Study Information

Title

Provide the working title of your study. It is helpful if this is the same title that you submit for publication of your final manuscript, but it is not a requirement.

Learning complex questions from the input

Authors

The author who submits the preregistration is the recipient of the award money and must also be an author of the published manuscript. Additional authors may be added or removed at any time.

Ben Ambridge

Research Questions

Please list each research question included in this study.

It is sometimes claimed (e.g., Chomsky, 1980; Crain & Nakayama) that children cannot learn the structure of complex yes/no questions - i.e., questions containing a relative clause (e.g., Is the boy who is smoking crazy?) - from the input, because the input does not contain any questions of this form. However, some authors (e.g., Ambridge, Rowland & Pine, 2008) have argued that children can learn the structure of such questions from the input, by combining simple questions (e.g., Is the boy crazy?) and complex noun phrases (e.g., the boy who is smoking); both of which do occur in the input. This latter possibility will be investigated by means of a training study. An experimental and a control group (matched on a standardized test of grammar) will receive identical training on yes/no questions (e.g., Is the bird cleaning?; Is the fish swimming?). However, while the experimental group will be trained on complex noun phrases (e.g., the bird who's happy; the fish who's happy), the control group will be trained on semantically matched simple noun phrases (e.g., the happy bird; the happy fish). We predict that the experimental group will outperform the control group in a final test session, in which complex questions with new verbs and noun phrases (e.g., Is the crocodile who's hot eating?) will be elicited

Hypotheses

For each of the research questions listed in the previous section, provide one or multiple specific and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here.

1. The experimental group will produce significantly more correctly-formed complex questions (out of a maximum of 12) than the control group (as determined by maximal mixed effects models; e.g., Barr et al, 2013) 2. Significantly more children in the experimental than control group will produce at least one correctly formed question (as determined by chi-square test)

Sampling Plan

Existing Data

Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please do not hesitate to contact us if you have questions about how to answer this question (prereg@cos.io).

Registration prior to creation of data

Explanation of existing data

If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study. The purpose of this question is to assure that the line between confirmatory and exploratory analysis is clear.

NA

Data collection procedures

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.

All participants will be recruited from UK reception (age 4-5 years) and Year 1 (age 5-6 years) classes. Children will be recruited via their schools. Parents will be sent an information sheet and consent form, and children will be asked to give verbal assent. Children must be native speakers of English with no known language impairment. Children will receive training on 12 simple questions and 12 noun phrases per day, for five days (i.e., a total of 60 simple questions and 60 noun phrases). Children will repeat these questions/noun phrases spoken by the experimenter (in the context of a "bingo" game). Any child who does not correctly repeat at least half of the noun phrases and at least half of the questions on all five days will be excluded. We will endeavour to train children on five consecutive days and test on the fifth day. This may not always be possible because of absence. Any child who does not complete all five training sessions and the test session within a 12 day period (two Monday-Friday school weeks plus the 2-day weekend in between) will be excluded. We will verify that the experimental and control groups did not differ significantly in the total number of days taken to complete training.

no file selected

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?

122 children: 61 each in the experimental and control groups

Sample size rationale

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

Power calculation with d=0.3, Power = 0.5. Although a power of > 0.5 would have been desirable, a total sample size of 122 is our maximum in terms of time, money and personnel.

Stopping rule

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.

Data collection will not be terminated until the full sample has been recruited. Children who are excluded for not completing all five training days within a 12-day window will be replaced.

Variables

Manipulated variables

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 study has only one manipulated variable: Training Group, with two levels: Experimental (simple questions + complex noun phrases) and Control (simple questions + simple noun phrases).

no file selected

Measured variables

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.

We will elicit 12 questions from each child using the same elicitation procedure as in Ambridge et al (2008): http://pcwww.liv.ac.uk/~ambridge/Papers/Ambridge,%20Rowland%20&%20Pine%20(2008).pdf The DVs will be (a) Number of correct complex questions per child (out of a maximum of 12) (b) Number of children in each group producing at least one correctly formed question Covariates will be Age in months Score on standardized test of grammar Days taken to complete the training session Number of simple questions correctly repeated in training session Number of noun phrases correctly repeated in training session

no file selected

Indices

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.

NA

no file selected

Design Plan

Study type

Please check one of the following statements

Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

Blinding

Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.

For studies that involve human subjects, they will not know the treatment group to which they have been assigned.

Study design

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.

Between-subjects (unpaired) two-group design: Experimental vs Control Group NB: A pre-training baseline test is not possible, as this would involve exposing children to the complex questions that are elicited in the test session. We are therefore including a standardized test of grammar in lieu of a pre-test.

no file selected

Randomization

If you are doing a randomized study, how will you randomize, and at what level?

Children will be randomly allocated to the control vs experimental group by coin toss, before any training or testing begins

Analysis Plan

Statistical models

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.

To compare the number of correct questions produced by each group, we will run a linear mixed effects regression model in R with random intercepts for participant and item (test trial) and a by-participant random slopes for item (test trial). In case of non-convergence, we will simplify the model using the procedure outlined by Barr et al, 2013. Fixed effects will be Group (experimental vs control; coded as 1 vs 0) Age in months Score on the standardized grammar test Days taken to complete the five training blocks Number of simple questions correctly repeated during training Number of noun phrases correctly repeated during training No interactions will be included To compare the number of participants in the experimental and control groups producing at least one correct complex question, we will run a 2x2 chi square analysis: group (experimental/control) x children producing/not producing at least one correct complex question

no file selected

Transformations

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

Group will be coded as experimental (1) vs control (0). No other transformations will be used.

Follow-up analyses

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.

No

Inference criteria

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?

For the lmer, we will use the p values obtained by the model comparison (chi-square) procedure: i.e., does removing the predictor of Group from a maximal (converging) model result in significantly worse model fit at p<0.05 (one tailed, since we have a directional prediction).

Data exclusion

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

All children who complete the training and test to criterion (outlined above) will be included, and any who do not will be replaced.

Missing data

How will you deal with incomplete or missing data?

NA - All children who complete the training and test to criterion (outlined above) will be included, and any who do not will be replaced.

Exploratory analysis

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.

NA

Scripts

Upload an analysis script with clear comments

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.

no file selected

Other

Other

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

Socialize

 (http://twitter.com/OSFramework)  (https://www.facebook.com/OpenScienceFramework)
 (https://groups.google.com/forum/#!forum/opensciencefra mework) 
(https://www.github.com/centerforopenscience) 
(https://plus.google.com/103557785986281627994)
