

A freshly baked perspective on how we measure risk propensity in children

Emily S. Sumner, Ryan C. Stokes, Percy K. Mistry, Susanne M. Jaeggi, and Barbara W. Sarnecka

(sumnere, stokesr, pkmistry, smjaeggi, sarnecka@uci.edu)

University of California – Irvine, Department of Cognitive Sciences
Social and Behavioral Sciences Gateway, Irvine, CA 92697-5100 USA

Abstract

Children are notoriously fearless; running across slippery grass, eating bugs, and climbing up to the top of trees while their anxious parent yells at them to get down. Research on the development of risk propensity suggests that as we get older, risk aversion increases. However, there is little research on risk aversion within the preschool age range. The current study introduces the Child Risk Utility Measure (CRUM), a developmentally sensitive, dynamic task that measures risk propensity. We develop a novel cognitive modeling perspective to interpret our behavioral data and segregate the effects of risk propensity, probability learning, contaminant processes, and age effects within a unified framework. Unlike previous studies of risk propensity, we found reasonable evidence suggesting that preschool-aged children do not become more cautious as they get older. Any effect of age was attributed to confounds due to contaminant behavior. We suggest this as an alternative explanation for age effects in risk propensity during early childhood.

Keywords: risk propensity; cognitive development; decision making; cognitive modeling

Introduction

Children are notoriously fearless; running across slippery grass, eating bugs, and climbing up to the top of trees while their anxious parent is screaming at them to get down. Research on the development of risk propensity suggests that as we get older, we become more cautious (Paulsen, Platt, Huetzel, & Brannon, 2012). However, there is little research on risk aversion within the preschool age range. There are few laboratory procedures that assess young children's risk-taking preferences.

In Kerr and Zelazo (2004), authors made a simplified version of the Iowa Gambling Task, a standardized measure of risk aversion for adults, for preschool-aged children. Authors found that 3-year-olds chose more from the deck that had higher gains, but larger losses, than from the deck that had lower gains but lower losses. Weller, Levin, and Denburg (2011) had participants aging from 5-85 years complete a computerized version of The Cups task, where participants have to choose between a sure bet, or a risky gamble to win money. They found that younger children had riskier performance on this task, and that risk aversion increased throughout development.

However, a limitation of these studies is that they use two alternative forced choice tasks, which allows children who may not understand the task to participate. After the data is collected, it is then difficult to tell who was able to comprehend the task because the primary thing being assessed is proportion of risky choices. A more dynamic task would enable us to measure participants' comprehension of the task, as well as get more detailed information about individual differences in risk propensity.

An example of a dynamic measure is the Balloon Analogue Risk Task (BART), which is often used to assess an adult's propensity for risk (Lejuez et al., 2002). In this virtual task, participants press a button to pump air into a balloon. A participant's earnings are proportional to the size of the balloon, so there is incentive to pump as much air into the balloon as possible. However, as the balloon gets larger, the likelihood of the balloon popping increases. If the balloon pops, the participant loses everything from that trial. Riskier performance on the BART has been linked to real-world risky behaviors including smoking and drug use (Lejuez et al., 2002). Given children's difficulty with interpreting certain types of symbolic representation (DeLoache, 2004) and well-documented difficulties tracking numerosities higher than three (Feigenson, Dehaene, & Spelke, 2004), the current version of the BART task may be too abstract for young children.

The first aim of the current study is to create and implement methods that enable us to assess individual differences of younger children's risk propensity. Understanding this will give us a complete picture of individual differences in risk preferences, particularly whether risk propensity is trait-like, or simply a function of age. To do this, we developed a new application for touchscreens, the Child Risk Utility Measure (The CRUM), that respects the cognitive limitations of three- to six-year-old children. A secondary aim is to contribute a cognitive modeling framework that allows making inferences about children's underlying risk propensity and about how they might learn and adapt in this task. We identify limitations of some existing models and introduce a model that explicitly incorporates censored behavior that is typically observed in such tasks, implements a contaminant process to model a possible lack of understanding that kids might have, and infer age-effects using Bayesian inference. All of this is done within a single modeling framework.

The CRUM: Child Risk Utility Measure

Participants

We tested 24, three- to six-year-old children at preschools throughout Orange County, CA (Range: 39.8 - 70.2 months, average: 52.7 months). Eight additional participants were tested, but excluded due to the inability to complete the task or unresponsiveness.

Procedure

Children are seated at a table with a plate and a tablet directly in front of them (see Figure 2). The experimenter sits to the

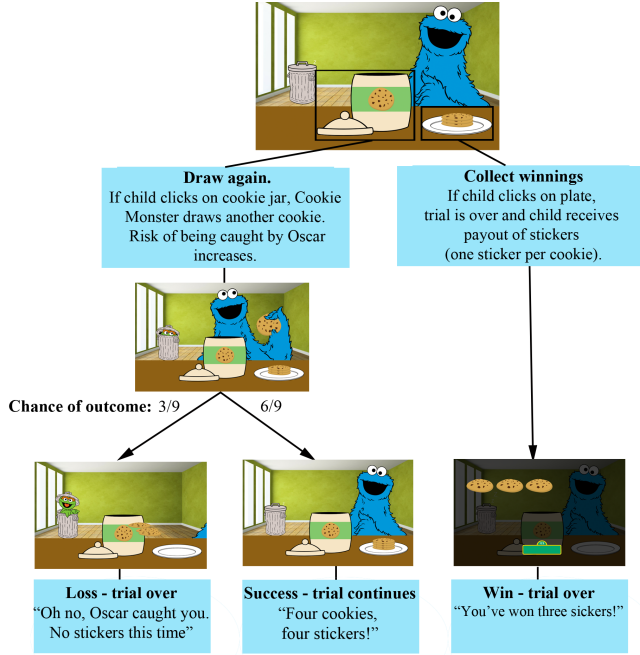


Figure 1: Design of the CRUM task.

left of the participant and explains that they are going to play a game called “Sneaky Cookies” where they can win stickers. The experimenter demonstrates the game first and says: “In this game, Cookie Monster tries to take cookies from the cookie jar without Oscar catching him. When Oscar is far away, it is harder for him to catch us. For every cookie Cookie Monster gets, you get a sticker. You can click the cookie jar to try and get another sticker, or click the plate to keep the stickers.”

The experimenter plays three trials, demonstrating both a win and a loss. To control for understanding of the game, the experimenter asks the child “If I want another sticker, what should I click on?” and “If I want to keep my stickers, what should I click on?” Each time a cookie is successfully taken, the experimenter places a sticker on the plate right in front of the child, and says “N cookies, N stickers!” This reduces the cognitive load needed to complete the task because it removes the step of remembering that each cookie collected is equivalent to a sticker reward.

When the child clicks on the plate to collect the stickers, the experimenter says, “You won N stickers!” The stickers are then put into an envelope for the child to take home. If Oscar wakes up, the experimenter says, “Oh no, Oscar caught you!” and dumps the stickers into the box that is out of reach and sight of the participant. The risk of getting caught by Oscar increases as the number of cookie increases (0/9, 1/9, 2/9,..., 9/9). For example, if the plate has two cookies on it, the chance that Oscar will catch Cookie Monster is 2/9. As a visual cue to the increasing risk, the garbage can moves closer to the table each time a cookie is taken from the jar.

Subjects complete two blocks of 9 trials for a total of 18 tri-

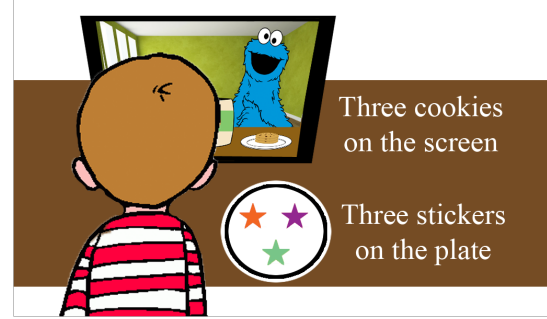


Figure 2: The set up of the experiment. Participants sit with a tablet. The number of cookies that Cookie Monster currently has is also represented physically in front of the participant on a plate as stickers. This reduces the abstractness of the task and allows the child to evaluate their potential winnings before making their next decision.

als. The probabilities of each trial were predetermined using a random number generator at each potential click. This give us 18 trials total, each with a maximum number of clicks before getting caught. Trials were presented in a random order for each participant.

Results

Table 1: Summary statistics, Means (standard deviation)

	Stickers won	Times caught	Risk incl
With Contaminants			
– Block 1	9.0 (7.20)	5.42 (2.90)	0.76 (0.23)
– Block 2	11.88 (8.30)	4.42 (3.09)	0.72 (0.24)
Without Contaminants			
– Block 1	12.0 (5.65)	4.22 (2.31)	0.68 (0.22)
– Block 2	15.83 (5.19)	2.89 (1.74)	0.62 (0.20)

To look at risk inclination in this task, we use the ratio of clicks a child made over the total number of potential clicks that could have been successfully made. Since all children completed the same trials, they can be compared to each other.

$$\text{Risk Inclination} = \frac{\text{Clicks before getting caught}}{\text{Possible successful clicks}}$$

Values closer to 1 indicate that the participant took more risks on the task and got caught more often. A child who got caught by Oscar each trial would have a risk inclination value of 1. Lower numbers indicate that the child took fewer risks. A child who stopped after just one cookie on each trial would have a risk propensity measure of about 0.27 (see Figure 3).

Of course, the CRUM task only measures risk propensity if the participant has the goal of collecting as many stickers as possible. During testing, we observed that some children seemed to enjoy getting caught by Oscar. Such children might click every time, either because getting caught by Oscar was

more fun than winning stickers, or because they did not understand how the task worked. Six children showed this pattern. We decided to treat always-clicking as a contaminant behavior (i.e., as a behavior that interferes with our measurement of the child’s risk-taking propensity) and to conduct separate analyses of the data with and without those six participants. When we included those six participants in our analysis, we found that risk inclination decreased with age ($\eta = -0.013$, $p < 0.0105$). However, when we exclude those six participant, age was neither with cookies won ($\eta = 0.263$, $p > 0.05$), nor risk inclination ($\eta = -0.009$, $p > 0.05$). This suggest that the age effect is driven by the children who were not comprehending the task. When data representing contaminating behaviors are excluded from the analysis, there is clear evidence of learning between Block 1 and Block 2. Children won significantly more stickers ($t = -2.61$, $p < 0.05$) and were caught significantly fewer times ($t = 2.91$, $p < 0.01$) in second block than in the first. However, we found no evidence of a difference in risk propensity ($t = 0.81$, $p > 0.05$) between Blocks 1 and 2. That is, there was no hint that children either took more or fewer risks overall as they learned more about the task.

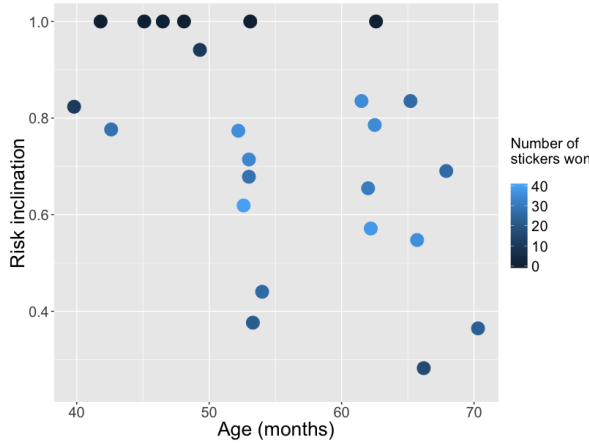


Figure 3: Age and risk inclination. Each dot represents one child. The darker the dot, the fewer stickers the child collected. The Y-axis shows risk inclination, with higher values indicating more risk taking.

Inferring components of the cognitive process using a cognitive model

To further analyze the CRUM data, we implement an adapted version of the 2-parameter Bayesian hierarchical model described in van Ravenzwaaij, Dutilh, and Wagenmakers (2011). Using cognitive modeling enables us to comprehensively look into individual differences in behavior on this task. Below, we first provide a summary of the original model and then detail the changes made in our implementation.

1. Summary of the original 2-parameter model:

The optimal number of clicks (ω) is based on the individual’s

propensity for risk taking (γ) and belief about the probability of getting caught on a particular click (p_{belief}). This probability is assumed to remain constant regardless of the number of clicks already made on a particular trial, and across all trials.

$$\omega = \frac{-\gamma}{\ln(1 - p_{belief})}$$

The probability of clicking for the n_{th} time on any particular trial is based on ω and the behavioral consistency of the individual (β) :

$$p_n^{click} = \frac{1}{1 + e^{\beta(n-\omega)}}$$

Finally, the likelihood function (probability of the data conditional on the parameters), $p(D|\omega, \beta)$, depends on whether a particular trial k was a success (participant stopped voluntarily, $d=1$) or failure (participant got caught, $d=0$), and is given by the equation below, where n_k is the number of clicks on trial k .

$$p(D = n_k | \omega, \beta) = \prod_{i=1}^t ((\prod_{i=1}^{n_k} p_i^{click})(1 - p_{n_k+1}^{click})^d)$$

The model has 2 free parameters per individual: risk-taking propensity (ω) and behavioral consistency (β). The believed probability of getting caught on a particular click (p_{belief}) is assumed to be constant and told to the participant as part of the experimental procedure.

2. Beliefs about the probability of getting caught:

We propose a compromise between the 2-parameter model described above and the 4-parameter model also proposed in van Ravenzwaaij et al. (2011). The latter assumes that participants might update their belief about the probability of getting caught after each trial based on some form of learning. However the paper suggests that this approach does not allow for robust recovery of parameters of the model and hence recommend using the 2-parameter model. Since the probability of getting caught is not explicitly taught to the children in our experiment (unlike in some other BART experiments), we treat this (the child’s belief about the probability of getting caught) as a free parameter. Secondly, since the experiment is carried out in two, clearly distinct blocks, we assume that children may learn and update their beliefs between blocks, and hence treat the beliefs as two free parameters per child (one for each block).

3. Treatment of failure trials: Right censoring of data:

Trials on which participants get caught do not provide information on the upper bound for risk propensity, since we do not know how much longer the participants would have persevered for if they had not been caught. However, they do provide information about the lower bound for risk propensity. In traditional BART analysis (Lejuez et al., 2002), these trials are excluded. van Ravenzwaaij et al. (2011) includes the tri-

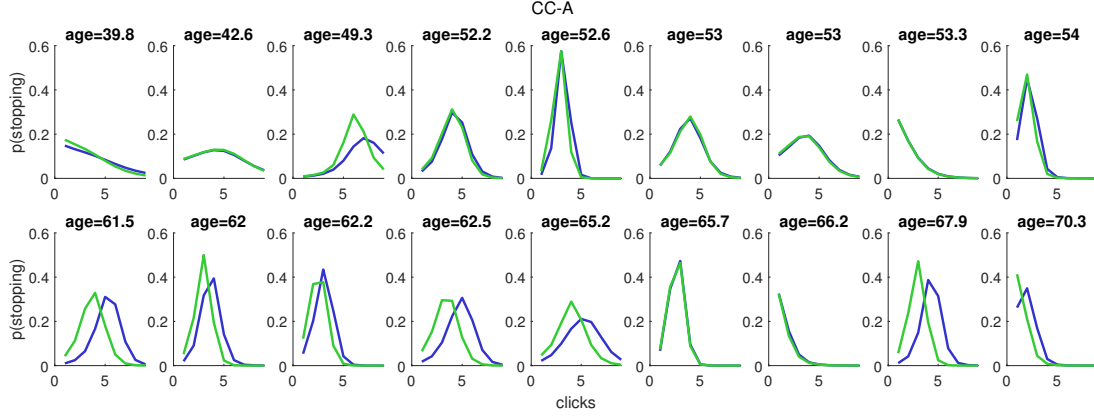


Figure 4: Each graph represents an individual's probability of when they will end the trial. Blue lines indicate the probabilities of stopping in block 1 and green lines indicate the probabilities of stopping in block 2.

als, but the formulation described earlier implicitly assumes that the participant would have stopped voluntarily even if they had not been caught, since $(1 - p_{n+1}^{click})^0 = 1$. In our view, this amounts to inferring an artificial upper bound on risk propensity for these trials, and may lead to an underestimation of the risk propensity of individuals. To counter this, but still include information about the lower bound that these trials provide on risk propensity, we propose an alternative approach that uses right censoring. In this approach, for failure trials we provide information to the model that constrains inference about the risk propensity only in one direction. This can be easily implemented within a Bayesian inference framework. In effect, when $d=1$, the likelihood function remains the same as before. When $d=0$ however, the likelihood function is defined differently. This is implemented in JAGS using the *dinterval* distribution and is an established method to model censored data.

$$p(D = n_k | \omega, \beta) = \prod_{k=1}^t ((\prod_{i=1}^{n_k} p_i^{click})(1 - p_{n_k+1}^{click}))^d (X_k)^{1-d}$$

$X_k = 1$ if the predicted value of the number of clicks conditional on the parameters on censored (failure) trials is greater than the maximum number of clicks possible on that trial k without getting caught, and 0 otherwise. X_k is always equal to 1 for successful trials.

4. Testing the effects of age on risk propensity

Previous experiments have demonstrated that risk aversion increases with age (Kerr & Zelazo, 2004; Paulsen et al., 2012; Weller et al., 2011). To investigate this hypothesis, we test a model where the risk propensity is modeled as being dependent on the age of children. The subscript p indicates individual participants.

$$\begin{aligned} \gamma_p &\sim N(\mu_{p\gamma}, \sigma_\gamma) \\ \mu_{p\gamma} &= s_0 (age_p) + z_0 \end{aligned}$$

Here, s_0 , z_0 , and σ_γ are free parameters in the model; s_0 and z_0 are the slope and intercept for a regression of risk

propensity against age. If σ_γ is large it indicates that the regression is not a good fit and that age might not be a strong predictor of risk propensity. This would be corroborated by a poor predictive capability of this model. On the other hand, a small inferred value of σ with good predictive capability of the model would indicate that the inferred values of s_0 and z_0 are robust. Inference about whether age is a strong predictor of risk propensity depends on the inferred value of s_0 . Note that when working with kids of this age, it is quite feasible to have a relatively large number of contaminants.

If indeed there are contaminants in the subject pool, this would affect the inferences we make at a hyperparameter level, and may lead to incorrect or exaggerated inferences about how aspects such as age affect risk propensity. To detect contaminants, we implement a mixture model, where the behavior of an individual is assumed to be generated either by the cognitive process described earlier, or by a contaminant process. We define a contaminant process as one where participants keep clicking till they get caught. A participant could be identified as a contaminant if the data has a much higher likelihood of being generated by the contaminant process than the cognitive account, even if not all trials for that participant are failure trials).

Modeling Results

1. Model comparison:

A comparison (see Table 2) of posterior predictives with the data shows that the model including censoring (CENS), and including both censoring and contaminants (CCA) perform significantly better than a basic model that does not account for these factors.

2. Individual differences in beliefs about the probability of getting caught:

Figure 4 shows the individual differences in each child's propensity to stop (collect the cookies) after each click. In-

terestingly, older children appear to shift towards collecting earlier during the second block. Figure 5 shows that on average, the learning seems to indicate a change in belief between blocks, with an increase in the probability of collecting after 1, 2 or 3 clicks, and a reduction for clicks greater than 3. A Bayesian t-test reveals a Bayes factor (BF) of 6.7 in favor of the mean increase between blocks in the probability of stopping within the first 3 clicks being more than 0. The observation that children begin to collect cookies earlier in the second block can be explained based on a change in their belief about the underlying probability of getting caught. The effect of age on this increase is inconclusive, with a BF of 2.5 in favor of a positive correlation between age and increased probability ($r = 0.434$).

Table 2: Model comparison: Lower bias, RMSE, DIC preferred; higher correlation indicate better results)

Model	Bias	RMSE	Correlation	DIC
Basic	-0.460	1.297	0.716	1464
CENS	0.107	0.754	0.897	570
CCA	0.104	0.740	0.900	575

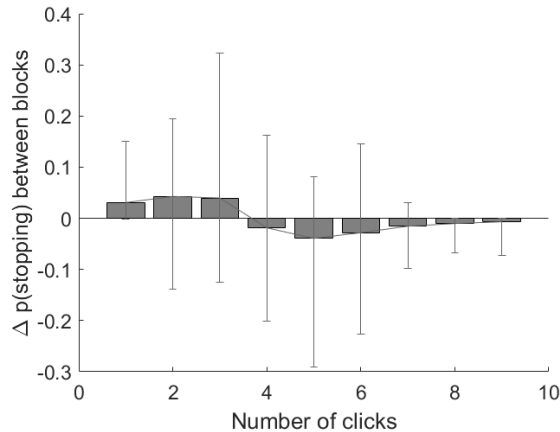


Figure 5: Bars indicate the mean change in probability of stopping (collecting cookies) from the first to the second block, across all participants. Positive values indicate an increase in the probability of stopping after that click.

3. Impact of including censoring mechanisms:

Including censoring of failed trials provides more accurate and less biased posterior predictives. Through including we are able to use data that have typically been excluded from a BART data analyses. We suggest that those using a BART-like task in the future consider incorporating censoring into their analyses.

4. Risk propensity and effects of age:

Modeling age effects without excluding contaminants results

in a large inferred value of $z_0 = 12.8$ vs $z_0 = 5.8$ when we factor in a mixture model that includes a contaminant process. The contaminant process identifies children that misunderstand the objective of the game (or perhaps have a different set of objectives). The behavior of these children is thus attributed to a contaminant process rather than high risk propensity. Apart from a much lower intercept, inferences about the relation between age and risk propensity are also affected. Without excluding contaminants, the 95% interval for the inferred value of s_0 is $[-0.36, -0.04]$ which does not include 0. If we explicitly model contaminants, the inferred interval is $[-0.18, 0.06]$, which includes 0. We test the null hypothesis that $s_0 = 0$ (no effect of age), which is inconclusive when we do not exclude contaminants ($BF = 1.2$), but has reasonable evidence in favor of the null ($BF = 7.7$) when contaminants are explicitly modeled. Most of the identified contaminants are younger kids, that are otherwise identified as having a high risk propensity in the CENS model, and contribute to a large negative s_0 coefficient. Modeling the contaminant process excludes these participants from the regression, and results in a realistic coefficient. Finally, the model inferred parameter for risk propensity (γ) has a correlation of 0.54 with the raw risk inclination measure.

5. Behavioral Consistency:

Older children were slightly more consistent (higher β) in their behavior than younger children, although a Bayes factor in favor of a correlation ($r = 0.426$) was not conclusive ($BF = 2.36$).

Discussion

We introduce a novel, developmentally sensitive, task which provides us with detailed information regarding an individual's risk aversion and behavioral consistency. We also contribute to the literature on cognitive modeling of BART-like tasks. Incorporating censoring provides more realistic inferences about risk propensity that allow us to improve the quality of predictions. Adding a regression model within the Bayesian inference framework allows us to infer hyperparameters that directly capture and compute Bayes factors for age-effects. Unlike previous studies of risk propensity, we found reasonable evidence suggesting that preschool-aged children do not become more cautious as they get older. Although our data did show an apparent age effect, the effect disappeared when we excluded data representing contaminant behaviors (clicking consistently and losing every trial). It seems likely that this behavior arises either because children do not understand how the task works, don't have the goal of maximizing stickers (perhaps because getting caught by Oscar is more fun), or even because they persevere on clicking on the cookie jar. We substantiate this claim about contaminants by explicitly incorporating a mixture model that models a contaminant process and provides a likelihood estimate of whether observed behavior is more likely to have been generated by excessively high risk propensity or by

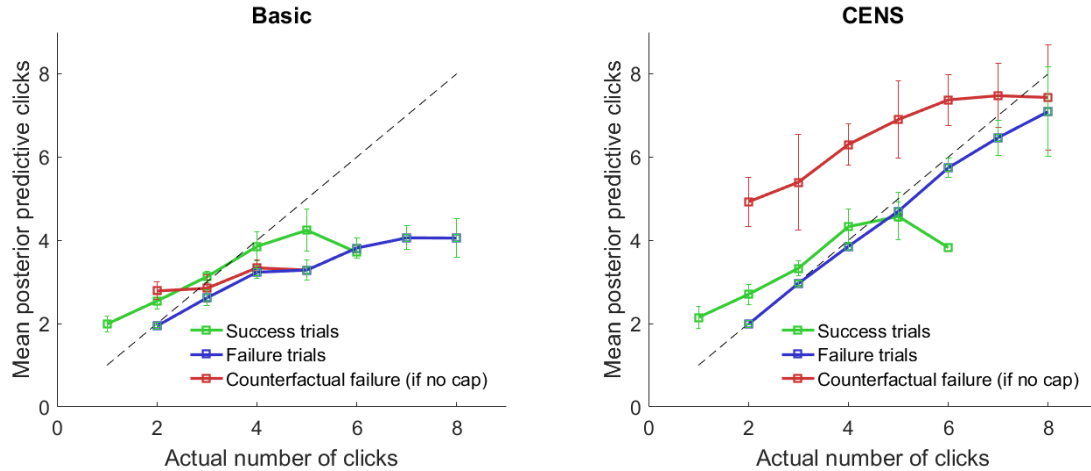


Figure 6: Mean posterior predictive number of clicks based on a basic model and a model that incorporates censoring. The basic model (left) significantly under-predicts failure trials. The censored model (right) generates realistic counterfactual value of the number of clicks that kids might have made on failure trials if there was no cap (i.e. they did not get caught). These predictions are supported by improved bias, RMSE, correlation and DIC values using the censored model.

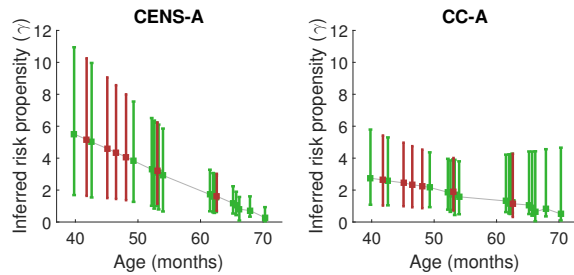


Figure 7: Inferred risk propensity as a function of age. Error bars capture the 95% CI. Bars in red indicate the contaminants identified by the CC-A model.

the contaminant process. We find evidence for the latter. Excluding these contaminant children explains away the age effects. Applying our methods to reanalyze studies that have found age effects could reveal whether these are real or caused by some confound, and provide a Bayesian estimate of the age-effects.

A key finding is that most children seemed to explicitly learn and adjust their beliefs about the probability of getting caught on any click. We also found individual differences in the children's risk propensity levels. Differences in risk propensity as well as the level of learning could stem from personal traits, environmental factors, or both. In the future, we plan to test older children, and to add measures of executive function, in order to determine the particular factors that give rise to individual differences in risk propensity and learning.

Acknowledgments

We would like to thank the administrators, parents, and children at University Montessori Preschool for their participa-

tion, members of the Sarnecka Lab and the Working Memory & Plasticity Lab for useful discussion, and the Research Assistants of the Sarnecka Lab for recruiting and running subjects. We'd also like to thank Chelsea Parlett & Mac Strelloff for useful discussion.

References

- DeLoache, J. S. (2004). Becoming symbol-minded. *Trends in cognitive sciences*, 8(2), 66–70.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, 8(7), 307–314.
- Kerr, A., & Zelazo, P. D. (2004). Development of hot executive function: The children's gambling task. *Brain and cognition*, 55(1), 148–157.
- Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: the balloon analogue risk task (bart). *Journal of Experimental Psychology: Applied*, 8(2), 75.
- Paulsen, D. J., Platt, M. L., Huettel, S. A., & Brannon, E. M. (2012). From risk-seeking to risk-averse: the development of economic risk preference from childhood to adulthood. *Frontiers in psychology*, 3, 313.
- van Ravenzwaaij, D., Dutilh, G., & Wagenmakers, E.-J. (2011). Cognitive model decomposition of the bart: assessment and application. *Journal of Mathematical Psychology*, 55(1), 94–105.
- Weller, J. A., Levin, I. P., & Denburg, N. L. (2011). Trajectory of risky decision making for potential gains and losses from ages 5 to 85. *Journal of Behavioral Decision Making*, 24(4), 331–344.