CONFIDENTIAL - FOR PEER-REVIEW ONLY

Perceptions of Choice Over Fact and Value-Based Beliefs (#47837)

Created: 09/16/2020 02:08 PM (PT)

1) Have any data been collected for this study already?
It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid pre-registration nevertheless.

2) What's the main question being asked or hypothesis being tested in this study?
Questions: Do we think that others are free to believe whatever they want or do we think that other's beliefs are determined by the availability of evidence? Do children and adults differ in how they perceive the controllability of others' evidence-based beliefs?

3) Describe the key dependent variable(s) specifying how they will be measured.
The dependent variable is control over belief, a binary response (yes/no) indicating whether the individual in the vignette had to possess a belief (no, no control over belief) or whether they could believe otherwise (yes, control over belief). A secondary question also asks participants why they think that the beliefs are controllable or uncontrollable. We will categorize our responses into three categories: no answer, I don't know, and referencing the presence or absence of evidence

Example stimuli and response question (No evidence; single trial):
This is a story about Mike and what Mike thinks! On Christmas, Mike and his family opened all the presents beneath the tree except for one box. The unopened box has Mike's name on it. Mike wonders if there is an action figure or a toy train in the box. The box is heavily wrapped, and Mike has not picked it up or looked in it. There are no clues for what is inside. Mike believes there is an action figure in the box. Instead, of believing there is an action figure in the box, could Mike believe there is a toy train in the box? Answers: Yes, No. Why do you think that? __

4) How many and which conditions will participants be assigned to?
We will apply a within-subjects design for our 5 conditions: 3 experimental conditions (No evidence, Strong evidence, Strong counterevidence) and 2 control conditions (Action-intentional, Action-impossible). Children and adults will participate in all conditions. For each experimental condition, participants will see 2 trials (out of 6 possible stories for that condition). For each control condition, participants will see all 2 trials. Participants will see 10 total trials.

Adults will be tested over Amazon Mechanical Turk using Qualtrics software. Children will be tested live over Zoom using a PowerPoint presentation.

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.
Analysis of only the experimental conditions:
We will first analyze only analyze the experimental conditions.
Therefore, we will use a Generalized Linear Mixed Model with the following model equation:

```
full <-
glmer(controllability ~ condition * age +
(1 + conditiondummy coded and centered + trialz-transformed | SubjectID),
family = binomial(link=logit), data = dataframe )
```

If the model does not converge, we will follow the steps here: https://rpubs.com/bbolker/lme4trouble1. If the procedure does not fix convergence issues or some of the random effects aren't identifiable, we stepwise simply the model. First, we will drop the correlations of random intercepts and slopes, and then the random slope of trial.

We will first compare the full model with a null model only containing the random intercept and random slopes. If there is a significant difference, we will fit reduced models by dropping one predictor at a time. These reduced models will then be compared with the full model to see whether the inclusion of one particular predictor improves the model significantly. If the interaction between condition and age, or the main effects of condition or age are significant, we will run follow up pairwise comparisons to investigate specific differences between conditions or age groups (e.g., no evidence vs strong evidence conditions; 5-6-year-olds vs adults in the strong evidence condition).

To test whether the model estimates are significantly different from chance level we calculate the confidence intervals using 1000 parametric bootstraps and check whether 0.5 is included in the interval.

Comparison with the reference conditions:
Additionally, we will also compare each belief condition to two reference conditions (action possible, action impossible). This allows us to, for example, test whether the controllability ratings differ between the strong evidence belief condition and the impossible action condition. Therefore, we will subset our data frame, so that only the data of the conditions of interests are included and fit the following above-mentioned model equation again.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.
Individuals that do not pay attention during their trials will be excluded. Children will be excluded that are distracted and are not able to look and follow along with the stories, or do not respond to questions. Adults will be excluded if they compete the survey too quickly (limit is 2 minutes, indicated by the experimenter reading through the survey as fast as possible).

7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.
We conducted a simulation to evaluate the power of our planned analyses assuming different individual variability levels for in total 120 participants (40 per group). We simulated the following performance levels: Probability that participants judge beliefs based on no evidence as controllable: 5-6-year-old children: 60%, 7-8-year-old children: 70%, adults: 90%, probability that participants judge beliefs based on strong evidence as controllable: 5-6-year-old children: 35%, 7-8-year-old children: 20%, adults: 5%, probability that participants judge beliefs with strong counter evidence as controllable: 5-6-year-old children: 60%, 7-8-year-old children: 75%, adults: 90%. We simulated small, medium, and large random intercepts. We simulated the random intercepts of subject ID by drawing random numbers (one per subject) from a normal distribution with a mean of zero and a standard deviation of 0.1, 0.2, or 0.4, respectively. The size of the standard deviations corresponds approximately to 0.25, 0.5, or 1 times the coefficient for the intercept (± 0.4 in link space). Finally, we simulated small and medium random slopes of condition within subject by drawing random numbers from a normal distribution with a mean of zero and a standard deviation of approximately 0.25 and 0.5 times the coefficient for the differences between conditions. This resulted in random slopes of 0.25 and 0.5 for the strong evidence condition, 0.025 and 0.05 for the strong counter evidence condition. We do not expect a significant effect of trial. Therefore, we did not simulate this fixed effect or any interaction effects including trial, but we still accounted for them by including the predictor variable in the model. We then combined the fixed and random effects into the linear predictor and generated the response variable by sampling from a binomial distribution using the inverse logit transformed linear predictor as the probability of picking the rational agent. For each combination of test condition performance level and random intercepts, we simulated 100 datasets (with 120 subjects, 3 conditions per subject, two trial per condition) and fitted the below mentioned full model including the two-way-interaction of condition*age group, the respective reduced models to determine the lower-level effects (main effects) and a null model only comprising the random effects. For each combination, we evaluated the models based on whether the likelihood ratio test for the full-null model comparison as well as the likelihood ratio test for the interaction condition*age group turned out to be significant. Overall, the power (percentage of simulations that lead to a significant effect) for finding this significant interaction was very good (above 0.9) for all variations of random intercepts and random slopes. Thus, we decided to test 120 participants in this study.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)
Only 5/80 children have been tested.