Abstract Rule Learning in Gain and Loss Frames

Jay Gandhi

Thesis Advisor(s): Daniel Kimmel, M.D., Ph.D, and Daphna Shohamy, Ph.D

Undergraduate Thesis for the Psychology-Neuroscience Honors Program

Columbia University in the City of New York

Submitted in May 2020

Acknowledgements

I would like to express appreciation for my thesis advisors, Dr. Daniel Kimmel and Dr. Daphna Shohamy for their continued support, insight and guidance. Without their mentorship and commitment at each stage of the process, this thesis could not have been realized.

I would also like to extend gratitude to Dr. Nim Tottenham, whose encouragement and guidance kept this thesis on schedule. Her enthusiasm kept me and fellow students motivated through the two-year program.

Heartfelt admiration must additionally go out to everyone at the Shohamy Learning Lab, especially those who stomached ill-rehearsed practice talks, and provided insightful feedback on the science and presentation.

Finally, a special thanks to my fellow members of the Psychology-Neuroscience Honors Program for their moral support and constructive recommendations along the way.

***

Abstract

The ability to abstract rules from prior experiences and apply them in new contexts is critical to adaptive behavior. Prior research has established that the way in which options are framed (i.e. in terms of gains or losses), can shape decisions, and possibly learning; however, the extent to which such framing influences the ability to learn and apply rules remains unknown. In this study, participants learned the correct response to each of several images that were linked by a hidden rule. Though not necessary, this rule could be exploited to maximize performance. As feedback, participants received positive and negative reinforcement (monetary gains and losses) in alternating sessions. We found a significant impact of gain session history on performance, and but a marginal impact of the current frame. To investigate whether these framing sensitivities were related to those classically shown in the decision-making literature, participants also completed a canonical financial decision-making task. We replicated classical findings – participants were more risk-averse in the gain frame – but the relationship with framing effects on inference was equivocal in the present sample. Our results suggest that rule-acquisition and application might be differentially sensitive to framing, such that gain framing appears to improve *acquisition* more than *application*. Further investigation is required to determine the extent to which framing sensitivities in the contexts of inference and risky-choice relate.

Introduction

The question of how humans learn to behave in novel situations is central to understanding adaptive behavior. Consider, for example, how one might infer what is appropriate to wear to a new workplace. Though one has never experienced this workplace directly, one can *infer*, based on experiences with *prior* workplaces, what is appropriate to wear to this *new* one. This differs from trial-and-error learning, in which one would need to try out various attires to gauge what is appropriate in the new context. The process at stake here involves first learning an abstract rule from a set of related experiences, and subsequently, applying this rule so as to guide behavior in novel situations (see e.g., Gluck et al., 2002; Schapiro et al., 2013; McKenzie et al., 2014; Wikenheiser et al., 2017).

Whether the goal is reinforcing a target behavior or acquiring an abstract rule, learning generally is a response to positive or negative feedback (Skinner, 1963). Through operant associations of action and outcome, one can learn to predict likely outcomes of behaviors, and thereby behave in ways that maximize one's chances of favorable outcomes.

In light of decision-making, the frame of options, i.e., whether they are described as gains (rewards) or losses (penalties), has been shown to have profound effects on choices. Classically, this has been shown in single-shot decision-making, wherein subjects are more likely to choose a safe, low-paying option over a risky, high-paying option when the choice is framed as a potential gain rather than loss, even when the net payout is equivalent between frames (Tversky & Kahneman, 1979, 1981, 1986; De Martino, Kumaran, Seymour, & Dolan, 2006).

Critically, however, these single-shot decision-making studies did not investigate learning. In fact, recent studies have suggested that trial-and-error learning might be equally sensitive to positive and negative feedback (Palminteri et al., 2015). Using neural and computational approaches, the researchers proposed that participants integrate choice values with the overall subjective value of a learning environment so that rewards and punishments are relatively, rather than absolutely, encoded.

What remains unknown, then, is whether the way that outcomes are framed influences other kinds of learning, such as inference. The findings here could be clinically relevant, for several psychiatric illnesses (e.g., PTSD, GAD) manifest disruptions in rule-acquisition or -application (see e.g., Rubin et al., 2008). Further, several are typified by loss-minimizing outlooks, rather than gain-maximizing ones (Maner et al., 2007; Charpentier et al., 2017).

To investigate the extent to which gain-loss framing influences the ability to abstract and apply rules, we conducted an experiment in which participants made decisions about images that were linked by a latent rule. Though not necessary, abstracting and applying the rule led to better performance and higher payoffs. In alternating “Gain” and “Loss” sessions, participants either started with no money and sought to gain monetary rewards for correct responses or received an endowment and sought to avoid monetary penalties. Next, in order to examine whether individual differences in framing effects on inference were related to classical framing effects shown for decision-making under risk, we also administered a single-shot decision-making task that assessed each participant’s risk-aversion under Gain and Loss frames.

Methods

Healthy adult participants were invited to participate for pay or course credit at Columbia University. Since we did not expect gender-based differences, we recruited equally across genders. After providing consent, participants completed two computer-based tasks programmed using PsychoPy (Pierce, 2008) and run on 23-inch iMac computers (Apple Inc., Cupertino, CA).

Inference Task

The first task was a modified reversal task, in which participants were instructed to learn the correct responses to a set of images. The image set included two exemplars from each of four image categories: faces, scenes, body parts and tools. Participants had to indicate their response to a displayed image on every trial with the left (F) or right (J) key within 6 seconds, following which, they learned the outcome of their choice. Between trials, there was an inter-trial interval (1.5s), and a fixation screen (0.5s). Trials were randomly interleaved such that on any given trial, a participant was equally likely to see any of the eight images.

Participants were forewarned that the correct responses could change during the task. However, they were not informed that these changes occurred in reliable ways. Nonetheless, if participants abstracted the latent rule, they could infer correct responses so as to earn more money than would be possible by simple trial-and-error learning. The rule entailed a hierarchy such that pairs of image categories shared either the same correct response or the same outcome magnitude. For example, in one session, faces and scenes might share a leftward response, while faces and body parts might share a low outcome magnitude of 5¢ (see Figure 1).

To test whether participants learned the rule, the contingencies change without notice after blocks of about 25 trials (exponentially distributed; minimum 10 trials/block). Essentially, the correct responses reverse such that images that previously had a leftward response (faces and scenes) now have a rightward response, while images that previously had a rightward response (faces and body parts) now have a leftward response. The outcome contingencies change too such that they are preserved for one pair and reversed for the other (see Figure 1). Thus, while faces and scenes still share a response (rightward in the current block), faces and tools now share an outcome magnitude. Participants thus undergo 4-8 contingency switches, alternating between the two contingency sets or *contexts*, to complete a session.

Each image in the task is defined on three variables: (1) its image category, (2) the correct response, and (3) the outcome magnitude possible (image value). Since pairings are distinct, knowing two of these variables, one could determine the third. The interrelationships between images, i.e. the temporally-defined context, is implicit as a hidden fourth variable. Moreover, from a *single* image-response pairing, it is possible to determine the current context and the correct responses to *all* the images at the given time point.

Participants completed 4 sessions of the task. From one session to the next, the images within each image category changed. In addition, the ways in which image categories were linked changed, so participants could not use the precise relationships between image categories from one session on the next. However, the general structure of the rule remained the same across sessions (i.e., two image categories shared the same response, the other two shared the opposite response, etc.), thus participants could learn this meta-level structure in one session and apply it to acquire the specific instance of the rule in a later session.

Critically, the sessions alternated between versions in which feedback was framed in terms of monetary rewards (Gain frame) and versions in which feedback was framed in terms of penalties (Loss frame). Session orders were counterbalanced across participants such that half started with a Gain session (Group A) and the other half with a Loss session (Group B). During sessions in the Gain frame, participants saw the message “Win X¢” for a correct response and “Don’t win X¢” for an incorrect one. Conversely in the Loss frame, participants saw the message “Lose X¢” for a correct response and “Don’t lose X¢” for an incorrect one. The amount that they did *not* win (or lose) was reported so that feedback on each trial was equally informative (i.e., whether the response was correct or incorrect and whether the frame was Gain or Loss).

Ahead of each Loss session, participants were endowed with $30 from which penalties were deducted (Gain sessions had no endowment). The endowment was selected such that, for equivalent accuracy, Gain and Loss sessions resulted in the same net payout. At the end, the participant received a bonus payout based on one randomly selected session.

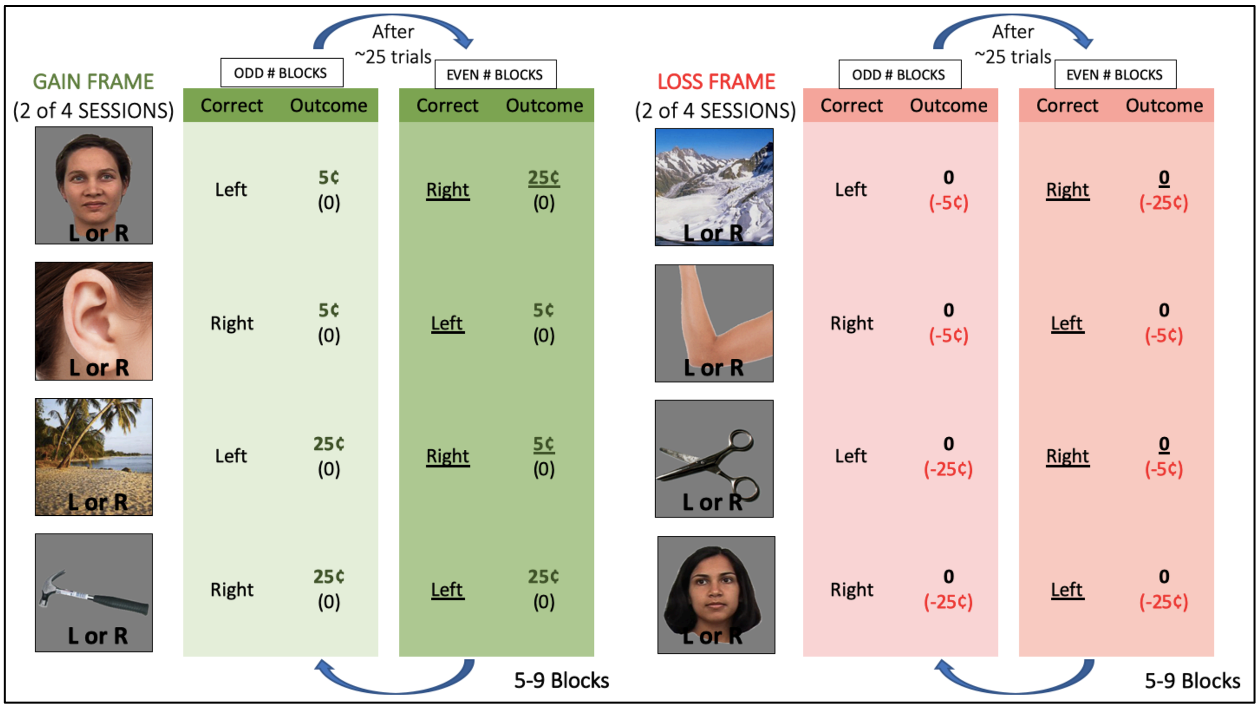

Figure 1. The image categories, correct responses, and outcomes in Gain and Loss frames. Values in parentheses indicate the outcome of the incorrect option. From one session to the next, the image exemplars change, as do the precise pairings of image categories. Thus, a participant cannot use the precise relationships learned in a previous session, but can use the general structure of the rule.

Modeling Inference Performance

In this task, the goal for participants was to learn the correct response to each of several images. Since the correct responses changed block-wise, participants had to update their beliefs after each block transition based on feedback. They could do this independently for each image (by trial-and-error learning), or they could exploit the abstract rule, such that after detecting a change for one image, they could infer the correct responses to all the other images. To test for inference, we looked specifically at the first encountered images of the four image categories within each block. These trials were considered specifically because trial-and-error learning and inference would predict divergent responses on these trials.

Irrespective of approach, performance on the first trial of each new block should be poor since block changes were un-cued. On subsequent trials of interest, if a participant were updating beliefs independently, performance would continue to be low. This is because with an independent updating mechanism, feedback on any trial could only improve performance on subsequent encounters with images of the same category. Thus, if a participant were only using trial-and-error learning, their responses on these trials would be based on beliefs that were last updated in the previous block, i.e., when the correct response was different. However, if participants learned the rule, they could infer beliefs about *other* images in the absence of *direct* reinforcement. This could occur after feedback to just the first image within each new block. Consequently, on the first encounter with the *next* category, performance could be above chance.

To assess performance on these trials, we fit a hierarchical logistic regression model to accuracy data (treating participants as random effects), with terms representing trial order, current session frame, and frame history. To achieve this, we used MATLAB's *nlmefitsa* function that uses the stochastic expectation-maximization algorithm to estimate parameters.

$$P(correct) = \delta\ logit^{-1}(ImOrder + isCurrGain + isPrevGain + 1) + \frac{1-\delta}{2} \quad (1)$$

Where $ImOrder$ is an indicator for the first encountered image of the image categories within each block, with 0 signifying the first image (on which inference is unlikely) and 1 denoting each of the 3 subsequent images (on which inference is *required* for correct responses); $isCurrGain$ is a binary variable whose value is 1 when the current session is a Gain session and 0 when it is a

Loss session; $isPrevGain$ is a binary variable that stores 1 when at least one prior session has been a Gain session; and δ, constrained to the interval [0, 1], represents the maximal proportion of correct responses. It is used since a participant's performance may maximize at a value lower than 1 due to unintentional errors (at a rate specified by $\frac{1-\delta}{2}$), and for the same reason, their accuracy may not be exactly 0 on the first trial of each block. We also fit this model using participants' *cumulative* prior Gain session experience, but the Bayesian Information Criterion (BIC), estimated by Importance Sampling, revealed $bitPrevGain$ to be more informative.

Single-Shot Decision-Making Task

Participants next completed a classical frame-sensitivity task (De Martino et al., 2006) to assess the influence of framing on single-shot decision-making under risk. On each trial, participants first received an endowment. Next, they chose between a sure-bet, which guaranteed a portion of the endowment, and a gamble, which provided the entire endowment but with a stated probability. Thus, if the gamble is selected, there also remains a chance of losing the entire endowment. On each trial, the expected values of the sure-bet and gamble options were equal, such that if the participant received an endowment, E, and the gamble offered a probability, p, of keeping the full endowment, then the sure-bet was set at Ep. Critically, the frame was manipulated such that on half the trials, the sure-bet option was framed as the choice to *keep* Ep of the endowment, and on the other half, it was framed as the choice to *lose* $E(1-p)$ of the endowment. Participants did not receive feedback after their choice. At the end of the experiment, one randomly selected trial was selected for a bonus payout.

Four endowments ($25, $50, $75 and $100) were used with each of four gamble probabilities (0.2, 0.4, 0.6 and 0.8) and two frames (gain and loss). Participants repeated each trial-type for a total of 64 trials. We included 16 additional "catch" trials, in which one option was clearly better than the other to ensure that participants were responding diligently. On these, the gamble probability was set at 0.05 or 0.95, and the sure-bet was the choice to keep (or lose) an amount equal to half the endowment. These trials were excluded from analyses.

Assessing Framing Effects on Single-shot Decisions

To assess framing effects on decision-making under risk, we modeled the probability that a participant selected the sure-bet on each trial type using a hierarchical logistic regression model. Again, we used MATLAB's *nlmefitsa* function, treating participants as random effects.

$$P(sure) = logit^{-1}(isLoss + Endow + gambleP + 1) \tag{2}$$

Where $isLoss$ is a binary variable for the current frame that has a value of 0 when the trial is in the Gain frame and 1 when the trial is in the Loss frame; $Endow$ encodes the endowment value; and $gambleP$ encodes the gamble probability (which is equal to the proportion of the endowment kept in the sure-bet on experimental trials). Models that included any to all interactions possible were considered, but the above model was determined most suitable based on the Bayesian Information Criterion (BIC).

Results

Participants and Inclusion/ Exclusion Criteria

We excluded from the *entire* study any participant who did not demonstrate evidence of learning on the inference task. This was revealed by the absence of an upward trend in accuracy within blocks. Participants who learned diligently, whether by trial-and-error learning or inference, had a characteristic drop in performance on the trial right after a block transition and improved, at varying rates, through the block.

We separately excluded participants who appeared to be guessing on the single-shot decision-making task. We qualitatively observed a cluster of participants who had both a low accuracy on catch-trials (these trials had a clear better option) and a very quick average reaction time on critical trials. One participant elected to end the study before completion of the single-shot decision-making task but consented to our using their inference task data.

Our final sample comprised of participants (n = 24, age M = 21.46, SD = 2.34), of which 11 identified as female. Most were included in the analyses for both experiments (n = 23).

Inference Task

Participants randomly assigned to Group A (n = 13) began their four alternating sessions with the Gain frame and those in Group B (n = 11) started with the Loss frame. In Session 1, we see that participants in Group A (Gain frame) outperformed those in Group B (Loss frame). At this stage, we cannot not distinguish if the Gain frame facilitated acquisition, or acquisition *and* application of the rule. In Session 2, both groups have improved. However, the participants who were previously in a Gain frame now are in a Loss frame (Figure 2: Group A, purple lines) and yet their inference performance is greater than participants who are currently in a Gain frame (Figure 2: Group B, yellow lines). Thus, it does not appear that a Gain frame is necessary to *apply* the abstract rule. Of course, it could be that Group B's capacity for inference is fundamentally lower, or that one's initial experience with the rule determines all future performance. However, by Session 3, when both groups have now had at least one experience with a Gain session, we see that performance is comparable between the groups. It appears then that rule-acquisition, but not rule-application, is responsive to framing.

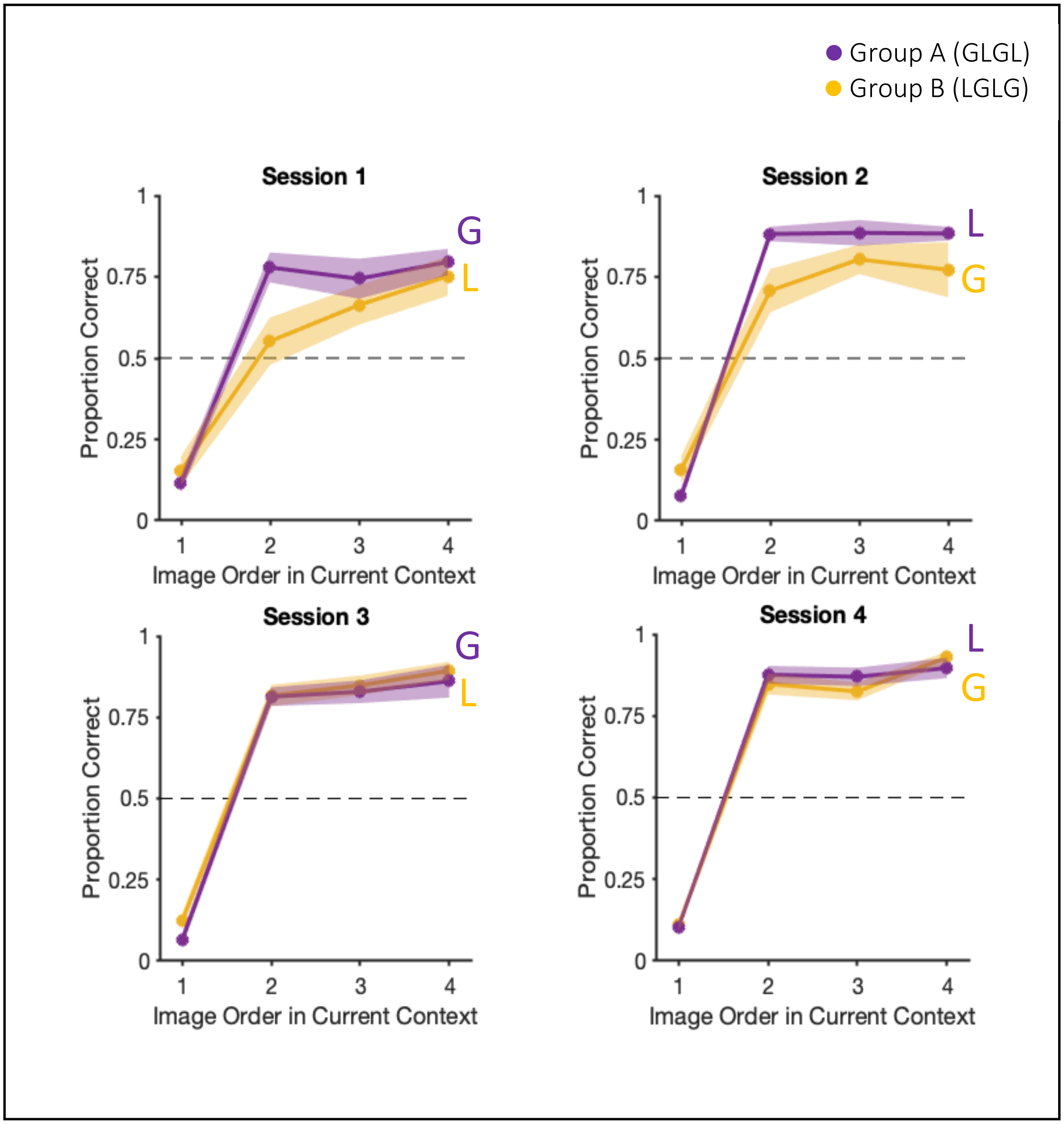

Figure 2. Inference performance by session (panels) and group (colored lines, see legend). In each panel, the average proportion correct is shown for the first encounter with each image category after a context change (i.e., inference trial), numbered by the order in which the images were encountered in a given context, then averaged across contexts and participants. Shaded error regions represent +/- 1 SEM, calculated at the participant level.

To test these ideas, we applied a hierarchical logistic model with terms for the image order within the current context, current session frame, and prior gain frame history (refer to Methods). First, we found a robust influence of image order ($\beta_{imOrder} = 7.03,\ CI_{95} = 2.59$), confirming that inference occurred on the second, third and fourth novel image encounters within the block. Second, we found a significant effect of the presence of previous Gain experience ($\beta_{isPrevGain} = 1.74,\ CI_{95} = 0.85$), i.e. a Gain frame prior to the current session improved performance on the current session. Finally, the influence of current frame was also significant, though smaller ($\beta_{isCurrGain} = 0.62,\ CI_{95} = 0.48$). Refer to open circles in Figure 3 for remaining group-level statistics.

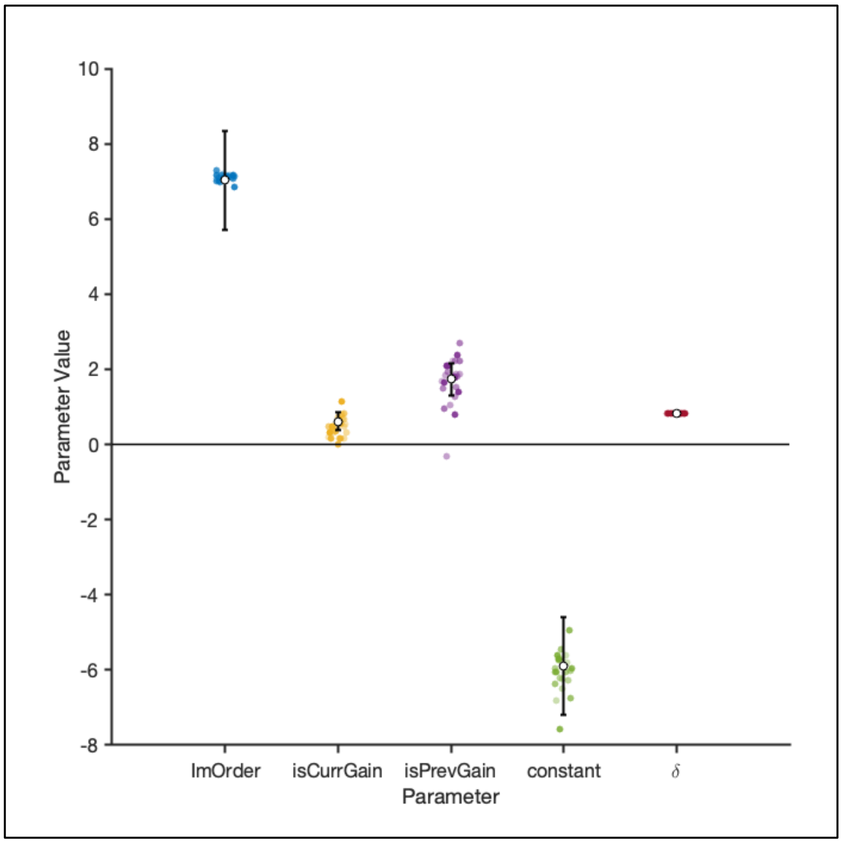

Figure 3. Parameters obtained from the hierarchical logistic regression, showing a positive effect of having encountered at least one session in the Gain-frame. Comparatively, the influence of the current session was less positive. Each colored point denotes a parameter for a single participant, while white points signify fixed-effects. Error bars represent estimated standard errors on the fixed-effects.

Single-Shot Decision-Making Task

In the ancillary task, a participant made a series of decisions between a sure-bet and a gamble under different frames. As per prior literature, we hypothesized that participants would be more risk-averse in the Gain frame and risk-seeking in the loss frame. To test this idea, we modeled the proportion of sure-bets with a hierarchical logistic regression model with terms for the frame, endowment value, and gamble probability. We replicated classical findings: experiencing a decision in the Loss frame appeared to lower the probability that the sure-bet was selected, as indicated by a negative $isLoss$ parameter ($\beta_{isLoss} = -1.51,\ CI_{95} = \ 0.52$). Refer to open circles in Figure 4 for remaining group-level stats.

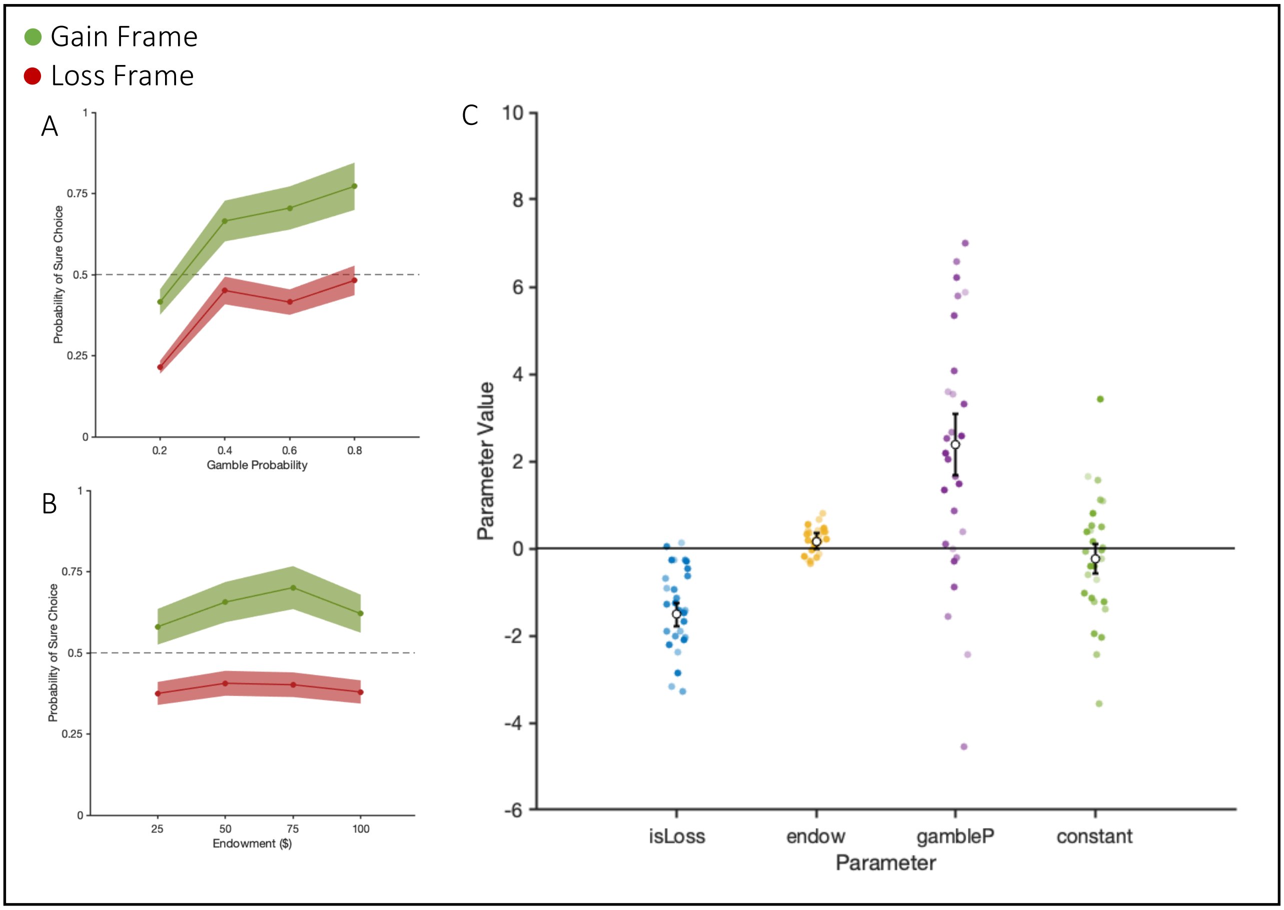

Figure 4. (A) Data from the Risky-Choice Framing Task split by frame for increasing gamble probabilities (averaged over the four endowments). (B) Data from the Risky-Choice Framing Task split by frame for increasing endowments (averaged over the four gamble probabilities). Shaded error regions represent +/- SEM, calculated at the participant level. (C) Parameters obtained from the hierarchical logistic regression, verifying that Loss framing decreased sure-bet choice. Catch trials are excluded in plots and the model. Each colored point denotes a parameter for a single participant, while open circles signify fixed-effects. Error bars represent estimated standard errors on the fixed-effects.

Relating Frame Sensitivities

Our next objective was to consider whether frame sensitivity in the context of inference performance was related to frame sensitivity for single-shot decision-making under risk at the individual level. From a converging evidence standpoint, we considered two hypotheses. First, framing-sensitivities in the contexts of inference and risky-choice originate from a shared mechanism. This would predict a correlation between the influence of framing on inference performance and that on risky-choice. Second, these framing-sensitivities could originate from independent mechanisms. This would predict no relationship between the influences of framing on the respective phenomena.

To test these ideas, we first measured the correlation between the $bitPrevGain$ parameter from the Inference Task model and the $isLoss$ parameter from the single-shot decision-making task. These parameters were correlated, albeit weakly and non-significantly (r = -.27, p = .22). Though below the threshold of significance, the correlation is not inconsistent with the shared mechanism hypothesis. In other words, Gain framing in the context of inference (especially rule-acquisition) and risky choice *may* be related, though findings here are inconclusive.

Next, we addressed the relationship between framing effects on the *current* session and those on single-shot decision-making under risk by correlating the $isCurrGain$ parameter from the Inference Task model with the $isLoss$ parameter from the single-shot decision-making task. These parameters did not appear to be related (r = .05, p = .83), consistent with the independent mechanism hypothesis.

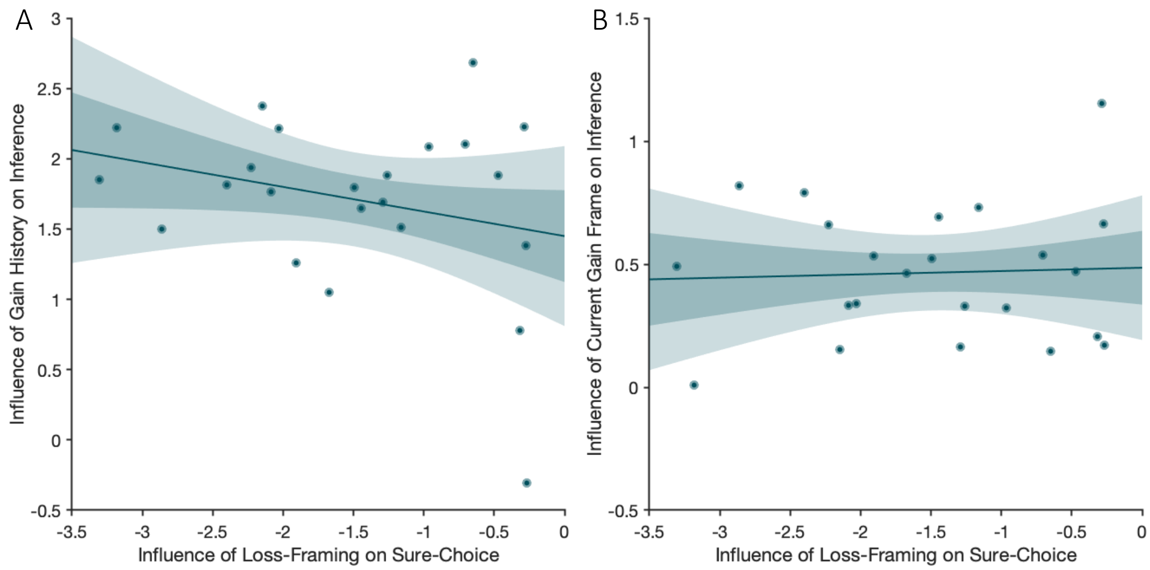

Figure 5. (A) Scatter plot of the $isPrevGain$ parameter from the Inference Task with the $isLoss$ parameter from the single-shot decision-making task, $r = -.27$, $p = .22$. (B) Scatter plot of the $isCurrGain$ parameter from the Inference Task with the $isLoss$ parameter from the single-shot decision-making task, $r = .05$, $p = .83$. In both plots, error regions represent ± 1SD and the ± 2 SD.

Discussion

In the present study, we found a significant impact of a single prior Gain experience on inference performance. Following that gain experience, however, performance appeared to not be influenced by frame to the same extent. Our results thus indicate that rule-acquisition and rule-application might be differentially sensitive to frame. Specifically, gain framing seems to improve rule-acquisition; however, framing may not have a commensurate effect on the application of a rule, once acquired.

By leveraging individual variation, we also tested the link between the influence of framing on inference and that on single-shot decision-making under risk. For prior gain framed experience, our results suggest a common source of sensitivity to frame, but the correlation fell below the threshold of statistical significance. This, however, does not preclude the possibility of a shared mechanism, since the sample size in this pilot study was limited. Further investigations should thus consider replicating these findings in a larger sample.

The hypothesis that outcome values are integrated with the subjective value of the context (Palminteri et al., 2015), thus might not readily extend to rule learning. In fact, in the context of framing, rule-acquisition appears to align better with decision-making under risk than with trial-and-error learning, and the opposite appears true for rule-application. These findings resonate with prior research that posits differing neural mechanisms for rule-learning and trial-and-error learning (Shohamy, 2008). A step further, it seems that the processes of rule-acquisition and -application might be dissociable.

In the present work, our assessment of inference is far from perfect. For one, to evaluate inference, we could only consider trials on which inference and trial-and-error learning led to different predictions. Our approach makes it difficult to address the question of how rule learning and trial-and-error learning mechanisms might interact or coexist within a session. Second, by collapsing choices into accuracy, we essentially ignore outcome magnitudes, and perhaps other enlightening aspects of the task's hierarchy.

To address such issues, our first proposition would be to develop of more quasi-mechanistic/ computational models to assess inference, perhaps by accommodating abstraction into reinforcement learning paradigms. In addition, even with a descriptive model fit to critical trials as in the present study, our task has an extensive hierarchy that may be better incorporated using Multilevel Bayesian Models, which we will be at the core of future work.

Our second proposition is to conduct imaging studies to better ascertain the neural signatures of inference, especially in terms of the dual processes of rule-acquisition and rule-application. As a working hypothesis, it might be that framing informs neural populations involved in the acquisition of rules, but not so much downstream areas that facilitate application once the rule has been learned. It could be that the areas involved in acquisition interact with others, such as the amygdala, which has been shown to be preferentially activated by loss frames in the context of decision-making under risk (De Martino et al., 2009). Research in the monkey has revealed that contingencies and variables like those in our task may be represented in (and abstracted by) the hippocampus, alongside the dlPFC and ACC (Bernardi et al., 2018 *preprint*). Interactions between each of these regions and the amygdala are noteworthy. Further, the amygdala has been shown to exert a particularly relevant influence on the hippocampus with regards to processing motivationally salient information (Zheng et al., 2017).

One implication of the present study is that it is important to acknowledge that framing seems to impact different cognitive processes differently even within the domain of decision-making. Risk assessment, incremental belief updating through experience (trial-and-error learning), learning of abstract knowledge (rule-acquisition), belief updating through abstract knowledge (rule-application), etc. all appear to be influenced by frame in different ways. Next, it might be helpful to think of inference in terms of the distinct mechanisms of rule-acquisition and -application, for even though application depends on acquisition (McDaniel et al., 2014), these phenomena seem to be governed – or at least influenced – by different processes.

References

Bernardi, S., Benna, M. K., Rigotti, M., Munuera, J., Fusi, S., & Salzman, C. D. (2018). *The geometry of abstraction in hippocampus and pre-frontal cortex* [Preprint]. Neuroscience. https://doi.org/10.1101/408633

Charpentier, C. J., Aylward, J., Roiser, J. P., & Robinson, O. J. (2017). Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. *Biological Psychiatry*, *81*(12), 1014–1022. https://doi.org/10.1016/j.biopsych.2016.12.010

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, *313*(5787), 684–687. https://doi.org/10.1126/science.1128356

Gluck, M. A., Shohamy, D., & Myers, C. (2002). How do people solve the "weather prediction" task? : Individual variability in strategies for probabilistic category learning. *Learning & Memory*, *9*(6), 408–418. https://doi.org/10.1101/lm.45202

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185

Maner, J. K., Richey, J. A., Cromer, K., Mallott, M., Lejuez, C. W., Joiner, T. E., & Schmidt, N. B. (2007). Dispositional anxiety and risk-avoidant decision-making. *Personality and Individual Differences*, *42*(4), 665–675. https://doi.org/10.1016/j.paid.2006.08.016

McDaniel, M. A., Cahill, M. J., Robbins, M., & Wiener, C. (2014). Individual differences in learning and transfer: Stable tendencies for learning exemplars versus abstracting rules. Journal of Experimental Psychology: General, 143(2), 668–693. https://doi.org/10.1037/a0032963

McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, *83*(1), 202–215. https://doi.org/10.1016/j.neuron.2014.05.019

Palminteri, S., Khamassi, M., Joffily, M., & Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, *6*(1), 8096. https://doi.org/10.1038/ncomms9096

Peirce, J. W. (2008). Generating images for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. https://doi.org/10.3389/neuro.11.010.2008

Rubin, D. C., Berntsen, D., & Bohni, M. K. (2008). A memory-based model of posttraumatic stress disorder: Evaluating basic assumptions underlying the PTSD diagnosis. *Psychological Review*, *115*(4), 985–1011. https://doi.org/10.1037/a0013397

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, *16*(4), 486–492. https://doi.org/10.1038/nn.3331

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. *Neuroscience & Biobehavioral Reviews*, 32(2), 219–236. https://doi.org/10.1016/j.neubiorev.2007.07.008

Skinner, B. F. (1963). Operant behavior. *American Psychologist*, *18*(8), 503–515. https://doi.org/10.1037/h0045185

Tversky, A, & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Tversky, Amos, & Kahneman, D. (1986). Rational choice and the framing of decisions. *The Journal of Business*, *59*(S4), S251. https://doi.org/10.1086/296365

Wikenheiser, A. M., Marrero-Garcia, Y., & Schoenbaum, G. (2017). Suppression of ventral hippocampal output impairs integrated orbitofrontal encoding of task structure. *Neuron*, *95*(5), 1197-1207.e3. https://doi.org/10.1016/j.neuron.2017.08.003

Zheng, J., Anderson, K. L., Leal, S. L., Shestyuk, A., Gulsen, G., Mnatsakanyan, L., Vadera, S., Hsu, F. P. K., Yassa, M. A., Knight, R. T., & Lin, J. J. (2017). Amygdala-hippocampal dynamics during salient information processing. *Nature Communications*, *8*(1), 14413. https://doi.org/10.1038/ncomms14413