Template taken from the OSF Pre-registration Challenge
PhotoDraw Pregistration, 7/20/2018

Research Questions (required)

When children are asked to “draw a dog”, older children tend to produce more recognizable drawings (Long, Fan & Frank, 2018). However, younger children may have a harder time remembering what a dog “looks like”, and therefore include fewer details -- and this may lead to them producing less recognizable drawings. To what degree are children’s drawings influenced by their ability to remember what things look like?

1. Will children produce more recognizable drawings when they are shown a canonical photograph of an object category *before* being asked to draw this category?
2. Will children produce more recognizable drawings when they are shown a canonical photograph of an object category *while* being asked to draw this category (and asked to draw the object in the picture)?

Hypotheses:

We predict that showing children a canonical exemplar of a category will lead to more recognizable drawings, both when it is shown for 5 seconds before they draw (e.g., “working memory” condition) and when it remains on the screen the entire time (e.g. “perception” condition) relative to when they are asked to draw from their memory.

Sampling Procedure

Please describe the process by which you will collect your data. If you are using human subjects, this should include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline. For studies that don’t include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples, or batch numbers you will use.

Participants will be recruited from the floor of the Children’s Discovery Museum in San Jose, CA. Children who have documented learning disabilities or impaired vision/hearing will be not be recruited. Subjects will not be paid for their time but may receive a sticker for participating.

Existing Data

None of the data has yet been collected for this study. We have piloted the study at Bing Nursery School in a within subjects paradigm to ensure feasibility in our youngest children. We found that a within-subjects design had potential task carry-over effects and are now completing this pre-registration prior to beginning data collection at the San Jose Children's Discovery Museum.

Sample Size

Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If

the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?

We will recruit participants until we have usable data from 20 children in each condition (3 conditions, 60 children total). See "Exclusion criterion" for how individual trials/children will be excluded from analyses.

Sample Size Justification *This could include a power analysis or an arbitrary constraint such as time, money, or personnel.*

This sample size will allow us approximately 100 drawings in each condition for our primary analyses, giving us what we hope is reasonable power.

If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.
(n/a)

Manipulated variables (required)
Describe all variables you plan to manipulate and the levels or treatment arms of each variable. For observational studies and meta-analyses, simply state that this is not applicable.

The primary experimental manipulation will be the type of cue used to prompt children to produce a figurative drawing (i.e., a drawing that represents a real-world object). This manipulation will be performed between-participants. Children will be psuedo-randomly assigned to one of the three experimental conditions depending on their age.

Specifically, there will be three cue types: (1) a verbal cue provided by the experimenter in a 5s video clip ("semantic" condition) (2) a photo cue presented for 5s accompanied by the audio from the same video clip, which then disappears; ("working memory" condition) (3) a photo cue which then remains on screen continuously while children produce their drawing (30s maximum, "perception" condition). In the perception condition, children will hear, "What about this [x]? Can you draw the [x] in the picture?". In the two other trials, they will hear, "What about a [x]? Can you draw a [x]?"

Measured variables (required)
Describe each variable that you will measure. This will include outcome measures, as well as any predictors or covariates that you will measure. You do not need to include any variables that you plan on collecting if they are not going to be included in the confirmatory analyses of this study.

(1) The child's age.
(2) Each stroke made by each child (x and y coordinates, along with the timestamp of the stroke, saved in svg format). These strokes will be rendered out into a .png of the final

drawing. This same type of data is collected on the tracing trials and the drawing trials. Each stroke will be sent to a secure database linked to a session ID and the child's age.

(3) Recognizability of each drawing, assessed behaviorally in an AFC task. Three foils will be chosen for each of the five target items (cup, train, shoe, cat, rabbit) and will be: arm, bottle, spoon, car, bus, couch, hat, book, lamp, dog, mouse, horse, bird, bear, lion). 30 participants will be recruited from amazon mechanical turk as assessed in Long, Fan, & Frank (2018) (see paper at http://langcog.stanford.edu/papers_new/long-2018-cogsci.pdf).

(4) Deep-CNN Features for each drawing in Layer 7 of VGG-19 (see also Long, Fan, & Frank for a detailed description of how these are extracted). Features will be extracted using a pre-trained, python implementation of VGG-19 in PyTorch (see code at https://github.com/brialorelle/kiddraw/blob/master/analysis/museumdraw/python/extract_features.py). The model has frozen weights and has not been tuned on drawings or sketches in any way and is publicly available.

Indices (required)

If any measurements are going to be combined into an index (or even a mean), what measures will you use and how will they be combined? Include either a formula or a precise description of your method. If your are using a more complicated statistical method to combine measures (e.g. a factor analysis), you can note that here but describe the exact method in the analysis plan section.

(1) Drawing time will be calculated at the length of time from the first stroke to the last stroke.

(2) A drawing will be counted as "recognized" if a rater choses the correct label for the image.

Study design (required)

Describe your study design. Examples include two-group, factorial, randomized block, and repeated measures. Is it a between (unpaired), within-subject (paired), or mixed design? Describe any counterbalancing required. Typical study designs for observation studies include cohort, cross sectional, and case-control studies.

All children will complete three tracing trials in order to familiarize them with the tablet and to assess their motor production abilities. This will yield a within-subjects measure which will be included as a covariate. Afterwards, they will complete five figurative drawing trials. The whole experient should take less than 3 minutes. The experimenter will not provide explicit feedback on the drawings, but will hold the tablet for the child as needed.

The main design is a between-subjects manipulation with three conditions: drawing from perception, drawing from working memory, and drawing from semantic memory. As noted below, children will be pseudo-randomly assigned to one of these three conditions.

See https://rxdhawkins.me:8882/index.html for a demonstration of the task. (Note that you need to input the condition (W,S,P) into the text box on the first page of the task. The task is currently formulated for a 12.9 inch iPad pro and does not adapt to every monitor screen size; however it easily be viewed within the developer console of Google Chrome by right-clicking and selecting "Inspect".)

Randomization (required) *If you are doing a randomized study, how will you randomize, and at what level?*
Children will be recruited and assigned to a drawing condition the based on their age (i.e., the experimenter will attempt to recruit an even distribution of children across this age range in each condition). If a child had to be excluded post-hoc for not meeting our data inclusion criterion (see below), then a child may be assigned a specific condition in order for there to be an equal number of children from each drawing condition.

The five items (cup, train, shoe, cat, rabbit) will be randomized within each task group automatically each time the experiment is run (implemented in javascript).

Statistical models (required)
What statistical model will you use to test each hypothesis? Please include the type of model (e.g. ANOVA, multiple regression, SEM, etc) and the specification of the model (this includes each variable that will be included as predictors, outcomes, or covariates). Please specify any interactions that will be tested and remember that any test not included here must be noted as an exploratory test in your final article.

1. **Analysis 1 (primary):** Does the recognizability of children's drawings differ in the perception, working memory, vs. semantic conditions? We expect drawings in the perception condition to be the most recognizable. We will model recognizability scores for each drawing as a function of these three conditions in a generalized non-linear mixed-effect model, as in Long, Fan, & Frank (2018).
    a. **Dependent variable:** The proportion of adults who correctly recognized each drawing derived from the behavioral experiment.
    b. **Independent variables**: The drawing condition (semantic condition, working memory condition, or perception condition) for each drawing and its interaction with children's age.
    c. **Random effects:** We will include random intercepts and random slopes for both items (e.g., cup, cat) and participants. If the model fails to converge, we will eliminate random slopes for items and then participants. We will strive to keep the model as maximal as possible.

In lme4, this model would be specified as:
glmer(cbind(count_recognized, count_not_recognized) ~ (drawing_condition)
*scale(childs_age) + (1 + drawing_condition | child_id) + (1+ drawing_condition |category)

2. **Analysis 2 (secondary):** In a secondary analysis, we will rerun the same analysis but include covariates related to the amount of effort spent drawing (e.g. time, number of strokes) and children's motor control abilities as assessed by tracing score performance (pixel-wise F1 score on "shape" trials).
    a. **Covariates**: (1) We will include children's motor abilities performance, as assessed by the tracing trials, as a covariate in these analyses. A "motor score" will be derived by taking a pixel-wise F1 score on the "shape" tracing trials only and applied to all drawings made by this child. (2) We will also include the amount of time spent on each drawing, as well as (3) the number of strokes made. All covariates will be scaled.

3. **Analysis 3 (exploratory):** Are drawings made in the perception / working-memory conditions more similar to "the" photo of the [cat] that they drew vs. "[cats]" in general?
    a. *Specific-photo RDM:* We will construct a representational dissimilarity matrix (RDM) based off of the feature representations (in VGG-19, Layer 7) from the specific 5 photos that were used as cues in the experiment.
    b. *Category-average RDM:* We will construct a representational dissimilarity matrix (RDM) based off of the feature representations (in VGG-19, Layer 7) from 20 photos from each category -- photos will be full-color, isolated pictures of the objects on white backgrounds in canonical viewpoints, similar to the specific photos that are used in the experiment (and the previous RDM).
    c. We will then compare the RDMs made from the sketches in each condition (working memory, semantic, and perception) with these two matrices (a,b) using Spearman's rank correlations (as in Long, Fan & Frank, 2018), yielding a total of 6 correlations. Spearman's rank correlation will be assessed using the upper triangle of each matrix and by excluding the diagonal.
    d. We expect that correlations will be higher between the perception and the specific-photo RDM than with the category-average RDM. We expect that this difference in correlation may be attenuated for the working memory condition, and lowest (or reversed) for drawings from the semantic condition.

Transformations (required)

If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process.

For analysis 1 & 2, as noted above, covariates will be scaled for the GLM analyses. As we expect recognizability rates to be lowest in the "semantic" condition, this will be the reference condition for the "drawing_condition" categorical variable.

For analysis 3, all VGG features will be normalized across the entire image set. All RDMs will be computed using correlation distance (1 - pearson's correlation).

Follow-up analyses (required)
If not specified previously, will you be conducting any confirmatory analyses to follow up on effects in your statistical model, such as subgroup analyses, pairwise or complex contrasts, or follow-up tests from interactions? Remember that any analyses not specified in this research plan must be noted as exploratory.
None planned.

Inference criteria (required)
What criteria will you use to make inferences? Please describe the information you'll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?
Analysis 1 & 2: We predict that both the working memory condition and the perception condition will have coefficients in the GLMM that explain significant variance relative to the reference condition (the semantic drawing condition). This will be assessed by examining whether the p-value corresponding to the model coefficient is < .05.
Analysis 3: We do not have specific predictions about the exact magnitude of the correlations. We will interpret the resulting correlation coefficients as part of an exploratory analysis.

Data exclusion (required)
How will you determine which data points or samples (if any) to exclude from your analyses? How will outliers be handled?

Trial-level criterion for exclusion:
—If zero strokes are made, the trial will be excluded
—If two objective, condition-blinded observers BOTH agree that a drawing is a "scribble" in the sense that the child is non-compliant with the task instructions, this drawing will be excluded from all analyses.

Child-level criterion for exclusion:
–If a child contributes fewer than 3 drawings (excluding tracing trials) after these exclusion criteria, their data will not be included in analysis.
–If the child does not follow instructions by either:
(1) scribbling three times in a row (round and rounds or back-and-forth) on any kind of trial (including tracing trials or task practice trials) *OR*
(2) explicitly stating that they are drawing something other than the cued item more than twice (i.e. 3 times).
If either of these three things happen, the experimenter will end the session and thank the child for participating, and their data will NOT be included in analyses.

Missing data (required) *How will you deal with incomplete or missing data?*

For our primary analysis, data will be modeled at the level of the individual drawing. Our child-level exclusion criteria will ensure that we have a comparable number of drawings in each condition.

Exploratory analysis (optional)

If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time.

(n/a)

Upload an analysis script with clear comments (optional)

This optional step is helpful in order to create a process that is completely transparent and increase the likelihood that your analysis can be replicated. We recommend that you run the code on a simulated dataset in order to check that it will run without errors.

(n/a)

Other (optional)

If there is any additional information that you feel needs to be included in your pre-registration, please enter it here.

(n/a)