

# How to Make a Proceedings Paper Submission

Anonymous CogSci submission

## Abstract

Include no author information in the initial submission, to facilitate blind review. The abstract should be one paragraph, indented 1/8 inch on both sides, in 9 point font with single spacing. The heading ‘Abstract’ should be 10 point, bold, centered, with one line of space below it. This one-paragraph abstract section is required only for standard six page proceedings papers. Following the abstract should be a blank line, followed by the header ‘Keywords’ and a list of descriptive keywords separated by semicolons, all in 9 point font, as shown below.

**Keywords:** Add your choice of indexing terms or keywords; kindly use a semi-colon; between each term.

## Introduction

### Models

#### Learning problem

##### Discrete Time Model

We formalized the learning problem that participants face in our experiments as a form of Bayesian concept learning (Tenenbaum, 1999; Goodman, 2006), represented graphically in Fig. X. The goal is to learn a concept  $\theta$ , which is a set of probabilities for independent binary features  $\theta_{1,2,\dots,n}$ , where  $n$  is the number of features. Over the course of a block, the learner receives information about  $\theta$  by observing exemplars  $y$ : instantiations of  $\theta$ , where each feature  $y_{1,2,\dots,n}$  is either on or off. Each feature  $\theta_i$  and its corresponding exemplar  $y_i$  form a Beta-Bernoulli process:

$$p(\theta_i) \sim \text{Beta}(\alpha_i, \beta_i) \quad (1)$$

$$p(y_i|\theta_i) \sim \text{Bernoulli}(\theta_i) \quad (2)$$

Since the features are independent, this relationship holds for the entire concept  $\theta$ . In previous work, two information-theoretic quantities, surprisal and Kullback-Leibler (KL) divergence, resulting from the stimulus were shown to be linked to looking behavior (Kidd et al., 2012; Poli et al., 2020). Surprisal, calculated as  $-\log(p(y|\theta))$ , intuitively refers to how surprising a stimulus  $y$  is given the model’s beliefs about  $\theta$  - the intuition that surprising events should result in longer looking times has served as a foundational assumption in developmental psychology. KL-divergence measures how much a model needs to change to accommodate a new stimulus  $y$ . If an observation causes a large change, we speculated that a proportionally long looking time is necessary to integrate the new information.

##### Continuous Time Model

However, to model the time course of attention, we did not want to assume that stimuli are encoded perfectly and instantaneously. Instead, we suggest that participants gather repeated noisy samples  $\bar{z}$  from the exemplars. For any sample  $z$  from an exemplar  $y$  there is a small probability  $\epsilon$  to misperceive the feature as off when it was actually on, and vice versa. Therefore, by making noisy observations  $\bar{z}$ , the learner obtains information about the true identity of the exemplar  $y$ , and by extension, about the concept  $\theta$ . By Bayes’ rule:

$$P(\theta|\bar{z}) = p(\bar{z}|y)p(y|\theta)p(\theta)/p(\bar{z}) \quad (3)$$

where  $p(\bar{z}|y_i)$  is fully described by  $\epsilon$ , and  $p(y|\theta)$  by Bernoulli processes as in Eq. 2.

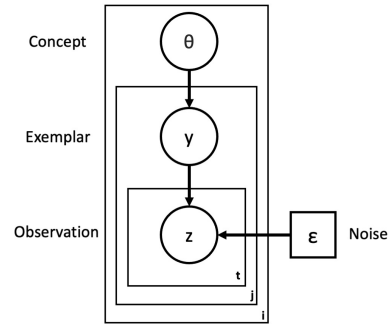


Figure 1: Graphical representation of our model. Circles indicate random variables. The squares indicate fixed model parameters.

**#Sampling** The formulation of the model in continuous time allows us to do two things: First, we can explicitly model the learner’s decision on when to stop sampling by asking the model to decide, after every sample  $z$ , whether it wants to continue sampling from the same stimulus or not. This is in contrast to the discrete time models presented here and in previous work (Kidd et al., 2012; Poli et al., 2020), where one has to resort to correlating information-theoretic measures to looking data, rather than generating it. Second, a consequence of making a decision at every time step allows us to study the behavior of another information-theoretic measure:

the expected information gain (EIG). EIG is commonly used in rational analyses of information-seeking behavior - that is to assess whether information-seeking is optimal with respect to the learning task (Oaksford & Chater, ; Gureckis, ). Importantly, EIG is a forward-looking measure that considers the potential for learning from the next sample. Since discrete time models operate on the level of a whole stimulus, rather than individual samples, EIG would look forward to the next stimulus in these models, rather than the next sample, and therefore not be able to capture the decision of whether to keep looking. EIG to describe looking time is therefore only possible in the continuous time models.

We compute EIG by weighing the information gain from each possible next observation by the probability of that observation. We defined information gain as the KL-divergence between the hypothetical posterior after observing a sample  $z_{t+1}$  and the current posterior:

$$EIG(z_{t+1}) = \sum_{z_{t+1} \in [0,1]} p(z_{t+1}|\theta_t) * KL(\theta_{t+1}, p(\theta_t)) \quad (4)$$

Finally, to get actual sampling behavior from the model, it has to convert EIG into a binary decision about whether continue looking at the current sample, or to advance to the next trial. The model does so using a luce choice between the EIG from the next sample and a constant EIG from looking away.

$$p(\text{look}) = \frac{EIG(z_{t+1})}{EIG(z_{t+1}) + EIG(\text{world})} \quad (5)$$

We also studied the behavior of the model when replacing EIG with continuous time versions of the other linking hypotheses, surprisal (the probability of a given  $z$  under the  $P(\theta_t)$ ) and KL-divergence between the posterior  $p(\theta_t)$  and the prior  $p(\theta_{t-1})$ .

## Experiment

### Methods

**Participants** We recruited 449 participants (Age  $M = 30.83$ ;  $SD = 17.44$ ) on Prolific. They were randomly assigned to one of the three conditions of the experiment (Curiosity:  $N = 156$ ; Memory:  $N = 137$ ; Math:  $N = 156$ ). Participants were excluded if they showed irregular reaction times or their responses in the filler tasks indicates low engagement with the experiment. All exclusion criteria were pre-registered. The final sample included  $N$  participants (Curiosity  $N = 143$ ; Memory:  $N = 98$ ; Math:  $N = 139$ ).

**Procedure** This is a web-based self-paced visual presentation task. Participants were instructed to look at a sequence of animated creatures at their own pace and answer some questions throughout. At the end of the experiment, participants were asked to rate the similarity between pairs of creatures and complexity of creatures they encountered on a 7-point Likert Scale. Each participant saw eight blocks in total, half of which used creatures with high perceptual complexity, and half of which used creatures with low perceptual complexity.

On each trial, an animated creature showed up on the screen. participants can press the down arrow to go to the next trial whenever they want after a minimum viewing time of 500 ms.

Each block consisted of six trials. A trial can be either a background trial (B) or a deviant trial (D). A background trial presented a creature repeatedly, and the deviant trial presented a different creature from the background trial in the block. Two creatures in the blocks were matched for visual complexity. There were four sequences of background trials and deviant trials. Each sequence appeared twice, once with high complexity stimuli and once with low complexity stimuli. The deviant trial can appear at either the second (BDBBBB), the fourth (BBBDBB), or the sixth trial (BBBBBD) in the block. Two blocks do not have deviant trials (BBBBBB). The creatures presented in the deviant trials and background trials were matched for complexity.

Participants were randomly assigned to one of the three conditions: Curiosity, Memory, and Math. The three conditions only differed in the type of questions asked following each block. In Curiosity condition, participants were asked to rate “How curious are you about the creature?” on a 5-point Likert scale. In Memory condition, a forced-choice recognition question followed each block (“Have you seen this creature before?”). The creature used in the question in both conditions was either a creature presented in the preceding block or a novel creature matched in complexity. In Math condition, the participants were asked a simple arithmetic question (“What is  $5 + 7$ ?”) in multiple-choice format.

**Stimuli** The animated creatures (Fig 1) were created using Spore (a game developed by Maxis in 2008). There were forty creatures in total, half of which have low perceptual complexity (e.g. the creatures do not have limbs, additional body parts, facial features, or textured skin), and half of which have high perceptual complexity (i.e. they do have the aforementioned features). We used the “animated avatar” function in Spore to capture the creatures in motion.

### Results

**Analytic Approach** The sample size, methods, and main analyses were all pre-registered and are available at [LINK]. Data and analysis scripts are available at [LINK].

**Manipulation Check** The complex animated creatures were rated as more perceptually complex ( $M =$ ;  $SD =$ ) than the simple animated creatures ( $M =$ ;  $SD =$ ). Pairs of background creature and deviant creature were rated as moderately dissimilar to one another ( $M =$ ;  $SD =$ ).

### Evaluating the Paradigm

```
Linear mixed model fit by REML. t-tests use Satterthwaite's
lmerModLmerTest]
Formula: log(trial_looking_time) ~ I(exp(1)^(-trial_number))
      block_type + (1 | subject)
Data: d
```

REML criterion at convergence: 37415.2

Scaled residuals:

|  | Min     | 1Q      | Median  | 3Q     | Max    |
|--|---------|---------|---------|--------|--------|
|  | -3.7455 | -0.6692 | -0.0499 | 0.5968 | 5.4843 |

Random effects:

| Groups  | Name        | Variance | Std.Dev. |
|---------|-------------|----------|----------|
| subject | (Intercept) | 0.2752   | 0.5246   |
|         | Residual    | 0.4250   | 0.6519   |

Number of obs: 18198, groups: subject, 380

Fixed effects:

```
(Intercept)
I(exp(1)^(-trial_number))
trial_typedeviant
block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant
I(exp(1)^(-trial_number)):block_typesimple_dissimilar
trial_typedeviant:block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant:block_typesimple_dissimilar

(Intercept)
I(exp(1)^(-trial_number))
trial_typedeviant
block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant
I(exp(1)^(-trial_number)):block_typesimple_dissimilar
trial_typedeviant:block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant:block_typesimple_dissimilar

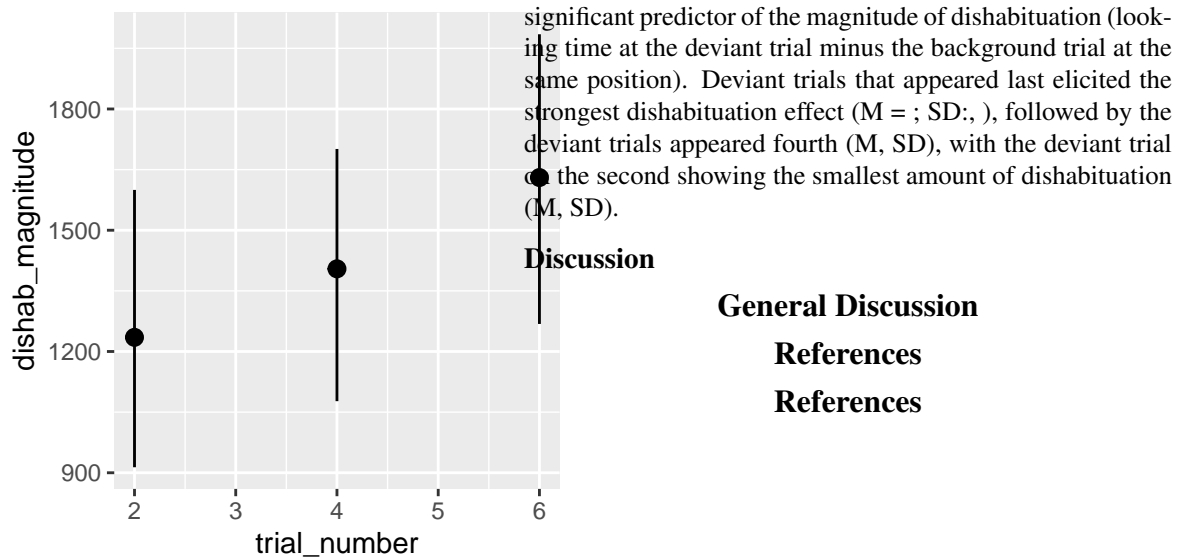
(Intercept)
I(exp(1)^(-trial_number))
trial_typedeviant
block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant
I(exp(1)^(-trial_number)):block_typesimple_dissimilar
trial_typedeviant:block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant:block_typesimple_dissimilar
```

```
trial_typedeviant:block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant:block_typesimple_dissimilar

(Intercept)
I(exp(1)^(-trial_number))
trial_typedeviant
block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant
I(exp(1)^(-trial_number)):block_typesimple_dissimilar
trial_typedeviant:block_typesimple_dissimilar
I(exp(1)^(-trial_number)):trial_typedeviant:block_typesimple_dissimilar

---
Signif. codes:  0 '***' '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1.0
              7.531e+00
Correlation of Fixed Effects:
      (Intr) 6.052e-01) trl_ty blk__ I(x(1)^(-_))
I(x(1)^(-_)) -0.196-5.175e-02
trl_typedvnt -0.108-2.020e+00
blk_tyspm_ -0.228-4.020e-01      0.237
I(x(1)^(-_)):_ 0.032-1.099e-01 -0.647 -0.070
blk_tyspm_ 0.038 5.037e-01 -0.144 -0.606 0.116
trl_typedvnt 0.077 0.144 -0.708 -0.336 0.458
I((1)^(-_)):_ -0.023 2.041e-02 0.458 0.099 -0.708
trl_typedvnt 5.415e-02
I(x(1)^(-_)) 2.737e-02
trl_typedvnt 1.299e-02
blk_tyspm_ 3.310e-01
I(x(1)^(-_)):_ 7.652e-02
I((1)^(-_)):_ 3.865e-02
blk_tyspm_ 4.679e-01
I((1)^(-_)):_ -0.647      df
              4.432e+02
              1.781e+04
              1.781e+04
              1.781e+04
              1.781e+04
              1.781e+04
              1.781e+04
              1.781e+04
              t value
              264.857
              39.240
              22.848
              -3.984
              -6.724
              1.22e-10

Three criteria were selected to evaluate whether the
paradigms successfully captured the characteristic looking
time patterns observed in infant literature: habituation (the
decrease in looking time for a stimulus with repeated pre-
sentations), dishabituation (the increase in looking time to a
new stimulus after habituated to one stimulus), and complex-
ity effect (longer looking time for perceptually more complex
stimuli). To evaluate the phenomenon quantitatively, we ran a
linear mixed effects model with maximal random effect struc-
ture. [DESCRIBE THE MODEL] [REPORT THE MODEL
RESULTS]
```



significant predictor of the magnitude of dishabituation (looking time at the deviant trial minus the background trial at the same position). Deviant trials that appeared last elicited the strongest dishabituation effect ( $M =$  ;  $SD =$  , ), followed by the deviant trials appeared fourth ( $M$ ,  $SD$ ), with the deviant trial the second showing the smallest amount of dishabituation ( $M$ ,  $SD$ ).

## Discussion

## General Discussion

## References

## References

### Order Effect

```
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: dishab_magnitude ~ trial_number + (1 | subject)
Data: dishab_d
```

REML criterion at convergence: 44965.3

Scaled residuals:

| Min     | 1Q      | Median  | 3Q     | Max    |
|---------|---------|---------|--------|--------|
| -7.5761 | -0.3319 | -0.0903 | 0.2873 | 7.5993 |

Random effects:

| Groups   | Name        | Variance | Std.Dev. |
|----------|-------------|----------|----------|
| subject  | (Intercept) | 2450681  | 1565     |
| Residual |             | 21340805 | 4620     |

Number of obs: 2272, groups: subject, 380

Fixed effects:

|              | Estimate | Std. Error | df      | t value | Pr(> t )     |
|--------------|----------|------------|---------|---------|--------------|
| (Intercept)  | 1029.26  | 268.80     | 2225.33 | 3.829   | 0.000132 *** |
| trial_number | 99.05    | 59.39      | 1890.38 | 1.668   | 0.095529 .   |

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Correlation of Fixed Effects:

|             | (Intr) |
|-------------|--------|
| trial_numbr | -0.884 |

While visualizing the data, we unexpectedly found that the position in which the deviant trial appeared in the sequence had an effect on the shape of the habituation and dishabituation curves. To explore this phenomenon quantitatively, we operationalized the magnitude of dishabituation as the difference between the looking time at the deviant trial minus the background trial at the same position. Then, we fit a mixed effect model with the position of deviant as fixed effect and [??] as a random effect. We found that the position was a