

TITLE**:** **The influence of input structure on learning**

PROTOCOL VERSION DATE**:** August 20, 2023

VERSION**:** 1

# **PRINCIPAL INVESTIGATOR (PI)**:

Name: Lei Yuan

Telephone: 303-492-7487

Email**:** lei.yuan@colorado.edu

# KEY PERSONNEL

**Name**: Lucile Vleugels

**Role in project**: Graduate Student

**Name**: Christina Collins

**Role in project**: Lab Coordinator

**Name**: Chloe Nowak

**Role in project**: Undergraduate research assistant

# GENERAL RESEARCH STAFF

Five undergraduate/graduate research assistants will assist with this protocol throughout the time it takes to conduct the study described here. The PI will ensure that appropriate CITI and protocol specific training is maintained and DEPA reporting is conducted annually. General research staff responsibilities will include, for example, distributing recruitment flyers, contacting potential participants and their families, scheduling, administering survey instruments, conducting human-subject studies, and entering data.

# OBJECTIVES

This study aims to investigate how the structure of different language systems affect learning and attention patterns during learning.

# BACKGROUND AND SIGNIFICANCE

Variations in language structures impact learning significantly. For example, in English, 11 and 12 are named "eleven" and "twelve". In contrast, they are named as "ten one" and "ten two" in Chinese. The Chinese names are arguably more “systematic” to a new learner, as the base-10 structure of our number system is more directly represented in the names. Correlational evidence has shown that Chinese-speaking children acquire the Arabic written multi-digit system earlier than their English-speaking counterparts. This prompts our question: Do more systematic languages lead to more efficient learning, as compared to less systematic languages? How might these learning differences unfold over time during learning?

The current study aims to explore whether a language’s number naming system contributes to the efficiency of learning new concepts. If so, how does the impact of language structure impact learning in real-time contexts? Does a learner's eye movement and behavior change depend on whether a language's structure is more or less systematic? By collecting empirical data, we seek to uncover how a language’s structure may affect the process of learning new concepts. This inquiry advances our understanding of the intricate relationship between language systems, attentional patterns, and statistical learning to optimize instructional strategies for future education practices.

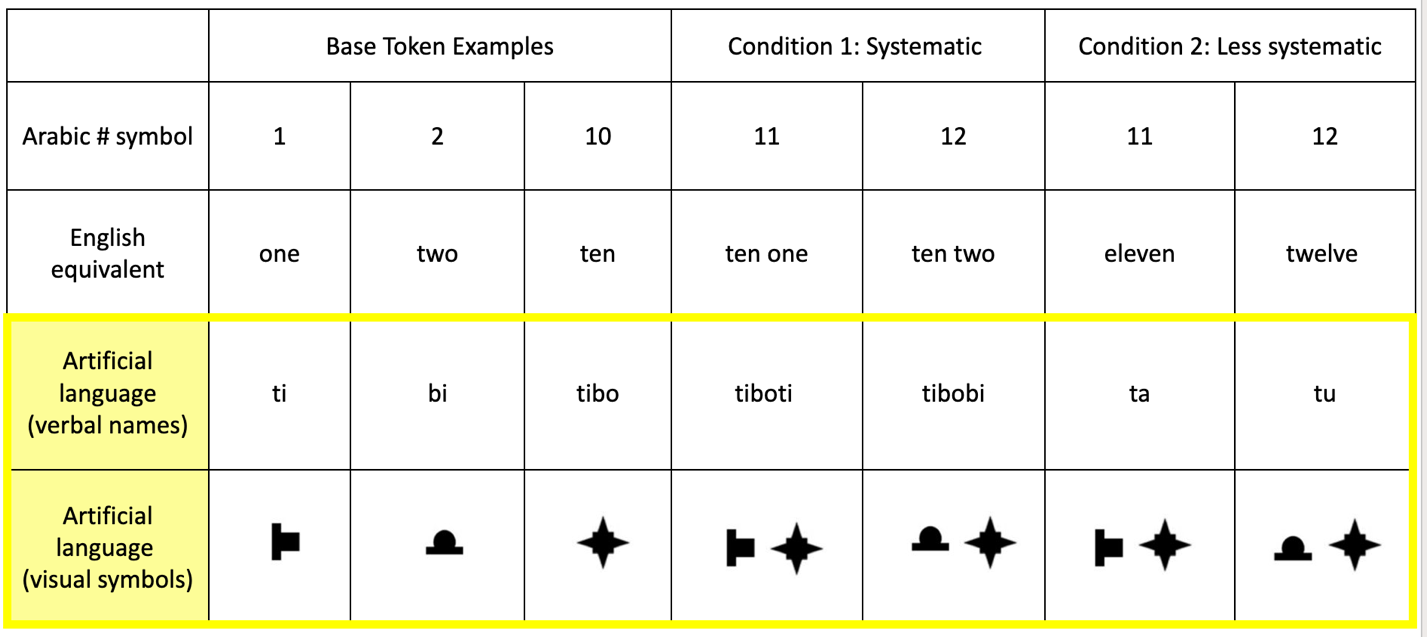
# PRELIMINARY STUDIES

N/A

# RESEARCH STUDY DESIGN

**Design:** We will have two between-subject conditions: 1) Systematic language condition and 2) Less systematic language condition. Participants will be randomly assigned to one of the conditions. Both conditions use artificial stimuli and none of our participants will have had any prior experience with our stimuli.

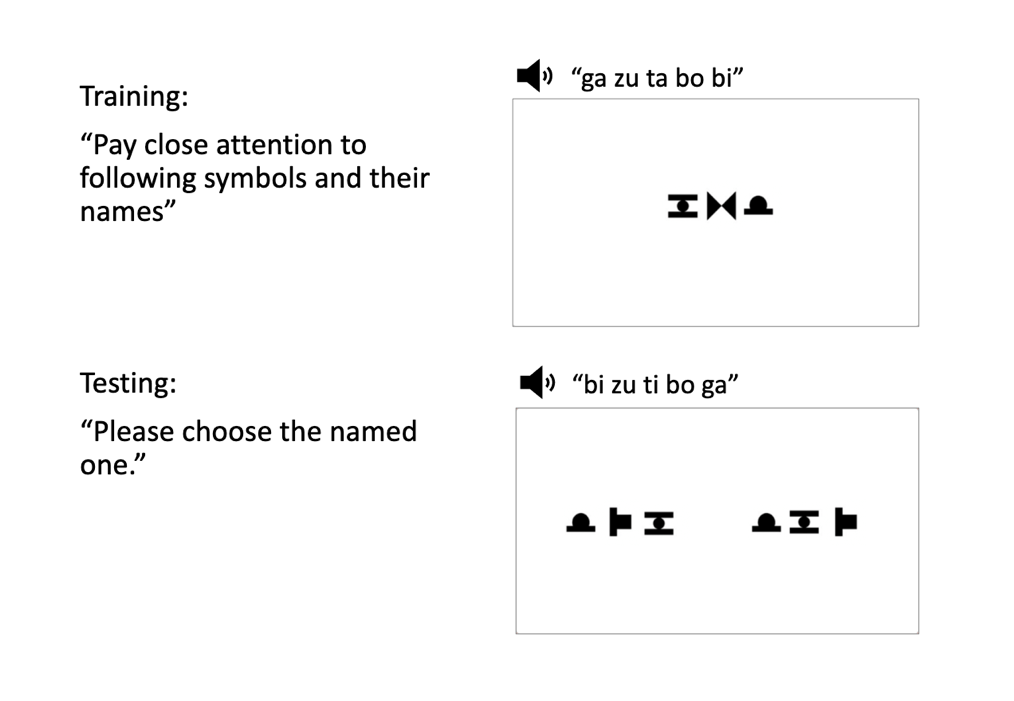
Figure 1 illustrates the main difference between the two conditions: displaying the Arabic symbol for the number, the English or Chinese name, and the verbal and symbolic artificial name in the equivalent language condition. The artificial language systems that we created are inspired by the multi-digit number naming system. As shown in Figure 1, each visual symbol maps onto its corresponding verbal names like the Arabic system. For example, the first artificial visual symbol from the left is called “ti”, and it is equivalent to “1” and the English name “one”. The important difference between the two conditions is in the systematicity of the mapping between visual symbols and their verbal names. In the Systematic language condition, the left symbol pair is named as “tiboti” and mimics the Chinese naming of “ten one”. In contrast, for the Less Systematic language condition, the same visual symbol is named with a new unique verbal name “ta” as “eleven” in English.



(Figure 1: Overview of differences between Systematic and Less Systematic conditions)

**Describe the task**: Participants' ability to recognize named configurations between two choices will be assessed following their training on the respective artificial language condition. Crucially, participants will undergo four training blocks, each succeeded by a testing round. This design enables the examination of learning speed across the study duration and potential divergences between conditions.

Throughout the training phases, participants will be exposed to pairs of verbal names and visual symbol stimuli derived from the artificial language. Notably, participants will be passive observers during training, solely tasked with absorbing the verbal names and visual symbol presented together. Subsequently, the testing phases will present participants with two sets of visual symbols alongside a verbal name. Upon hearing the verbal cue, such as "bi zu ti bo ga," participants will be prompted to select the corresponding visual symbol. Responding will involve using the "left" or "right" arrow keys on a keyboard. Figure 2 outlines the general task procedures, showing the researcher-spoken prompts and the accompanying figures and verbalization played on the screen.



(Figure 2: Examples of training and testing phases)

**Data to be collected:** We will measure how the different language structures affect learning by comparing ***behavioral performance*** outcomes during testing from each language condition. Performance will be measured in terms of ***accuracy and reaction time***. We also collect ***eye tracking*** data to investigate if the different structural regularity in the languages leads to differential attentional patterns during learning.  Finally, we will collect 3rd person video footage of the trials, to provide trial record and in case of eye tracker camera failure.

**Data analysis methods:** We will compare how participants’ accuracy varies between the two experimental conditions and with participants’ prior knowledge. Eye-movement data will be analyzed using time-series analysis methods.

**Sample size:** A power analysis conducted in G\*power indicates that, for a GLM with two predictors (condition, training block), a sample size of 50 is needed to achieve 80% power with a median effect size of 0.20. We will recruit 60 participants to ensure the power of our study and potential problems with data collection.

**Duration:** This study is expected to take approximately 1 year from recruitment to study closure.

|  |  |
| --- | --- |
| Name of procedure/instrument/tool | Purpose (i.e., what data is being collected?) |
| Eye Tracker | Measures participants gaze data and produces data mapping where they are looking on the screen. |
| Computer and Keyboard | Used to present experiment stimuli and measure participants behavioral responses during testing (“left” vs. “right” choices) |
| 3rd Person Camera | Films a 3rd person view of participants |

# FUNDING

None.

# ABOUT THE SUBJECTS

1. For this study, we plan to enroll 60 adult undergraduate students from the SONA System Research Participant Pool. Subjects will not be recruited nor discriminated against based on gender, race, ethnicity, or age.

|  |  |
| --- | --- |
| Subject Population(s) | Number to be enrolled in each group |
| CU Boulder Undergraduates recruited from PSYC1001 via the SONA System Research Participant Pool | 60 |

# VULNERABLE POPULATIONS

No vulnerable populations will be used in this study.

# RECRUITMENT METHODS

Subjects will be drawn from the SONA System Research Participant Pool. SONA subjects are enrolled in the pool via an accredited psychology class. Subjects voluntarily participate in the pool based on class requirements to obtain research points.

|  |
| --- |
| List recruitment methods/materials and attach a copy of each in eRA |
| 1. SONA System Research Participant Pool |

# COMPENSATION

Participants will be given 2 credits at the conclusion of the study session.  Participants will be compensated regardless of their completion of this study. In cases where a scheduled participant does not show up for their study time, they will not be compensated.

# INFORMED CONSENT

Consent will be obtained in the DEL lab in the Muenzinger Psychology building. Consent documents will be printed and manually signed by each participant if they choose to participate. A copy will then be given to the participants for them to keep.

# PROCEDURES

**Consent & Set-up**

Upon arrival at the DEL lab, participants will be guided to the testing room by a lab research assistant. Consent forms for the experiment will be provided for review and signature. Once consent is obtained, participants will receive an orientation session regarding the data collection instruments including eye tracking and the computer keyboard. For eye tracking setup, a sticker will be affixed to the participant's forehead, which is not intrusive and allows the eye tracker to accurately calibrate and record their eye movements.

**Task Orientation & Conditions**

After the setup is finalized, participants will be introduced to the task structure. The experiment encompasses two phases—training and testing—conducted sequentially. All participants will be randomly assigned to either the "Systematic" or "Less Systematic" condition, tasked with learning a new language (as described above).

**Training & Testing**

During training, participants were told to pay attention to the screen as visual symbols appear accomplished by verbal names. The verbal names will be presented to participants auditorily. Participants will wear headphones to ensure that the auditory information is clearly presented without external noise interference. They are not directed to focus on any particular language features in either condition; their goal is simply to attend to the visual and auditory information as they are presented.

There will be four blocks of training and testing. During testing, visual stimuli will be presented, and participants will be prompted to select the stimulus corresponding to the heard verbal cue. Response input will be recorded via the "left" or "right" keyboard keys. Throughout the training and testing phases, eye movement will be captured and analyzed to track the evolution of gaze patterns.

|  |  |  |  |
| --- | --- | --- | --- |
| Visit # | Procedures/Tools | Location | How much time the visit will take |
| Visit 1 | * Artificial language learning task: Participants will be presented with visual and auditory information and will be asked to pay attend and they will later be tested on what they have remembered and learn about them. | DEL Lab, Muenzinger | 40 minutes |

# SPECIMEN MANAGEMENT

N/A

# DATA MANAGEMENT

The data will be managed according to the University’s highly confidential data standard. To protect human subjects and maintain confidentiality, a unique ID number will be assigned to each participant for data recording, storage, and analysis. The only name identification will be on the consent form that the participants will have filled out. Only participants who provide written consent will participate. All consent forms will be stored and locked away in a secure filing cabinet in a locked room in the PI’s laboratory, accessible only to authorized project members who have passed tests on the protection of human subjects and responsible conduct of research and have been proven and listed on relevant IRB protocols.

Digital data will be stored on physical servers housed in the Space Science Center (SPSC) located in the Research Park on the East Campus of the University of Colorado Boulder. Data is directly accessible (read/write) from computational resources within the CU research computing environment for users who are students, faculty, or staff members, have a valid university login ID, have passed Human subjects training (CITI), and have been approved both by the PI and by the university’s Office of Information Technology. Data is accessible from outside the CU research computing environment only via secure data transfer protocols (scp, sftp, gridftp/GlobusOnline) through designated gateway nodes.  De-identified data will be retained indefinitely on the servers.

# PROVISIONS TO PROTECT THE PRIVACY INTERESTS OF PARTICIPANTS

Participants will be tested in a private room, and only trained researchers will have access to that room and the collected data. Participants will only interact with the trained researchers during the informed consent and data collection process.

# WITHDRAWAL OF PARTICIPANTS

No participants will be withdrawn without their consent. Data for any subjects who choose to withdraw will be destroyed.

# RISKS TO PARTICIPANTS

There are no foreseeable risks to participants.

# MANAGEMENT OF RISKS

The consent form will provide general descriptions of the activities that participants will engage in during the study. The experimenter will explain to the participant that they may withdraw from the study without any penalty.

# POTENTIAL BENEFITS

There are no direct benefits to the participant.

# PROVISIONS TO MONITOR THE DATA FOR THE SAFETY OF PARTICIPANTS

N/A

# MEDICAL CARE AND COMPENSATION FOR INJURY

N/A

# COST TO PARTICIPANTS

There is no financial cost for participating in the study.

# DRUG ADMINISTRATION

N/A

# INVESTIGATIONAL DEVICES

N/A

# WORKING WITH OTHER INSTITUTIONS

N/A

# SHARING OF RESULTS WITH PARTICIPANTS

The results of this research will not be shared with participants.

# REFERENCES

N/A