeing (Rutledge et al. 2014) and subjective estimation of one's experience (Keren et al. 2021)

Reference for this answer

R. B. Rutledge, N. Skandali, P. Dayan, R. J. Dolan, A computational and neural model of momentary subjective well-being. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12252–12257 (2014).

Keren H, Zheng C, Jangraw D, Chang K, Vitale A, Rutledge R, Pereira F, Nielson DM, Stringaris A. The temporal representation of experience in subjective mood, *eLife* (2021) 10:e62051

Reviewer #2 (Remarks to the Author):

R2.1. The study titled "Experience and advice consequences shape information sharing strategies" by Hernán Anlló and colleagues aimed to explore how individuals decide to give advice within a learning context, particularly focusing on the roles of accumulated knowledge, social anxiety, and prosocial tendencies, as well as the impact of different consequences (monetary costs, reputational stakes) on advising behavior. Using a reinforcement learning framework, the researchers found that individuals were more likely to share advice as they gained experience and moved from exploratory to exploitative decision-making. Additionally, the introduction of consequences significantly influenced advice-giving: monetary costs reduced it, while reputational stakes increased it. Socially anxious individuals were cautious about giving advice during uncertain, early stages but more likely to advise once confident in their choices. Prosocial individuals were generally more inclined to give advice, especially under reputational conditions. The study concluded that advice-giving is a complex behavior influenced by a combination of personal experience, psychological traits, and external incentives.

The paper is well written and addresses a significant gap in the literature. I have a few concerns regarding the modeling framework and the interpretation of the results.

We thank the Reviewer for the interest they show in our study. We have now addressed all of these concerns to the best of our ability. We hope the Reviewer will be happy with these amendments, but remain at your full disposal for any possible further corrections.

MAJOR Criticisms:

R2.2. 1. Reputation feedback: The absence of real-time feedback on the adviser's reputation means participants were not influenced by immediate consequences of their advice on a trial-by-trial basis. This could have impacted their decision-making process, as they did not experience the direct and immediate repercussions of their advice on their reputation. Real-time feedback could potentially lead to different advising behaviors, as participants would continuously adjust their strategies based on the immediate feedback received after each trial. This may be acknowledged in the Discussion of the results.

We thank the Reviewer for this remark, and we are very sorry about the confusion in the description of the reputation condition. In fact, participants had trial-by-trial information about their adviser ratings (but not from other people using their advice, which is what we meant to say). After sharing their advice, participants observed the outcome of their own choice, and were also presented with the evaluation of their advice based on simulation, and the update to their adviser ratings (see Figure 1 panel B for the visual presentation of this information). Participants were told that their advice and their adviser ratings would be presented to future players of the game (but this did not actually happen during the experiment). Participants could therefore evaluate the consequence of their advice on their ratings, and decide whether to share more advice and or avoid sharing.

We now clarify this in the text in different places:

(Page 24 - Main manuscript - Figure 1 LEGEND)

Figure 1. A. Trial outline. In all experiments, participants had to choose between two lakes with the goal of maximising their fishing yield. Before receiving feedback participants could choose to publish their choice as advice for future participants undergoing the same learning task. **B. Advice consequences.** Study 1 allowed participants to give advice without specifying advice consequences. Study 2 included a replication of this condition ("No cost [replication]"), and additional between-participant conditions where advice came with different consequences. In the "Monetary" consequence context, participants had to pay 1 fish (£0.003) for publishing advice. **In the "reputation" consequence context, participants were told that an adviser rating would be calculated on a trial-by-trial basis based on their advice, and that their adviser ratings would be made public with their advice for future participants.**

(Page 9 - Main manuscript)

[...]

The “Monetary” condition introduced a small monetary cost (1 fish, equivalent to £0.003 reduction in bonus payment) which had to be paid in exchange for broadcasting advice . The “Reputation” condition introduced reputational consequences by telling participants that their advice would be presented to future participants along with their adviser-ratings. These adviser-ratings were calculated on a trial-by-trial basis, based on their current advice, and presented to the participants after each advice they gave (Figure 1 B). Thus, giving advice might increase or reduce the participants public adviser-ratings. Finally, the “Responsibility” condition explained to participants that their advice would have a direct effect on the decisions made by future participants, as these future participants would be compelled to follow their advice. Task instructions are included in the supplementary materials.

R2.3. 2. The impact of uncertainty on decision-making: While the RL model in the study effectively captures the overall dynamics of learning and decision-making under uncertainty, it does not explicitly differentiate between informational/epistemic and irreducible uncertainty. Participants’ decisions to give advice might be influenced differently by their ability to reduce uncertainty through learning (informational uncertainty) versus their strategies for handling inherent randomness (irreducible uncertainty). I was wondering if the authors considered comparing this RL model to Bayesian models that explicitly differentiate the roles of uncertainty in decisions to offer advice.

We agree that different types of uncertainty could affect the decision to share advice differently. At the time of conceiving the present experiment we did acknowledge the contrast between informational and irreducible uncertainties, and aimed to separate them by design. On the one hand, we targeted informational uncertainty by default, by studying the progressive learning that takes place inherently during RL. On the other hand, we targeted irreducible uncertainty by introducing three different blocks of variable difficulty, where difficulty level was directly determined by increasing inherent randomness behind choices. Namely, for each option reward was randomly drawn from a gaussian distribution with a standard deviation of 1.7, and a different mean value for each difficulty condition: Easy (6.5/3.5), Medium (6/4); Hard (5.5/4.5). This allowed us to understand the impact of increasing irreducible uncertainty in the learning process. What we observed was that advice-giving mirrored choice behaviour, and thus as performance decreased with difficulty, so did advising frequency. Still, we acknowledge that novel Bayesian models targeting these dimensions separately could further disentangle their influences and add informative insight on analyses. We have now added mention of this in our general discussion.

(Page 15 - Main manuscript)

4.General discussion

[...]

It is important to note that the present work was constrained by certain limitations. [...]. Future work should target these elements, possibly by including race models, drift

diffusion models and subjective metacognitive trial-by-trial checks. Additionally, future efforts targeting these effects would also benefit from considering Bayesian modelling alternatives that separate inherent and informational uncertainty, and their distinct impact on the decisional process behind the act of sharing information as advice.

R2.4. 3. Over-Advising Complications: A significant portion of participants engaged in over-advising (giving advice indiscriminately), which complicates the interpretation of the data. How did the authors debrief participants, and to what extent was over-advising driven by specific task instructions?

This is an important point, and we now acknowledge in the text the complication of evaluating over-advising in our task. The instructions to participants regarding the advice procedure in experiment 1 were:

“After you choose a lake, you will have the opportunity to send your choice as an advice to players that will participate in a follow-up experiment. If you choose to publish your advice, it may be displayed to all participants in a follow-up experiment, involving the same lakes/vacations. Accurate advice can greatly accelerate learning, leading to greater bonus for other players. However, inaccurate advice can hinder learning, resulting in lower bonuses for other players.”

We emphasised both the positive consequences of good advice, and the negative consequences of inaccurate advice. We don't think that this text could bias participants to over-share, as sharing your choice in the first few trials of the learning block is clearly non-informative. As we say in the manuscript, there could be other reasons for oversharing, such as normative pressure to share, or wanting to appear helpful, and low cost of sharing information which we directly test in the follow-up experiment. Of note, all instructions for every task are now included in the supplementary materials.

R2.5. 4. Parameter recovery and test-retest reliability: Two parameters were central to this model: (i) Learning Rate: This parameter reflects how quickly participants adjusted their expectations based on new information, and (ii) Inverse Temperature: This parameter indicates the consistency of participants' choices, acting as a measure of decision-making stochasticity. Were these parameters well recovered from simulations (i.e., if you simulated behavior based on the empirical estimates of these parameters, did you recover the same values? Typically, intraclass coefficients are computed to evaluate this). Since parameter recovery is a lower bound on test-retest reliability, have the authors also explored whether both the key behavioral variables as well as the parameter estimates were reliable when testing the participants a second, third time, etc.? This would also clarify whether the cluster of over-advisers is a reliable, stable effect as well.

This is indeed an important point. Following this request and the similar request from another reviewer, simulation and parameter recovery (with ICC checks) have been added to the manuscript. The procedure consisted of 4 distinct steps.

(1) We simulated per-trial artificial behaviour using our model and the parameters fitted from the observed behavior (see figure S8).

(2) We compared participant and simulated behaviour, and observed through Bayes factor analysis that there was no difference of substance between the two, which we interpreted as a sign of good fit.

(3) For parameter recovery, rather than just focusing on the parameters that came from fitting the empirical data, we generated novel simulated agents to solve our task, with predefined random assortments of parameter values. We preferred this approach due to its stringency: it pushes on the flexibility of the model by asking it to fit combinations of parameters that may lay outside the scope of human performance.

(4) We fitted our model to this simulated behaviour, and observed whether the parameters obtained matched the random original parameters used in the first place (i.e. if said parameters were “recovered” succesfully). Here, robustness was determined by the quality of recovery (i.e. a statistically significant correlation where R at least > 0.5 , and ICC analysis). For both the learning rate and the inverse temperature the diagnostics were good and reliable.

The results of these steps, and the ensuing modifications of our manuscript, are outlined below.

(Page 11 - Supplementary materials)

COMPUTATIONAL MODELING

[...]

To evaluate how adequate the fitting procedure was in regard to our sample, we implemented a tried and tested approach^{7,8}, that consists of fitting our model to each individual participant's behaviour, recover the value of each individual's parameter, and then use said model and parameters to generate new simulated behaviour at the individual level. Crucially, this produces simulations that are comparable to specific participant behaviour. Results of this computation are shown in Figure S8.

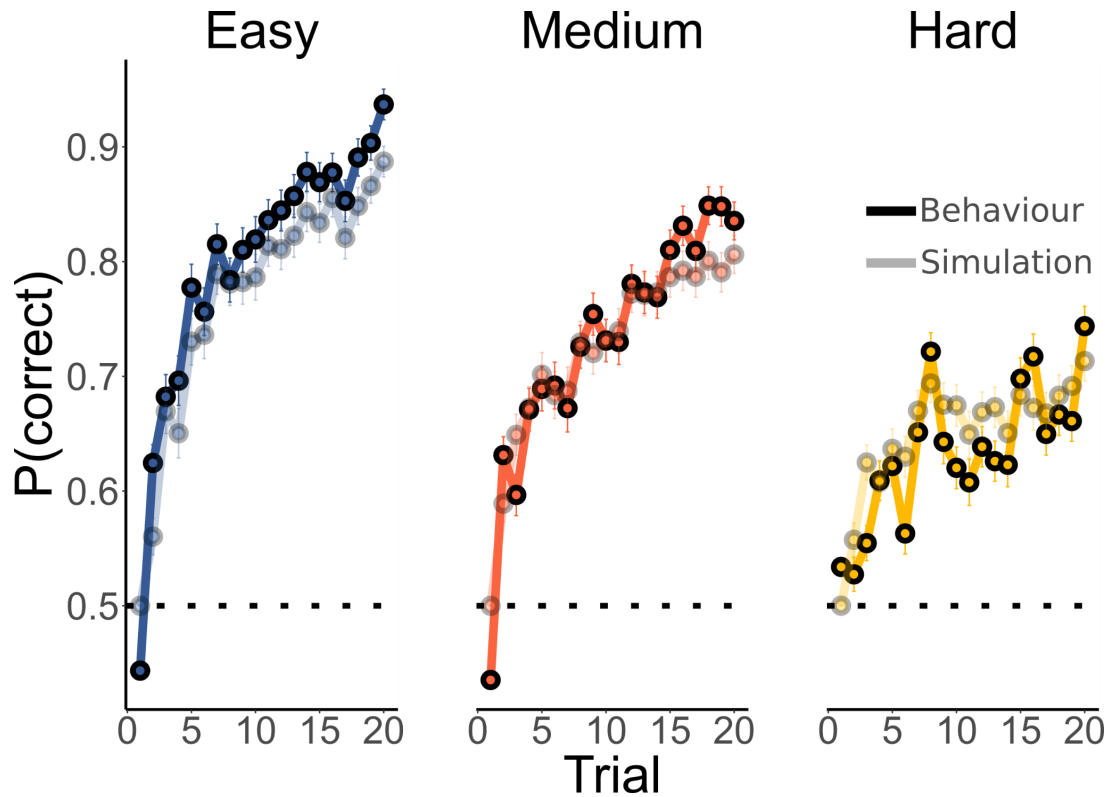


Figure S8. Simulations vs. participant behavior. Mean performance per block difficulty on a trial basis, for both participants and simulations. (Bars are standard error of the mean; $P(\text{correct})$ is correct choice ratio, where correct choice is the choice that maximizes value; $N = 373$, simulated agents $N = 373$).

Visual inspection of the results shows a simulated performance that closely mirrors actual behaviour. To confirm this similarity, we computed the Bayes Factor for an effect of a difference between real and simulated behaviour. We did so by using the Bayesian Information Criterion approximation to the Bayes Factor approach⁹, fitting two identical regressuib models that differed only on a “Data Type” categorical factor (levels: real behaviour vs. simulation). Overall, results greatly favored the model without the “data type” factor ($P(\text{correct}) \sim \text{Difficulty}$), regardless of whether we considered it as a main effect ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type}$; $\text{BF} = 0.0024$), an interaction ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type}:\text{Difficulty}$; $\text{BF} = \text{BF} < 0.0001$), or both ($P(\text{correct}) \sim \text{Difficulty} + \text{Data Type} + \text{Data Type}:\text{Difficulty}$; $\text{BF} < 0.0001$).

We finished our robustness tests by completing the implementation of the “simulation plus parameter recovery” approach^{10,11}. Following the successful use of our model to simulate participant behaviour above, we proceeded to generate artificial agents with predefined random assortments of parameter values. Then, we presented these agents with our default task exactly as it was presented to human participants (i.e. three blocks of increasing difficulty, 20 trials per block), and refitted our model on their behaviour to evaluate if parameters could be succesfully recovered. The model of our task relied fundamentally on two parameters (i.e. learning rate $[\alpha]$ and inverse choice temperature

[□]). Thus, we picked 9 random values for each of these parameters, and conceived one simulated agent for every possible parameter combination (for a total of $N = 81$). As shown in Figure S9.A, trial-by-trial simulated performance was coherent with task design (i.e. learning and decreasing performance with increased difficulty were observed). Figure S9.B shows that the correlation between the randomly-generated parameters and the recovered parameters was acceptable both for learning rates (Pearson's product-moment correlation $t = 85$, $p < 0.001$, $R = 0.772$ (95% confidence interval: 0.760, 0.783) and inverse temperatures ($t = 92$, $p < 0.001$, $R = 0.800$ (0.788, 0.808)).

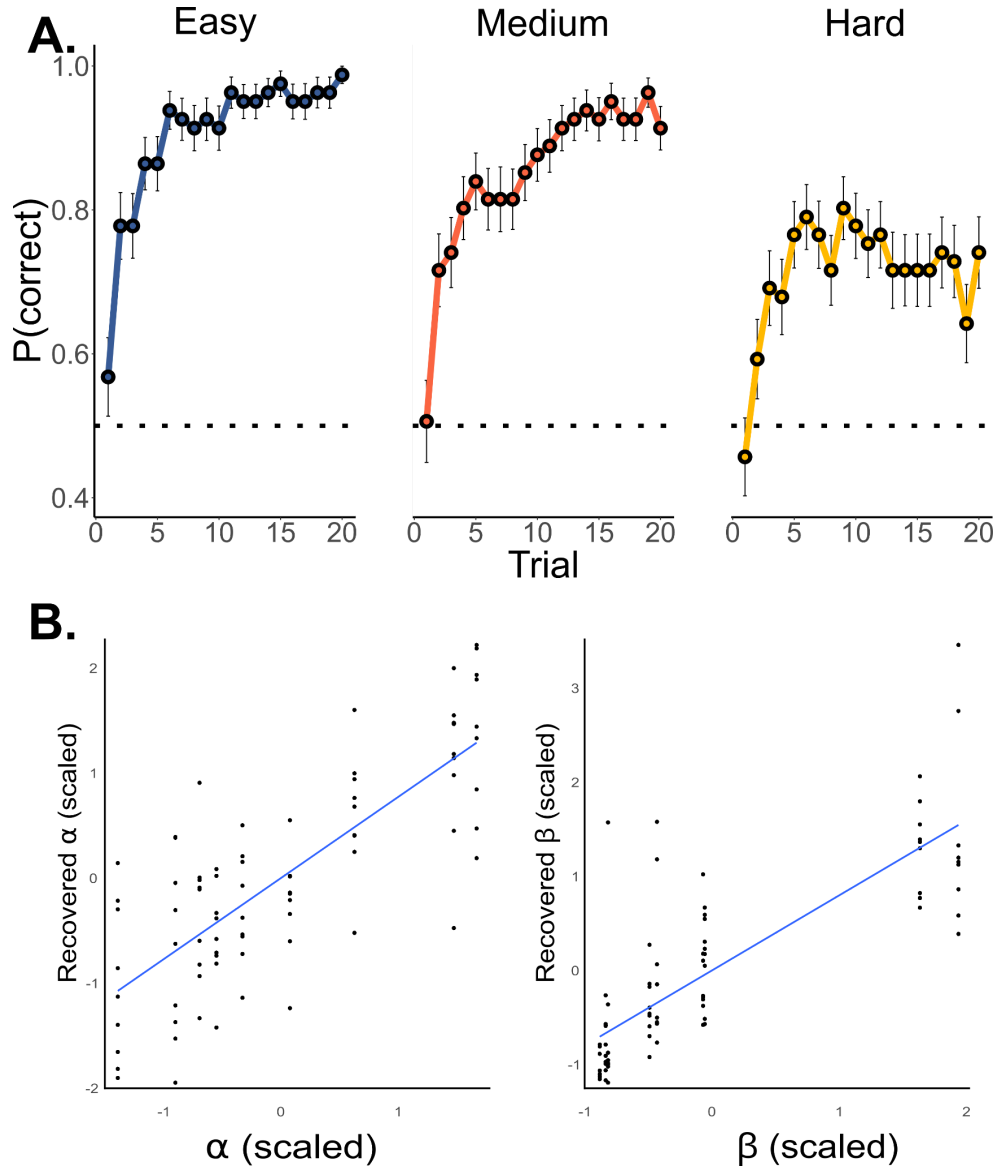


Figure S9. Parameter recovery. **A.** Per trial performance of simulated agents. **B.** Correlations of original and refitted learning rates and inverse temperatures (scaled). (Bars are standard deviation; $P(\text{correct})$ is correct choice ratio, where correct choice is the choice that maximizes value; simulated agents $N = 81$).

Following the obtention of the recovered parameters, we conducted an Intraclass Correlation Coefficient analysis (ICC), for both learning rate and inverse temperature, and observed significant moderate-high to high reliability¹². Results of this procedure are outlined in Table S8 below.

ICC Learning rate (α)							
	ICC	F	DF1	DF2	CI (lower)	CI (higher)	P
Single raters (absolute)	0.7	5.8	80	81	0.58	0.80	<.001
Single random raters	0.7	5.8	80	80	0.58	0.80	<.001
Single fixed raters	0.7	5.8	80	80	0.57	0.80	<.001
Average raters (absolute)	0.83	5.8	80	81	0.73	0.89	<.001
Average random raters	0.83	5.8	80	80	0.73	0.89	<.001
Average fixed raters	0.83	5.8	80	80	0.73	0.89	<.001
ICC Inverse temperature (β)							
Single raters (absolute)	0.65	4.7	80	81	0.51	0.76	<.001
Single random raters	0.65	4.9	80	80	0.51	0.76	<.001
Single fixed raters	0.66	4.9	80	80	0.52	0.77	<.001
Average raters (absolute)	0.79	4.7	80	81	0.67	0.86	<.001
Average random raters	0.79	4.9	80	80	0.67	0.86	<.001
Average fixed raters	0.80	4.9	80	80	0.68	0.87	<.001

Table S8. Intraclass correlation coefficient analysis for recovered vs randomly-generated parameters. *Single raters absolute: single raters using an absolute agreement definition. Each participant is rated by a different, randomly selected rater. Single random raters: single raters under the assumption that raters are randomly selected. Reflects consistency of ratings, not just agreement. Single fixed raters: raters are fixed, consistency is measured within this fixed set of raters. Average raters (absolute): the average of multiple raters using an absolute agreement definition. Average random raters: average of multiple raters assuming that raters are randomly selected. Average fixed raters: average of multiple raters with the assumption that the set of raters is fixed. ICC: intraclass correlation coefficient value. DF1: degrees of freedom for numerator. DF2: degrees of freedom for denominator. CI: 95% confidence interval bounds. P: p-value.*

Overall, these results confirmed that for the purpose of estimating choice probability, which was the key element of interest here given its role as a proxy for the accumulation of knowledge on each option, our model provided reliable estimations.

Concerning the last part of the Reviewer's question, we did not perform a test-retest experimental design, and therefore we cannot establish whether participants' behavior such as over-advising is consistent over time for the exact same sample, but rest assured we do not argue nor claim that this is the case. We observed that when we repeated the experimental design from Experiment 1 in the second experiment (No Cost condition) similar results and similar portion of over-advisers in our new sample appeared, which indicates in our view that this *may* be a common behavior under our experimental conditions, and one that is sensitive to experimental manipulations, as was shown in our second experiment.

References for this answer

7. Bavard S., Rustichini A., Palminteri S. Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Sci. Adv.* 7, eabe0340 (2021). DOI:10.1126/sciadv.abe0340
8. Anlló, H., Bavard, S., Benmarrakchi, F. et al. Comparing experience- and description-based economic preferences across 11 countries. *Nat Hum Behav* 8, 1554–1567 (2024). DOI:10.1038/s41562-024-01894-9
9. Wagenmakers, E. J. A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, 14(5), 779-804 (2007).
10. Palminteri, S., Wyart, V., & Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. *Trends in cognitive sciences*, 21(6), 425–433. <https://doi.org/10.1016/j.tics.2017.03.011>
11. Wilson R.C., Collins A.G.E. (2019) Ten simple rules for the computational modeling of behavioral data *eLife* 8:e49547, <https://doi.org/10.7554/eLife.49547>
12. McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <https://doi.org/10.1037/1082-989X.1.1.30>

MINOR Criticisms:

R2.6. 1. It would be helpful to use exact p values when reporting the results.

We understand the Reviewer's remark, but P-values are reported according to Comms Psych guidelines. We have little choice on the matter, unfortunately.

R2.7. 2. It is unclear whether the study was adequately powered for the mixed-effects linear and logistic regression models. Perhaps this could be clarified.

We are sorry for this important omission. All the information regarding sample size was provided in the preregistration, and is now included in the main paper.

Study 1:

(Page 3 - Main Manuscript)

We then used simulations of the effects of the mixed-effects linear regression using simr package (Green and MacLeod 2016). We found a partial eta square of 0.07 for PTM (i.e. smallest eta square of all anticipated effects), and simulated the model to evaluate its power in different sample sizes. We found that a sample size of N=250 would allow detection of partial eta-squared of 0.07, which was expected to be the lowest effect size in our design for individual differences effects, with alpha=0.05 and power of 90%. We set the sample size to 300 participants, as we expect that some of the participants will not meet our exclusion criteria.

Study 2:

Following pilot studies and simulations of our experimental design (Green and MacLeod 2016), we found that a sample size of $N=200$ (50 per condition) would allow detection of a partial eta-square of 0.05, which was the lowest effect size expected for individual differences, with $\alpha=0.05$ and power of 90%. We set the sample size to 300 participants, as we expected that some of the participants would not meet our exclusion criteria.

References for this answer

Green, P. & MacLeod, C. J. SIMR : an R package for power analysis of generalized linear mixed models by simulation. *Methods Ecol. Evol.* 7, 493–498 (2016).

Reviewer #3 (Remarks to the Author):

(First of all, I apologize for the delayed response due to teaching commitments and conference travels.)

R3.1. Anlló and colleagues examined the question of whether information sharing (i.e., advice) is shaped by reinforcement learning processes (Study1) and how such relationship is modulated by advice consequences (Study2; ie, whether sharing information is costly, or is related to participants' reputation/responsibility). Across both studies, individual differences associated with social anxiety and prosociality are also examined. With appropriate statistical analyses and computational modelling, the authors reported that participants were indeed willing to share information on average, and the degree to share information is linked to knowledge of the task (as reflected in better choice accuracy, and computationally derived latent variables). Advice consequences modulated information sharing, and both social anxiety and prosociality played some role in modulating the extent of information sharing.

The manuscript is overall well-written and easy to follow, and the conclusions are supported by the data analysis approach and the results. Some aspects, however, could be unpacked, improved or clarified. I list my detailed comments below and hopefully they will help revise the paper.

(ps, It would be helpful if the authors could include page numbers and line numbers in their manuscript in the future, which will make reviewers' lives a lot easier. Thank you.)

We thank the reviewer for the time and effort invested in their review, and apologize for the lack of line numbering. We will make sure to tend to this in future submissions.

R3.2. (1) Despite a generally good introduction, it lacks detail regarding social anxiety, prosociality, and advice consequences. Indeed, all aspects are the main focuses of the study and are preregistered. But among other potentially interesting traits, why did the current study

focus on social anxiety and prosociality? What are the justifications? Some brief background literature would be helpful. Similarly, the choice of advice consequences (costly, regulation, responsibility) among other possible consequences are not motivated/justified by literature.

Our previous work on advice-giving, along with other studies examining social motivations, indicated that individual differences in social anxiety and prosocial tendencies can shape information sharing and advice giving. Specifically, the work of Zaatri et al. 2022 examined systematically the relations between levels of social anxiety and advice-giving, showing that it shaped the participants' advice-giving strategies from those aimed at maximising influence to those aimed at maximizing conformity. In addition, advice-giving is often framed as a prosocial behavior, especially when no immediate monetary benefit to the adviser is presented, and therefore prosocial tendencies may be relevant to advice-giving.

Regarding the second experiment, we introduced a number of potential consequences that directly arise from our previous work. Responsibility and reputation were both implicated in Zaatri et al. and many others (Linstrom et al. 2021 for example) as important factors shaping advice giving, and monetary cost was also shown to shape advice-giving strategies (Hertz et al. 2020).

However, it is clear that our choices are not exhaustive, both the decisions regarding individual traits to examine and those regarding potential cost conditions.

We revised the introduction to highlight the notions of influence, reputation, prosociality and social anxiety:

(Page 2 - Main Manuscript)

These explanations for why individuals follow advice rely on the assumption that advice is useful. However, prioritising information quality is not the sole motivation driving delivery of advice²². One important incentive for sharing information, particularly in competitive communication settings, is the desire to enhance one's influence and reach^{2,4}. For instance, social media posts often aim to maximise social rewards obtainable from other users, and to shape their impressions^{4,23}. These patterns are shown to be flexible, as advisers are known to adapt their sharing strategy to the incentive structure, and to changes in the consequences of information sharing²⁴. In the context of interpersonal advice-giving, where information is shared with specific individuals, normative factors can come into play: people may feel pressure to share information and give advice as a signal of social commitment^{25,26}. In such situations, the fear of appearing impolite or uncooperative may drive individuals to share information, even when the information may not be useful²⁷. In addition, individual traits may shape advice-giving behaviour. For example, prosocial tendencies may shape the willingness to exert effort to help others (Lockwood et al. 2016), and high levels of social anxiety may lead individuals to overweight social risks and responsibility associated with information sharing and opt for conformity over providing informative advice¹⁸. Collectively, these findings bring into question the rationality and ecological validity of advice-seeking as a practice that

maximises the receiver's benefits (and not just the adviser's). Some recent efforts addressed this question indirectly, by showing that people do consider others' goals and perspectives in the context of sharing private information^{29,30}. Further evidence shows that advisers can be reluctant to impose their own advice over alternative advice of better quality^{2,31}. However, the extent to which advising behaviour is driven by the adviser's level of knowledge, and the contribution of other goals to this behaviour, remains unclear.

(Page 15 - Main Manuscript)

Finally, we focused here only on two individual traits, prosociality and social anxiety, and on a limited number of potential consequences of advice giving. While these decisions were informed by previous works on information sharing and advice-giving, they are not exhaustive and other traits and consequences may be important but were beyond the scope of this work.

References for this answer

- Zaatri, S., Aderka, I. M. & Hertz, U. Blend in or stand out: social anxiety levels shape information-sharing strategies. *Proc. Biol. Sci.* 289, 20220476 (2022).
- Lindström, B. et al. A computational reward learning account of social media engagement. *Nat. Commun.* 12, 1311 (2021).
- Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E. & Roiser, J. P. Neurocomputational mechanisms of prosocial learning and links to empathy. *Proceedings of the National Academy of Sciences* 113, 9763–9768 (2016).

R3.3. (2) In Study 1, and also “no cost” condition of Study 2, how exactly were participants instructed regarding their advice sharing? Were all participants well-informed with their bonus was unrelated with their decision of broadcasting? It is vital, if participants linked the results of advice broadcasted with incentives, than it was merely a non-social reward learning game. This might also explain the “over-advising” pattern observed in the task. It would be helpful to clarify this.

We acknowledge that the Reviewer is right, as this is indeed an important point. We now include all instructions to all blocks in the supplementary materials.

In particular, the instructions to participants regarding the advice procedure in Experiment 1, and in the no-cost condition, were as follows:

“After you choose a lake, you will have the opportunity to send your choice as an advice to players that will participate in a follow-up experiment. If you choose to publish your advice, it may be displayed to all participants in a follow-up experiment, involving the same lakes/vacations. Accurate advice can greatly accelerate learning, leading to greater bonus for

other players. However, inaccurate advice can hinder learning, resulting in lower bonuses for other players.”

We emphasized both the positive consequences of good advice, and the negative consequences of inaccurate advice. Importantly, advice had no implication on the participants' bonus, only their own choices regardless of their decisions to broadcast them as advice.

R3.4. (3) A few questions regarding the RL modeling:

(3a) How was the 'outcome' coded in the model? Is it 0/1 for no-win/win, or is it the actually amount (e.g., 6) used? This is important and relevant for the initialization of Q values. I checked the code on OSF, and it seems that Qs are initialized to be 0.5. This is fine if 0/1 were used for outcomes, but this might not accurately reflect the task if actual amounts were used.

Thank you for this remark. Concerning the initialization of Q-values in the model, indeed they are initialized to be 0.5, and rewards, i.e. the number of fishes caught in each trial, were divided by 10 to range between 0-1. This makes the initial 0.5 indicating an initial belief that reward is at the middle of the range.

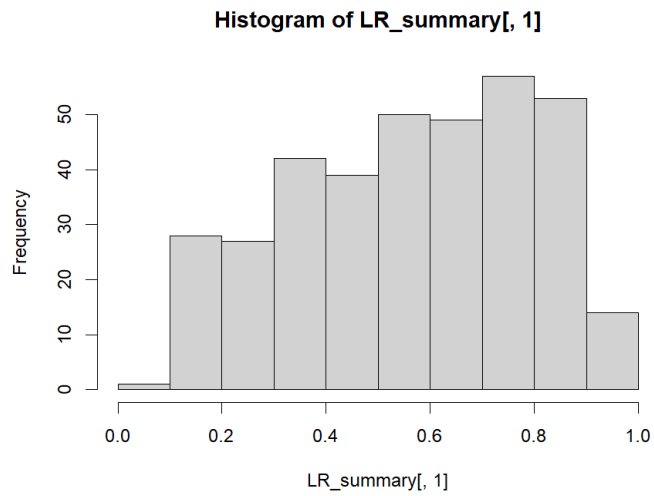
R3.5. (3b) Related to “Our models included within-subjects effects related to the task, [...], and effects that varied between participants [...]”. I looked at the Stan model code on Github, but I cannot find how this covariates were included. Also, it would be very helpful to report the parameter results (learning rate, temperature) related to the task design and individual differences. Especially I am curious to see if there is any relationship between the inverse temperature parameter and SPIN and PTM, as the authors were indeed trying to interpret some of the results in the vein of exploration v. exploitation. (see also my comment #4).

The model was fitted using hierarchical approach, where group level parameters were estimated for learning rate and inverse temperature, and individual level parameters that account for individual parameters. For each individual, the learning rates were the sum of group level parameter and individual level parameter. Individual level parameters were initialized as Gaussian with zero mean and sigma of 1, and group level parameters priors were non-informative gaussians with zero mean and sigma of 1. Learning rates parameters were calculated as logit of the free parameters to make them range between 0-1, and inverse temperature parameters were calculated as the exponent of the free parameters to make the positive.

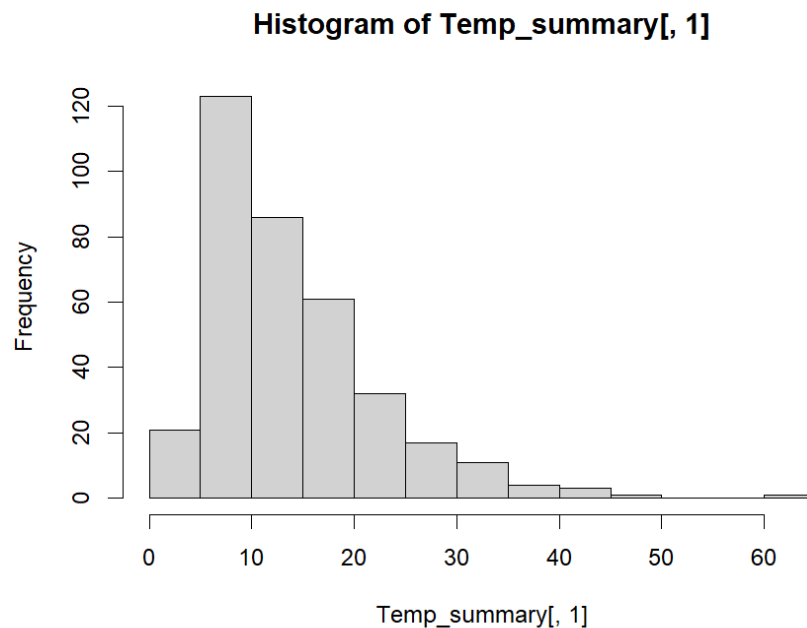
For estimation of p(choice) and PE, we used the mean of the posterior distribution of each individual parameter as a point-estimate.

We now include the individual level parameters and posterior distributions in the supplementary materials. We also report the mean and distribution of these parameters. We do not copy the edit to the manuscript here for length, but below we present the corresponding figures:

Learning Rates: mean $\pm$ std 0.56 ± 0.23

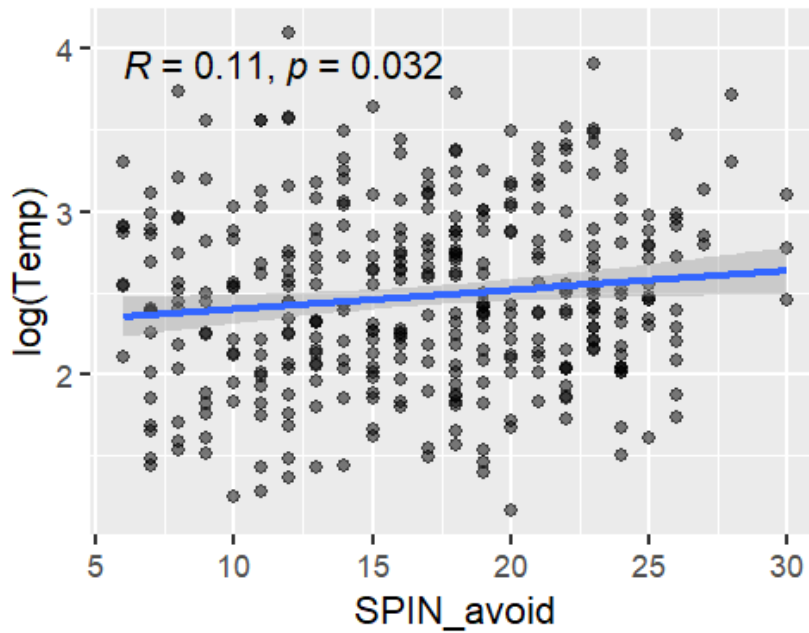


Inverse Temperature: mean $\pm$ std 13.96 $\pm$ 8.36

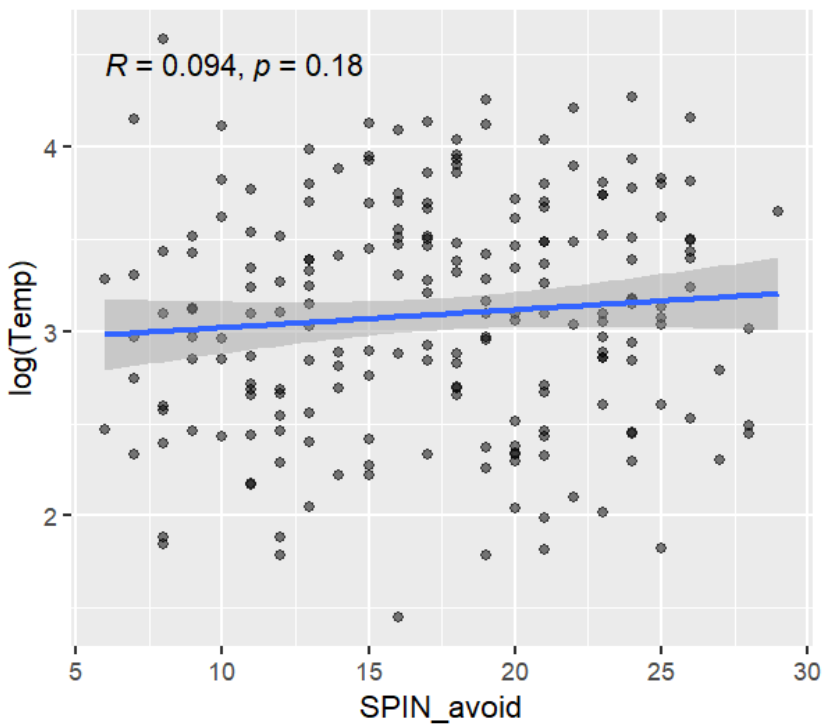


We also examined the relation between inverse temperature and social anxiety levels as an exploratory analysis in experiments 1 and 2. We found that participants with higher levels of SPIN, specifically in the avoidance subscale, showed higher levels of inverse temperature, which correspond with lower exploration rates in experiment 1, but we did not find such overall relation in experiment 2 (see below). While we present this analysis here in attention to the reviewer's comment, we consider that this exploratory analysis is beyond the scope of this paper, as it merits an independent lengthier investigation, with its own set of hypotheses.

Experiment 1 (log of the temperature vs SPIN, fear dimension):



Experiment 2 (log of the temperature vs SPIN, fear dimension):



R3.6. (3c) In SI, the description of computational modeling seems to be grabbed directly from elsewhere. I have no problem with this practice per se. But as this study did not include any

alternative models, also as the authors clarified, it is not necessary to include any description of WAIC or other model evaluation related information.

We thank the Reviewer for catching this. It was not copied from elsewhere, but rather an ancillary element from a previous analysis plan. The Reviewer is indeed correct, we agree that as a standalone measure the WAIC is not very informative. It has now been removed.

R3.7. (3d) It might be helpful to consider some work on RL in social decision-making, e.g., Olsson et al., 2020; Zhang et al. 2020; Lockwood & Klein-Flügge, 2020.

We thank the Reviewer for their very helpful literature recommendation.

R3.8. (4) Results around SPIN traits are very interesting, but some aspects are somewhat confusing.

(4a) I am not sure why PE is also included in the GLM. In the formula, before Section 2, PE is included to test H1.3; but judging from H1.3, I am not sure what role is PE playing here.

Regardless, now that PE is tested, I wonder why there is no interaction term with SPIN/PTM?

In our pre-registration we hypothesized that recent prediction error, i.e., how negatively or positively surprised the participants were with the outcome of their choice, would affect their likelihood to share their current choice as advice, in addition to the learned likelihood of receiving reward $p(\text{choice})$. This is based on previous literature showing how prediction errors affect reward-related wellbeing (Rutledge et al, 2014) and subjective estimation of one's experience (Keren et al, 2021). However, we did not predict an interaction between PE and individual traits, as our main hypothesis was that expertise, captured in the learned $p(\text{choice})$ variable, will interact with individual traits to shape advice giving behavior.

We now added this motivation to the main text:

(Page 6 - Main manuscript)

We fitted the model using uninformative priors, and used the mean of the posterior as the point estimate of the individual parameters. We then used these parameters to estimate trial-by-trial probability of choice $p(\text{choice})$ and prediction error. Probability of choice represents the accumulated knowledge about the relative rewards of the two options, and is expected to be high when choosing an option with much higher expected reward from the other option (see eq. 3). It therefore captures the knowledge of the participants about expected rewards. We therefore used this value, and its interaction with individual traits, to predict advice-giving behaviour. In addition, we hypothesised that recent prediction error, i.e., how negatively or positively surprised the participants were with the outcome of their choice, would affect their likelihood to share their current choice as advice. This last hypothesis was based on previous literature showing how prediction errors affect reward-related wellbeing (Rutledge et al. 2014) and subjective estimation of one's experience (Keren et al. 2021)

Reference for this answer

R. B. Rutledge, N. Skandali, P. Dayan, R. J. Dolan, A computational and neural model of momentary subjective well-being. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12252–12257 (2014).

Keren H, Zheng C, Jangraw D, Chang K, Vitale A, Rutledge R, Pereira F, Nielson DM, Stringaris A. The temporal representation of experience in subjective mood, *eLife* (2021) 10:e62051

R3.9. (4b) The interaction between SPIN and P(choice) seems to be one of the most interesting results in this work. But I do not fully agree with the interpretation of “[...] probability to give advice increased with SPIN (fear) score for exploitative trials (i.e., later trials)”. Indeed, judging from the data, individuals with high SPIN tend to give more advice when P(choice) is high; but it is necessarily linked to exploitative trials (later trials)? This interpretation relied heavily on the assumption that in later trials P(choice) is high – this might be true, but not obvious. I wonder if trial number can be included in this GLM? Or alternatively, can there be a way to better define/differentiate explorative and exploitative trials? This will make the interpretation stronger.

We agree that framing choices as exploratory or exploitative is not a trivial matter, and that a lot of literature is devoted to this topic. In our case, we determined whether trials were primarily exploratory or exploitative based on the classic conceptualization by Daw and others (see for example Beeler et al, 2010), where by definition exploitation consists of choosing the reward-maximising option with a probability consistently higher than chance. This seems like an appropriate decision in the context of our experimental design, where reward distributions are fixed. However, we attenuated our conclusion and rewrote it to better reflect the technical measure and be less about exploration/exploitation.

We very much agree with the Reviewer’s idea that this necessitates that in later trials P(choice) should be higher. We are thus happy to confirm this notion by fitting a regression that considers all participants across both experiments, and uses trial number and SPIN to predict the value of P(choice). Results are presented in the table below:

P(choice) ~ Trial number + SPIN + Trial number:SPIN				
	Estimate (for intercept = 0.6)	χ^2	DF	P
Trial number	0.02	1628	1	<0.001
SPIN (scaled)	-0.003	1	1	0.243
Trial number x SPIN (scaled)	0.0001	7	1	0.008

Confirming our expectations, the regression showed that choice probability increased with trial progression, and while it was not affected by SPIN directly, SPIN did have an incremental effect

as it interacted with trial number. Following the Reviewer's suggestion, we have added this operation to our work, to make our interpretation stronger:

(Page 1 - Supplementary materials)

[...]

Finally, more socially-anxious participants were more likely to share choices with high $p(\text{choice})$, i.e., when the chosen option's predicted reward was higher than the unchosen option's, which could be interpreted as more advise in exploitative trials but less likely to advise in exploratory trials(Daw et al. 2006), suggesting that more socially anxious individuals are more conservative when advising others.

INDEX

[...]

Choice probability, exploitation and trial progression.....p08

(Page 8 - Supplementary materials)

One of the main findings of the present work indicated that individuals with high SPIN tended to give more advice when $P(\text{choice})$ was high. We interpreted this phenomenon as a tendency of socially-anxious individuals to share more information when manifesting an exploitative rather than exploratory behaviour. We determined whether trials were primarily exploratory or exploitative based on a classic conceptualization in Reinforcement Learning literature⁵, where by definition exploitation tendencies consist of choosing the reward-maximising option with a probability consistently higher than chance.

This interpretation, however, relied heavily on the assumption that in later trials $P(\text{choice})$ must have been increasingly higher. To verify this assumption, we fitted a regression that considered all participants across both experiments, and used trial number and SPIN to predict the value of $P(\text{choice})$. Results are presented in the Table S4 below:

P(choice) ~ Trial number + SPIN + Trial number:SPIN				
	Estimate (intercept = 0.6)	χ^2	DF	P
Trial number	0.02	1628	1	<0.001
SPIN (scaled)	-0.003	1	1	0.243
Trial number x SPIN (scaled)	0.0001	7	1	0.008

Table S4. Impact of SPIN and trial progression on choice probability (as obtained through model fitting). Type 2 Wald χ^2 test for main effects and interactions. SPIN (fear) dimension was centered (rescaled) for convergence purposes. χ^2 = Chi square; DF = degrees of freedom; P = p-value; SPIN = Social phobia Inventory - fear dimension. N = 583.

Confirming our expectations, the regression showed that choice probability increased with trial progression, and while it was not affected by SPIN directly, SPIN did have an incremental effect as it interacted with trial number.

References for this answer

-Beeler JA, Daw N, Frazier CRM, Zhuang X, Tonic dopamine modulates exploitation of reward learning, *Front. Behav. Neurosci.*, Sec. Motivation and Reward, Volume 4 - 2010 |

<https://doi.org/10.3389/fnbeh.2010.00170>

-Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B. & Dolan, R. J. Cortical substrates for exploratory decisions in humans. *Nature* 441, 876–879 (2006).

R3.10. (4c) Relatedly, the PE results on advice-giving are somewhat counterintuitive. The data says when PE_{t-1} is high, participants are more likely to share advice. But does this symbolize an explorative advising strategy: supposedly, in earlier trials, participants hadn't learnt well, so they had higher PEs, and then in later trials, PEs were reduced. So, is it inconsistent with the finding that individuals tend to advise more in exploitative trials? As I commented above, it might be interesting to see the interaction between PE and SPIN, pending PE should indeed be included. And as a side note, Q and PE are negatively correlated by nature in RL models, so I also wonder if there may be a collinearity problem.

We understand the Reviewer's concerns. While indeed one could expect PE to diminish as learning increases, it is important to consider that PEs have a polarity. Namely, positive PEs represent a positive degree of surprise that reflects unforeseen reward maximization, and negative PEs represent a mismatch with the expected results detrimental to maximisation. With this logic in mind, it wouldn't be so surprising in our view that participants who are positively surprised with unexpected rewards would share more advice, and those who are penalized with lesser rewards decide not to share. As for further exploring the relation between PE and SPIN, while a deeper understanding of this possible relation would indeed be interesting, for the same reasons exposed in [4.a] above, we feel like this would overcrowd the models and compromise fit quality (all the more so if considering possible collinearity issues). We therefore chose to keep PE novel analyses to a minimum, enough to conform with the preregistration plan and good research practices. We would much rather keep the analyses this way, unless the Reviewer and Editor decide otherwise.

Finally, PE contributes to Q values ($Q(t+1) = Q(t) + LR * PE(t)$), and through it to p(choice). It is true that p(choice) is expected to increase as learning progresses, if the participant is always following a reward maximization process, and PE is expected to be reduced as one learns. However, due to the stochasticity of the task and participants' choices, participants sometimes chose options with low p(choice), i.e., chose the option with lower expected reward, and experienced both negative and positive prediction errors throughout the learning process. We therefore felt that a limited addition of PE (as presently included) was an important control variable.

R3.11. (5) For an interpretation of Study 3, “In the responsibility consequence condition, we observed that advice and performance results were very similar to the No cost condition, which may contribute to the idea that responsibility may be implicitly driving advising even when not explicitly mentioned.” This is true in some aspects, however, it is not very convincing in another aspect, e.g., the ratio of over-advising, and the asymmetry between the relationship between PTM and advice calibration. Thus, it is reluctant for readers to be convinced that they are “very similar”, because if advising is largely driven by responsibility, then there would not be so many over-advisors. Besides, based on this similarity to point out the idea that responsibility driving advising is not reasonable. Also, relevant statistical results showing non-significant differences between these two conditions were not provided.

The Reviewer is correct in all accounts, and some clarifications on our behalf are in order. Concerning the first point raised by the Reviewer, that the ratio of over-advising indicates that free advice is not inherently responsible, this is indeed a fair point, and we updated the text to reflect this. What we should have said was that *within people who do not over-advise*, unmotivated advising appears to be close to responsible advising. This has now been specified in the manuscript (see below).

In addition, it is true that PTM and advice calibration appear to reveal differences between these conditions, at least on some accounts. Therefore, we have now amended our manuscript to clarify the actual reach of our affirmation and its limitations. In particular, we have also added a Bayes Factor analysis (through the Bayesian Information Criterion approximation to the Bayes Factor) to estimate the strength of evidence for this difference, as an evidence strength check. Interestingly, while the difference between these conditions is indeed significant from a frequentist point of view ($\chi^2(1) = 6$, $p = 0.014$), the Bayesian Information Criterion approximation to the Bayes Factor indicates that in terms of evidence strength said difference is barely worth mentioning ($BF < 0.001$). For reference, the strength of the evidence for a difference between the No Cost and the Monetary conditions is $BF = 151259$. Overall, we interpret this to mean that despite significant finer-grained differences in the calibration of advice, driven by PTM scores, the general advising behavior between both conditions is demonstrably similar.

(Page 13 - Main manuscript)

3.3 Discussion

[...]

In the responsibility consequence condition, **we observed that within individuals who did not over-advise**, advice and performance results were very similar to the No cost condition (see Supplementary Materials - Differences between No Cost and Reputation conditions (Experiment 2)). Despite significant finer-grained differences between these conditions (i.e. the effects driven by prosociality), the broad similarity between them may contribute to the idea that responsibility may be implicitly driving advice even when not explicitly mentioned for those participants who were mindful of not sharing advice indiscriminately.

(Page 1 - Supplementary materials)

INDEX

[...]

[...]

Differences between No Cost and Reputation conditions (Experiment 2).....p09

(Page 9 - Supplementary materials)

Our findings show that within people who do not over-advise, unmotivated advising appears to be close to responsible advising. We posit it is possible that the awareness or default responsibility that led said participants to not over-advise in the first place may have also influenced the degree of responsibility with which they decided to give advice every trial. At the same time, we did observe that the relation between PTM and advice calibration appear to reveal that these conditions are indeed different. Therefore, we decided to conduct a Bayes Factor analysis (BIC approximation to BF, as before) evaluating the strength of the general difference between the two conditions. Interestingly, while the difference is indeed significant ($\chi^2(1) = 6$, $p = 0.014$), the Bayes Factor indicated that in terms of evidence strength it was barely worth mentioning when all model predictors were considered ($BF < 0.001$). For reference, the strength of the evidence for a general difference between the No Cost and the Monetary conditions was $BF = 151259$. Overall, we interpret this to mean that despite significant finer-grained differences in the calibration of advice, driven by PTM scores, the general advising behavior between both conditions is demonstrably similar.

Minor:

R3.12. - Introduction: “[...] will often rely more on advice than even their own experiential information”. This might not be always true, as a large body of literature has shown that advisees tend to weigh their own beliefs over external advice (i.e., the well-known 'egocentric bias', Morin et al., 2021; Rader et al., 2017; Yaniv et al., 2000).

This is a very fair point. We now acknowledge this when the topic is presented in the introduction (note the added citations 6-8 are the one the Reviewer provided).

(Page 2 - Main manuscript)

Introduction

[...]

Advice distinguishes itself from social observation in that it is deliberately shared by an adviser, who takes on the role of influencing the receiver's behavior and beliefs. In many well-explored settings, it has been observed that advisees tend to favour their own beliefs over external advice^{6,7,8}. However, recent research has shown that on occasion,

when advice is conveyed in the context of a two-armed bandit task, individuals will often preferentially follow advice rather than imitating observed behaviour^{9,10}, and will often rely more on advice than even their own experiential information^{11–13}.

R3.13. - The notion of the softmax inverse temperature is not consistent between the main text (called beta) and the SI (called lambda).

We thank the Reviewer very much for this very important remark. It appears that there has been an error in the typesetting of the document, which has now been corrected.

(Page 11 - Supplementary materials)

COMPUTATIONAL MODELING

We operationalized our Reinforcement Learning (RL) models through STAN⁵, which facilitates Bayesian statistical inference because it allows for the specification of complex statistical models together with their corresponding posterior inferences. Our RL model quantified how individuals updated the perceived value (Q-value) of different options based on choice outcomes (rewards). Thus, it was structured around two critical parameters: the learning rate (α) and the inverse temperature (β).

R3.14. - Fig1B, how did the “reputation” task work? Why the adviser is rated by 3 people? This is not explicit in the description.

We revised the figure legend to include the relevant information:

Figure 1. A. Trial outline. In all experiments, participants had to choose between two lakes with the goal of maximising their fishing yield. Before receiving feedback participants could choose to publish their choice as advice for future participants undergoing the same learning task. **B. Advice consequences.** Study 1 allowed participants to give advice without specifying advice consequences. Study 2 included a replication of this condition (“No cost [replication]”), and additional between-participant conditions where advice came with different consequences. In the “Monetary” consequence context, participants had to pay 1 fish (£0.003) for publishing advice. In the “reputation” consequence context, participants were told that an adviser rating would be calculated on a trial-by-trial basis based on their advice, and that their adviser ratings would be made public with their advice for future participants. In the “Responsibility” consequence context, participants were made aware that their advice would affect the choices of future participants.

References:

- Lockwood, P. L., & Klein-Flügge, M. C. (2021). Computational modelling of social cognition and behaviour—a reinforcement learning primer. *Social Cognitive and Affective Neuroscience*, 16(8), 761-771.
- Morin, O., Jacquet, P. O., Vaesen, K., & Acerbi, A. (2021). Social information use and social information waste. *Philosophical Transactions of the Royal Society B*, 376(1828), 20200052.

Olsson, A., Knapska, E., & Lindström, B. (2020). The neural and computational systems of social learning. *Nature Reviews Neuroscience*, 21(4), 197-212.

Rader, C. A., Larrick, R. P., & Soll, J. B. (2017). Advice as a form of social influence: Informational motives and the consequences for accuracy. *Social and Personality Psychology Compass*, 11(8), e12329.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, 83(2), 260-281.

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, 15(6), 695-707.

Please find below our point-by-point responses (**in black**, in-text modifications **in red**) to all of the Editor's and the Reviewers' remarks (copied **in blue**).

Best regards,

Hernán Anlló, Gil Salamander, Stefano Palminteri, Nichola Raihani, Uri Hertz

Reviewer #1 (Remarks to the Author):

I appreciate the authors' efforts to revise the manuscript in response to my concerns. However, the description of the computational modeling remains unclear, and certain parts are even contradictory. I would like to request that the authors include the mathematical descriptions (i.e., equations) to clarify the following points.

R1.1. > "We used non-informative priors, where the learning rate was taken as the logit of a parameter α_{logit} with a prior of Gaussian ..., and the beta was an exponential of a parameter β_{exp} with a prior of Gaussian..."

Logit or exponential transformations of the Gaussian are not "non-informative." Please double-check the definition of the non-informative prior.

R1.2. > "For individual-level noise variables, we used a Gaussian prior with a mean of 0 and a sigma of 1."

It is unclear what the "individual-level noise variables" refer to, as I could not locate relevant information in the manuscript.

R1.3. The authors stated that a hierarchical modeling approach was used, but the structure remains difficult to understand. For instance, are the parameters, α_{logit} and β_{exp} , individual-level or group-level parameters? How are the individual-level parameters assumed to be drawn from their population-level parameters? Where and how is the Cholesky decomposition applied within the hierarchical model? There are many ambiguities in the current description of the computational modeling. Once again, I request that the authors provide the mathematical descriptions (i.e., equations) of the hierarchical model (e.g., <https://www.princeton.edu/~ndaw/d10.pdf> and <https://osf.io/preprints/psyarxiv/vj2wp>).

We thank the reviewer for this opportunity to further clarify our process and improve our manuscript. The reviewer has requested further clarifications and corrections on three main issues concerning the model fitting process: a correction on the use of the term "non-informative priors" and exponential transformations, the use of individual-level variables, and a lack of clear equations to outline our model structure (along several other related queries). These points all revolve around our use of Bayesian hierarchical model fitting, and the way in which we present our models and analyses. Therefore, rather than addressing each question individually and atomizing our answer, we think it will be more clear to answer everything in tandem, by amending our analysis section accordingly (see revised and elaborate description now part of the methods section below). We are certain that the reviewer will easily identify each of their queries as they go over the newly-amended text. We hope that this time we have managed to produce sufficient clarifications, but we remain at your full disposal should you require further amendments or corrections.

To analyse learning computationally, model fitting was carried out hierarchically for all participants using Bayesian parameter optimization methods, implemented in STAN⁴⁵. We fitted a classic Q-learning model to choice data^{46,47}. Participants learn the Q value of each option by trial and error, updating their estimation of the value of an option after observing its outcome. Initial Q values were set to 0.5, and rewards values divided by 10 to range between 0 and 1. Value update required the estimation of a prediction error, which is the difference between the currently experienced reward and the current Q value:

$$[1] \quad PE_t = Reward_t - Q(choice)_t$$

The Q value is then updated using an individual learning rate α_i , where i is the individual index:

$$[2] \quad Q(choice)_{t+1} = Q(choice)_t + \alpha_i \cdot PE_t$$

Decisions are made according to a softmax function, resulting in an estimation of the probability of the participant to make a choice, governed by individual inverse temperature parameter β_i :

$$[3] \quad p(choice)_t = \frac{\exp(\beta_i \cdot Q(choice)_t)}{\exp(\beta_i \cdot Q(choice)_t) + \exp(\beta_i \cdot Q(unchosen)_t)}$$

(i.e. high β value indicates a deterministic decision rule, while values that are close to zero indicate chance-level decision).

Using hierarchical model fitting⁴⁷, individual parameters were derived from group level parameters, with a covariance matrix Σ . As α should range between 0 and 1, for the estimation we used a logit transformation $\alpha = \frac{1}{1 + \exp(-\alpha^r)}$ that allows α^r to be normally distributed. As β should be positive, we used an exponential transformation $\beta = \exp(\beta^r)$ that allows β^r to be normally distributed:

$$[4] \quad \begin{pmatrix} \alpha_i^r & \beta_i^r \end{pmatrix} \sim N\left(\begin{pmatrix} \alpha_{group} & \beta_{group} \end{pmatrix}, \Sigma\right)$$

The covariance matrix includes the standard deviation of the learning parameter σ_α and the temperature parameter σ_β and their correlation ρ :

$$[5] \quad \Sigma = \begin{pmatrix} \sigma_\alpha & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta \end{pmatrix}$$

To calculate the individual level parameters based on a covariance matrix, we used Cholesky decomposition⁴⁸ to derive the lower triangular square root L w of a correlation matrix C :

$$[6] \quad C = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

To derive the contribution to individual level parameters, we used individual parameters z_α, z_β which are independently drawn from $N(0, 1)$ distribution, and multiplied them by matrix L and the diagonal matrix that includes the group level standard deviations of learning parameters:

$$[7] \quad \begin{pmatrix} \alpha_i^r & \beta_i^r \end{pmatrix} = \begin{pmatrix} \alpha_{group} & \beta_{group} \end{pmatrix} + (diag(\sigma_{alpha}, \sigma_\beta) L(z_\alpha, z_\beta))^T$$

The hierarchical model therefore aims to estimate the posterior distribution of parameters set θ which included the group level parameters $\{\alpha_{group}, \beta_{group}, \sigma_\alpha, \sigma_\beta, \rho\}$ and all individual level parameters $\{z_\alpha, z_\beta\}$, which minimizes the posterior probability of the data given these parameters. This is calculated by using the likelihood of participants choices:

$$[8] \quad p(choices_1, ...choices_n | \theta) = \prod p(choices_i | \theta)$$

where individual choice probabilities are calculated according to equation 3.

The posterior probability distribution of all parameters is calculated:

$$[9] \quad p(choices_1, ..., choices_n) \propto p(choices_1, ...choices_n | \theta) p(\theta)$$

And poster probability distribution of individual level parameters is estimated:

$$[10] \quad p(choices_i, \theta) = p(\alpha_i, \beta_i) p(\alpha_i, \beta_i | \theta)$$

We set the priors for the group level learning rate and inverse temperature to be normally distributed $\alpha_{group} \sim N(0, 2)$, $\beta_{group} \sim N(0, 2)$. The standard deviations were exponentially distributed $\sigma_{alpha} \sim Exp(1)$, $\sigma_{beta} \sim Exp(1)$. The Cholesky parameters were distributed according to a Choleski LKJ distribution $\rho \sim ljkCholesky(2)$. Individual random effects were Z distributed.

We used the mean of the posterior probability distribution of individual parameters (equation 10) as the point estimate of the individual parameters. We then used these parameters to estimate trial-by-trial probability of choice $p(choice)$ and prediction error.

Reviewer #2 (Remarks to the Author):

R2.1. The authors have adequately addressed all of my comments.

We thank the Reviewer for the interest they showed in our study, and their truly valuable comments.

Reviewer #3 (Remarks to the Author):

The authors have done a very thorough job at addressing my previous comments. They appear to explain the literature much clearer, make the results stronger, and support the conclusion and interpretation better. I have some additional questions for clarification.

We thank the Reviewer for their key contributions in improving our manuscript.

R3.1. re: R3.4. Thanks for clarifying the details of initial Q values, as motivated by dividing reward magnitudes by 10 to range between 0-1. It'd be helpful to include this information in the main text (e.g., page 6).

We have now added this information as requested.

(Page 6 - Main Manuscript)

Our starting point was to evaluate the Q-values for chosen and unchosen options, and its prediction error (PE), on a trial-by-trial basis. **Model fitting was facilitated by dividing rewards by 10, as to keep reward values between 0 and 1.**

R3.2. re: R3.10. “[...], it wouldn’t be so surprising in our view that participants who are positively surprised with unexpected rewards would share more advice, and those who are penalized with lesser rewards decide not to share.” This made me wonder if it may make sense to test a learning model with double learning rates (for positive/negative PE)? And generally why in the first place did the authors decide only to test one model without comparing it to other alternatives? I am aware that including the double learning rates model deviates from the pre-reg, but it would be helpful to justify in the main text why only one model is tested.

This suggestion is indeed very good. Incidentally, we are currently examining this dimension in a different project, where learning about different social information sources is examined.

As for the present work, we decided not to fit novel models that lied outside of our preregistered hypotheses. The main focus of our paper was not the evaluation of the specific learning processes in the task, which are well established, but to use said processes to uncover how learning shaped advice-giving. Since we had no prior cognitive hypothesis related to how or why multiple learning rates would or should affect the process of advice-giving, we decided against testing multiple learning models as not to distract from the main hypotheses.

R3.3. I appreciate the new results presented in SI, but most supplementary Tables and Figures are not yet referenced in the main text.

We have now added references to the supplementary information as requested.

(Page 8 - Main Manuscript)

We fitted the model to participants' choices, and recovered individual learning rates and temperature parameters (see *Computational Modelling in Supplementary Materials* for full details of the model fit and parameter distribution).

(Page 12 - Main Manuscript)

We proceeded to fit the Q-learning model for estimation of choice probability ($P(\text{choice})$) and prediction error (PE) for each trial of each experiment (see *Computational Modelling in Supplementary Materials* for full details of the model fit and parameter distribution).

Please find below our point-by-point responses (**in black**, in-text modifications **in red**) to all of the Editor's and the Reviewers' remarks (copied **in blue**).

Best regards,
Hernán Anlló, Gil Salamander, Stefano Palminteri, Nichola Raihani, Uri Hertz

Reviewer #1 (Remarks to the Author):

The authors have adequately addressed all the concerns I raised.

We thank the Reviewer for their commentaries. The manuscript is better thanks to them.