### Table of Contents

[Study Information 1](#_Toc73046309)

[Design Plan 2](#_Toc73046310)

[Sampling Plan 4](#_Toc73046311)

[Variables 6](#_Toc73046312)

[Analysis Plan 6](#_Toc73046313)

[References 11](#_Toc73046314)

### Study Information

1. Title

Investigating the causal effects of presentation duration, sequence, and format on binary risky choice

1. Authors

Felix Molter, Peter N. C. Mohr

1. Description

Visual gaze is associated with decision making, such that alternatives that are looked at longer are more likely to be chosen. The causal direction of this association, however, is the subject of ongoing debates, with increasing evidence for a causal effect of gaze on choice. Here, we test multiple facets of this causal effect in an incentive-compatible binary risky choice task, where we externally manipulate presentation duration and sequence of choice alternatives and their attributes.

1. Hypotheses

**H1:** Presentation *duration* affects choices.

H1a: Longer presented *alternatives* are chosen more frequently.

H1b: Alternatives with higher values on longer presented *attributes* are chosen more frequently.

**H2:** Presentation *sequence* affects choices.

H2a: Alternatives presented *last* are chosen more frequently.

H2b: Alternatives with higher values on the attribute presented *last* are chosen more frequently.

**H3:** Presentation *format* affects choice processes.

H3a: Alternative-wise and attribute-wise presentation formats elicit decision processes more influenced by within-alternative and between-alternatives attribute integration, respectively.

### Design Plan

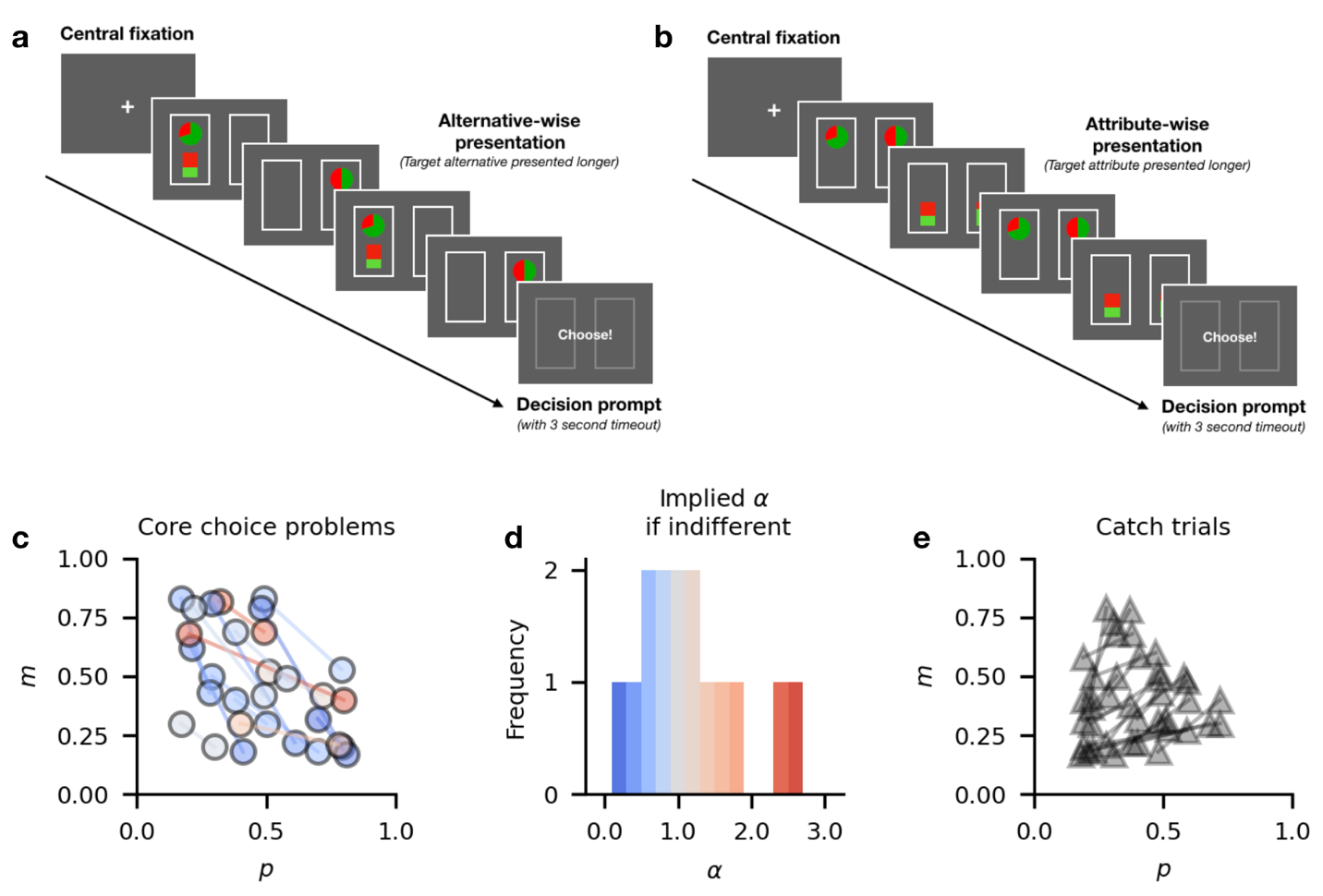


Figure 1: **Sequential presentation risky choice task and choice problems**. **a)** Alternative-wise sequential presentation. **b)** Attribute-wise sequential presentation. **c)** 15 core choice problems. Each circle represents one choice alternative, described by its probability p and outcome m. Each connected pair of circles indicates one pair of choices used to construct the 120 experimental trials. Color indicates the α value for which the expected utilities of a pair are equal. **d)** Distribution of indifference implied α values of the 15 core choice problems from **a)**. **e)** Catch trials. Each connected pair of triangles represents one of 20 catch trials, where one alternative is dominated by the other on both attributes.

1. Study type

Experiment: We perform an online experiment, where participants make repeated choices between two risky gamble alternatives. Presentation duration, sequence and format are manipulated across trials (see Study design).

1. Blinding

No blinding is involved in this study. Participants might or might not be aware of the duration manipulation in the task.

1. Behavioral task & Stimuli
   1. Task:

We will collect choice and response time data from a binary sequential presentation risky choice task, where participants make repeated incentivized choices between two graphically displayed all-or-nothing gambles (Figure 1). Each gamble represents the chance to win a monetary amount *m* with a probability *p* and nothing otherwise. Winning amounts *m* are represented by partially filled bars, and winning probabilities *p* are represented by pie charts. Each trial is divided into a presentation phase and a choice phase. During the presentation phase, participants learn about the two available gambles' attributes. Information is presented sequentially. There are two types of presentation: *Alternative-wise* and *attribute-wise* presentation: In trials with alternative-wise presentation (Figure 1**a**) both attributes of one gamble are shown simultaneously, followed by both attributes of the other gamble. In trials with attribute-wise presentation (Figure 1**b**), one attribute (e.g., winning probability *p*) of both gambles is presented simultaneously, followed by the other attribute (e.g., amount *m*) of both gambles. Presentation always alternates two times between alternatives or attributes (i.e., A-B-A-B). Crucially, one alternative (in trials with alternative-wise presentation) or attribute (in trials with attribute-wise presentation) in each trial is selected to be the target, and is shown longer than the other one. Target attributes or alternatives are always presented for 1500 ms, whereas other alternatives or attributes are presented for 1000 ms, resulting in a final presentation advantage for the target of 1000 ms (2x 1500 ms vs. 2x 1000 ms).

After the presentation phase, stimulus information is hidden and participants are prompted to make a choice between the two alternatives within 3 seconds.

After completing all choice trials, one gamble chosen by the participant in a randomly determined trial is played out for a real bonus payment.

The task is implemented in jsPsych (de Leeuw, 2015) and can be run at https://moltaire.github.io/causality\_task.

* 1. Stimuli and trials:
     1. Experimental trials: We created a set of 15 core choice problems including one alternative with a higher probability of winning a lower amount (*Hp*) and one alternative with a lower probability of winning a higher amount (*Hm*). Choice problems were created algorithmically to cover a large space of the attribute space, be maximally different from each other, and be diagnostic of different risk attitudes. For this last criterion, we controlled the *α* values for which two alternatives in a pair would have equal expected utility (using a standard power utility function). The distribution of these indifference-implied α values is shown in in Figure 1d. Core choice problems are illustrated in Figure 1c.

Then 8 trials are created for each core problem by fully crossing the factors *(i) presentation format, (ii) target alternative / attribute, and (iii) presentation order (target first and third vs. second and last).* This results in a total of 120 experimental choice trials.

* + 1. Catch trials: 20 catch trials with one dominant alternative are added for a total of 140 trials. In catch trials, individual presentation durations are set to 1250 ms, resulting in the same overall presentation duration (5000 ms), but no presentation advantage for any alternative or attribute. Catch trials are illustrated in Figure 1e.

1. Study design
   1. The study uses a fully crossed factorial within-subjects design with three factors (presentation format, duration, and order) that have two levels each:
      1. *Presentation format*: Gambles are presented alternative-wise or attribute-wise.
      2. *Target (longer shown) alternative / attribute*: Presentation duration can favor the *Hp* or *Hm* alternative, either by showing an alternative longer (in alternative-wise presentation) or by showing the attribute longer on which an alternative has a higher value (in attribute-wise presentation).
      3. *Presentation order*: The longer shown alternative or attribute can be shown first (and third) or second (and last).

For each of 15 core choice problems, 8 trials are created by fully crossing these factors, resulting in a total of 120 experimental choice trials. 20 catch trials with one dominant alternative are added for a total of 140 trials.

1. Randomization

All participants make choices for the same 140 choice problems divided into two blocks. Trial order in each block will be randomized for each participant using jsPsych's (de Leeuw, 2015) method of type *without-replacement*. Horizontal position of the alternatives is randomly determined in each trial.

Block order is counterbalanced between participants.

### Sampling Plan

1. Existing data

Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized. Pilot data have been collected and used to develop analyses and models, but will not be part of the main sample.

1. Data collection procedures

The experiment is run online. Participants will be recruited via the platform [Prolific.co](https://www.prolific.co/). Participants will be paid £3.75 for completing the study estimated to take about 30 minutes. Compensation will be raised to £5 if our sample size is not reached within 7 days of beginning recruitment. Participants have the chance to win a bonus amount ranging from £0 to £9.85, as one of their chosen gambles will be played out after the experiment. The experiment includes webcam-based eye tracking, so only participants with a working webcam can participate. The experiment requires participants to use Google Chrome or Mozilla Firefox browsers. Data collection will begin in May 2021. The task including instruction and training procedure can be run at <https://moltaire.github.io/causality_task>.

1. Sample size

Our target sample size is 200 participants. We will accept entries on the Prolific platform until 200 complete submissions are reached.

1. Sample size rationale

Sample size was determined to exceed that of previous laboratory-based experiments addressing related questions, accounting for the possibility of lower quality data due to the online setting.

1. Stopping rule

We will stop data collection 14 days after starting to accept submissions on Prolific.

### Variables

1. Manipulated variables

We manipulate the *presentation format* (alternative-wise vs. attribute-wise), longer shown *target alternative / attribute* (*Hp* longer vs. *Hm* longer in alternative-wise presentation; *p* longer vs. *m* longer in attribute-wise presentation) and *presentation order* (target alternative or attribute first and third vs. second and last).

1. Measured variables

The main variables of interest are participants' choices (indicated by key presses) and their response times (the time between decision prompt and key press) in each trial.

* 1. Webcam-based eye tracking

Participants' gaze will be recorded using webcam-based eye tracking implemented in the JavaScript library webgazer.js (Papoutsaki et al., 2016; Yang & Krajbich, 2020). To this end, the task includes a webcam setup at the start of the experiment, and 13-point calibration and validation routines before the start of each block.

* 1. Additional measures

After completion of the behavioral task, we ask participants to report their age and gender, whether they are red-green color blind, whether they had trouble distinguishing red and green colors in the task, describe their decision strategy in a few sentences, report whether they performed the task seriously or not (Aust et al., 2013). Participants are informed that their response to this item only serves assurance of data quality and does not result in loss of benefits for them. Optionally, participants can leave additional comments to the researcher.

### Analysis Plan

1. Statistical models
   1. Bayesian regression analysis of choice behavior

We will analyze choice behavior using a Bayesian mixed logistic regression implemented in bambi (Capretto et al., 2021), where the dependent variable is choice (*Hp* vs. *Hm*) in each trial. The predictors are:

* Expected-value-difference ()
* Presentation-format (effects-coded: -0.5 = by-attribute; +0.5 = by-alternative),
* Duration-favors (effects-coded: -0.5 = duration favors *Hm*; +0.5 = duration favors *Hp*),
* Last-stage-favors (effects-coded: -0.5 last stage favors *Hm*; +0.5 = last stage favors *Hp*),
* the interaction terms (duration-favors by presentation-format) and (last-stage-favors by presentation-format).

The model will include random intercepts and slopes for all predictors across participants.

* 1. Hypotheses H1, H1a, H1b, H2, H2a, and H2b will be tested explicitly with paired Bayes Factor *t*-tests implemented in the R package BayesFactor (Morey & Rouder, 2011). The analyses will be complemented with a Bayesian estimation analysis of the effect size *d* (Kruschke, 2013). In each case the we test the directed alternative hypothesis that a difference in choice probability between conditions is greater than zero.
     1. H1: To test the marginal effect of presentation duration on choice (across presentation formats), we will compute the probability of choosing *Hp* when it was favored by presentation duration (because either the alternative *Hp* or the attribute *p* was shown longer) and when it was not, and perform a test of their difference.
        1. H1a: To test the duration effect for *alternative*-wise presentation we will compute the probabilities of choosing *Hp* in trials with *alternative*-wise presentation when *Hp* was shown longer, and when *Hm* was shown longer, and perform a test of their difference.
        2. H1b: To test the duration effect for *attribute*-wise presentation we will compute the probabilities of choosing *Hp* in trials with *attribute*-wise presentation when *p* was shown longer, and when *m* was shown longer, and perform a test of their difference.
     2. H2: To test the marginal effect of presentation *order* on choice (across presentation formats), we will compute the probability of choosing *Hp* when it was favored by the last presentation stage (because either the alternative *Hp* or the attribute *p* was shown last) and when it was not, and perform a test of their difference.
        1. H2a: To test the order effect for *alternative*-wise presentation we will compute the probabilities of choosing *Hp* in trials with *alternative*-wise presentation when *Hp* was shown last, and when *Hm* was shown last, and perform a test of their difference.
        2. H2b: To test the order effect for *attribute*-wise presentation we will compute the probabilities of choosing *Hp* in trials with *attribute*-wise presentation when *p* was shown last, and when *m* was shown last, and perform a test of their difference.
  2. H3: To test the hypothesis, that choice strategies are affected by presentation format, we fit two behavioral models to the choice and response time data of each participant, separately for trials with alternative- and attribute-wise presentation (see 18. Behavioral models). We then compute the relative fit of the models for each participant and presentation format by taking the difference of the models’ BIC. Finally, we perform directed paired Bayes factor *t-*tests of the differences, testing the directed hypothesis that the relative fit of the within-alternative integration model over the between-alternative model is increased in alternative-wise vs. attribute-wise presentation.

All statistical tests, behavioral models and procedures have been developed with pilot data and will be applied to the main sample without changes. Python code for all analyses is available as a versioned and timestamped release on GitHub under <https://www.github.com/moltaire/gaze-choice-causality/releases/preregistration>.

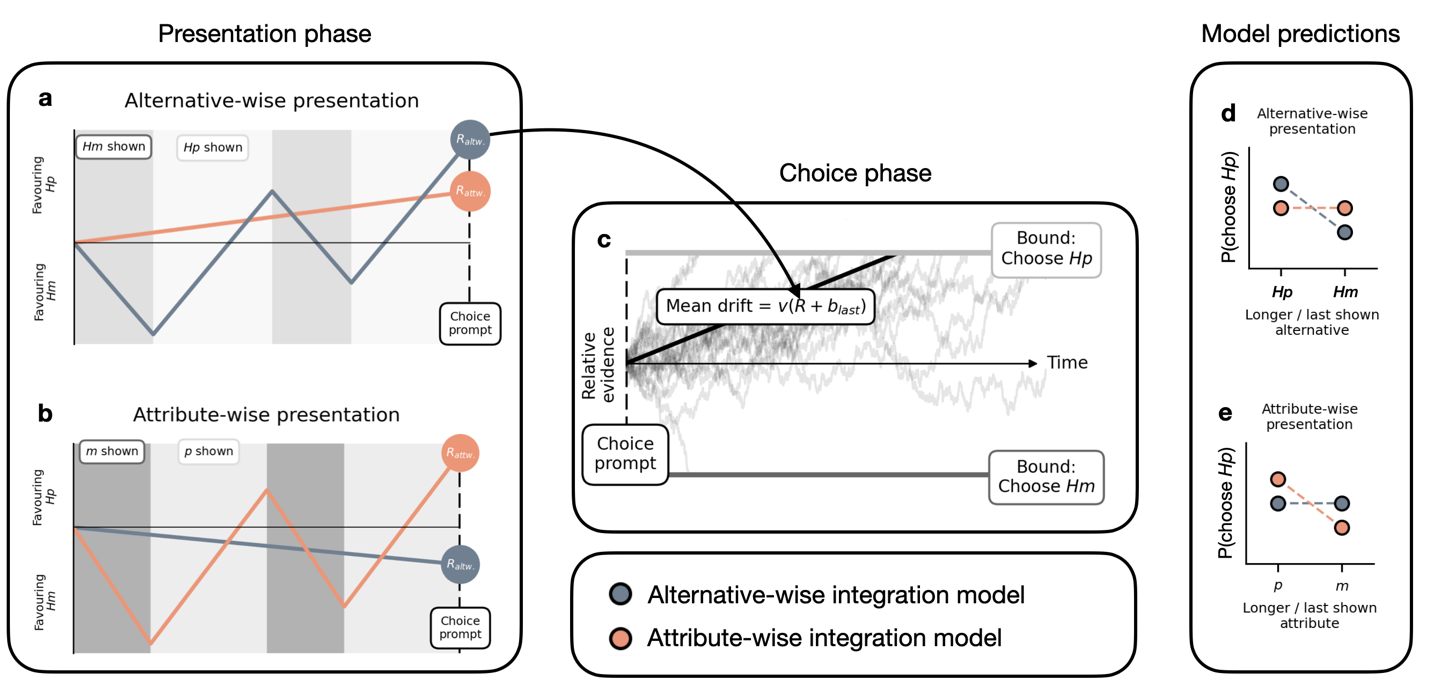


Figure 2. **Behavioral models and their predictions.** **a)** Construction of drift-rates in trials with alternative-wise presentation. Only the alternative-wise model’s drift rate is sensitive to effects of presentation duration. Analogously, only the alternative-wise model produces sequence effects in alternative-wise presentation, controlled by the parameter. **b)** Construction of drift rates in attribute-wise presentation. Here, only the attribute-wise model’s drift rate is sensitive to effects of presentation duration. Similarly, only the attribute-wise model produces sequence effects in attribute-wise presentation, controlled by the parameter. **c)** After construction of the drift rate in the presentation phase, choices and response times result from a diffusion process with the previously constructed drift rate. **d-e)** Both models’ predicted effects of alternative-wise **(d)** and attribute-wise **(e)** presentation duration. Slate-gray color refers to alternative-wise model. Orange color refers to attribute-wise model.

1. Behavioral modeling

We fit two different behavioral models to each participants’ choice and response time data, separately for trials with alternative- and attribute-wise presentation. Both models have a similar overall structure: During the presentation phase of the trial, a relative evidence signal *R* is assumed to be formed (Figure 2a-b), depending on the presentation format, duration and sequence. After the choice prompt, a diffusion process between two decision bounds (corresponding to choosing *Hp* and *Hm*, respectively) is initiated that elicits a choice at a time point *t* (Figure 2c). Crucially, the drift rate of the diffusion process is proportional to the relative evidence signal *R* at the end of the presentation phase. For both models, the diffusion process after the choice prompt is parameterized by a drift constant *v* (that scales the final relative evidence signal *R*), a noise parameter *s,* while the boundary separation is kept constant at a value of 1.

The two models only differ in the process calculating the relative evidence signal *R* during the presentation phase.

* 1. Alternative-wise integration model

The alternative-wise integration model (slate-gray in Figure 2) computes alternative-wise expected utilities, using a standard utility function (). During the presentation phase, the difference between expected utilities is assumed to accumulate over time. Critically, the momentarily not presented alternative’s utility is discounted by an *alternative-wise gaze-discount* *θ*. The final relative evidence signal is given by

Where and are relative presentation durations of *Hp* and *Hm*. Note that in trials with attribute-wise presentation, and are set to 0.5, as both alternatives’ attributes are presented equally long.

Additionally, the model uses a parameter to predict sequence effects in trials with alternative-wise presentation: Positive shift *R* towards the last-presented alternative, negative shift it to the *first-*presented alternative.

* 1. Attribute-wise integration model

The attribute-wise model (orange in Figure 2) assumes that the relative evidence *R* is computed through attribute comparisons between alternatives, weighted addition of attribute differences, and accumulation of differences over time. Importantly, it assumes that momentarily not presented attributes are discounted by an *attribute-wise gaze-discount η.* The final relative evidence signal is given by

Where and are attribute differences between *Hp* and *Hm* alternatives. and are relative presentation durations for *p* and *m* attributes, respectively. controls the relative weighting between probability and outcome attributes. Note that attributes are also normalized in each trial (by dividing by the sum of values on the attribute). During trials with alternative-wise presentation, relative gaze durations towards attributes are set to 0.5, since at every point during the presentation phase, information of both attributes is presented.

Additionally, the model uses a parameter to predict sequence effects in trials with attribute-wise presentation: Positive shift *R* towards the alternative with the higher value on the last-presented attribute, negative shift it to the alternative with the higher value on the *first-*presented alternative.

* 1. Model implementation

Both models are implemented in pyddm (Shinn et al., 2020) with a temporal resolution *dt* = 0.01, and a resolution of the evidence space *dx* = 0.01.

* 1. Parameter estimation

To address Hypothesis H5, both models will be fit separately to trials with alternative-wise and attribute-wise presentation using pyddm’s default differential evolution algorithm, minimizing BIC.

Additionally, we will fit models to each participant’s full data set, for a model-based characterization of their choice behavior.

* 1. Parameter recovery

To ensure interpretability and validity of the models’ parameter estimates, we will perform a parameter recovery study as follows: First, we estimate each model’s parameters from the empirical data of each participant. Then we simulate a synthetic data set of the same size as the empirical one, using the individual obtained estimates. Then we re-fit the models to the synthetic data and compare known generating to obtained recovered parameters by means of Bayesian linear regression (dependent variable: recovered parameter; independent variables: Intercept, generating parameter) and correlation analyses (Kruschke, 2014; Lee & Wagenmakers, 2013).

* 1. Model recovery

Similarly, we will perform model-recovery analyses by fitting each model to the synthetic data generated from all models. Then, for each generating model, we perform Bayesian model selection (Rigoux et al., 2014; Stephan et al., 2009) to identify the most likely generating model.

1. Transformations

Categorical variables will be effects-coded for the logistic regression analysis (see above).

1. Inference criteria
   1. Predictors in the GLM analysis and differences in BEST analyses are inferred to be credibly different from 0 if their 95% credible interval excludes 0 or at least 95% of their posterior mass is above (below) 0.
   2. Interpretation of Bayes factors will follow the classical categorization based on Jeffreys (1998): 1-3 Anecdotal evidence; 3-10 Moderate evidence; 10-30: Strong evidence; 30-100: Very strong evidence; >100: Extreme evidence.
   3. Individually best-fitting models will be identified by the lowest Bayesian Information Criterion (BIC; Schwarz, 1978).
   4. The best-fitting model on the group level will be identified by Bayesian Model Selection (Rigoux et al., 2014; Stephan et al., 2009)
2. Data exclusion

Participants will be excluded from the analysis without replacement if one of the following criteria is met:

* The participant chose the dominated alternative in catch trials more than 4 times (20% of catch trials).
* The participant reported to be red-green colorblind or having difficulties distinguishing the colors in the task.
* The participant reported having technical difficulties that prevented them from diligent execution of the task.
* The participant's self-reported decision strategy suggests that task instructions were misunderstood.
* The participant reported non-serious participation in the task (Aust et al., 2013).

1. Exploratory analyses
   1. Additional model-based analyses

We may conduct additional model-based analyses of choice behavior, that include a larger space of candidate models (e.g., Expected Utility Theory, Prospect Theory, Heuristic models).

* 1. Exploratory analyses of webcam-based eye tracking data

We will explore the data quality obtained from webcam-based eye tracking in a large sample collected on the Prolific.co platform. To this end, we plan to analyze effective sampling rate, compute mean deviations between predicted and actual gaze direction during eye tracking validation. If data quality is sufficient, we plan to use eye tracking data to test the extent to which presentation durations relate to gaze duration towards alternatives and attributes.

### References

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods*, *45*(2), 527–535. https://doi.org/10.3758/s13428-012-0265-2

Capretto, T., Piho, C., Kumar, R., Westfall, J., Yarkoni, T., & Martin, O. A. (2021). Bambi: A simple interface for fitting Bayesian linear models in Python. *ArXiv:2012.10754 [Stat]*. http://arxiv.org/abs/2012.10754

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Jeffreys, S. H. (1998). *The Theory of Probability* (Third Edition). Oxford University Press.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603. https://doi.org/10.1037/a0029146

Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*. Academic Press.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Morey, R. D., & Rouder, J. N. (2011). Bayes factor approaches for testing interval null hypotheses. *Psychological Methods*, *16*(4), 406–419. https://doi.org/10.1037/a0024377

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, 3839–3845.

Rigoux, L., Stephan, K. E., Friston, K. J., & Daunizeau, J. (2014). Bayesian model selection for group studies—Revisited. *NeuroImage*, *84*, 971–985. https://doi.org/10.1016/j.neuroimage.2013.08.065

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Shinn, M., Lam, N. H., & Murray, J. D. (2020). A flexible framework for simulating and fitting generalized drift-diffusion models. *ELife*, *9*, e56938. https://doi.org/10.7554/eLife.56938

Stephan, K. E., Penny, W. D., Daunizeau, J., Moran, R. J., & Friston, K. J. (2009). Bayesian model selection for group studies. *NeuroImage*, *46*(4), 1004–1017. https://doi.org/10.1016/j.neuroimage.2009.03.025

Yang, X., & Krajbich, I. (2020). *Webcam-based online eye-tracking for behavioral research*. PsyArXiv. https://doi.org/10.31234/osf.io/qhme6