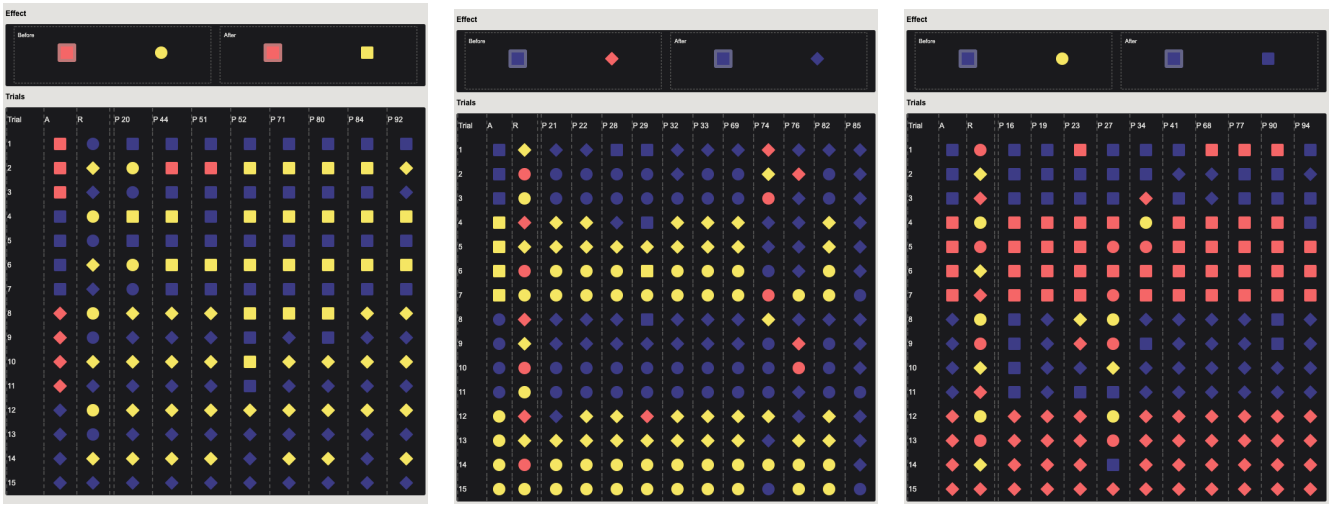
[Magic Stones] Report for Experiment 1

Bonan Zhao (b.zhao@ed.ac.uk)

January 10, 2020

1 Participant data



(a) To the same shape (b) To the same color (c) To the same object

(d) To a different shape (e) To a different color (f) To a different object

Figure 1: Each figure is for one learning condition. Each learning condition has 15 trials (15 rows).

2 Stats

- Age: min 24, max 67, mean 41.1639, sd 11.1238
- Gender: female 28 (45.9%), male 33 (54.1%)

Condition	Description	Count	Task dur. mean (min.)	S-Dfty. mean	Homogeneity
1	To the same shape	8	5.1873	3.38	74.69%
2	To a different shape	9	5.0942	3.45	57.10%
4	To a different color	12	4.1504	5.17	34.06%
5	To a different object*	11	5.2803	5.09	36.53%
6	To the same object	10	5.0351	3.00	65.50%
Total		61	4.8756 (sd 1.87)	4.65 (sd 3.16)	

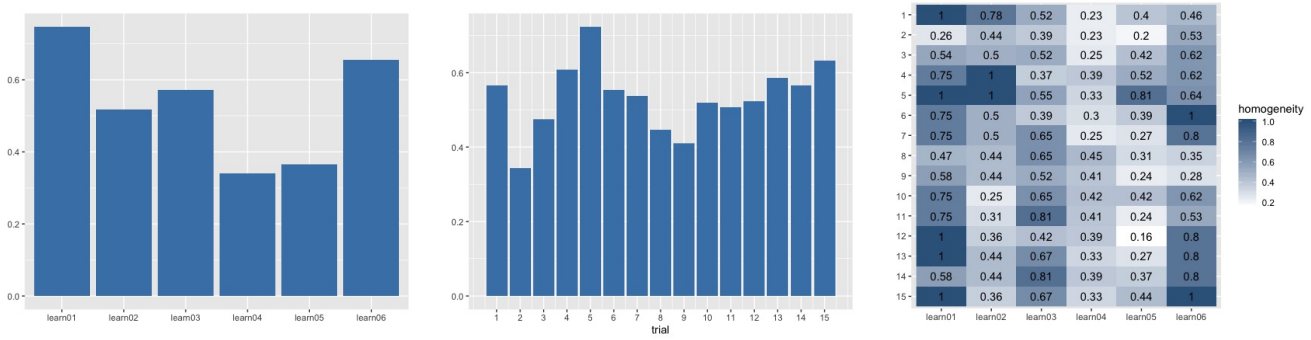
Table 1: Basic stats per condition. S-Dfty.: Self-report difficulty. *: Same color + diff shape.

Homogeneity

Homogeneity is defined as a scaled variance of selection frequency. Homogeneity H for a trial t_i for condition C is calculated by

$$var(\{\frac{n_{t_i}}{\sum_{i=1}^{15} n_{t_i}} | t_i \in C\})/var(M)$$

where $M = \{1, \underbrace{0, ..., 0,}_{14}\}$



(a) Aggregated by conditions. (b) Aggregated by tasks. (c) Overall breakdown

Figure 2: Homogeneity.

"To the same xx" V.S. "To a different xx"

Conditions 1, 3, and 6 can be classified as "to the same xx" group, where xx can be color, shape, or both (equivalent to the object), and conditions 2, 4, 5 can be classified as "to a different xx" group. Statistical test shows that compared to the "to a different xx" group, participant report the "to the same xx" group is significantly easier ($p < 0.0001$), and make more homogeneous predictions ($p = 0.0001$). However there is so significant difference for task duration between these two groups.

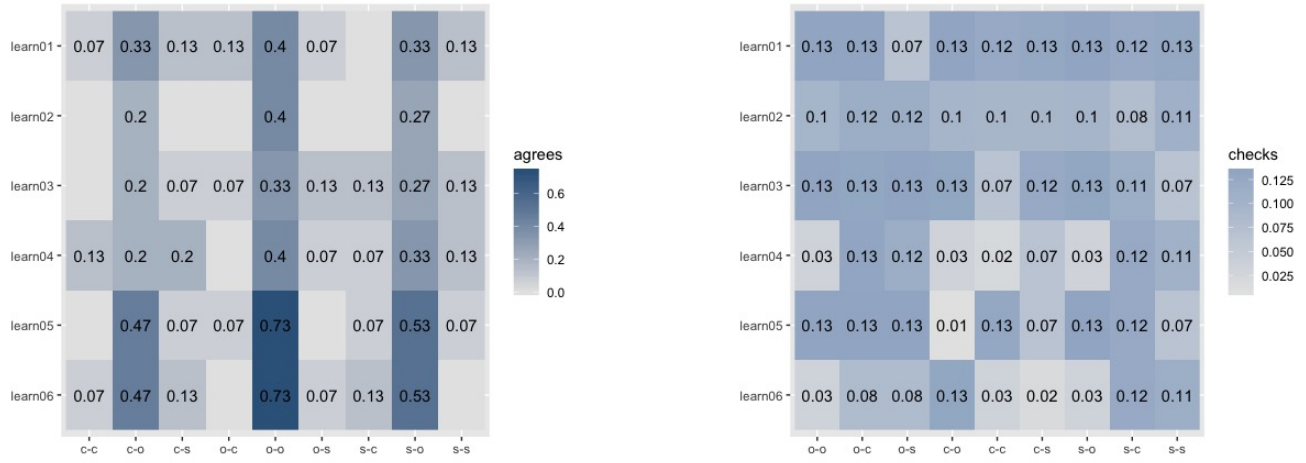
Arrow	$o \to o$	$o \to c$	$o \to s$	$c \to o$	$c \to c$	$c \to s$	$s \to o$	$s \to c$	$s \to s$
$\|f\|$	81	27	27	27	9	9	27	9	9
Total									225

Table 2: Number of causal power functions each arrow combination produces. o: $object$, c: $color$, s:$shape$.

3 Theories

Recall the nine theories in previous notes, as listed in Table 2. Figure 3 shows to what extent these nine theories predict the underlying rule or participant data separately.

Each cell is a summary of how well a theory predicts all the tasks for one condition, call it A (for *agreement*). For the "code setup" group, $A_C = N_C/15$ where N_C is the number of theory predictions that agree with condition C's underlying rule. When comparing with participant data, for each selection s in the theory predictions TP and participant selections PP, $A_C = (\sum_{i \in C} |P_{TH}(s_i) - P_{PP}(s_i)|)/15$.



(a) With code setup

(b) With participant selections

Figure 3: Theory predictions versus data.