

1 Experiment 2

1.1 Method

1.1.1 Participants

XX participants were recruited from the psychology undergraduate participant pool at the University of Connecticut (X Female, X Male, mean age = X).

1.1.2 Category Learning Tasks

This experiment used three different category learning tasks, each based on a different approach to category learning. We used these three tasks to investigate whether the paradigms used in different approaches engage category learning systems in a similar way. The order of category learning tasks was counterbalanced across participants. All category learning tasks were presented using PsychoPy v.1.84.2 (?).

Sloustky statistical density task. This task used the same procedure and stimuli as the task described in Experiment 1. However, instead of completing only two blocks, participants completed all four blocks. Because the previous experiment showed few significant order effects, the order of the four blocks was randomly generated for each participant.

Ashby perceptual category learning task. There were two versions to this task: Information-Integration (II) and Rule-Based (RB). Participants completed the II version and then the RB version. Prior research has shown that when participants are asked to switch between the declarative (hypothesis-testing) and implicit (associative) systems, they end up using rule-based strategies from the declarative system for all trials. Thus, by engaging the implicit system first, we aimed to reduce transfer effects between versions as much as possible.

In each version of the task, participants were told that they would be learning two categories and that perfect performance was possible. They were also told to be as quick and accurate as possible. In each trial, participants viewed a Gabor patch that belonged to one of the two categories. Each patch subtended 11° of visual angle. The stimuli were generated using category parameters from ?. The participant then had 5000ms to press a key, indicating which category they believed the stimulus belonged to. After a response, the participant received feedback ("Correct" or "Incorrect"). Feedback was presented for 1000ms, and then the next trial began. If the participant took more than 5000ms to respond, they saw "Too Slow" and proceeded to the next trial. Participants completed five runs of each version. Each run had 80 trials (40 from each category) presented in a random order. Thus, in total participants completed 400 II trials and 400 RB trials.

Taxonomic/thematic task. This task was adapted from ? and ?. There were also two versions of this task: one taxonomic and one thematic. Version order was counterbalanced across subjects, with some participants getting the taxonomic version first and others the thematic version first. Most versions of this type of task allow participants to choose the item that is most "semantically related," and thus do not ask participants to make either taxonomic or thematic choices on any given trial. As such, little research has looked at switching between taxonomic and thematic semantic judgments. Thus, counterbalancing was applied to control for order effects.

The stimuli were images taken from ?. We chose to use images in order to avoid automatic language processing. While participants likely did engage linguistic resources during the task, this should be due to how language relates to categorization rather than the features of the stimuli themselves. In each trial, four images were presented: a target, a taxonomically-related item, a thematically-related item, and an unrelated item. Taxonomically- and thematically-related items were chosen based on norms from ? where available. The ? norms were based on word stimuli rather than the images available from ?; as such, not all of the available images were normed. For images without norming information, we used our best judgment to pick items for each type of relation.

For each version, participants were told that they would be categorizing objects. They were told to pick the option that "goes best with" (thematic) or is "most similar to" (taxonomic) the target item. We chose these instructions based on previous research showing that slight differences in task instructions affect taxonomic and thematic judgments (?). After instructions, participants got five practice trials. In each trial, the images

were shown for 5000ms and participants had unlimited time to make a response. The practice trials were identical for the taxonomic and thematic versions of the task. After each response, participants received feedback ("Correct!" or "Oops!") for 1000ms. Once the practice trials were completed, participants received 24 test trials. While some images were seen in multiple trials, the 4-image combination for each trial was unique across the taxonomic and thematic versions of the task.

1.1.3 Executive Function Tasks

To measure executive function, we used three different tasks taken from the Psychology Experiment Building Language (PEBL) test battery (?). We chose three tasks to try and tap multiple aspects of executive function, including inhibition, planning, and task-switching. All three tasks were presented using the PEBL software.

Flanker task (inhibition). This task was an implementation of the ? flanker task, using a method similar to ?. In each trial, participants viewed a set of five arrows and were asked to respond based on the direction in which the center arrow was pointing (left or right). In congruent trials, all arrows faced the same way. In incongruent trials, the four distractor arrows pointed in the opposite direction of the target (center) arrow. In neutral trials, the four distractor arrows were just horizontal lines without arrowheads. Participants completed 20 trials for each condition in a 2 (direction; left vs. right) x 3 (condition; congruent vs. incongruent vs. neutral) design, for a total of 120 trials. **how many empty trials??** Each trial began with a 500 ms fixation, followed by the stimulus which appeared for 800ms. Participants were only allowed to respond during the 800ms that the stimulus was on the screen. After a response, there was an inter-trial interval of 1000ms. Participants received 12 practice trials before the actual experiment to get used to the timing of each trial. During practice trials, each response was followed by feedback ("Correct", "Incorrect") as well as a number indicating RT for that trial. This feedback was not provided for the test trials.

Switcher task (task-switching). This task was taken from ?. In this task, participants are presented with an array of colored shapes. Each colored shape has a single letter inside. For each trial, a single shape was indicated to be the target shape. Based on instructions at the top of the screen, participants were told to select a shape that matched the target shape on one of three dimensions (color, shape, or letter). Research from ? has shown that cueing a dimension using its entire name (e.g. "shape") does not require as many language resources as cueing a dimension using a single letter (e.g., "s"). Since one of the core hypotheses of this study was that language supports executive functions in the hypothesis-testing system, we used a version of the switcher task that cued dimension using just a single letter. We expect that this version of the task requires individuals to represent dimensions/selection rules internally, similar to how they might represent possible category rules when learning rule-based categories.

The task consisted of nine different arrays of ten shapes. For each array, participants made ten responses. In the first three arrays, participants switched between two of the three dimensions in a fixed order (e.g., C - S - C - S, etc.). The relevant dimensions were different for each array. For the second three arrays, participants switched between all three dimensions still in a fixed order (e.g., S - C - L - S - C - L, etc.). The specific order was different for each array. Finally, in the last three arrays participants switched between all three dimensions in a random order. Unlike previous arrays, in the last three participants were unable to anticipate the upcoming relevant dimension.

Tower of London task (planning). This task was a computerized version of the one described in ?. In this task, participants were shown a setup of colored disks in three stacks as well as a target setup. They were given a limited number of moves to make their setup match the target setup. Participants could only have one disk in their "hand" at a time, and they could only pick the top disk up off of any stack. The trials varied in the number of steps required to match the target setup from 2 to 5, with easier (2 step) trials at the beginning of the task and harder (5 step) trials at the end of the task. Participants were encouraged to take their time and plan out their moves before beginning each trial.

1.1.4 Behavioral Measures

Finally, we used four different behavioral assessments to measure vocabulary, syntax, and nonverbal IQ.. **Nelson-Denny vocabulary subtest.** To measure vocabulary, we used the same Nelson-Denny vocabulary subtest described in experiment 1.

Clinical Evaluation of Language Fundamentals recalling sentences and formulated sentences subtests. We used the CELF here to measure individuals differences in syntax production and perception. The recalling sentences subtest allowed us to look at receptive grammar, while the formulated sentences subtest provided a measure of expressive grammar. In the formulated subtest, participants view a scene and are asked to make a sentence containing a target word about that scene. Often, the target word encourages certain syntactic structures (e.g., "because").

Raven's Advanced Matrices. We used Raven's Advanced matrices to measure nonverbal IQ, as described in Experiment 1.

1.2 Procedure

1.3 Results

1.4 Discussion