

# Online Supplementary Materials

# **Supplementary Materials S1:**

## **Estimation Systems**

The materials in this supplementary chapter are relevant to Chapter 3 of the submitted thesis.

## .1 Estimation system workload capacity

### .1.1 Experiment 1: Fixed Item-Size, Mixed Item-Sets

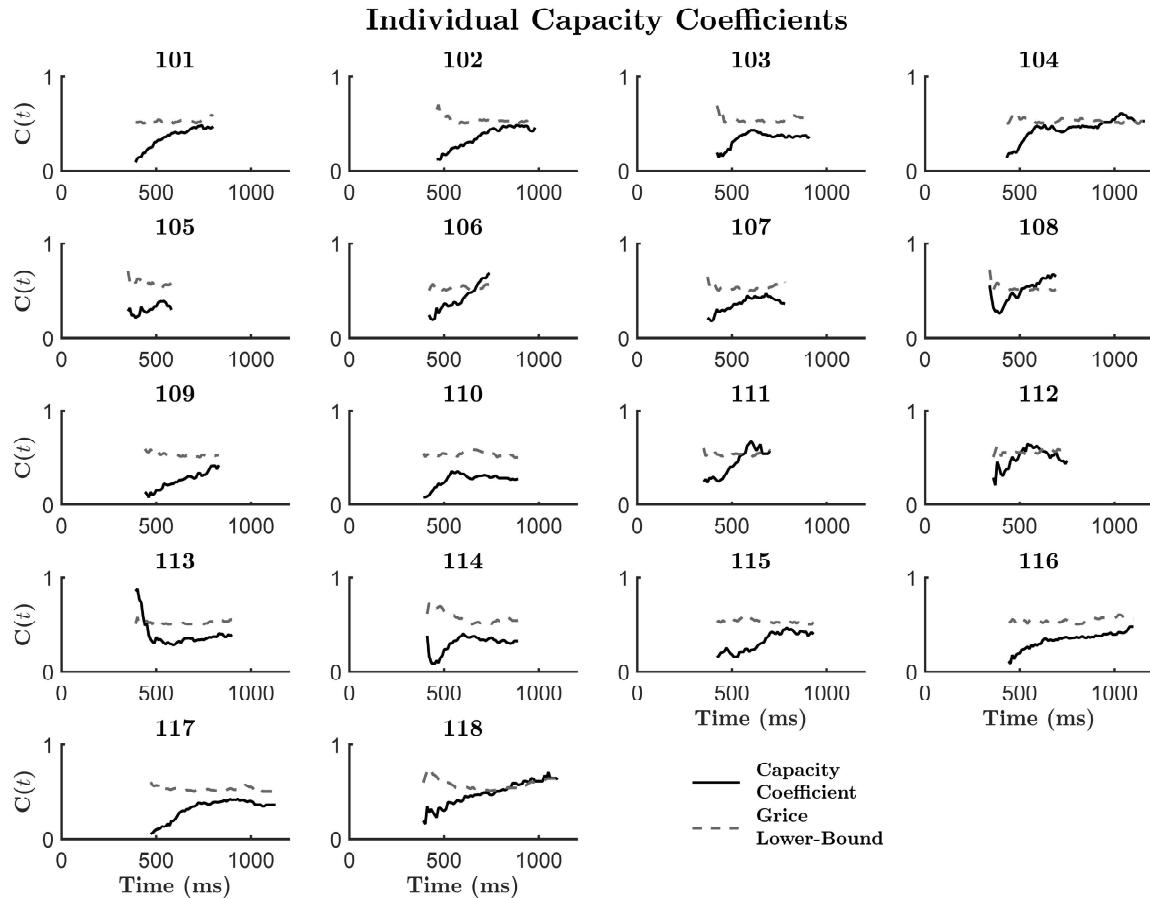


FIGURE S1.1.1: Capacity coefficient plots for each individual subject in Experiment 1. Dotted line depicts the Grice lower-bound. Black solid line illustrates the target capacity coefficient. All capacity coefficients are illustrated as on or below the lower-bound, suggesting severe workload capacity limitations in all subjects.

## .1.2 Experiment 2: Fixed Item-Size, Spatially Separate Item-Sets

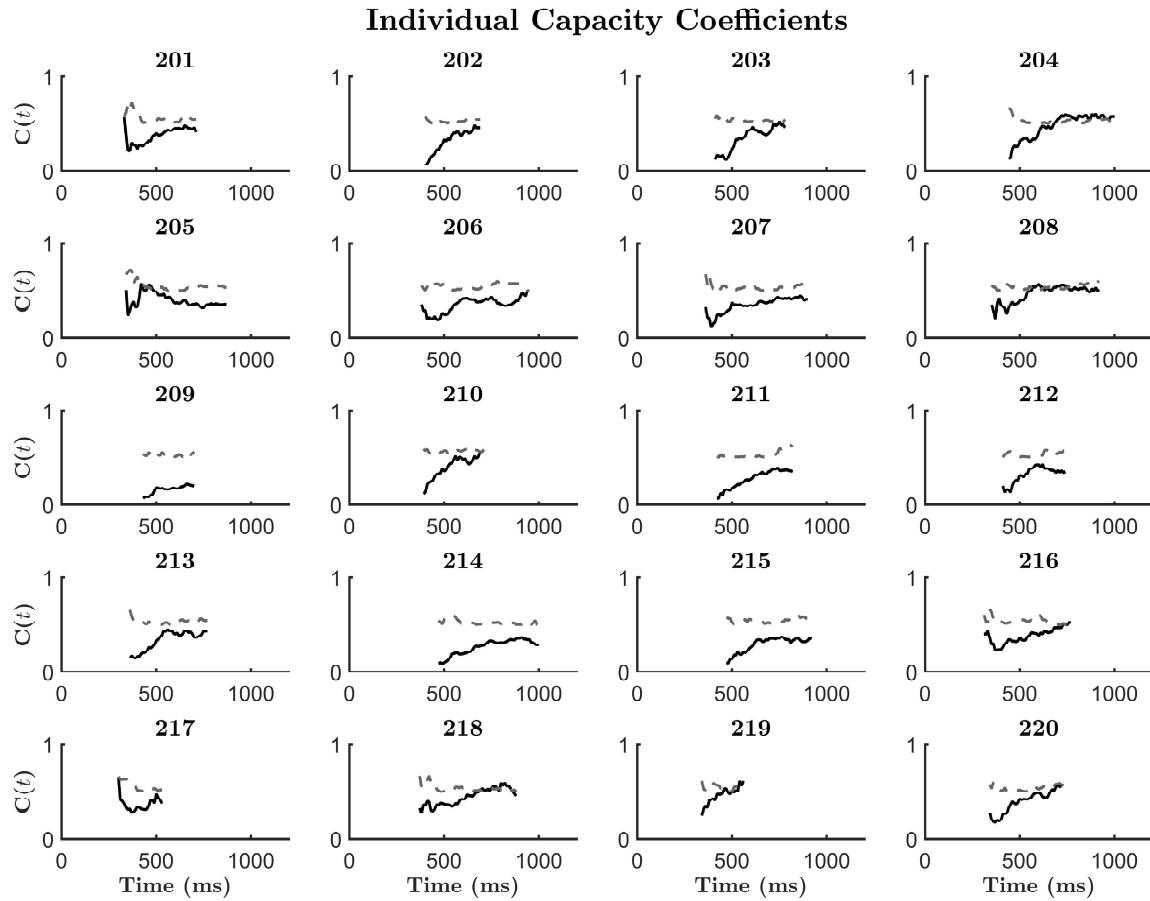


FIGURE S1.1.2: Capacity coefficient plots for each individual subject in Experiment 2. Dotted line depicts the Grice lower-bound. Black solid line illustrates the target capacity coefficient. All capacity coefficients are illustrated as on or below the lower-bound, suggesting severe workload capacity limitations in all subjects.

### .1.3 Experiment 3: Fixed Item-Set Area, Mixed Item-Sets

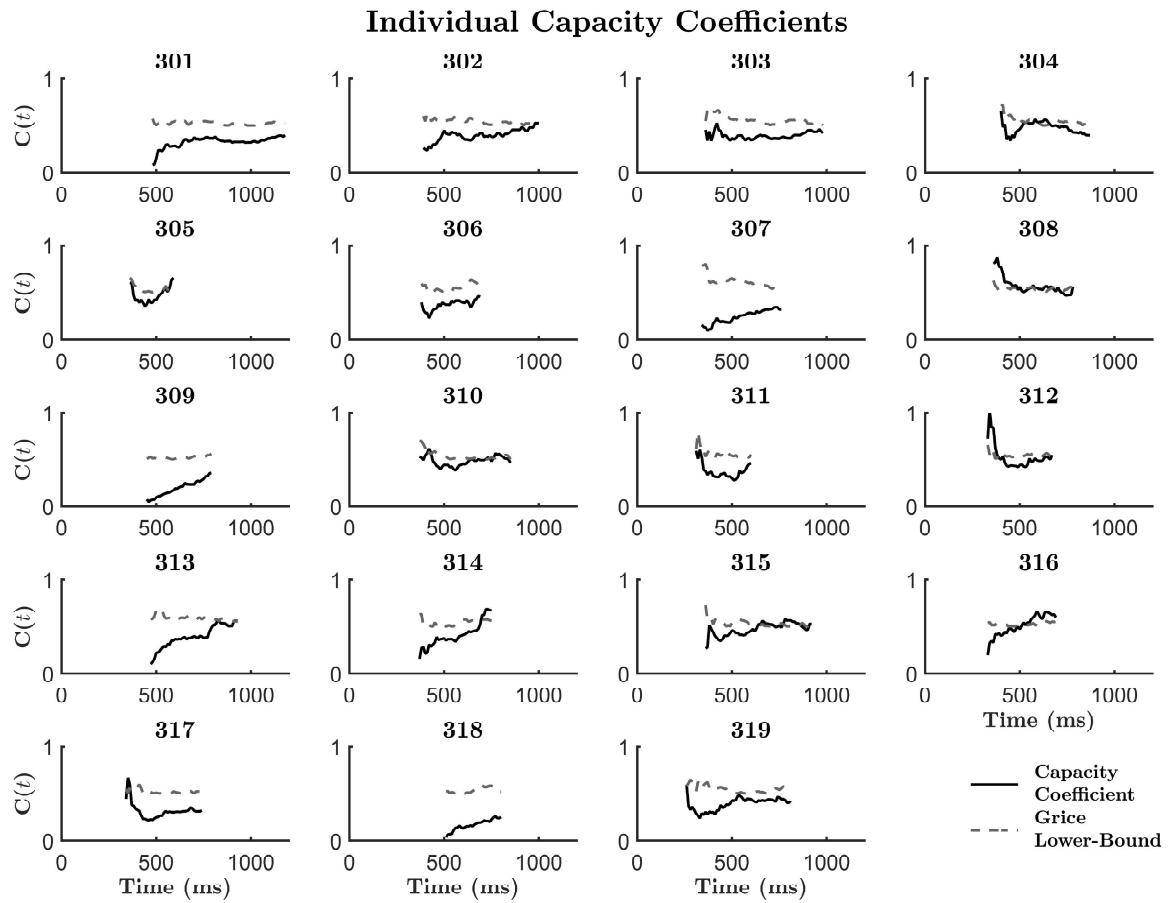


FIGURE S1.1.3: Capacity coefficient plots for each individual subject in Experiment 3. Dotted line depicts the Grice lower-bound. Black solid line illustrates the target capacity coefficient. All capacity coefficients are illustrated as on or below the lower-bound, suggesting severe workload capacity limitations in all subjects.

### .1.4 Experiment 4: Fixed Item-Set Area, Spatially Separated Item-Sets

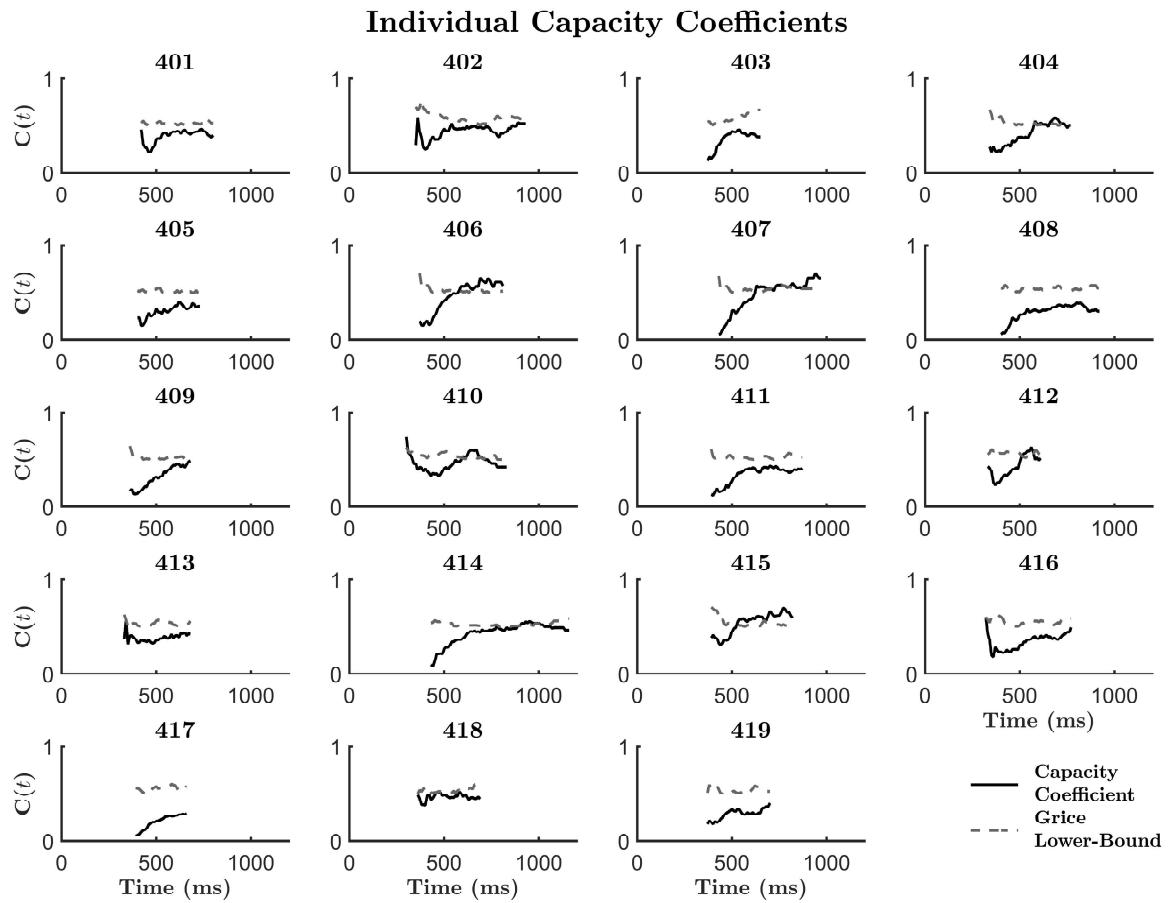


FIGURE S1.1.4: Capacity coefficient plots for each individual subject in Experiment 4. Dotted line depicts the Grice lower-bound. Black solid line illustrates the target capacity coefficient. All capacity coefficients are illustrated as on or below the lower-bound, suggesting severe workload capacity limitations in all subjects.

## .2 Resilience Difference

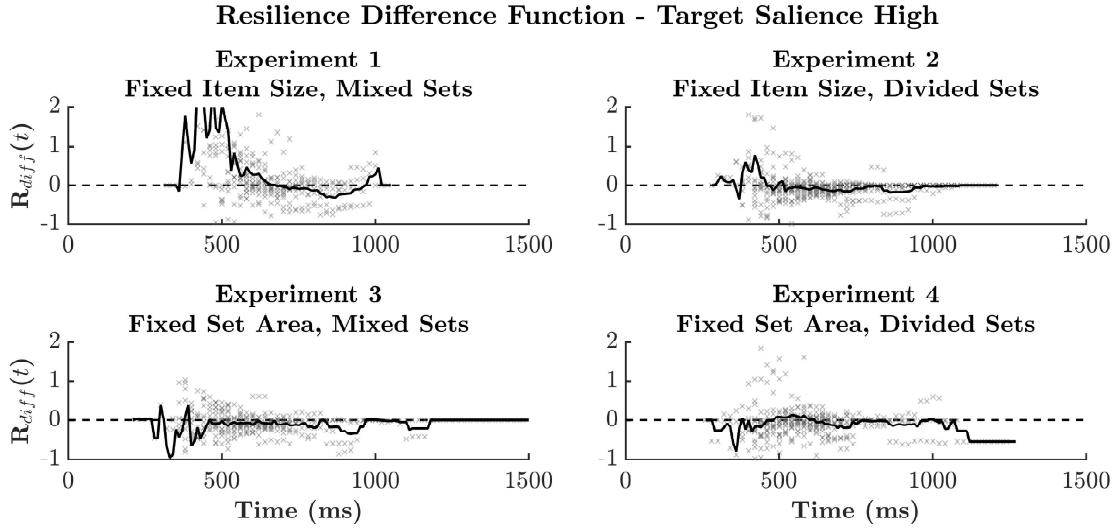


FIGURE S1.2.5: Resilience difference plots for Estimation Experiments 1 – 4 when target salience is high. Black lines depict the average resilience difference function across subjects, while grey markers illustrate individual subject's resilience difference functions. Average plots show a trend towards  $R_{diff}(t) = 0$ , supporting predictions made by an unlimited-capacity parallel minimum-time processing system.

### .3 Processing Architecture

#### .3.1 Experiments 2–4

TABLE S1.3.1: Table of individual double-target MIC and SIC results from Experiments 2–4. From left to right, ANOVA main effects for red target salience (R) and blue target salience (B), MIC value and interaction significance, Bayesian SIC model probability and associated architecture, and Bayes factor values from the hierarchical MIC analysis with associated processing architecture results: Parallel (P) and Serial (S).

Sub	Main Effects	MIC	SIC	P	SIC Model	$BF_{10}$	MIC Model
202	R & B	30*	<b>0.57</b>	P Min-Time	4.6	P Min-Time	
205	R & B	112*			<b>15.32</b>	P Min-Time	
212	R & B	8	0.44	P Min-Time	<b>3.18</b>	P Min-Time	
214	R & B	142**	<b>0.95</b>	P Min-Time	<b>18.75</b>	P Min-Time	
215	R & B	78*	<b>0.67</b>	P Min-Time	<b>10.42</b>	P Min-Time	
				<i>MIC Group Posterior</i>	<b>5.11</b>	P Min-Time	
303	R & B	128**	<b>0.95</b>	P Min-Time	<b>13.95</b>	P Min-Time	
307	R & B	-7	0.41	P Min-Time	2.27	P Min-Time	
310	R & B	26	<b>0.64</b>	P Min-Time	<b>3.35</b>	P Min-Time	
311	R & B	13			2.44	P Min-Time	
318	R & B	67*	<b>0.69</b>	P Min-Time	<b>6.9</b>	P Min-Time	
				<i>MIC Group Posterior</i>	<b>3.42</b>	P Min-Time	
402	R & B	-45*	<b>0.65</b>	P Max-Time	1.9	P Max-Time	
407	R & B	72**	<b>0.64</b>	P Min-Time	<b>8.73</b>	P Min-Time	
414	R & B	116**			<b>12.42</b>	P Min-Time	
416	R & B	-54*	<b>0.91</b>	P Max-Time	2.68	P Max-Time	
				<i>MIC Group Posterior</i>	2.68	P Min-Time	

MIC interaction significance tests were conducted at  $p < .33^*$ ,  $p < .05^{**}$  and  $p < .01^{***}$ . Conclusive model architectures were based upon an SIC model selection probability  $> 0.5$  and are displayed in bold font. MIC models with moderate evidence ( $BF_{10} > 3$ ) are also displayed in bold font.

### 3.2 Experiment 1: Fixed Item-Size, Mixed Item-Sets Double-Target SIC

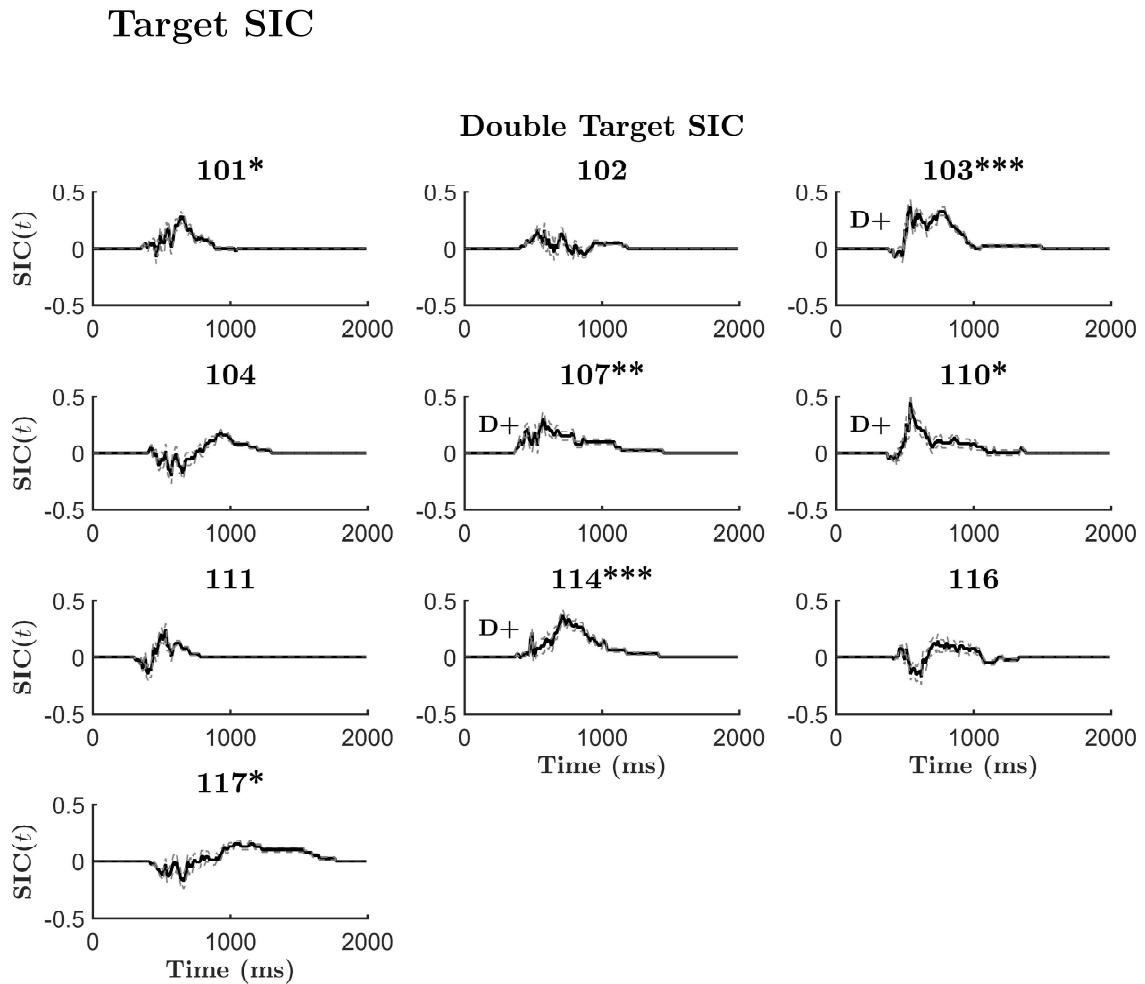


FIGURE S1.3.6: Double-Target survivor interaction contrast (SIC) plots for each individual subject in Experiment 1. Dotted line illustrates the bootstrap of the standard error.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$ . D-statistic, a measure of SIC deviance from  $SIC(t) = 0$  significance defined at  $\alpha < 0.15$ .

### .3.3 Experiment 2: Fixed Item-Size, Mixed Item-Sets Double-Target SIC

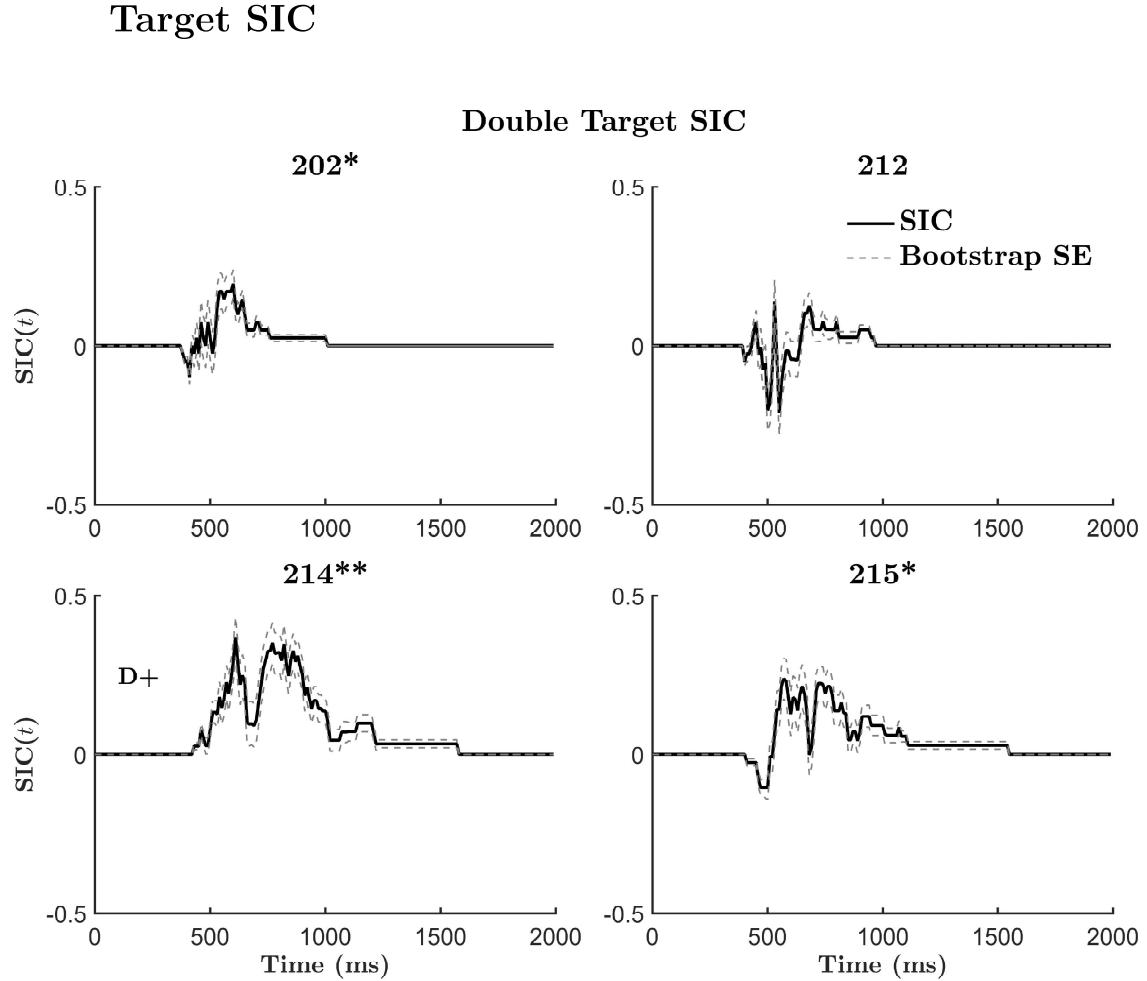


FIGURE S1.3.7: Double-Target survivor interaction contrast (SIC) plots for each individual subject in Experiment 2. Dotted line illustrates the bootstrap of the standard error.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$ . D-statistic, a measure of SIC deviance from  $SIC(t) = 0$  significance defined at  $\alpha < 0.15$ .

### .3.4 Experiment 3: Fixed Item-Size, Mixed Item-Sets Double-Target SIC

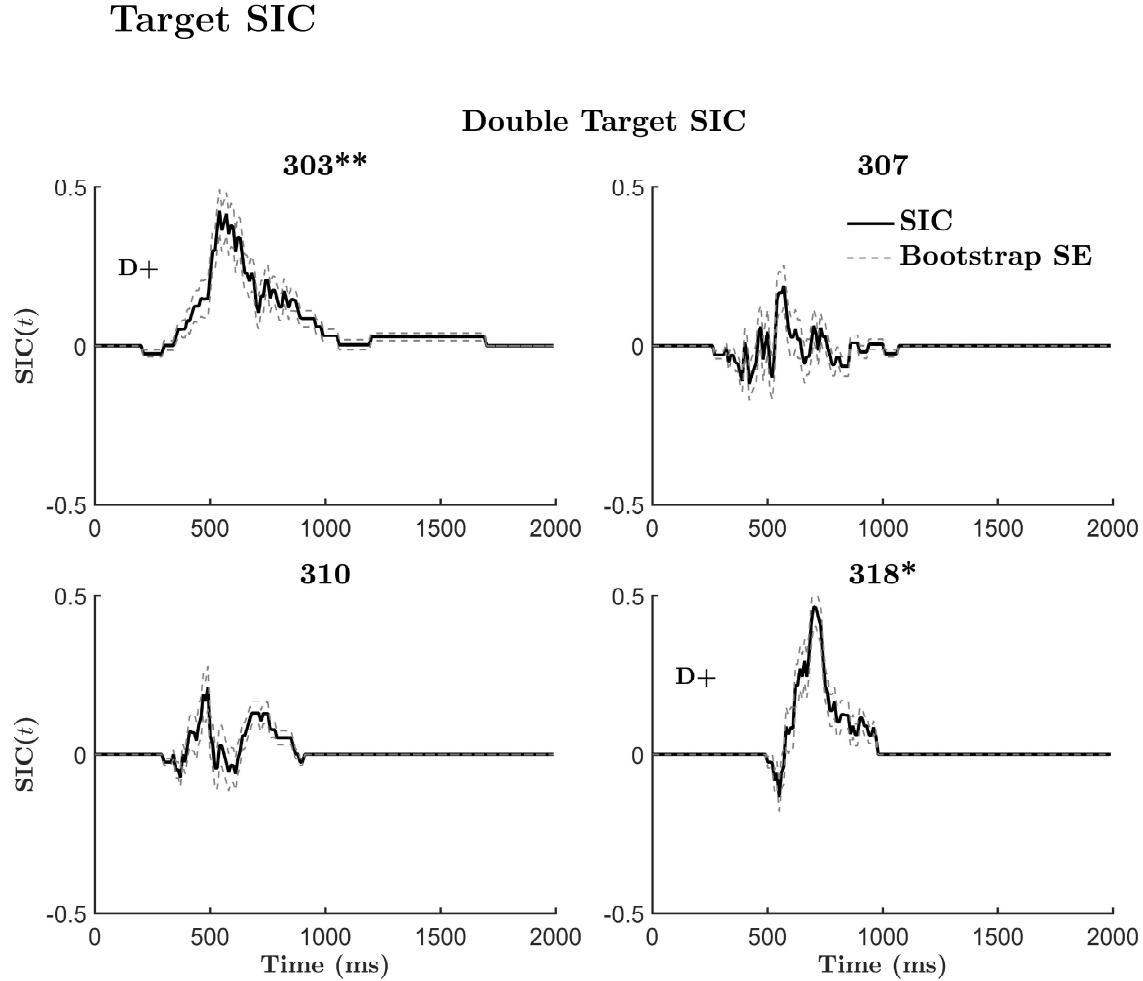


FIGURE S1.3.8: Double-Target survivor interaction contrast (SIC) plots for each individual subject in Experiment 3. Dotted line illustrates the bootstrap of the standard error.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$ . D-statistic, a measure of SIC deviance from  $SIC(t) = 0$  significance defined at  $\alpha < 0.15$ .

### .3.5 Experiment 4: Fixed Item-Size, Mixed Item-Sets Double-Target SIC

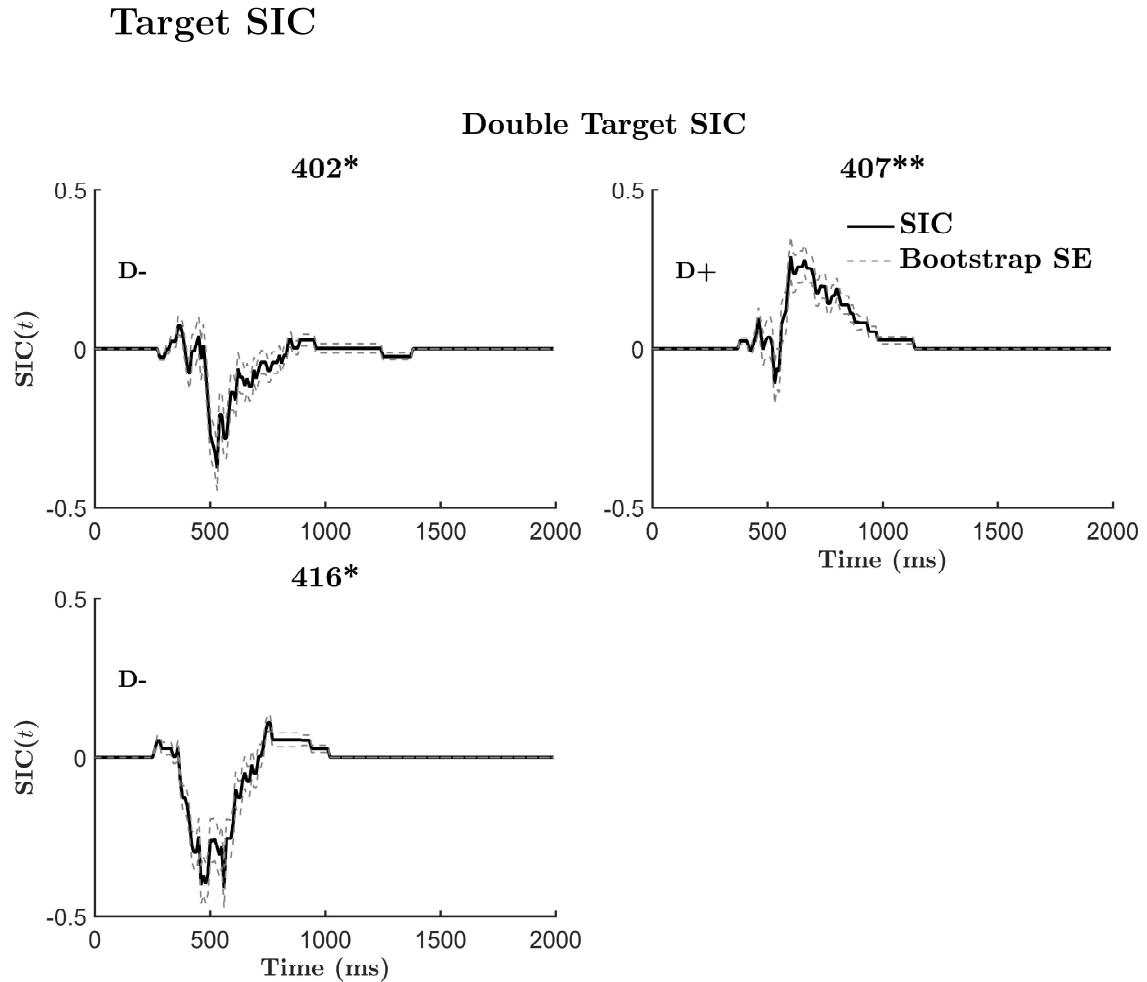


FIGURE S1.3.9: Double-Target survivor interaction contrast (SIC) plots for each individual subject in Experiment 4. Dotted line illustrates the bootstrap of the standard error.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$ . D-statistic, a measure of SIC deviance from  $SIC(t) = 0$  significance defined at  $\alpha < 0.15$ .

## .4 Individual MIC Plots

### .4.1 Experiment 1: Fixed Item-Size, Mixed Item-Sets Double-Target MIC Plots

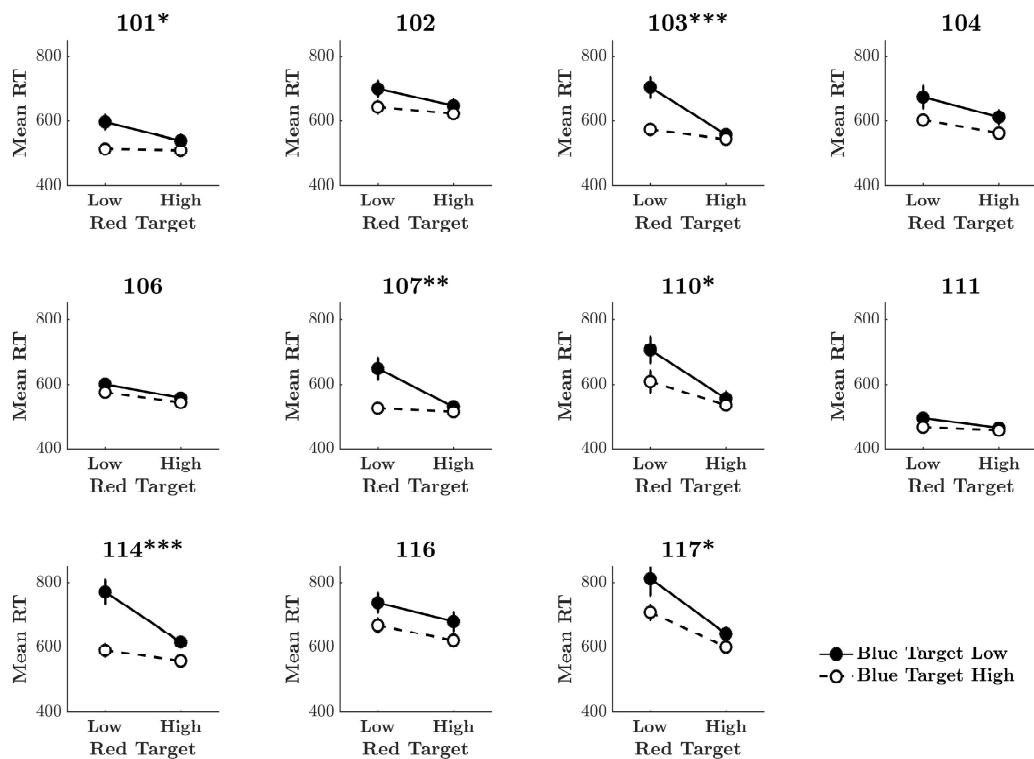


FIGURE S1.4.10: Double-Target mean interaction contrast (MIC) plots for each individual subject in Experiment 1. White-eggs indicate blue-target high-salience, black-eggs blue-target low-salience. Error bars are  $\pm 1$  standard error of the mean.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$

## .4.2 Experiment 2: Fixed Item-Size, Split Item-Sets Double-Target MIC Plots

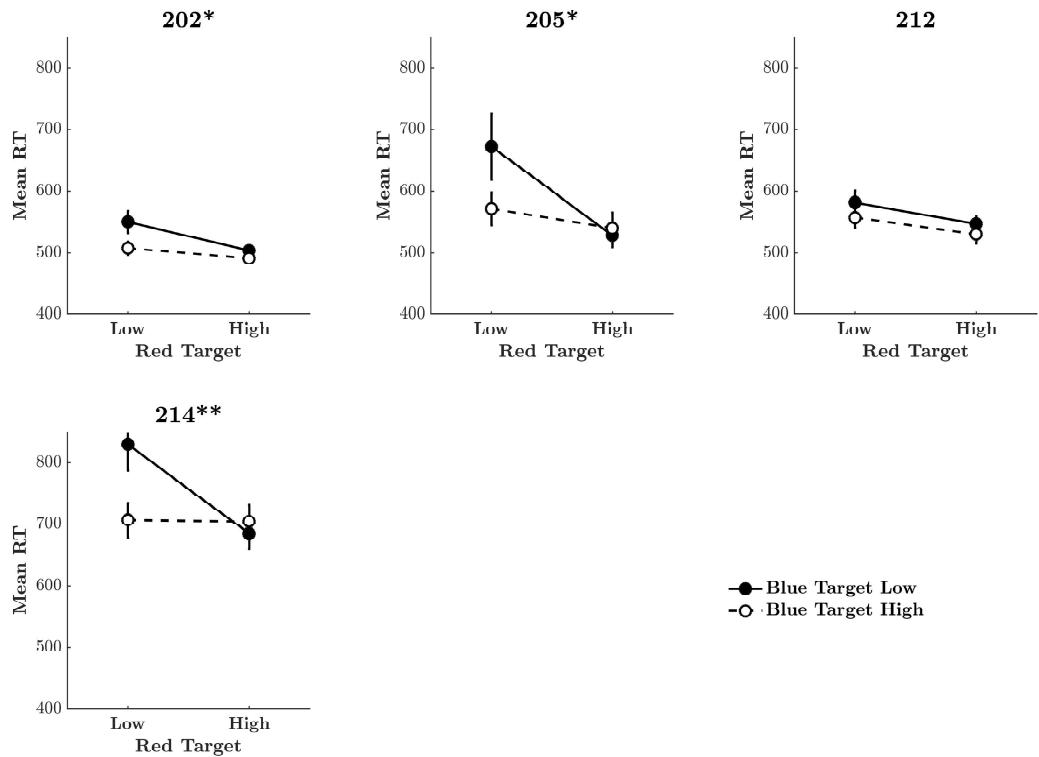


FIGURE S1.4.11: Double-Target mean interaction contrast (MIC) plots for each individual subject in Experiment 2. White-eggs indicate blue-target high-salience, black-eggs blue-target low-salience. Error bars are  $\pm 1$  standard error of the mean.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$

### .4.3 Experiment 3: Fixed Area, Mixed Item-Sets Double-Target

#### MIC Plots

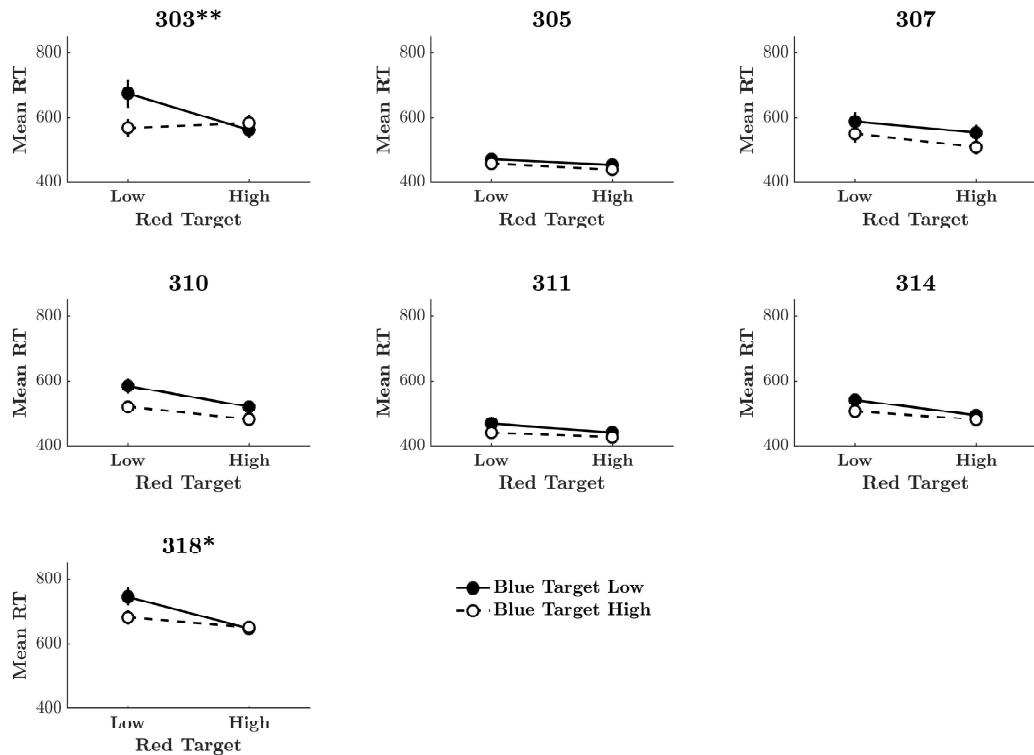


FIGURE S1.4.12: Double-Target mean interaction contrast (MIC) plots for each individual subject in Experiment 3. White-eggs indicate blue-target high-salience, black-eggs blue-target low-salience. Error bars are  $\pm 1$  standard error of the mean.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$

#### .4.4 Experiment 4: Fixed Area, Split Item-Sets Double-Target

##### MIC Plots

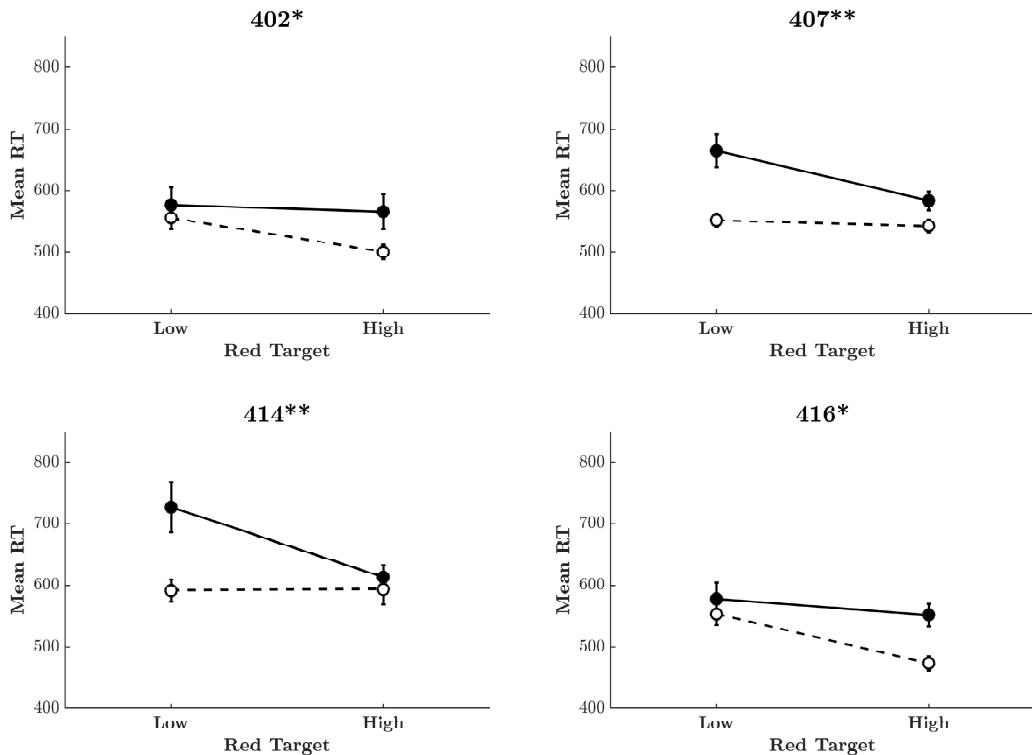


FIGURE S1.4.13: Double-Target mean interaction contrast (MIC) plots for each individual subject in Experiment 4. White-eggs indicate blue-target high-salience, black-eggs blue-target low-salience. Error bars are  $\pm 1$  standard error of the mean.

**Note.** MIC interaction contrast significance defined at  $p^* < .33$ ,  $p^{**} < .05$ ,  $p^{***} < .01$

# **Supplementary Materials S2:**

## **Subitizing Systems**

The materials in this supplementary chapter are relevant to Chapter 4 of the submitted thesis.

## .1 Target mean RT

TABLE S2.1.1: Paired *t*-test results (bonferroni corrected) for mean RT comparisons between double-target (Dt), red single-target (Rt) and blue single-target responses (Bt). Experiment 1 utilized a criterion of three, experiment 2 utilized a criterion of 4.

Exp	Description	Dt v Rt		Dt v Bt		Rt v Bt		
		<i>t</i> -test	(df)	<i>p</i>	<i>t</i> -test	<i>p</i>	<i>t</i> -test	<i>p</i>
1.1	Fixed Size; Mixed	5.084	(7)	< 0.01	3.346	< 0.05	-1.353	0.65
1.2	Fixed Size; Separate	2.889	(5)	0.1	5.877	< 0.01	-0.157	1.0
1.3	Fixed Area; Mixed	5.981	(7)	< 0.01	4.079	< 0.05	-1.724	0.39
1.4	Fixed Area; Separate	4.373	(7)	0.07	1.593	0.63	-1.341	0.82
2.1	Fixed Size; Mixed	3.433	(4)	0.08	3.528	0.08	-1.140	0.94
2.2	Fixed Size; Separate	12.433	(7)	< 0.001	10.806	< 0.001	-2.249	0.18
2.3	Fixed Area; Mixed	12.143	(3)	< 0.01	6.426	< 0.05	-2.311	0.31
2.4	Fixed Area; Separate	10.661	(7)	< 0.001	5.33	< 0.01	-2.217	0.19

## .2 Workload capacity

The following displays capacity coefficient workload functions for every individual reported in the subitizing task of Chapter 4. Each subplot displays a single participant, the black line denotes the capacity coefficient and the broken line denotes the Grice lower-bound. For every individual, capacity coefficients hover at or below the Grice bound for most of the response-time range indicating severely limited workload capacity. This result replicated across experiments, presented as separate figures.

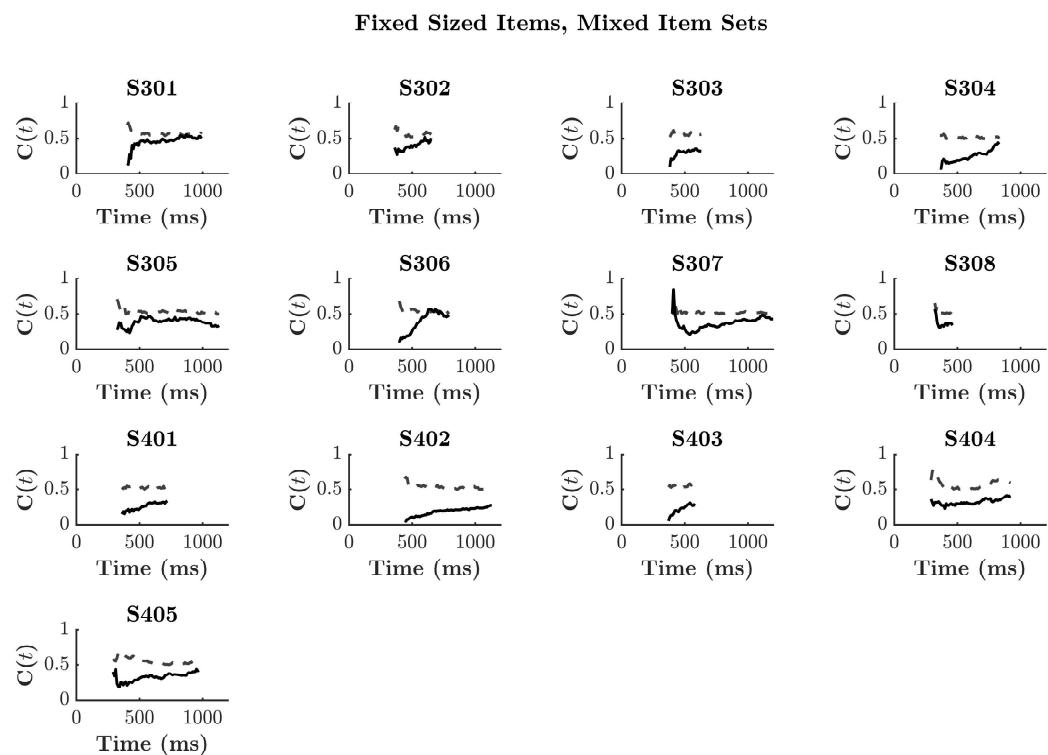


FIGURE S2.2.1: Capacity coefficient ( $C(t)$ ) plots for target-present responses when the item-sets were mixed and item-size was fixed. Subject IDs starting numbers denote the response criterion.  $C(t)$  is plotted in black, the Grice lower-bound is depicted by the dashed line.

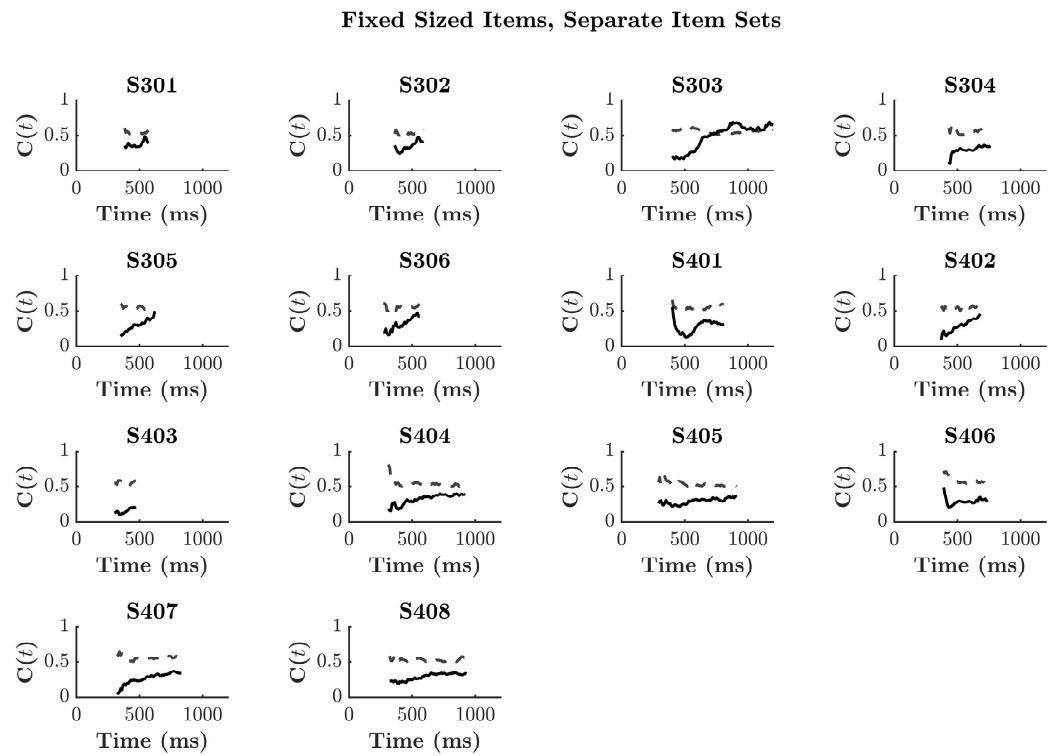


FIGURE S2.2.2: Capacity coefficient ( $C(t)$ ) plots for target-present responses when the item-sets were separated and item-size was fixed. Subject IDs starting numbers denote the response criterion.  $C(t)$  is plotted in black, the Grice lower-bound is depicted by the dashed line.

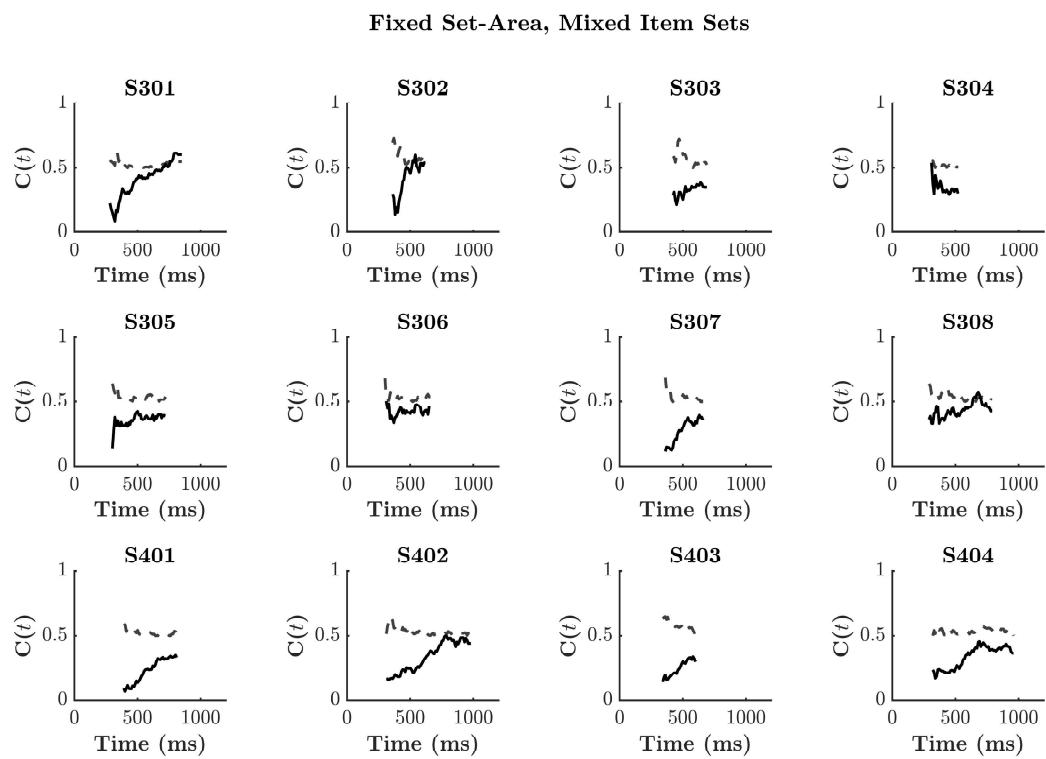


FIGURE S2.2.3: Capacity coefficient ( $C(t)$ ) plots for target-present responses when the item-sets were mixed and item-set area was fixed. Subject IDs starting numbers denote the response criterion.  $C(t)$  is plotted in black, the Grice lower-bound is depicted by the dashed line.

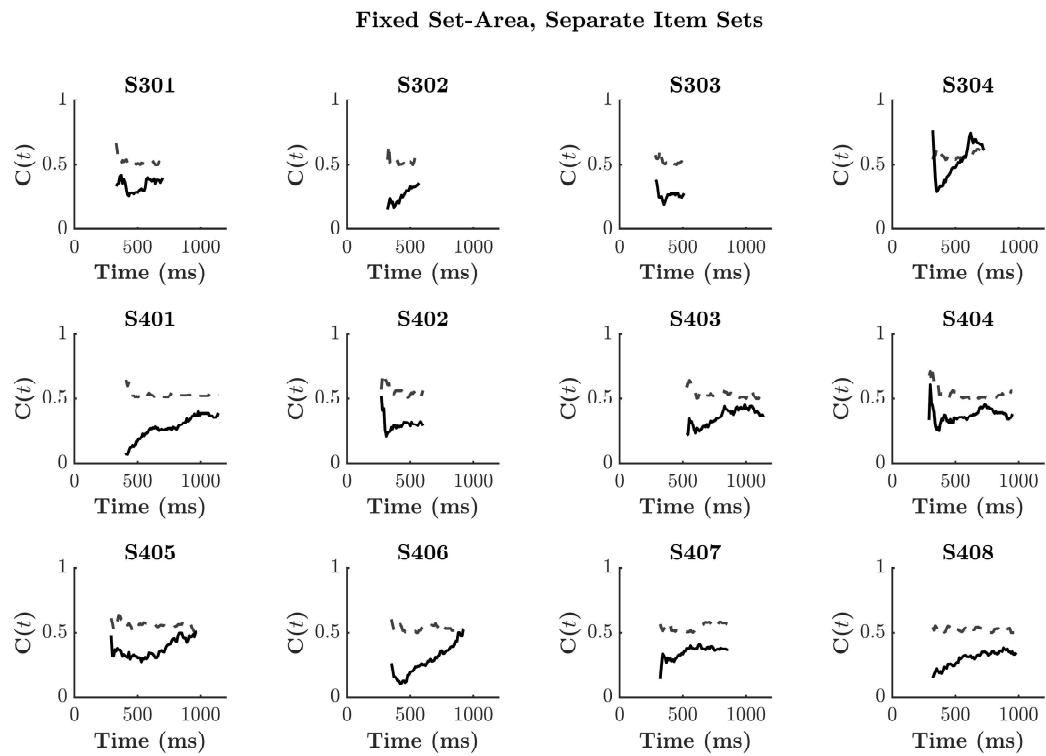


FIGURE S2.2.4: Capacity coefficient ( $C(t)$ ) plots for target-present responses when the item-sets were separated and item-set area was fixed. Subject IDs starting numbers denote the response criterion.  $C(t)$  is plotted in black, the Grice lower-bound is depicted by the dashed line.

### .3 Individual MIC plots

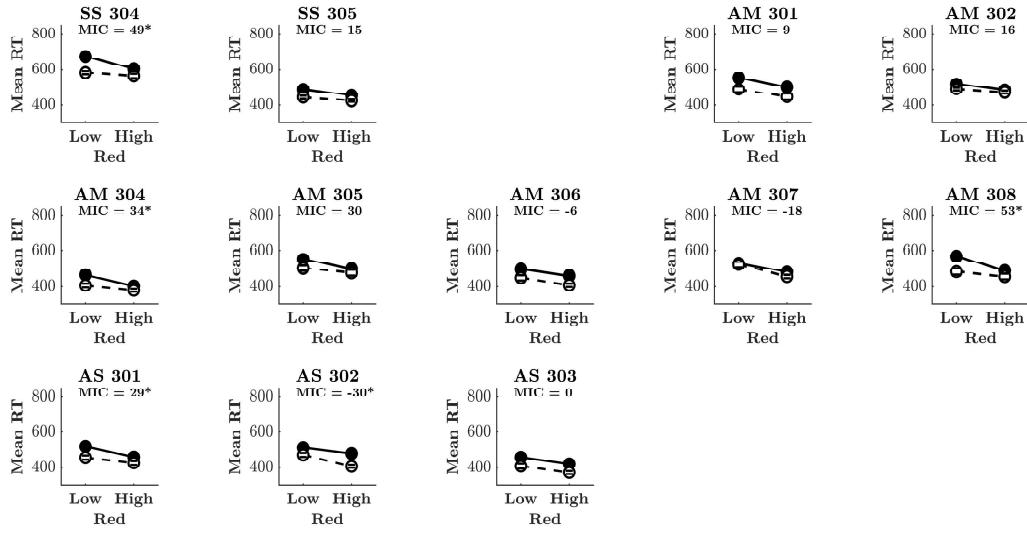


FIGURE S2.3.5: Individual MIC plots for participants across experiments when the criterion was three. Experiments displayed are when item-size was fixed and the item-sets were separate (SS), when the item-set area was fixed and the item-sets were mixed (AM), and when the item-set area was fixed and the item-sets were separate (AS).

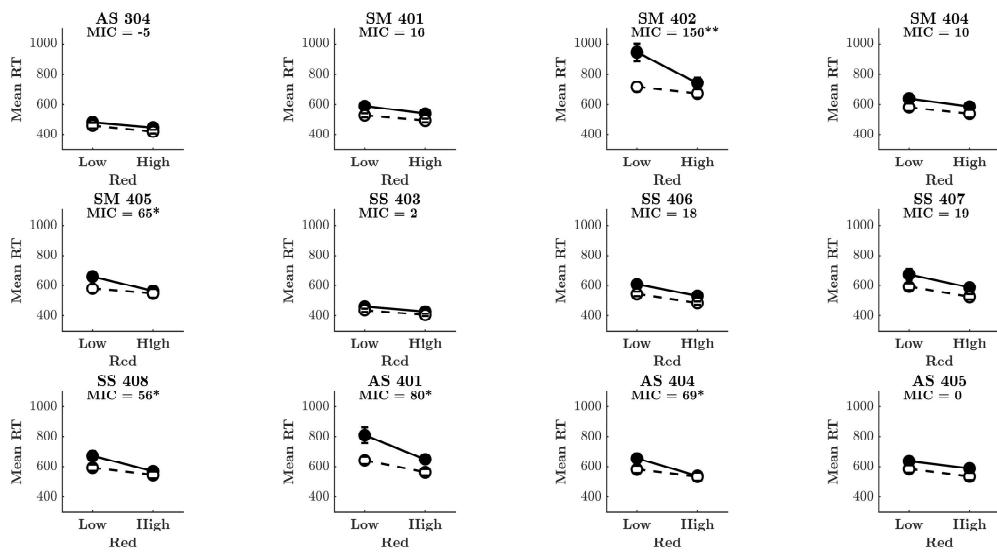


FIGURE S2.3.6: Individual MIC plots for participants across experiments when the criterion was three (subject ID 300) or four (subject ID 400). Experiments displayed are when item-size was fixed and the item-sets were mixed (SM), when item-size was fixed and the item-sets were separate (SS) and when the item-set area was fixed and the item-sets were separate (AS).

## **Supplementary Materials S3:**

### **Mixture processing models**

The materials in this supplementary chapter are relevant to Chapter 5 of the submitted thesis.

The following presents the analytic proofs used in the simulation of Poisson processing models, mixture models, mixture SICs and mixture capacity coefficients. The proofs herein will be described in terms of their application to the Poisson accumulator. Where appropriate we subscript the numerals 1 and 2 to indicate to which of the two channel components of the models we are referring. We superscript the model predictions with  $s$ ,  $p$ , or  $c$  to indicate reference to the serial, parallel, or coactive models, respectively. An additional superscript  $st$  or  $ex$  may be added to indicate self-termination or exhaustive processing, respectively.

## .1 Model simulation - Poisson accumulator

We simulated predictions for each of the independent channels of the serial and parallel models and the coactive model using pairs of Poisson accumulators. The number of accumulated counts is distributed as a Poisson random variable with a rate parameter  $\lambda$ :

$$P(U(t) = u|\lambda t) = \begin{cases} \frac{(\lambda t)^u e^{-\lambda t}}{u!}, & t \geq 0 \\ 0 & t < 0 \end{cases} \quad (1)$$

The processing time of a channel is the time it takes to reach a threshold number of counts,  $a$ . Thus, the processing time for an independent channel has an Erlang distribution,

$$f(t) = \frac{\lambda^a t^{a-1} \exp(-\lambda t)}{(a-1)!} \quad (1)$$

$$F(t) = 1 - \sum_{k=0}^{a-1} \frac{\exp(-\lambda t)(\lambda t)^k}{k!}. \quad (2)$$

### .1.1 Parallel self-terminating model

The survivor function predictions of the parallel self-terminating model were computed from Poisson distributions of each independent stimulus channels as:

$$S_{12}^{Pst}(t) = S_1(t) \times S_2(t). \quad (4)$$

In the case where only one channel carries the target information and allows for self-termination, under the assumption of stochastic independence, the parallel self-terminating processing time is determine solely by the target channel.

### .1.2 Serial self-terminating model

Serial self-terminating predictions were computed as:

$$f_{12}^{Sst}(t) = P(1, 2)f_1(t) + [1 - P(1, 2)]f_2(t) \quad (5a)$$

where  $P(1, 2)$  is the probability that channel 1 is processed before channel 2.

We are also interested in cases in which one dimension allowed for self-termination but the other dimension did not (in which case, both dimensions needed to be processed).

In this case, the density function is:

$$f_{12}^{Sst}(t) = P(1, 2)f_1(t) + [1 - P(1, 2)](f_2(t) * f_1(t)) \quad (5b)$$

Here,  $*$  refers to the convolution operation. This captures the assumption that channel 1 still needs to be processed when channel 2 is processed first.<sup>1</sup>.

### .1.3 Parallel exhaustive model

The cdf predictions for the parallel exhaustive model were computed from the Poisson distributions as:

$$F_{12}^{Pex}(t) = F_1(t) \times F_2(t). \quad (6)$$

### .1.4 Serial exhaustive model

The predictions for the serial exhaustive model were computed as:

$$f_{12}^{Sex}(t) = f_1(t) * f_2(t). \quad (7)$$

---

<sup>1</sup>The type of serial model adopted for the simulations below depended on the task context to which the measure applies. In an OR task, where redundant processing is possible, then the two formulations of the serial self-terminating model are identical in their predictions. Hence, for the  $SIC_{OR}(t)$  and  $C_{OR}(t)$  simulations, we utilized Equation 9a. For the remaining simulations, we use the exhaustive formulations.

### .1.5 Coactive model

To simulate a coactive processing model, we assumed that there was complete cross-talk between the processing channels. Following Johnson et al. (2010), we defined a model over the  $a_1 \times a_2$  state space, where  $a_i$  is the criterion for channel  $i$  (i.e., the number of accumulated counts needed before the channel can terminate). The probability that a single count was shared from channel  $j$  to channel  $i$  was defined as  $p_{ji}$ . The total amount of shared information was distributed as a binomial random variable. We utilized the analytic expression shown in Johnson et al. (their Appendix A.1) for the facilitatory parallel model with  $p_{ji} = p_{ij} = 1$  for the coactive processing model. This model is equivalent to a single Poisson accumulator channel with  $\lambda_{12} = \lambda_1 + \lambda_2$  and  $a_{12} = \max(a_1, a_2)$  or  $\min(a_1, a_2)$  depending on the stopping rule.

### .1.6 Mixture models

Along with the coactive model, equations 4-7 provide a formal description of the pure, non-mixture models. The mixture models were computed as additive mixtures of each of these components:

$$\begin{aligned}
 f_{Pst/C}^{MixOR}(t) &= pf^{Pst}(t) + [1 - p]f^C(t) \\
 f_{Pst/Sst}^{MixOR}(t) &= pf^{Pst}(t) + [1 - p]f^{Sst}(t) \\
 f_{Pex/C}^{MixAND}(t) &= pf^{Pex}(t) + [1 - p]f^C(t) \\
 f_{Pex/Sex}^{MixAND}(t) &= pf^{Pex}(t) + [1 - p]f^{Sex}(t)
 \end{aligned} \tag{8}$$

where  $p$  is the probability that the mixture model uses a parallel process, and  $1 - p$  is the probability that the mixture uses another process, coactive or serial (Equation 8). <sup>2</sup>

### .1.7 Capacity predictions

The capacity functions for mixture models follow a relatively complicated trade-off as a function of  $p$ . To determine an analytical prediction, we first note that the mixture of densities implies the mixture of CDFs,

$$F^{mix} = \int f^{mix} = \int pf^{(1)} + (1-p)f^{(2)} = p \int f^{(1)} + (1-p) \int f^{(2)} = pF^{(1)} + (1-p)F^{(2)}. \quad (9)$$

Likewise, because  $S = 1 - F$ ,

$$\begin{aligned} S^{mix} &= 1 - F^{mix} \\ &= 1 - (pF^{(1)} + (1-p)F^{(2)}) \\ &= 1 - pF^{(1)} - F^{(2)} + pF^{(2)} \\ &= p - p + 1 - pF^{(1)} - F^{(2)} + pF^{(2)} \\ &= p(1 - F^{(1)}) - p + 1 - F^{(2)} + pF^{(2)} \\ &= p(1 - F^{(1)}) + (1 - p)(1 - F^{(2)}) \\ &= pS^{(1)} + (1 - p)S^{(2)}. \end{aligned}$$

---

<sup>2</sup>We omitted the combination of serial and coactive because the results indicated that a mixture of serial and coactive processing would exhibit the same smooth mixture between a pure serial and pure coactive model as the other mixture models.

By substituting the mixture survivor functions and CDFs into Equations 5.1 and 5.2 respectively, we arrive at:

$$\begin{aligned} C_{OR}(t)^{mix} &= \frac{\log \left[ pS_{12}^{(1)} + (1-p)S_{12}^{(2)} \right]}{\log \left[ pS_1^{(1)} + (1-p)S_1^{(2)} \right] + \log \left[ pS_2^{(1)} + (1-p)S_2^{(2)} \right]} \\ C_{AND}(t)^{mix} &= \frac{\log \left[ pF_1^{(1)} + (1-p)F_1^{(2)} \right] + \log \left[ pF_2^{(1)} + (1-p)F_2^{(2)} \right]}{\log \left[ pF_{12}^{(1)} + (1-p)F_{12}^{(2)} \right]} \end{aligned} \quad (10)$$

The capacity predictions for two kinds of mixture models, parallel processing coupled with either serial or coactive processes, for both the OR and AND case, are shown in Figure 5.3. For the coactive/parallel mixture model, as  $p$  increases from 0 to 1, the predictions clearly reflect a smooth transition from the unlimited capacity parallel model predictions to the supercapacity coactive model predictions. Conversely, for the parallel/serial mixture model, the capacity predictions move smoothly from unlimited to limited capacity.

### .1.8 SIC predictions

The SIC predictions for a mixture model are relatively straightforward to determine analytically. Substituting Equation 9 into the SIC, we find the SIC for a mixture is a mixture of the SICs.

$$\begin{aligned}
SIC(t)^{\text{mix}} &= \left[ \left[ pS_{LL}^{(1)}(t) + (1-p)S_{LL}^{(2)}(t) \right] - \left[ pS_{LH}^{(1)}(t) + (1-p)S_{LH}^{(2)}(t) \right] \right] \quad (11) \\
&\quad - \left[ \left[ pS_{HL}^{(1)}(t) + (1-p)S_{HL}^{(2)}(t) \right] - \left[ pS_{HH}^{(1)}(t) + (1-p)S_{HH}^{(2)}(t) \right] \right] \\
&= p \left[ \left[ S_{LL}^{(1)}(t) - S_{LH}^{(1)}(t) \right] - \left[ S_{HL}^{(1)}(t) - S_{HH}^{(1)}(t) \right] \right] \\
&\quad + (1-p) \left[ \left[ S_{LL}^{(2)}(t) - S_{LH}^{(2)}(t) \right] - \left[ S_{HL}^{(2)}(t)S_{HH}^{(2)}(t) \right] \right] \\
&= pSIC(t)^{(1)} + (1-p)SIC(t)^{(2)}.
\end{aligned}$$

## **Supplementary Materials S4:**

### **The cost of errors**

The materials in this supplementary chapter are relevant to Chapter 6 of the submitted thesis.

## .1 Luce's choice model

Luce's (1963) choice model describes identification responses as probabilistic outcomes driven by the similarity of a stimulus to the others in the choice set, as well as a response-bias parameter — one for each stimulus. By estimating the parameters of the model, researchers can examine the theoretically meaningful similarity scores free from the effect of response-bias that can contaminate the observed data. Formally, the probability of making response  $j$  when presented with stimulus  $i$  can be expressed as:

$$C_{ij} = \frac{\eta_{ij}\beta_j}{\sum_{k=1}^N \eta_{ik}\beta_k} \quad (1)$$

where  $C_{ij}$  is the theoretical similarity matrix for  $i = 1, 2\dots N$ ,  $j = 1, 2\dots N$ . The similarity parameter  $\eta$  is symmetrical along the matrix diagonal i.e.,  $\eta_{ij} = \eta_{ji}$ , and  $\eta_{ii} = 1$  for all  $i$ . In the current study, we will employ nine unique numerals, resulting in  $[N(N + 1)/2] - 1 = 44$  free parameters to be estimated from the data.

We estimated the bias and similarity parameters of Luce's (1963) choice model using the combination of a custom Differential-Evolution Markov chain Monte-Carlo (DE-MCMC) process and maximum likelihood estimation (Myung, 2003). We initialised each of the 50 chains by estimating parameter values from Townsend's (1971) approximation of Luce's model:

$$\eta_{ij} = \sqrt{\frac{P(R_i|S_j)P(R_j|S_i)}{P(R_i|S_i)P(R_j|S_j)}} \quad (2)$$

$$\beta_j = \frac{1}{N} \sum_{k=1}^N \sqrt{\frac{P(R_j|S_j)P(R_k|S_j)}{P(R_j|S_k)P(R_k|S_k)}} \quad (3)$$

where  $R$  is the response probability given stimulus  $S$ ; then adding uniformly sampled noise.

On each iteration, each chain proposed updated parameter estimates by weighting the previous estimates with the estimates of two randomly selected chains using the weighting formula outlined by Turner, Sederberg, Brown, and Steyvers (2013). The log-likelihood of these new parameters and the previous ones were computed by generating an expected confusion matrix (using the estimated parameters and Luce's choice model) and comparing to the observed data, with the parameters that maximised the log-likelihood being kept.

After 500 iterations the parameters from the chain with the highest log-likelihood were used for further analysis.

## .2 Experimental accuracy by contrast level and participant

The following section examines accuracy during experimental trials, over the five levels of contrast. We show that our manipulation of contrast appropriately influenced accuracy, that accuracy was relatively stable across blocks, and that accuracy was close to 60% for all participants, across conditions of numeric-type (Arabic, Chinese, Thai and dot numerals).

During experimental trials, stimuli were presented at five signal-levels, one step below the critical contrast value (level 1: hardest), and three steps above (levels 3, 4 and

5: easiest). As shown in Figure S4.2.1.a., across numeric-types, mean accuracy increased linearly with the visibility of the contrast levels, being lowest at level 1 ( $\mu = .32$ ,  $\sigma = .02$ ) and highest at level 5 ( $\mu = .8$ ,  $\sigma = .03$ ). On average, accuracy was highest for Chinese numerals ( $\mu = .6$ ,  $\sigma = .21$ ), then Arabic ( $\mu = .59$ ,  $\sigma = .19$ ) and non-symbolic dots ( $\mu = .59$ ,  $\sigma = .19$ ), and finally, Thai numerals ( $\mu = .54$ ,  $\sigma = .19$ ).

A repeated-measures ANOVA found a significant main effect of contrast level on accuracy ( $F(4, 40) = 447.914$ ,  $p < .001$ ,  $\eta^2 = .99$ ), but not a main effect of numeric-type on accuracy ( $F(3, 30) = 2.134$ ,  $p = 0.12$ ,  $\eta^2 = .18$ ). There was no interaction effect between numeric-type and contrast level on accuracy ( $F(12, 120) = 0.951$ ,  $p = .5$ ,  $\eta^2 = .09$ ). Post-hoc pair-wise  $t$ -tests using the Bonferroni correction revealed significant differences between all combinations of contrast level ( $p < .001$ ). By contrast, pair-wise  $t$ -tests showed no difference in accuracy between numeric-types, except between familiar items, Arabic numerals and non-symbolic dots ( $p < .05$ ). All simple effects are reported in the supplementary materials, Tables S4.3.1 and S4.3.2. These results indicate our chosen signal levels appropriately influenced response accuracy. However, there appears to be no effect of numeric familiarity on response-accuracy. We will revisit this line of inquiry shortly.

Figure S4.2.1.b. depicts mean accuracy across experimental blocks, for each numeric-type. Mean accuracy was comparable between numeric-types, and increased marginally with block number, being lowest at block 1 ( $\mu = .50$ ,  $\sigma = .14$ ) and highest at block 13 ( $\mu = .62$ ,  $\sigma = .13$ ). A repeated-measures ANOVA found a significant main effect of block on accuracy ( $F(12, 120) = 14.733$ ,  $p < .001$ ,  $\eta^2 = .6$ ), and no main effect

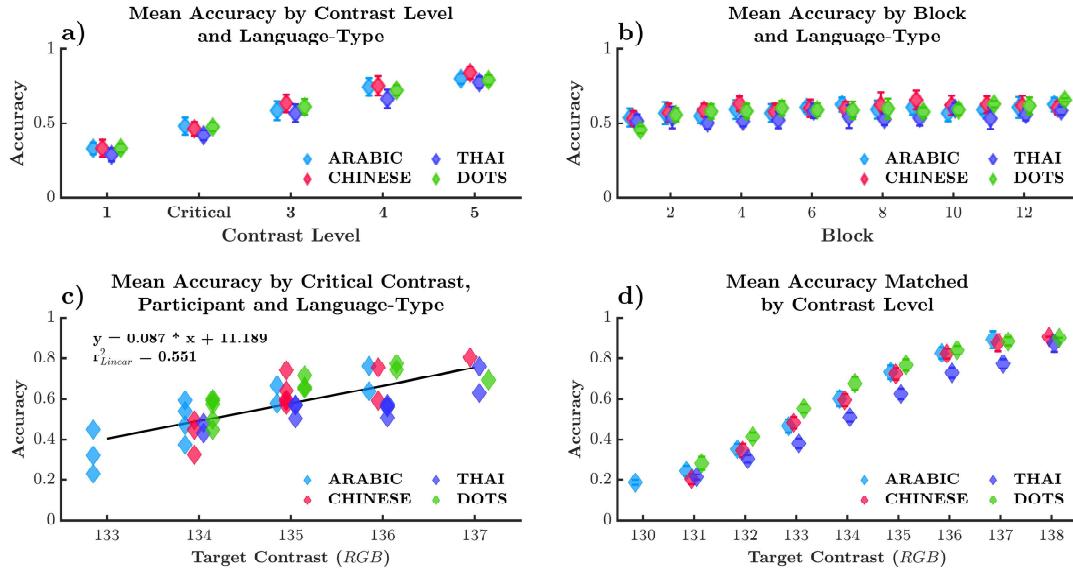


FIGURE S4.2.1: a) Mean accuracy across five signal contrast-levels, and four numeric-types. b) Mean accuracy across each experimental block. c) Mean accuracy for each participant by critical contrast level. d) Mean accuracy matched by contrast-level, across numeric-types. Error bars represent the standard-error of the mean.

of numeric-type on accuracy ( $F(3, 30) = 2.139, p = .012, \eta^2 = .18$ ). There was no interaction effect of numeric-type and block on accuracy ( $F(36, 360) = .975, p = .51, \eta^2 = .09$ ). Post-hoc pair-wise analysis revealed significant differences in accuracy between early and late experimental blocks. Block 1 differed significantly from blocks 5–13 ( $p < .01$ ), block 2 from blocks 12–13 ( $p < .01$ ) and block 3 from blocks 9, 11 and 13 ( $p < .05$ ). Simple effects are reported in the supplementary materials, Table S4.3.3. These results suggest a small practice effect, slightly boosting accuracy in later blocks.

Figure S4.2.1.c. presents mean experimental accuracy across critical contrast levels, separated by participant and numeric-type. A linear regression found a significant positive relationship between critical contrast and mean accuracy ( $r^2 = .551$ ), suggesting a dependency between contrast and accuracy. To disentangle the effect of numeric-type and contrast on accuracy, we assessed accuracy matched across RGB values from each participant's five signal-contrast levels (see S4.2.1.d).

Figure S4.2.1.d. presents mean accuracy matched across participant's five contrast-levels, separated by numeric-type. For example, if for Arabic numerals, participant S1 responded to RGB contrast values 130–134 and participant S2 responded to RGB contrast values 134–137, their accuracy at contrast value 134 would be averaged and depicted in Figure S4.2.1.d.

Figure S4.2.1.d. displays a positive relationship between contrast and matched accuracy. Matching accuracy for contrast levels when all numeric-types were presented, (i.e., excluding contrast values 130 and 138), accuracy was highest for non-symbolic dots ( $\mu = .63, \sigma = .22$ ), then Arabic numerals ( $\mu = .59, \sigma = .24$ ), then Chinese numerals ( $\mu = .58, \sigma = .25$ ) and lowest for Thai numerals ( $\mu = .51, \sigma = .21$ ).

We completed a two-way between-subjects ANOVA to assess the effect of numeric-type and contrast-level on matched accuracy (Figure S4.2.1.d). We found a main effect of numeric-type ( $F(3, 185) = 15.606, p < .001, \eta^2 = 0.04$ ), and a main effect of contrast-level ( $F(6, 185) = 148.814, p < .001, \eta^2 = 0.79$ ) on accuracy. There was no interaction effect between contrast level and numeric-type on accuracy ( $F(18, 185) = 0.003, p = .99, \eta^2 = 0.01$ ). Post-hoc pair-wise  $t$ -tests displayed significant differences between all contrast values ( $p < .001$ ), except between the highest RGB values, 136 and 137 (all pair-wise tests are reported in the supplementary materials, Table .3. Post-hoc  $t$ -tests displayed a significant differences in accuracy between all numeric-types ( $p < .05$ ), except for comparisons between Chinese and Thai, and Arabic and Thai numerals (all pair-wise tests reported in the supplementary materials, Table .3. These results show a clear effect of contrast-level on accuracy. After accounting for contrast level, trends indicate that accuracy was higher for familiar items (Dots and Arabic) compared to unfamiliar items

(Chinese and Thai), however, this was not borne out by the simple effects.

### .3 Simple effects: *t*-tests

TABLE S4.3.1: Post-hoc comparisons between contrast levels. Level 1 being the lowest signal contrast level (hardest) and level 5 being the highest (easiest). Level 2 is elsewhere referred to as the critical contrast level.

		Mean Difference	SE	t	Cohen's d	$p_{bonf}$
Level 1	Level 2	-0.113	0.009	-12.34	-3.720	< .001
	Level 3	-0.241	0.014	-17.13	-5.166	< .001
	Level 4	-0.355	0.016	-21.97	-6.623	< .001
	Level 5	-0.440	0.016	-26.89	-8.106	< .001
Level 2	Level 3	-0.128	0.009	-13.97	-4.211	< .001
	Level 4	-0.242	0.011	-21.72	-6.550	< .001
	Level 5	-0.327	0.015	-22.55	-6.798	< .001
Level 3	Level 4	-0.114	0.006	-18.67	-5.629	< .001
	Level 5	-0.199	0.009	-22.75	-6.860	< .001
Level 4	Level 5	-0.085	0.008	-10.29	-3.104	< .001

TABLE S4.3.2: Post-hoc comparisons between numeric-types.

		Mean Difference	SE	t	Cohen's d	$p_{bonf}$
ARABIC	CHINESE	-0.085	0.051	-1.678	-0.506	0.746
	THAI	-0.049	0.063	-0.766	-0.231	1.000
	DOTS	-0.120	0.036	-3.353	-1.011	0.044
CHINESE	THAI	0.037	0.058	0.640	0.193	1.000
	DOTS	-0.035	0.041	-0.839	-0.253	1.000
THAI	DOTS	-0.072	0.044	-1.616	-0.487	0.822

TABLE S4.3.3: Post-hoc comparisons of accuracy by block number.

		Mean Difference	SE	t	$p_{bonf}$
Block1	Block2	-0.032	0.014	-2.257	1.000
	Block3	-0.059	0.015	-3.992	0.199
	Block4	-0.072	0.015	-4.857	0.052
	Block5	-0.066	0.013	-5.209	0.031
	Block6	-0.082	0.013	-6.336	0.007
	Block7	-0.078	0.011	-6.756	0.004
	Block8	-0.080	0.013	-6.188	0.008
	Block9	-0.096	0.012	-8.144	< .001
	Block10	-0.092	0.012	-7.834	0.001
	Block11	-0.096	0.014	-6.623	0.005
	Block12	-0.096	0.014	-6.747	0.004
	Block13	-0.115	0.017	-6.841	0.004
Block2	Block3	-0.026	0.013	-1.961	1.000
	Block4	-0.040	0.014	-2.799	1.000
	Block5	-0.034	0.017	-2.004	1.000
	Block6	-0.049	0.013	-3.660	0.342
	Block7	-0.045	0.012	-3.891	0.234
	Block8	-0.048	0.016	-2.955	1.000
	Block9	-0.063	0.015	-4.334	0.116
	Block10	-0.060	0.016	-3.749	0.296
	Block11	-0.064	0.016	-3.870	0.243

**Table S4.3.3 continued from previous page**

		Mean Difference	SE	t	$p_{bonf}$
	Block12	-0.064	0.009	-7.064	0.003
	Block13	-0.082	0.013	-6.105	0.009
Block3	Block4	-0.014	0.012	-1.115	1.000
	Block5	-0.007	0.009	-0.806	1.000
	Block6	-0.023	0.009	-2.597	1.000
	Block7	-0.019	0.008	-2.338	1.000
	Block8	-0.021	0.010	-2.127	1.000
	Block9	-0.037	0.007	-5.200	0.031
	Block10	-0.034	0.011	-2.964	1.000
	Block11	-0.037	0.006	-5.826	0.013
	Block12	-0.038	0.008	-4.837	0.053
	Block13	-0.056	0.008	-7.089	0.003
Block4	Block5	0.006	0.014	0.456	1.000
	Block6	-0.009	0.012	-0.772	1.000
	Block7	-0.005	0.010	-0.527	1.000
	Block8	-0.008	0.010	-0.791	1.000
	Block9	-0.023	0.010	-2.409	1.000
	Block10	-0.020	0.014	-1.468	1.000
	Block11	-0.024	0.011	-2.203	1.000
	Block12	-0.024	0.009	-2.758	1.000
	Block13	-0.042	0.013	-3.359	0.566

**Table S4.3.3 continued from previous page**

		Mean Difference	SE	t	$p_{bonf}$
Block5	Block6	-0.016	0.011	-1.430	1.000
	Block7	-0.012	0.011	-1.104	1.000
	Block8	-0.014	0.009	-1.518	1.000
	Block9	-0.030	0.010	-3.108	0.865
	Block10	-0.027	0.010	-2.591	1.000
	Block11	-0.030	0.007	-4.418	0.101
	Block12	-0.030	0.012	-2.495	1.000
	Block13	-0.049	0.011	-4.255	0.131
	Block6	0.004	0.006	0.689	1.000
	Block7	0.002	0.009	0.162	1.000
	Block8	-0.014	0.010	-1.468	1.000
	Block9	-0.011	0.011	-1.020	1.000
	Block10	-0.014	0.009	-1.640	1.000
	Block11	-0.015	0.009	-1.636	1.000
	Block12	-0.033	0.007	-4.881	0.050
Block7	Block8	-0.003	0.007	-0.341	1.000
	Block9	-0.018	0.008	-2.217	1.000
	Block10	-0.015	0.012	-1.200	1.000
	Block11	-0.018	0.008	-2.332	1.000
	Block12	-0.019	0.008	-2.482	1.000
	Block13	-0.037	0.008	-4.783	0.058

**Table S4.3.3 continued from previous page**

		Mean Difference	SE	t	$p_{bonf}$
Block8	Block9	-0.016	0.008	-1.911	1.000
	Block10	-0.012	0.012	-1.019	1.000
	Block11	-0.016	0.005	-2.959	1.000
	Block12	-0.016	0.011	-1.506	1.000
	Block13	-0.035	0.011	-3.231	0.703
Block9	Block10	0.003	0.010	0.344	1.000
	Block11	-2.525e-4	0.007	-0.034	1.000
	Block12	-5.051e-4	0.009	-0.056	1.000
	Block13	-0.019	0.011	-1.684	1.000
Block10	Block11	-0.004	0.011	-0.312	1.000
	Block12	-0.004	0.011	-0.335	1.000
	Block13	-0.022	0.012	-1.815	1.000
Block11	Block12	-2.525e-4	0.010	-0.025	1.000
	Block13	-0.019	0.009	-2.119	1.000
Block12	Block13	-0.018	0.007	-2.729	1.000

#### .4 Scree analysis of bias-free MDS stress values

Scree analysis compares the multidimensional stress values (y-axis) against the number of MDS dimensions (x-axis). Scree analysis, such as this, is a subjective measure. A useful heuristic for identifying the correct number of dimensions is to look for the ‘elbow’

TABLE S4.3.4: Post-hoc comparisons of accuracy matched by contrast level, for RGB contrast values 131–136.

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
131	132	-0.118	0.028	-4.227	-1.572	< .001
	133	-0.235	0.027	-8.851	-2.504	< .001
	134	-0.361	0.026	-13.731	-3.528	< .001
	135	-0.476	0.027	-17.916	-5.024	< .001
	136	-0.568	0.029	-19.850	-7.684	< .001
	137	-0.621	0.034	-18.528	-8.921	< .001
	132	-0.117	0.021	-5.654	-1.240	< .001
132	133	-0.243	0.020	-11.925	-2.394	< .001
	134	-0.358	0.021	-17.265	-3.757	< .001
	135	-0.450	0.023	-19.309	-5.617	< .001
	136	-0.503	0.029	-17.277	-6.344	< .001
	137	-0.125	0.019	-6.764	-1.150	< .001
133	134	-0.240	0.019	-12.700	-2.306	< .001
	135	-0.333	0.022	-15.312	-3.505	< .001
	136	-0.386	0.028	-13.839	-3.947	< .001
	137	-0.115	0.018	-6.233	-1.054	< .001
134	135	-0.208	0.021	-9.728	-2.037	< .001
	136	-0.261	0.028	-9.452	-2.462	< .001
	137	-0.092	0.022	-4.262	-0.968	< .001
135	136	-0.146	0.028	-5.224	-1.478	< .001
	137	-0.053	0.030	-1.778	-0.675	1.000

TABLE S4.3.5: Post-hoc comparisons of accuracy matched by contrast level, across numeric-types.

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
ARABIC	CHINESE	-0.052	0.018	-2.857	-0.261	0.029
	DOTS	0.074	0.019	3.891	0.379	< .001
	THAI	-0.009	0.019	-0.472	-0.043	1.000
CHINESE	DOTS	0.126	0.019	6.779	0.674	< .001
	THAI	0.043	0.019	2.309	0.213	0.132
DOTS	THAI	-0.083	0.019	-4.271	-0.420	< .001

where an increase in dimensions does not meaningfully improve stress values. This elbow has been identified by a marker in each plot.

#### 4.1 Scree Plots

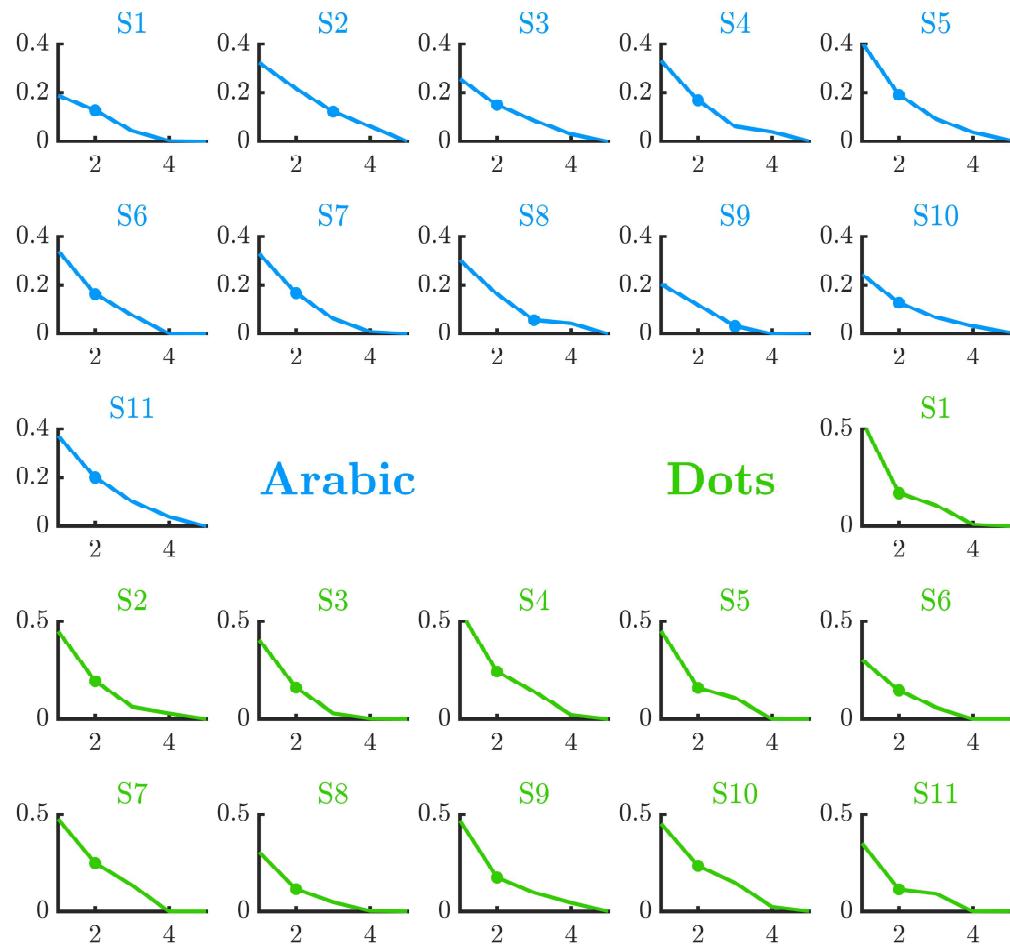


FIGURE S4.4.2: Bias-free MDS scree plots for Arabic digits (blue) and symbolic dots (green). The y-axis displays stress values, and the x-axis the number of dimensions. Markers identify the optimal number of dimensions in each scree plot.

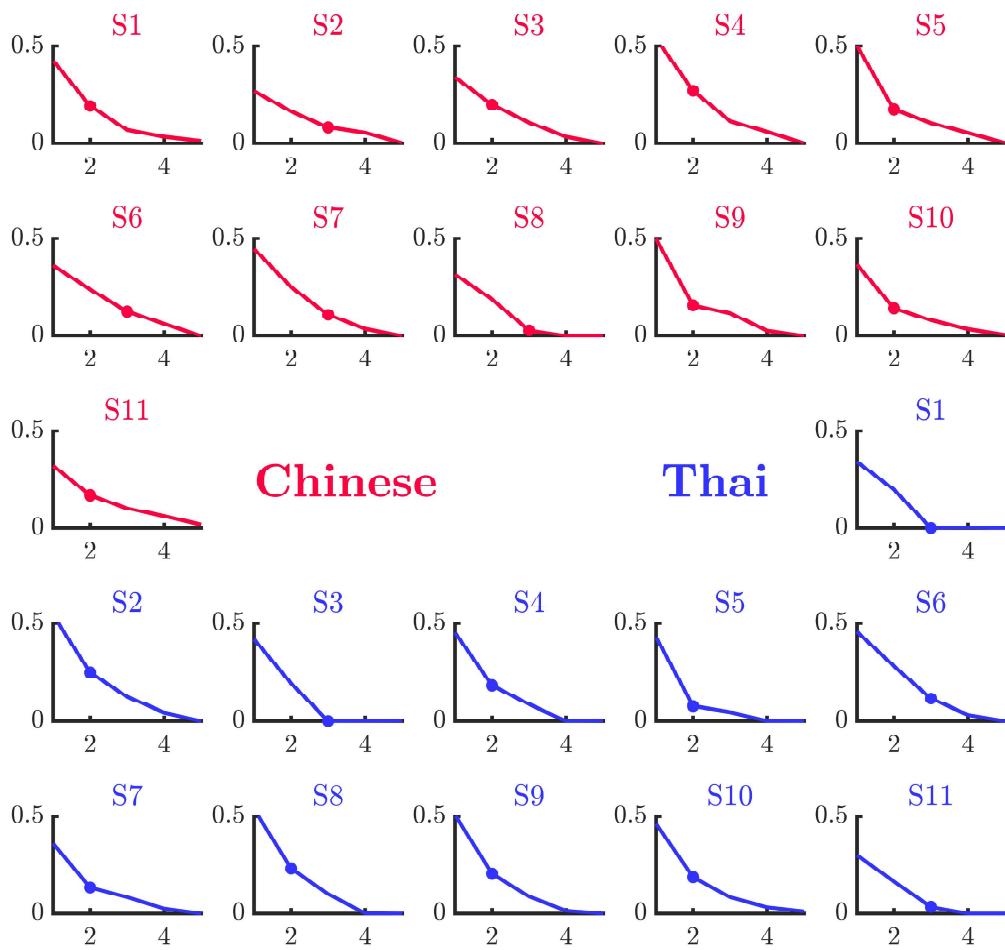


FIGURE S4.4.3: Bias-free MDS scree plots for Chinese (red) and Thai (purple) symbols. The y-axis displays stress values, and the x-axis the number of dimensions. Markers identify the optimal number of dimensions in each scree plot.

## 4.2 Individual MDS solutions

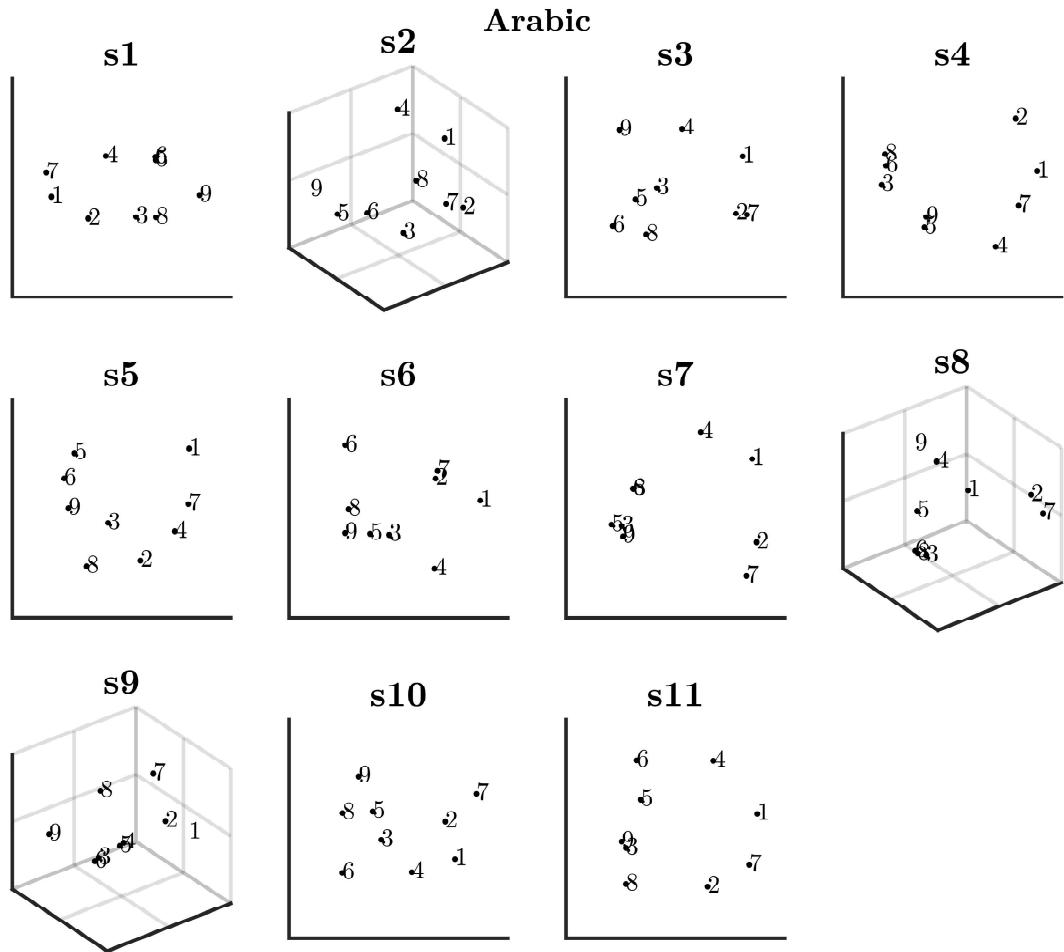


FIGURE S4.4.4: Individual bias-free MDS solutions for the Arabic digits.

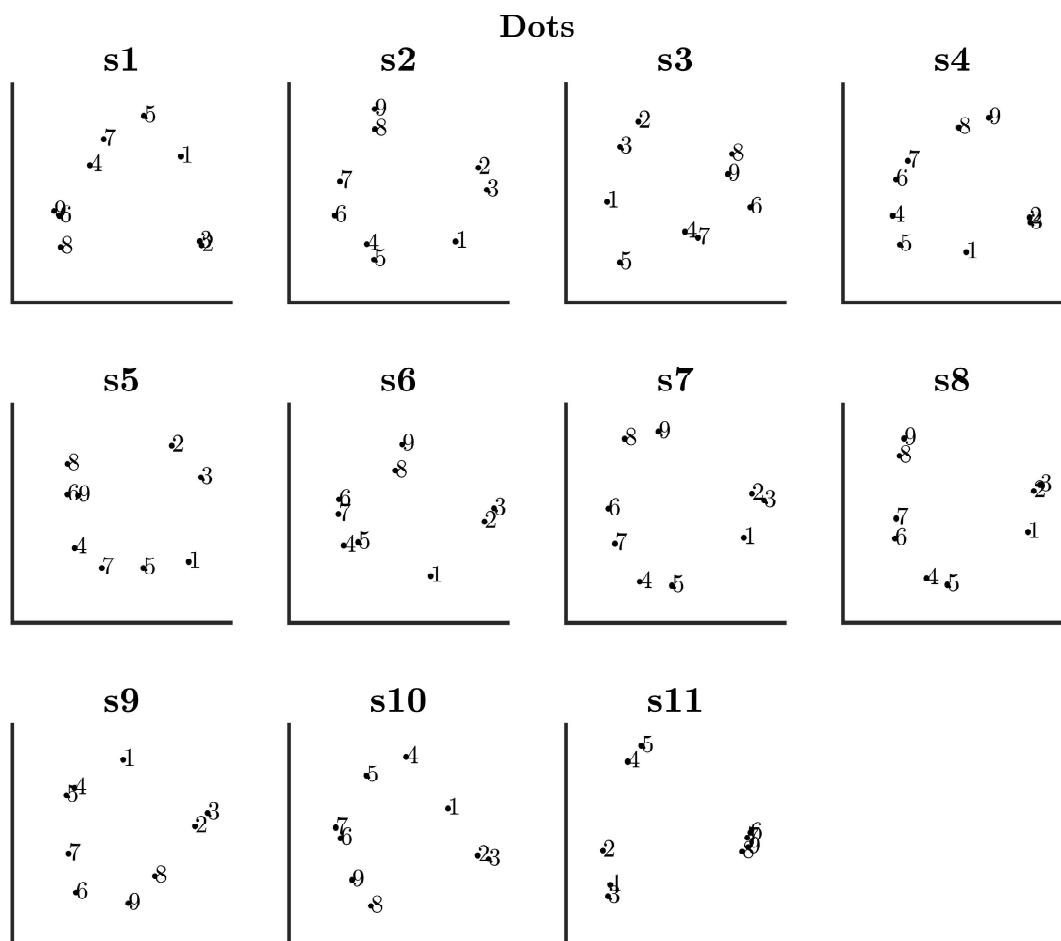


FIGURE S4.4.5: Individual bias-free MDS solutions for symbolic dots. Dots are represented by Arabic numbers for simplicity.

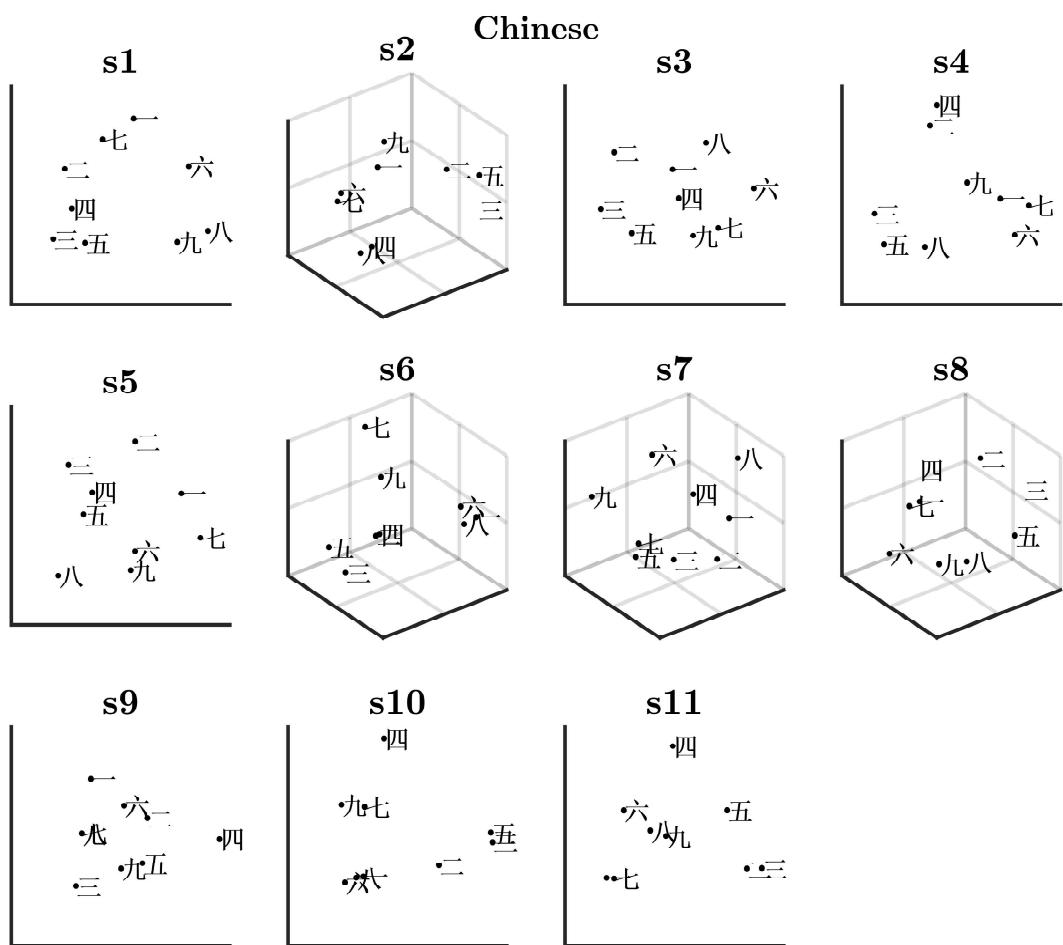


FIGURE S4.4.6: Individual bias-free MDS solutions for Chinese symbols.

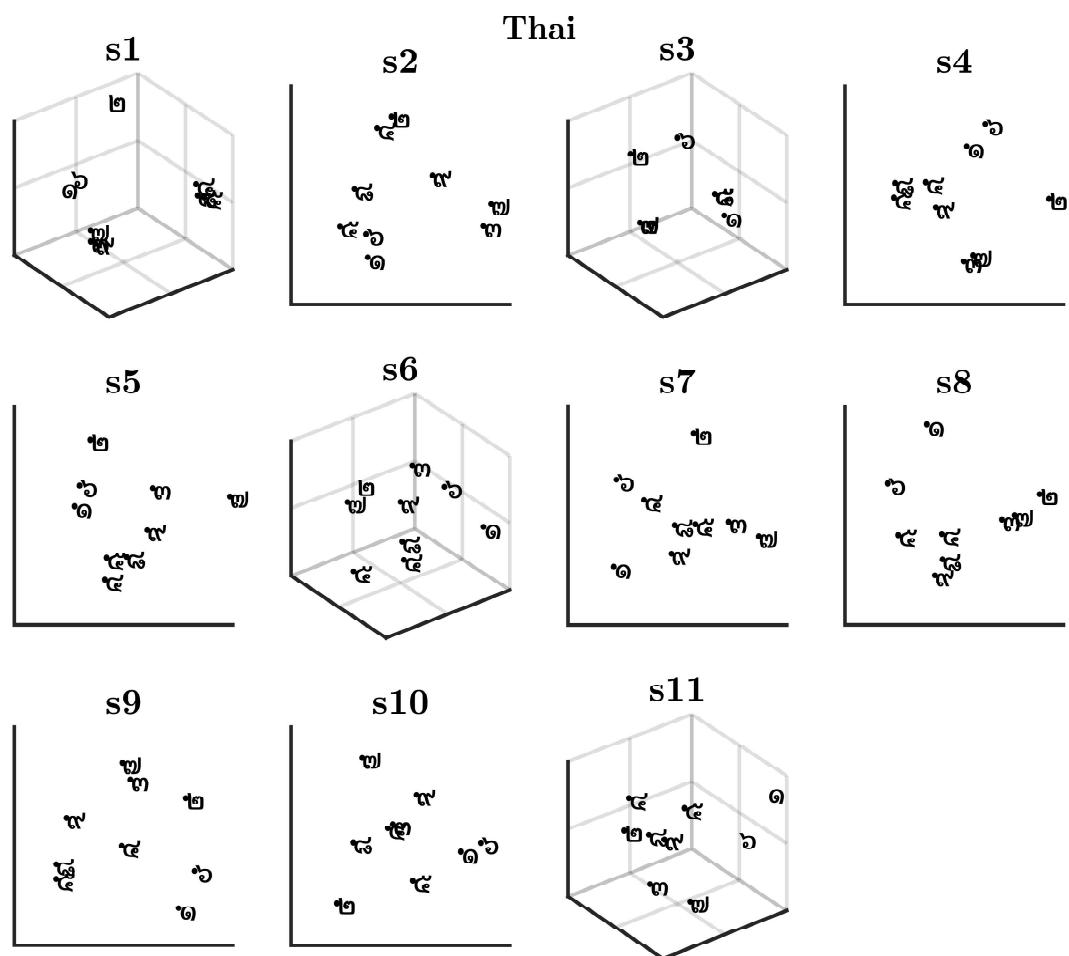


FIGURE S4.4.7: Individual bias-free MDS solutions for the Thai symbols.

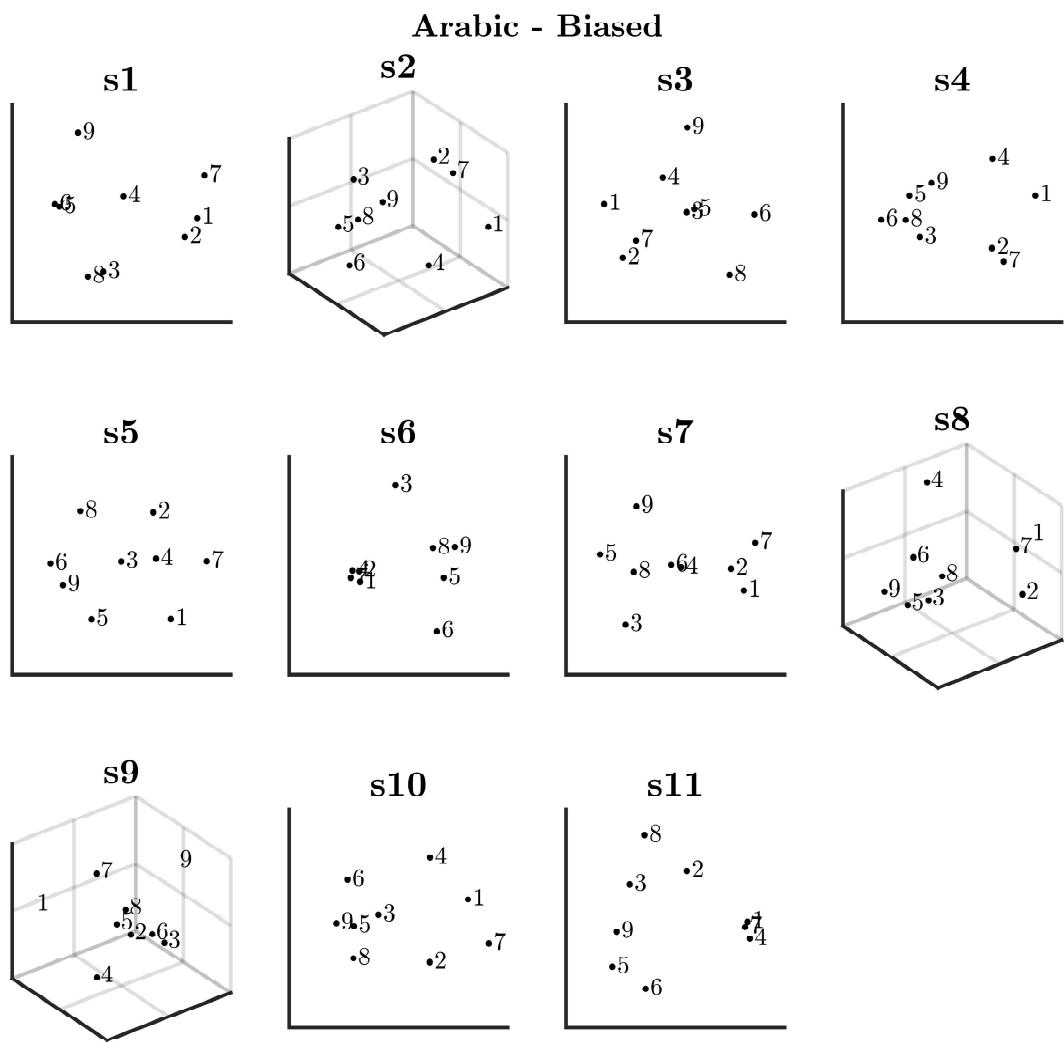


FIGURE S4.4.8: Individual biased MDS solutions for the Arabic digits.

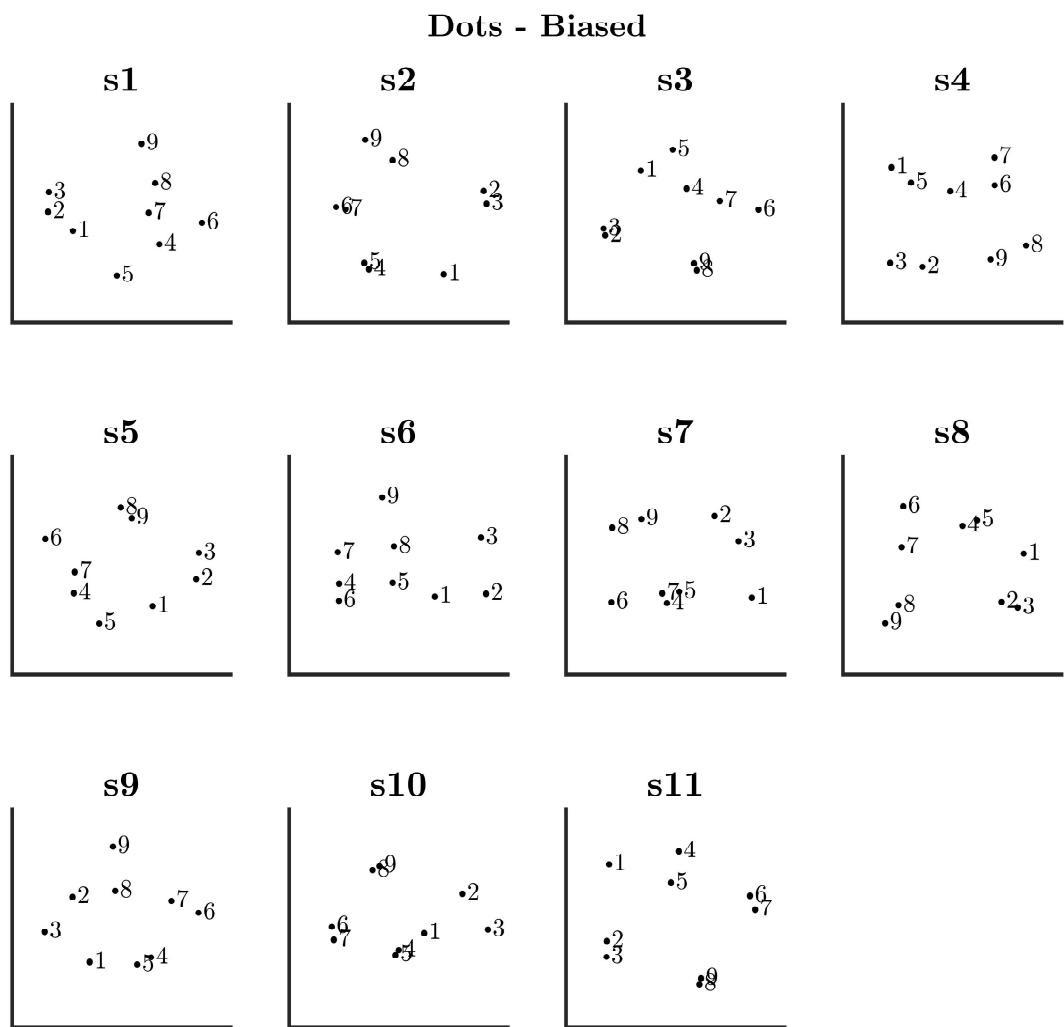


FIGURE S4.4.9: Individual biased MDS solutions for symbolic dots. Dots are represented by Arabic numbers for simplicity.

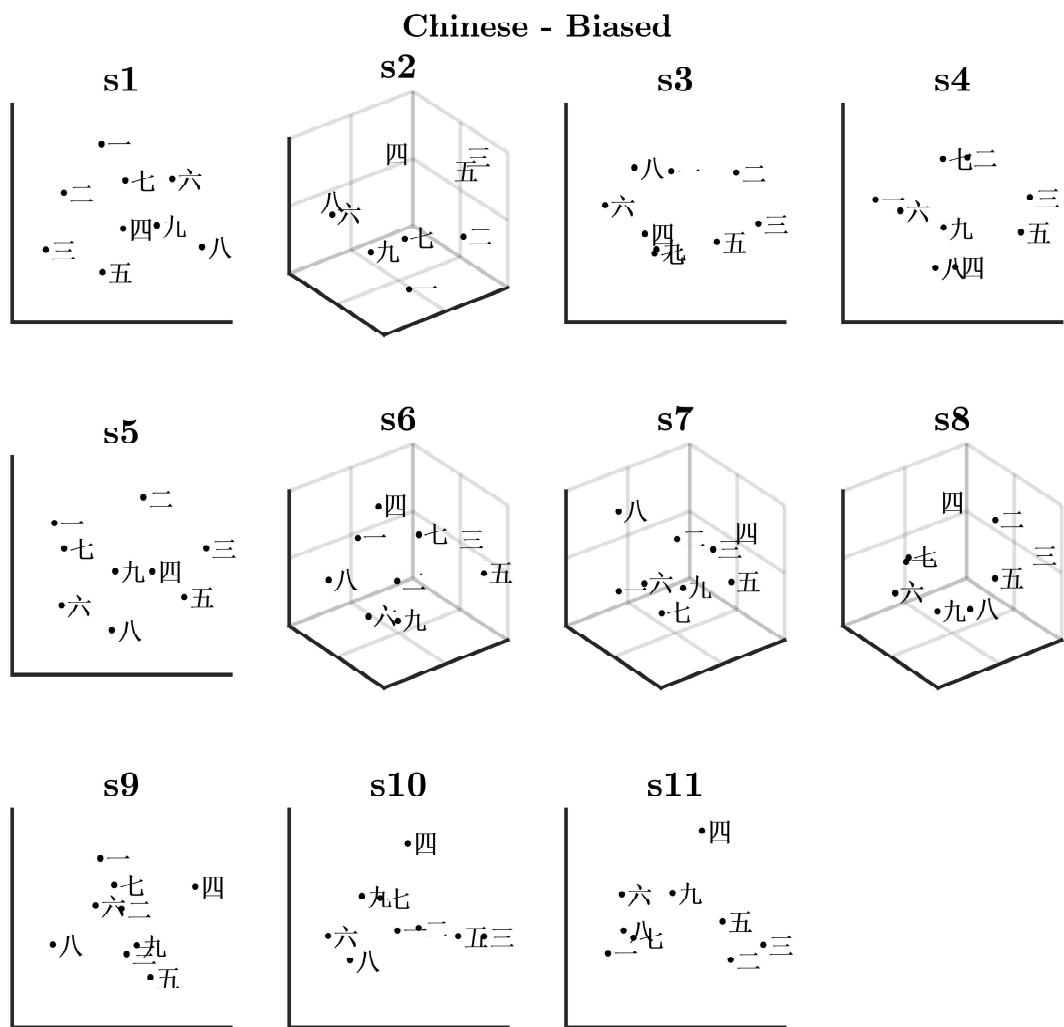


FIGURE S4.4.10: Individual bias-free MDS solutions for Chinese symbols.

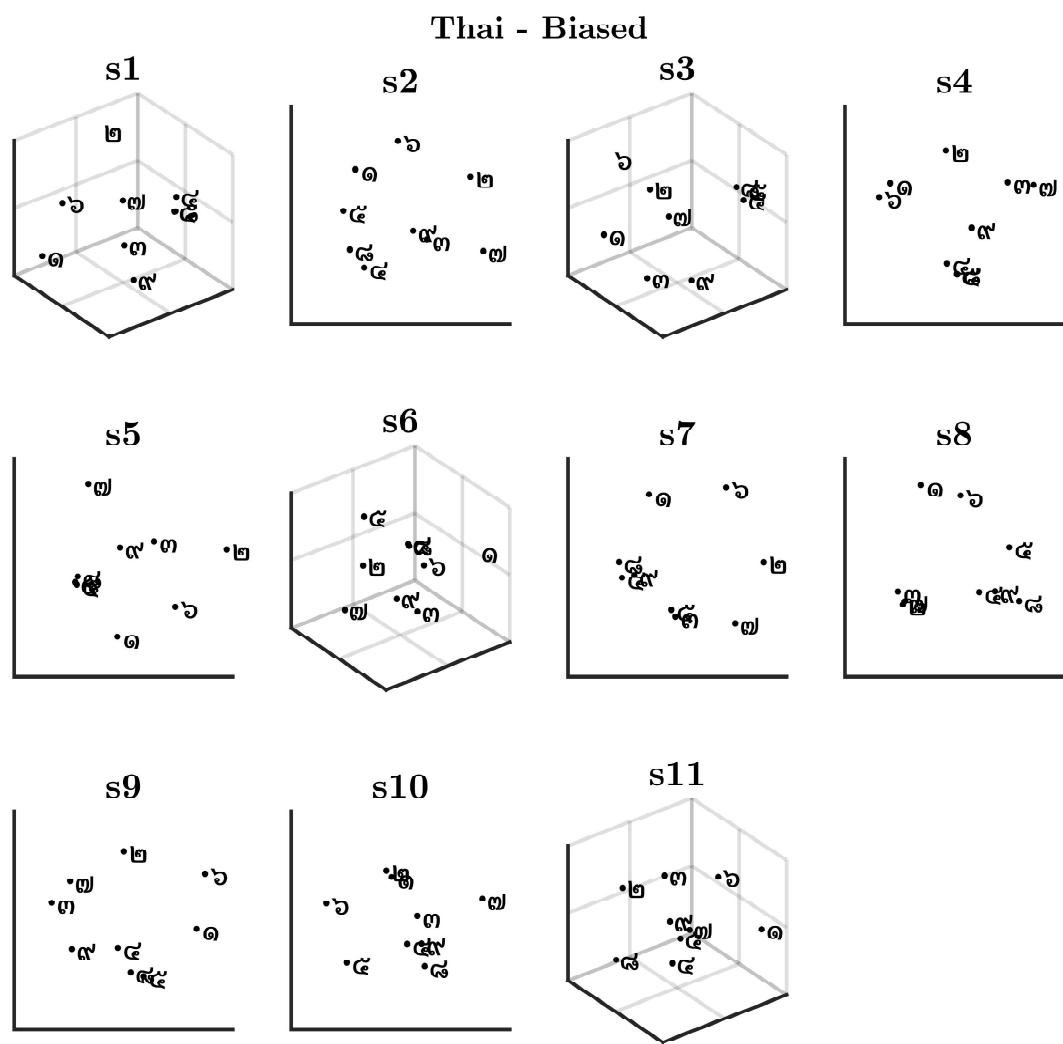


FIGURE S4.4.11: Individual bias-free MDS solutions for Thai symbols.

## .5 MDS cluster frequency heatmaps

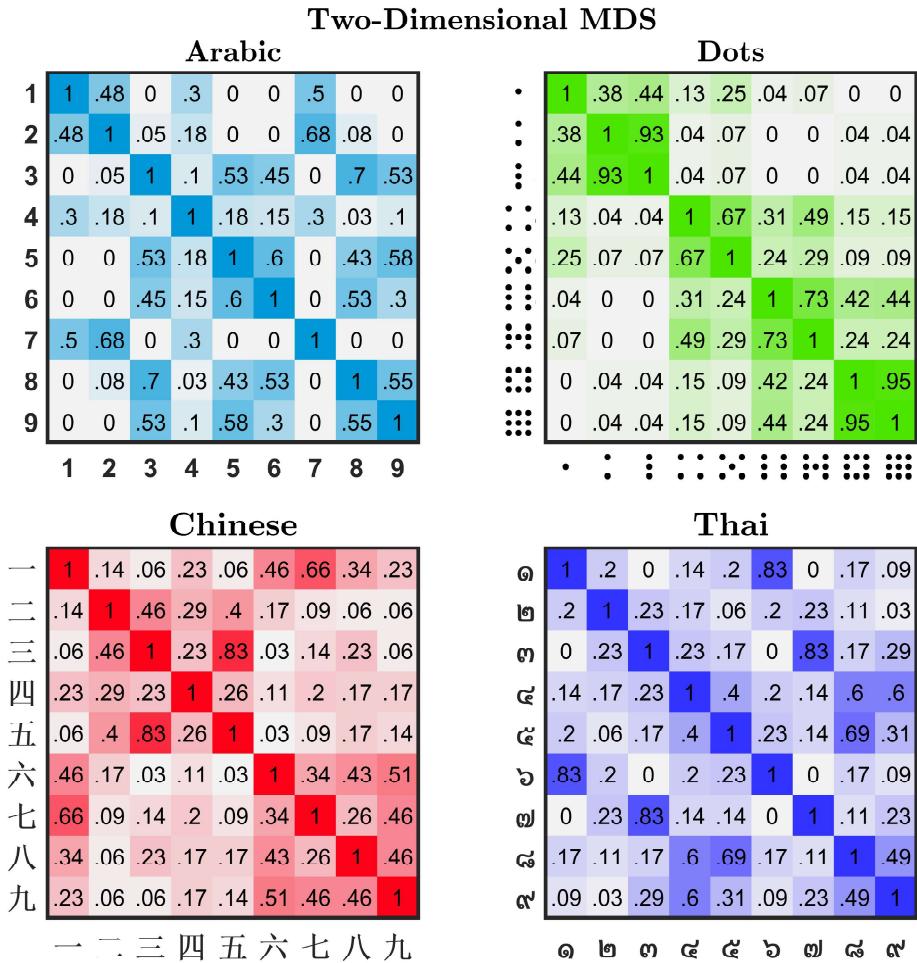


FIGURE S4.5.12: Proportional cluster-frequency heatmap for participants with two-dimensional MDS solutions, across 2–6 K-mean clusters. Larger proportions (darker colored squares) indicate items which most frequently cluster together.

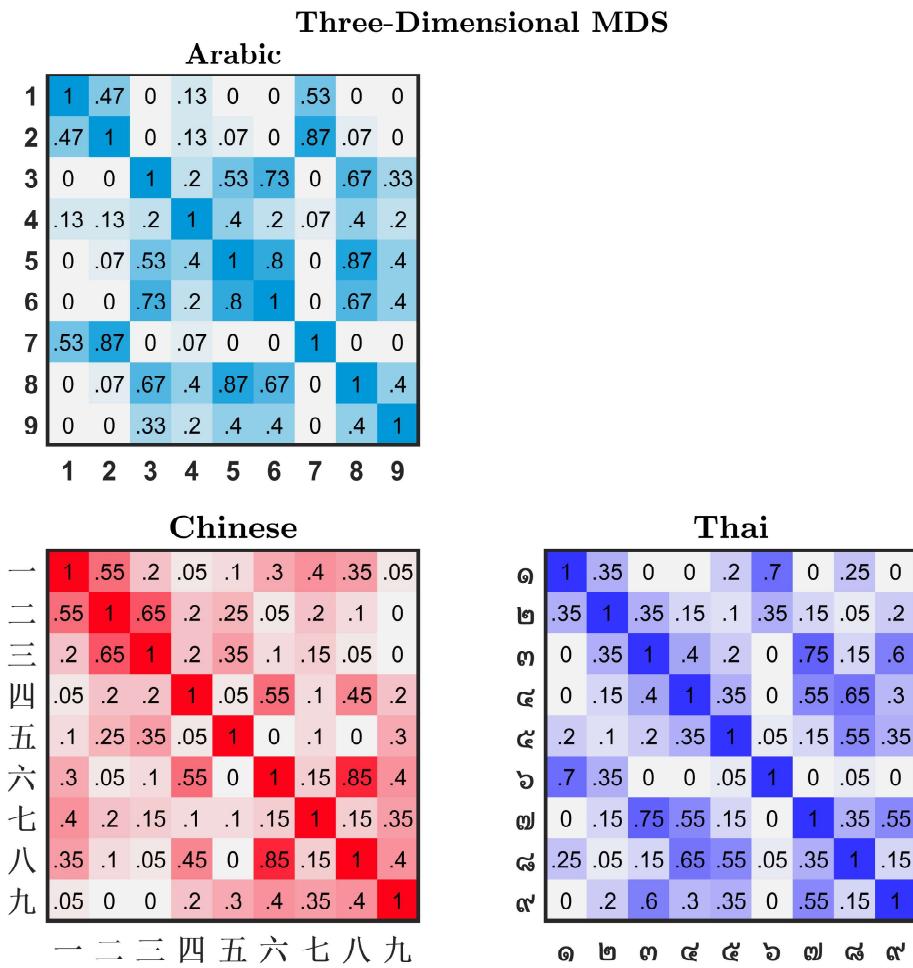
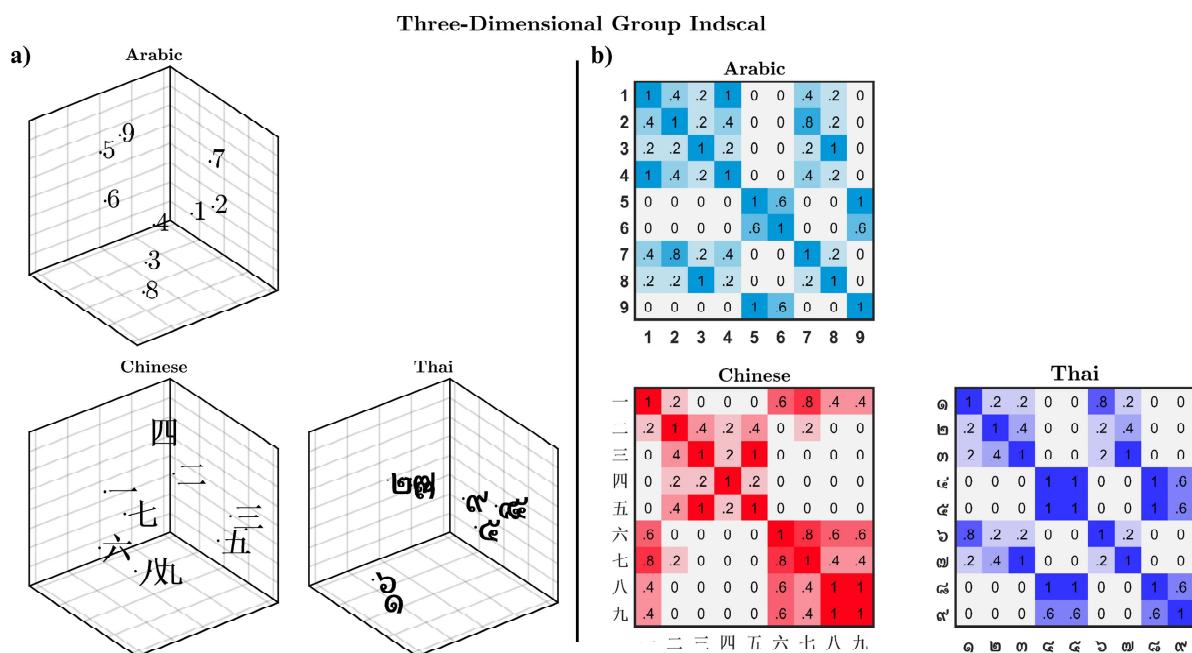


FIGURE S4.5.13: Proportional cluster-frequency heatmap for participants with three-dimensional MDS solutions, across 2–6 K-mean clusters. Larger proportions (darker colored squares) indicate items which most frequently cluster together.

## .6 Three dimensional group indscal solutions

Figure S4.6.14 displays the group indscal MDS and K-mean cluster frequency results for those participants identified with three MDS dimensions. No participants displayed a third MDS dimension in the symbolic-dot numeric-type. MDS and cluster frequency plots are comparable between three-dimensional and two-dimensional indscal results. Arabic items displayed similar clusters, however, item '9' shifts from being grouped with items '3' and '8', to being grouped with items '5' and '6'. Chinese results are comparable between two- and three-dimensional plots, except in the three-dimensional plot, item 四 shifts away from all items along the third-dimension. Finally, similar results were observed in the three-dimensional Thai MDS and cluster-frequency plots, except that item-numbers [1,6] move away from all other items along the third-dimension.



## **Supplementary Materials S5:**

### **The cross cultural cost of errors**

The materials in this supplementary chapter are relevant to Chapter 7 of the submitted thesis.

## .1 Analysis of the calibration staircase procedure

The following provides a full analysis of our successful staircase procedure. As the reader will soon learn, our staircase procedure successfully manipulated stimulus accuracy for all participants in all numeric types. Although one participant finished at a relatively high critical contrast level in the non-symbolic dot numeric type, the MDS results of this participant did not appear to differ from the other subjects. We summarised by saying this procedure was highly effective and successfully produced the necessary confusions

Figure S5.1.1 (top) depicts the calibration block for participant S1, responding to Arabic numerals. This staircase procedure was typical of most participants. The mean contrast level of the final 30 assessment trials (highlighted yellow) determined the critical contrast value — the value from which stimulus-signal levels were determined in the experimental session. Violin plots (bottom) depict the mean and standard-error of contrast values for assessment trials, for each participant and language-type. Critical contrast levels were stable across language-type and participants, except for participant S2 in the non-symbolic dots condition. This participant's staircase procedure resulted in the highest (easiest) critical contrast level, indicating their performance was particularly poor during the calibration block.

Colored ticks on the violin-plot (Figure S5.1.1, y-axis) show, on average, critical contrast levels were approximately equal for Chinese numerals ( $\text{RGB } \mu = 137.4, \sigma = 1.49$ ), Arabic numerals ( $\text{RGB } \mu = 137.6, \sigma = 1.15$ ) and non-symbolic dots ( $\text{RGB } \mu = 137.8, \sigma = 3.89$ ); and highest for Thai numerals ( $\text{RGB } \mu = 140.3, \sigma = 1.75$ ). A higher critical contrast level for Thai numerals suggests unfamiliar numeric items were more difficult to identify

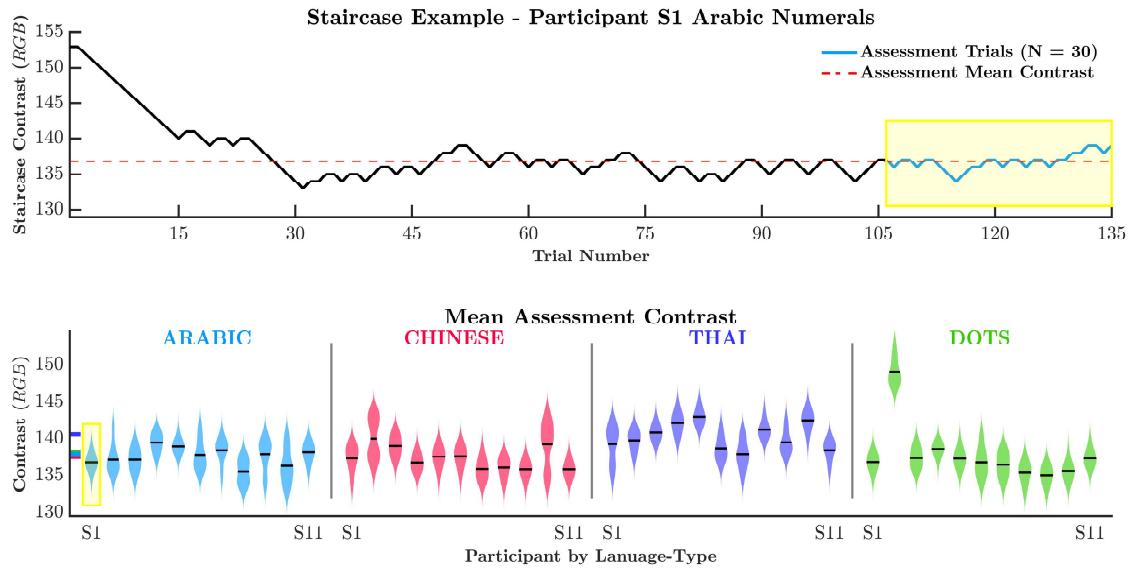


FIGURE S5.1.1: Plot of participant S1's Arabic numeral staircase procedure (top) and violin plots of individual participant's staircase assessment trials (bottom). Assessment trials (highlighted yellow for participant S1) determined the critical contrast value for the main experiment. For all language-types, participants displayed relatively stable contrast levels during the assessment window. Black lines on each violin plot represent the critical contrast value (mean RGB value over the assessment window). Colored ticks on the y-axis are the mean critical contrast values in each language-type.

than familiar numeric items (Chinese, Arabic and Dots). A one-way repeated-measures ANOVA displayed a significant main effect of critical contrast across language-types ( $F(3, 30) = 4.22, p < 0.05$ ). Post-hoc paired-sample  $t$ -tests displayed significant differences in critical contrast levels between Arabic and Thai numerals ( $t(10) = -4.217, p < .05$ ) and Chinese and Thai numerals ( $t(10) = -5.2, p < 0.01$ ); other comparisons did not reach statistical significance (summarized in Table S5.3.1).

## .2 Experimental accuracy by participant, contrast level and numeric type

During experimental trials, stimuli were presented at five signal-levels: one step below the critical contrast value (level 1: hardest) and three steps above (levels 3, 4 and 5: easiest). Across language-types, mean accuracy increased linearly with the visibility of the contrast levels (Figure S5.2.2.a). Accuracy was lowest at level 1 ( $\mu = .48$ ,  $\sigma = .17$ ) and highest at level 5 ( $\mu = .79$ ,  $\sigma = .11$ ). Across experimental trials, mean accuracy was highest for non-symbolic dots ( $\mu = .72$ ,  $\sigma = .14$ ), then Chinese numerals ( $\mu = .66$ ,  $\sigma = .12$ ), Arabic numerals ( $\mu = .62$ ,  $\sigma = .14$ ), and finally, Thai numerals ( $\mu = .59$ ,  $\sigma = .11$ ).

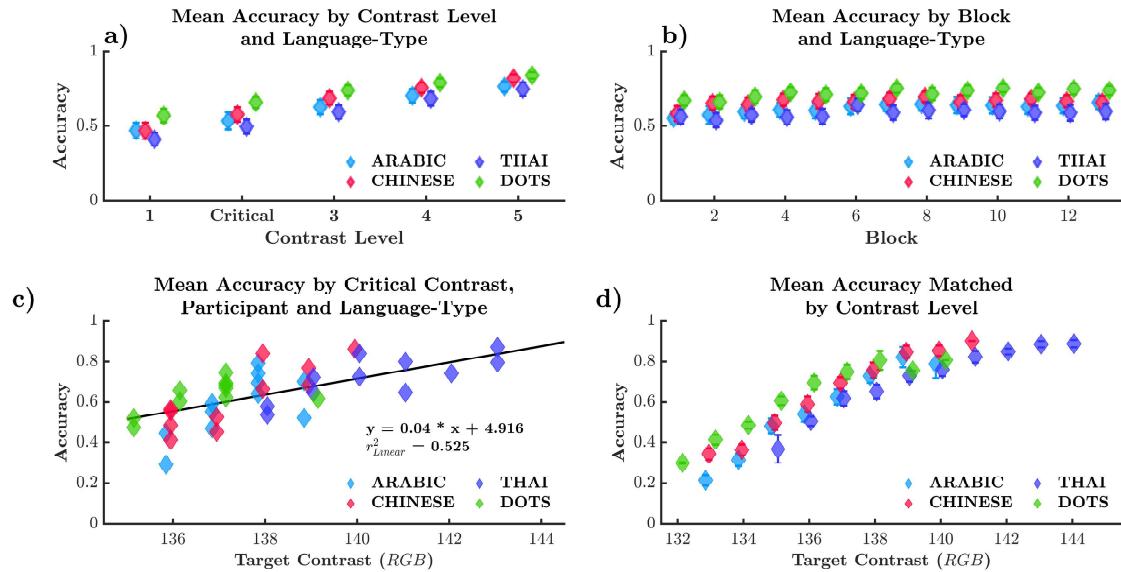


FIGURE S5.2.2: a) Mean accuracy across five signal contrast-levels, and four language-types. b) Mean accuracy across each experimental block. c) Mean accuracy for each participant by critical contrast level. d) Mean accuracy matched by contrast-level, across language-types. Error bars represent the standard-error of the mean.

A repeated measures ANOVA displayed a main effect of contrast level on accuracy ( $F(4, 40) = 227.262$ ,  $p < 0.001$ ,  $\eta^2 = 0.958$ ), but no main effect of language-type on accuracy ( $F(3, 30) = 2.042$ ,  $p = 0.13$ ,  $\eta^2 = 0.17$ ). There was a significant interaction

of language-type and contrast level on accuracy ( $F(12, 120) = 1.872, p < 0.05, \eta^2 = 0.158$ ). Post-hoc paired *t*-tests displayed significant differences between all combinations of contrast level ( $p < .001$ ), and no significant differences between comparisons of language-type (simple effects are reported in supplementary material .3, Tables S5.3.2 and S5.3.3. These results indicate our chosen signal levels appropriately influenced response accuracy, and that familiarity had no effect on response-accuracy. We will return to this shortly.

Figure S5.2.2.b. depicts mean accuracy across experimental blocks and language-type. Mean accuracy was comparable between language-types, and increased marginally with block number, being lowest at block 1 ( $\mu = .59, \sigma = .15$ ) and highest at block 7 ( $\mu = .67, \sigma = .15$ ), before plateauing to block 13 ( $\mu = .66, \sigma = .15$ ). A repeated-measures ANOVA displayed a significant main effect of block number on accuracy ( $F(12, 120) = 7.207, p < 0.001, \eta^2 = 0.419$ ), and did not display a significant interaction of block number and language-type on accuracy ( $F(36, 360) = 1.033, p = 0.42, \eta^2 = 0.094$ ). Post-hoc paired-sample *t*-tests displayed significant differences in accuracy between blocks 1 and 9–11 ( $p < 0.05$ ), and blocks 2 and 10 ( $p < 0.05$ ; simple effects reported in supplementary material D Table S5.3.4). These results suggest a small practice effect improved accuracy from early to late blocks.

Figure S5.2.2.c. presents mean experimental accuracy across critical contrast levels, separated by participant and language-type. A linear regression found a significant positive relationship between critical contrast and mean accuracy ( $r^2 = .525$ ), suggesting a dependency between contrast and accuracy. To disentangle the effect of language-type and contrast on accuracy, we assessed accuracy matched across RGB values from each participant's five signal-contrast levels (see S5.2.2.d).

Figure S5.2.2.d. presents mean accuracy matched across participant's five contrast-levels, separated by language-type. For example, if for Arabic numerals, participant S1 responded to RGB contrast values 130–134 and participant S2 responded to RGB contrast values 134–137, their accuracy at contrast value 134 would be averaged and depicted in Figure S5.2.2.d.

Figure S5.2.2.d. displays a positive relationship between contrast and matched accuracy. Matching accuracy for contrast levels when all language-types were presented (*i.e.*, excluding contrast values  $< 135$  and  $> 140$ ), accuracy was highest for non-symbolic dots ( $\mu = .74$ ,  $\sigma = .06$ ), then Chinese numerals ( $\mu = .71$ ,  $\sigma = .09$ ), then Arabic numerals ( $\mu = .66$ ,  $\sigma = .11$ ) and lowest for Thai numerals ( $\mu = .61$ ,  $\sigma = .09$ ).

We completed a two-way between-subjects ANOVA to assess the effect of language-type and contrast-level on matched accuracy (Figure S5.2.2.d). We found a main effect of language-type ( $F(3, 141) = 7.458, p < .001, \eta^2 = 0.08$ ), and a main effect of contrast-level ( $F(5, 141) = 24.252, p < .001, \eta^2 = 0.42$ ) on accuracy. There was no interaction effect between contrast level and language-type on accuracy ( $F(15, 141) = 0.512, p = .93, \eta^2 = 0.03$ ). Post-hoc *t*-tests displayed significant differences between all contrast values ( $p < .05$ ), except comparisons made between high RGB values: 137 and 138, 138 and 139, 138 and 140 and 139 and 140 (pair-wise tests are reported in the supplementary Table S5.3.5). This suggests a plateau in accuracy occurring at high RGB contrast values. Post-hoc *t*-tests displayed significant differences in matched accuracy between Arabic and Thai numerals ( $t(3) = 3.969, p < 0.001$ ) and non-symbolic dots and Thai numerals ( $t(3) = 3.988, p < 0.001$ ); other comparisons did not reach statistical significance (summarised in supplementary material D Table S5.3.6). When accuracy was matched by contrast level,

the familiar Arabic and Dot numeric-sets were more accurately reported than the Thai numeric-set. Familiar Chinese numerals tended to be more accurately reported than Thai numerals, however, this was not borne out at the statistical level.

### .3 Simple effects: *t*-tests

TABLE S5.3.1: Post Hoc Comparisons - CalibrationMeans

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
ENG	DOT	-0.215	1.203	-0.179	-0.054	1.000
	CHN	0.227	0.641	0.354	0.107	1.000
	THI	-2.691	0.638	-4.217	-1.272	0.011
DOT	CHN	0.442	0.974	0.454	0.137	1.000
	THI	-2.476	1.314	-1.884	-0.568	0.534
	CHN	-2.918	0.561	-5.200	-1.568	0.002

TABLE S5.3.2: Post Hoc Comparisons - Contrast Levels

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
L1	L2	-0.087	0.011	-7.985	-2.408	< .001
	L3	-0.182	0.010	-18.390	-5.545	< .001
	L4	-0.253	0.014	-18.701	-5.639	< .001
	L5	-0.313	0.017	-18.382	-5.542	< .001
	L2	-0.095	0.010	-9.846	-2.969	< .001
L2	L3	-0.166	0.012	-14.352	-4.327	< .001
	L4	-0.226	0.015	-14.633	-4.412	< .001
	L5	-0.072	0.008	-9.340	-2.816	< .001
L3	L4	-0.132	0.011	-11.762	-3.546	< .001
	L5	-0.060	0.007	-8.183	-2.467	< .001

TABLE S5.3.3: Post Hoc Comparisons - Numeric type

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
ARABIC	CHINESE	-0.034	0.068	-0.494	-0.149	1.000
	THAI	-0.133	0.064	-2.089	-0.630	0.380
	DOTS	-0.076	0.063	-1.202	-0.362	1.000
CHINESE	THAI	-0.100	0.048	-2.063	-0.622	0.397
	DOTS	-0.042	0.034	-1.241	-0.374	1.000
THAI	DOTS	0.058	0.057	1.006	0.303	1.000

TABLE S5.3.4: Post-hoc comparisons of accuracy by block number.

		Mean Difference	SE	t	Cohen's <i>d</i>	<i>p</i> <sub>bonf</sub>
Block 1	Block 2	-0.014	0.016	-0.878	-0.265	1.000
	Block 3	-0.034	0.016	-2.195	-0.662	1.000
	Block 4	-0.049	0.011	-4.452	-1.342	0.096
	Block 5	-0.043	0.016	-2.702	-0.815	1.000
	Block 6	-0.070	0.015	-4.839	-1.459	0.053
	Block 7	-0.074	0.016	-4.711	-1.421	0.065
	Block 8	-0.072	0.017	-4.142	-1.249	0.156
	Block 9	-0.068	0.014	-5.013	-1.512	0.041
	Block 10	-0.072	0.014	-5.141	-1.550	0.034
	Block 11	-0.063	0.013	-4.885	-1.473	0.050
	Block 12	-0.067	0.016	-4.156	-1.253	0.153
	Block 13	-0.070	0.015	-4.624	-1.394	0.074
Block 2	Block 3	-0.021	0.011	-1.817	-0.548	1.000
	Block 4	-0.036	0.012	-2.900	-0.874	1.000
	Block 5	-0.030	0.015	-1.959	-0.591	1.000
	Block 6	-0.057	0.016	-3.603	-1.086	0.376
	Block 7	-0.061	0.016	-3.851	-1.161	0.250
	Block 8	-0.059	0.016	-3.612	-1.089	0.371
	Block 9	-0.055	0.015	-3.728	-1.124	0.306
	Block 10	-0.059	0.011	-5.214	-1.572	0.031
	Block 11	-0.050	0.013	-3.732	-1.125	0.304

**Table S5.3.4 continued from previous page**

		Mean Difference	SE	t	Cohen's <i>d</i>	<i>p</i> <sub>bonf</sub>
	Block 12	-0.053	0.017	-3.129	-0.943	0.835
	Block 13	-0.056	0.015	-3.842	-1.158	0.254
Block 3	Block 4	-0.015	0.014	-1.062	-0.320	1.000
	Block 5	-0.009	0.020	-0.435	-0.131	1.000
	Block 6	-0.036	0.019	-1.943	-0.586	1.000
	Block 7	-0.040	0.019	-2.127	-0.641	1.000
	Block 8	-0.038	0.019	-2.025	-0.611	1.000
	Block 9	-0.034	0.017	-1.968	-0.593	1.000
	Block 10	-0.038	0.017	-2.217	-0.668	1.000
	Block 11	-0.029	0.018	-1.630	-0.491	1.000
	Block 12	-0.033	0.021	-1.563	-0.471	1.000
	Block 13	-0.035	0.019	-1.870	-0.564	1.000
Block 4	Block 5	0.006	0.009	0.688	0.207	1.000
	Block 6	-0.021	0.011	-1.982	-0.598	1.000
	Block 7	-0.025	0.010	-2.479	-0.747	1.000
	Block 8	-0.023	0.010	-2.347	-0.708	1.000
	Block 9	-0.019	0.009	-2.145	-0.647	1.000
	Block 10	-0.023	0.009	-2.451	-0.739	1.000
	Block 11	-0.014	0.010	-1.399	-0.422	1.000
	Block 12	-0.018	0.013	-1.383	-0.417	1.000
	Block 13	-0.020	0.012	-1.765	-0.532	1.000

**Table S5.3.4 continued from previous page**

		Mean Difference	SE	t	Cohen's <i>d</i>	<i>p</i> <sub>bonf</sub>
Block 5	Block 6	-0.027	0.010	-2.824	-0.851	1.000
	Block 7	-0.031	0.011	-2.905	-0.876	1.000
	Block 8	-0.029	0.011	-2.567	-0.774	1.000
	Block 9	-0.025	0.008	-3.105	-0.936	0.870
	Block 10	-0.029	0.009	-3.371	-1.016	0.555
	Block 11	-0.020	0.011	-1.901	-0.573	1.000
	Block 12	-0.024	0.012	-2.032	-0.613	1.000
	Block 13	-0.027	0.011	-2.330	-0.702	1.000
	Block 6	-0.004	0.010	-0.391	-0.118	1.000
	Block 7	-0.002	0.013	-0.141	-0.042	1.000
	Block 8	0.002	0.004	0.623	0.188	1.000
	Block 9	-0.002	0.009	-0.232	-0.070	1.000
	Block 10	0.007	0.011	0.673	0.203	1.000
Block 6	Block 11	0.004	0.012	0.305	0.092	1.000
	Block 12	7.576e-4	0.013	0.058	0.018	1.000
	Block 13	0.002	0.005	0.370	0.112	1.000
	Block 7	0.006	0.010	0.597	0.180	1.000
	Block 8	0.002	0.009	0.208	0.063	1.000
	Block 9	0.011	0.008	1.282	0.387	1.000
	Block 10	0.007	0.007	1.058	0.319	1.000
	Block 11	0.005	0.010	0.457	0.138	1.000
	Block 12					
	Block 13					

**Table S5.3.4 continued from previous page**

		Mean Difference	SE	t	Cohen's <i>d</i>	<i>p</i> <sub>bonf</sub>
Block 8	Block 9	0.004	0.012	0.329	0.099	1.000
	Block 10	-2.525e-4	0.010	-0.025	-0.008	1.000
	Block 11	0.009	0.010	0.888	0.268	1.000
	Block 12	0.005	0.009	0.623	0.188	1.000
	Block 13	0.003	0.010	0.247	0.075	1.000
Block 9	Block 10	-0.004	0.009	-0.500	-0.151	1.000
	Block 11	0.005	0.011	0.433	0.131	1.000
	Block 12	0.001	0.012	0.106	0.032	1.000
	Block 13	-0.002	0.012	-0.125	-0.038	1.000
Block 10	Block 11	0.009	0.005	1.695	0.511	1.000
	Block 12	0.006	0.011	0.516	0.156	1.000
	Block 13	0.003	0.009	0.311	0.094	1.000
Block 11	Block 12	-0.004	0.011	-0.328	-0.099	1.000
	Block 13	-0.006	0.008	-0.759	-0.229	1.000
Block 12	Block 13	-0.003	0.011	-0.246	-0.074	1.000

TABLE S5.3.5: Post Hoc Comparisons - Contrast Steps (RGB)

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
135	136	-0.095	0.030	-3.146	-0.780	0.030
	137	-0.184	0.029	-6.277	-1.638	< .001
	138	-0.248	0.030	-8.249	-2.221	< .001
	139	-0.302	0.039	-7.757	-2.787	< .001
	140	-0.315	0.043	-7.398	-3.005	< .001
136	137	-0.089	0.026	-3.482	-0.742	0.010
	138	-0.154	0.027	-5.784	-1.273	< .001
	139	-0.207	0.036	-5.716	-1.735	< .001
	140	-0.221	0.040	-5.490	-1.867	< .001
137	138	-0.064	0.026	-2.499	-0.573	0.204
	139	-0.118	0.036	-3.304	-1.077	0.018
	140	-0.131	0.040	-3.311	-1.231	0.018
138	139	-0.053	0.036	-1.471	-0.493	1.000
	140	-0.067	0.040	-1.664	-0.637	1.000
	139	-0.014	0.047	-0.288	-0.139	1.000

,

TABLE S5.3.6: Post Hoc Comparisons - Numeric type

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
CHINESE	DOT	-0.073	0.032	-2.259	-0.505	0.153
	ARABIC	-0.042	0.024	-1.735	-0.270	0.510
	THAI	0.059	0.026	2.271	0.393	0.148
DOT	ARIBIC	0.030	0.032	0.951	0.221	1.000
	THAI	0.132	0.033	3.988	1.046	< .001
ARABIC	THAI	0.101	0.026	3.969	0.702	< .001

TABLE S5.3.7: Post Hoc Comparisons - Numeric type

		Mean Difference	SE	t	Cohen's d	p <sub>bonf</sub>
ARABIC	DOTS	-0.098	0.036	-2.746	-0.586	0.073
	CHINESE	-0.059	0.042	-1.424	-0.304	1.000
	THAI	-0.091	0.045	-2.026	-0.432	0.334
DOTS	CHINESE	0.038	0.026	1.469	0.313	0.940
	THAI	0.007	0.038	0.183	0.039	1.000
CHINESE	THAI	-0.031	0.040	-0.794	-0.169	1.000

## .4 Response accuracy and response bias

Figure S5.4.3.a. shows the positive relationship (rank-order correlation  $r = .61$ ,  $p < .05$ ) between response-frequency (blue) and response-accuracy (orange) for Arabic numerals in participant S1. This figure shows that, as the frequency of responding with a specific numeral increases, so too does identification accuracy (e.g., Arabic numeral 7). Similarly, as response frequency decreases, response accuracy also decreases (e.g., Arabic numeral 5). This plot clearly shows the relationship between response-frequency (strength in Luce's model) and identification accuracy.

The dotted blue line in Figure S5.4.3.a represents a response-frequency matching the number of stimulus presentations. For example, a '6' response was made nearly as often as '6' was presented, however, these responses were correct only half of the time. By contrast, a '7' response was made nearly twice as often as it was presented with 50% accuracy, showing an effect of response-bias. It is unclear what effect this response-bias had on the identification accuracy of each stimulus. This ambiguity necessitates the use of Luce's choice model. The relationship between response-frequency and response-accuracy is clearest on a participant-by-stimulus basis and has been represented for each participant and numeric-type as a scatter plot in Figure S5.4.3.b.

This figure highlights the effect of response-bias on response-accuracy. As an example, the number '7' has above average conditional accuracy, however, this participant responds with the number '7' more often than any other number. Does the conditional accuracy for number '7' reflect numeric identifiability, or, merely a bias in responding? The relationship between response-frequency and response-accuracy is clearest

on a participant-by-stimulus basis, and has been represented for each language-type as a scatter plot in Figure S5.4.3.b.

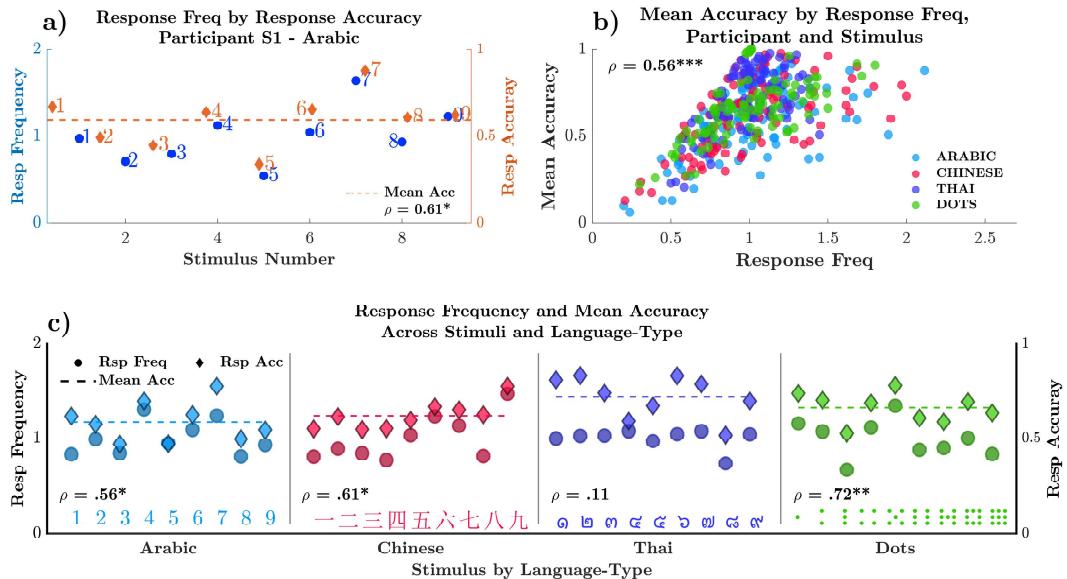


FIGURE S5.4.3: a) Response frequency by response accuracy for participant S1, Arabic numerals. b) Scatter plot depicting a positive correlation between mean stimulus accuracy and response frequency, across language-types. c) Response frequency by response accuracy for stimuli for the Arabic (left), Chinese (mid-left), Thai (mid-right) and Dot (right) language-types. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

Figure S5.4.3.b. depicts the response-frequency by response-accuracy for each stimulus in each language-type. There is a strong positive correlation ( $r = .56$ ,  $p < .001$ ) between response-frequency and response accuracy across all language-types. Figure S5.4.3.c shows the mean response-frequency and mean response-accuracy of each stimulus for each language-type. Averaging response-frequency and accuracy diminishes their correlation, however, clearly illustrates response patterns and accuracy for each stimulus. Together, these results show the positive relationship between accuracy and response-frequency, and motivate the need to remove response-bias before the assessment of the underlying mental space.

## .5 Scree analysis of bias-free MDS stress values

Scree analysis compares the multidimensional stress values (y-axis) against the number of MDS dimensions (x-axis). Scree analysis, such as this, is a subjective measure. A useful heuristic for identifying the correct number of dimensions is to look for the ‘elbow’ where an increase in dimensionality does not meaningfully improve stress values. This elbow has been identified by a marker in each plot.

### .5.1 Scree Plots

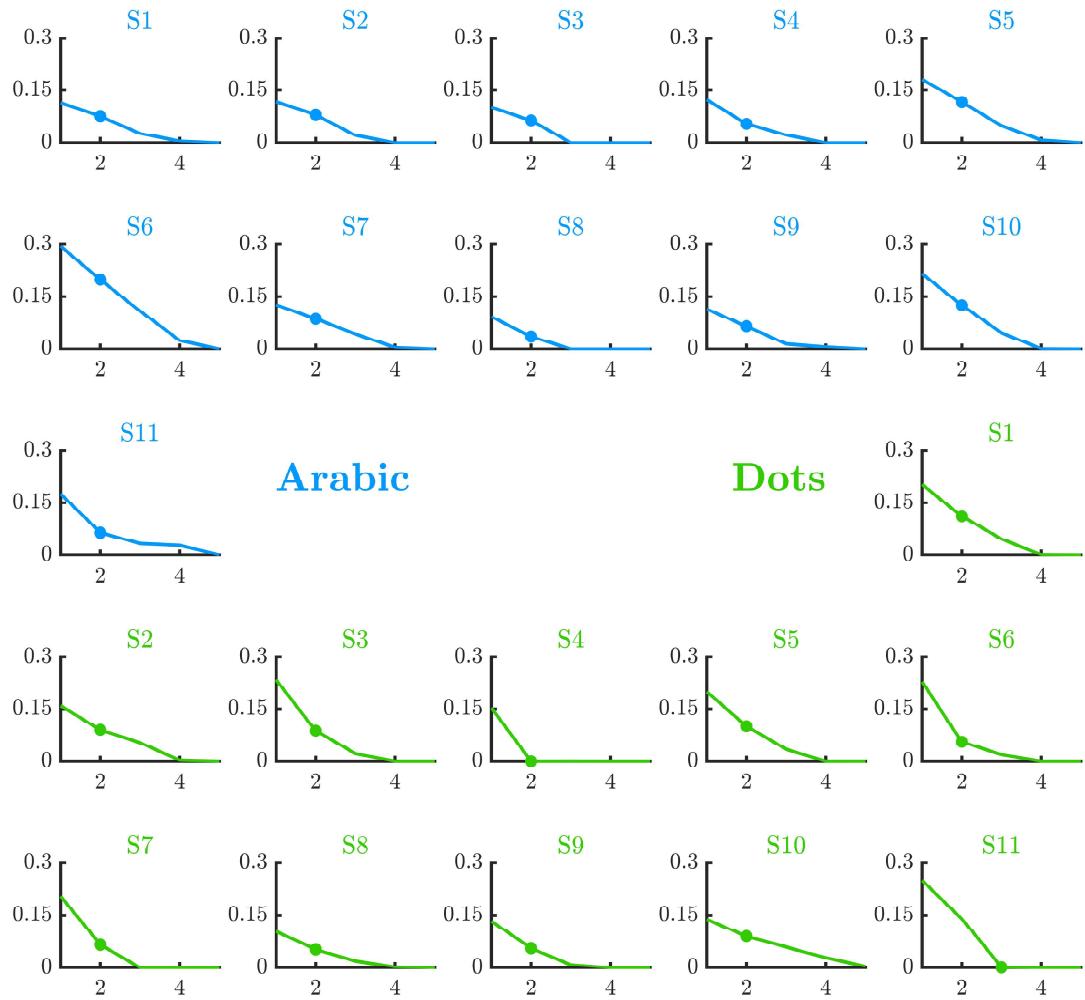


FIGURE S5.5.4: Bias-free MDS scree plots for Arabic digits (blue) and symbolic dots (green). The y-axis displays stress values, and the x-axis the number of dimensions. Markers identify the optimal number of dimensions in each scree plot.

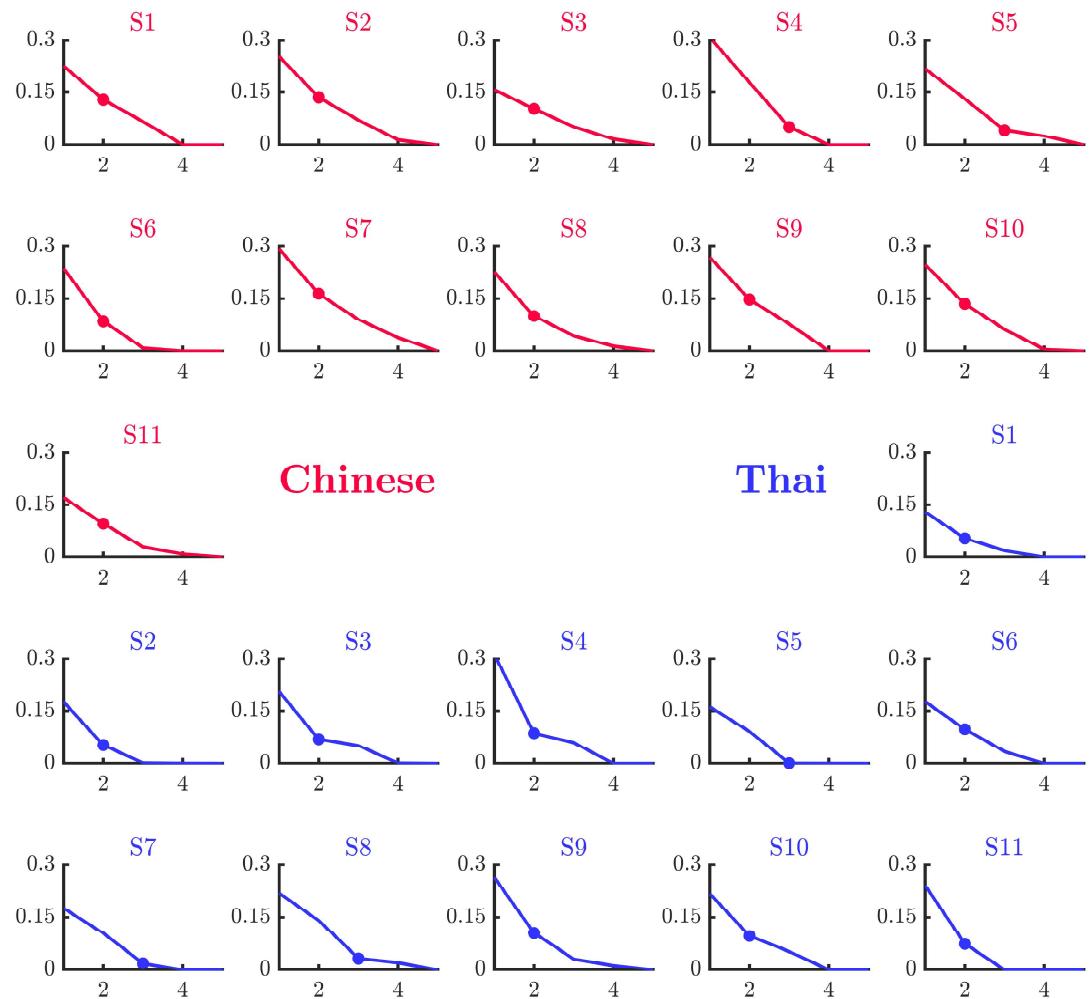


FIGURE S5.5.5: Bias-free MDS scree plots for Chinese (red) and Thai (purple) symbols. The y-axis displays stress values, and the x-axis the number of dimensions. Markers identify the optimal number of dimensions in each scree plot.

## .5.2 Individual MDS solutions

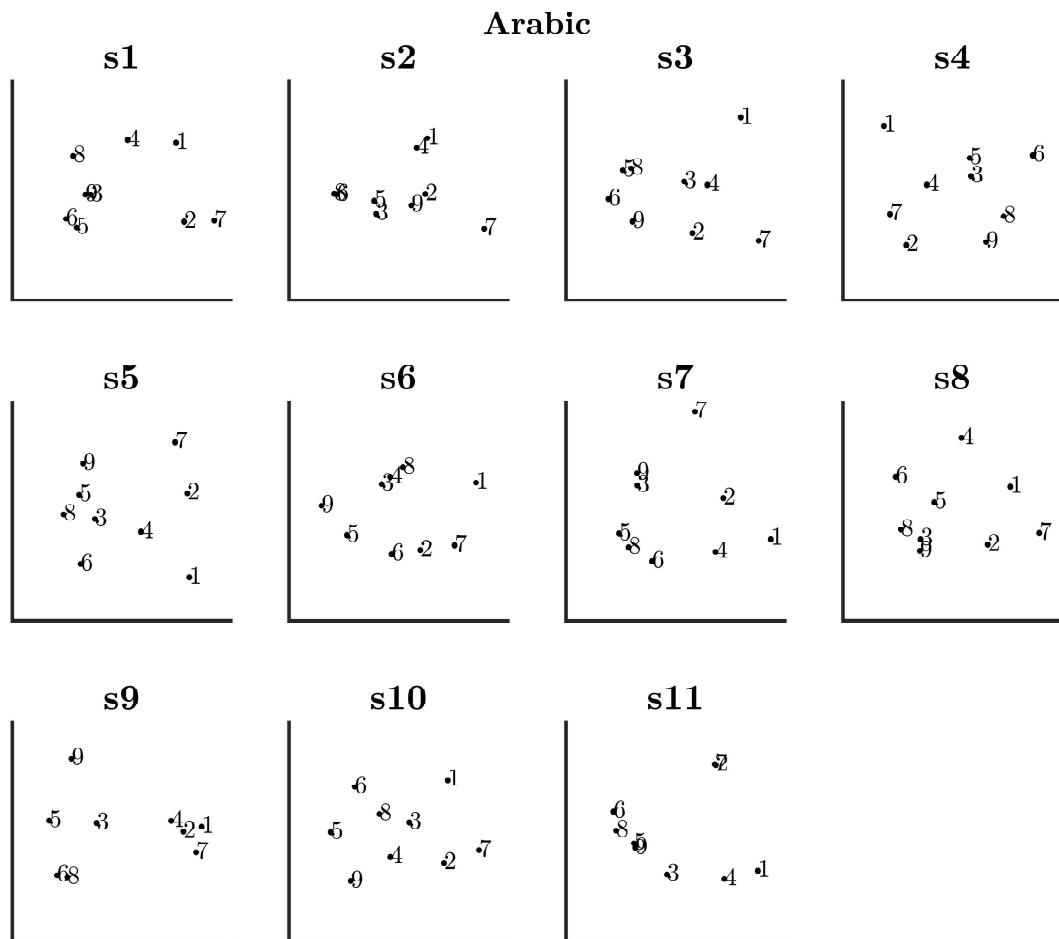


FIGURE S5.5.6: Individual bias-free MDS solutions for the Arabic digits.

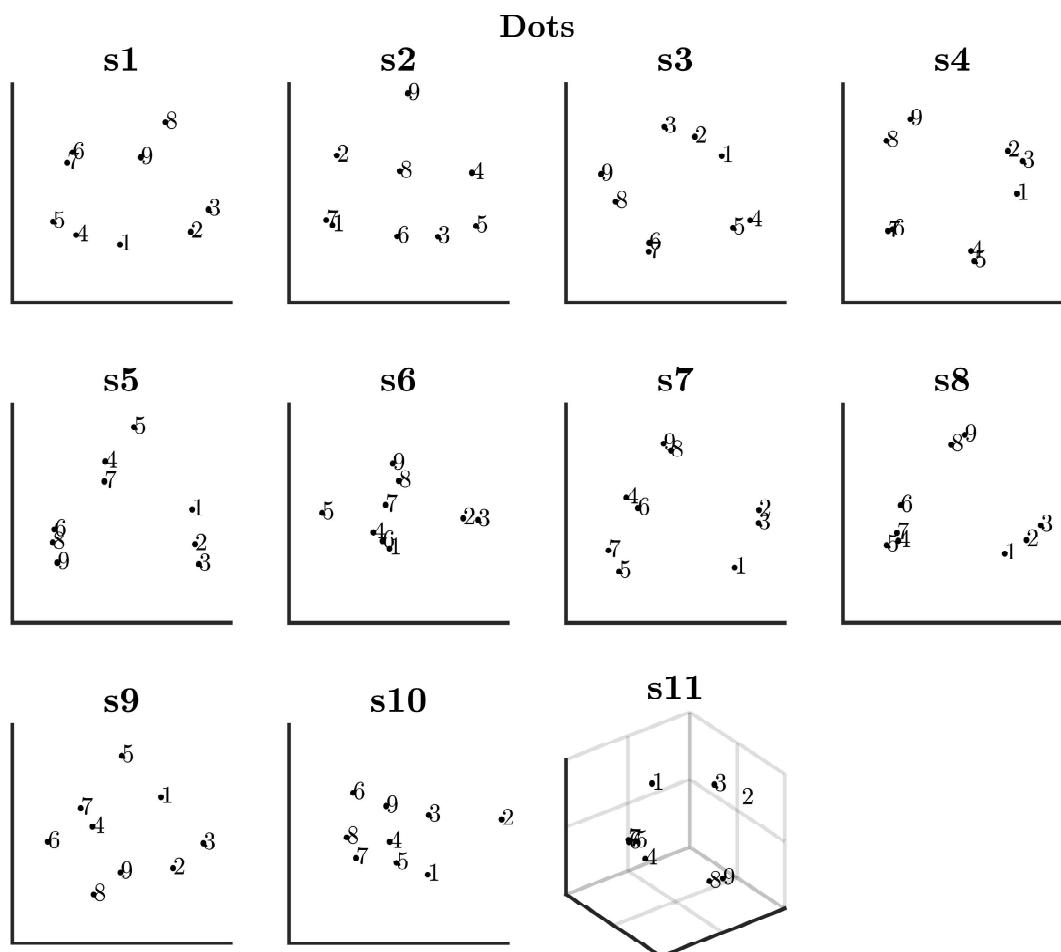


FIGURE S5.5.7: Individual bias-free MDS solutions for symbolic dots. Dots are represented by Arabic numbers for simplicity.

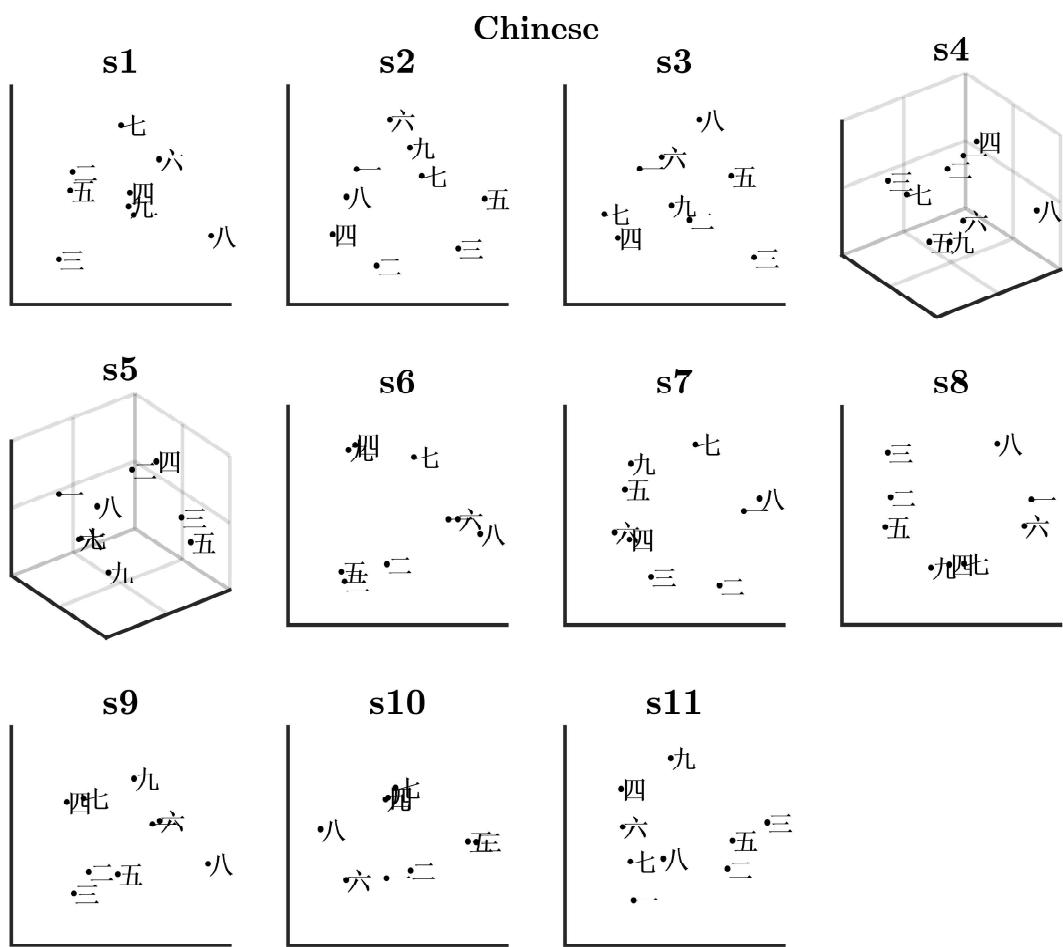


FIGURE S5.5.8: Individual bias-free MDS solutions for Chinese symbols.

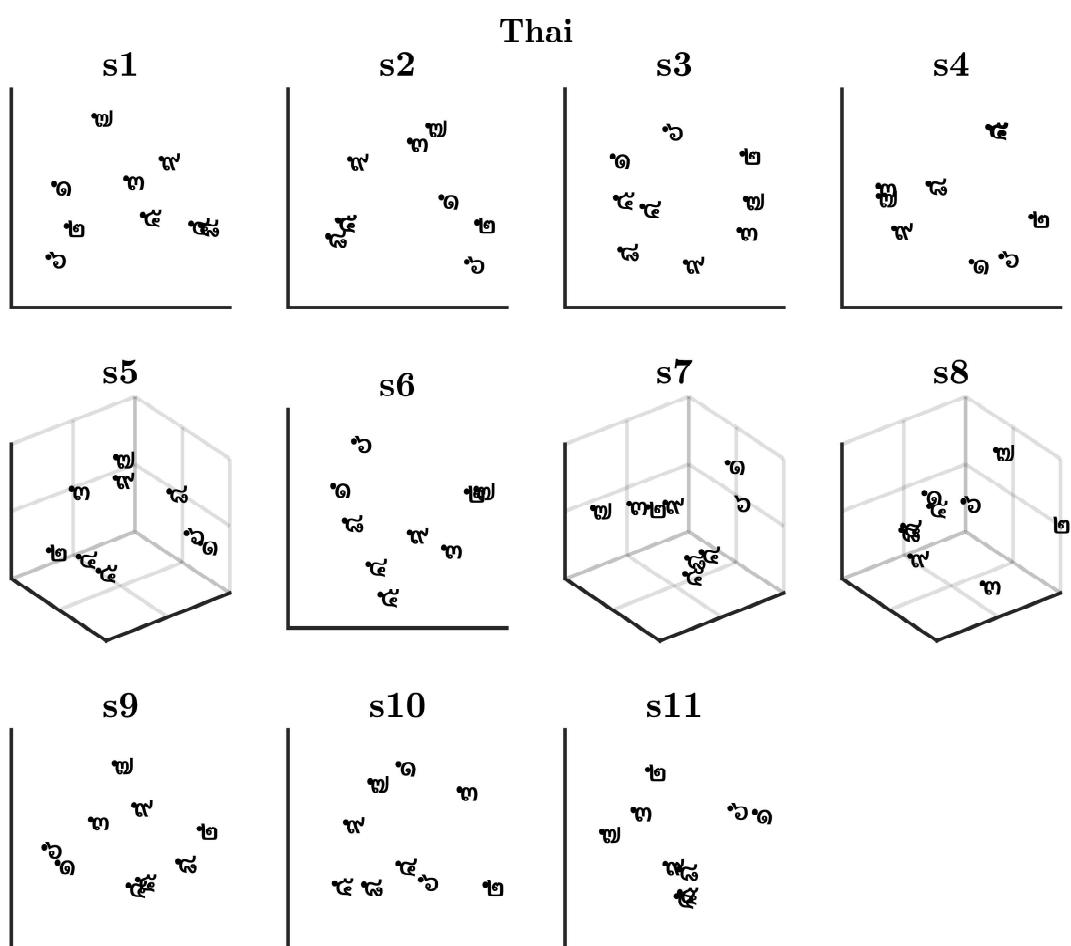


FIGURE S5.5.9: Individual bias-free MDS solutions for the Thai symbols.

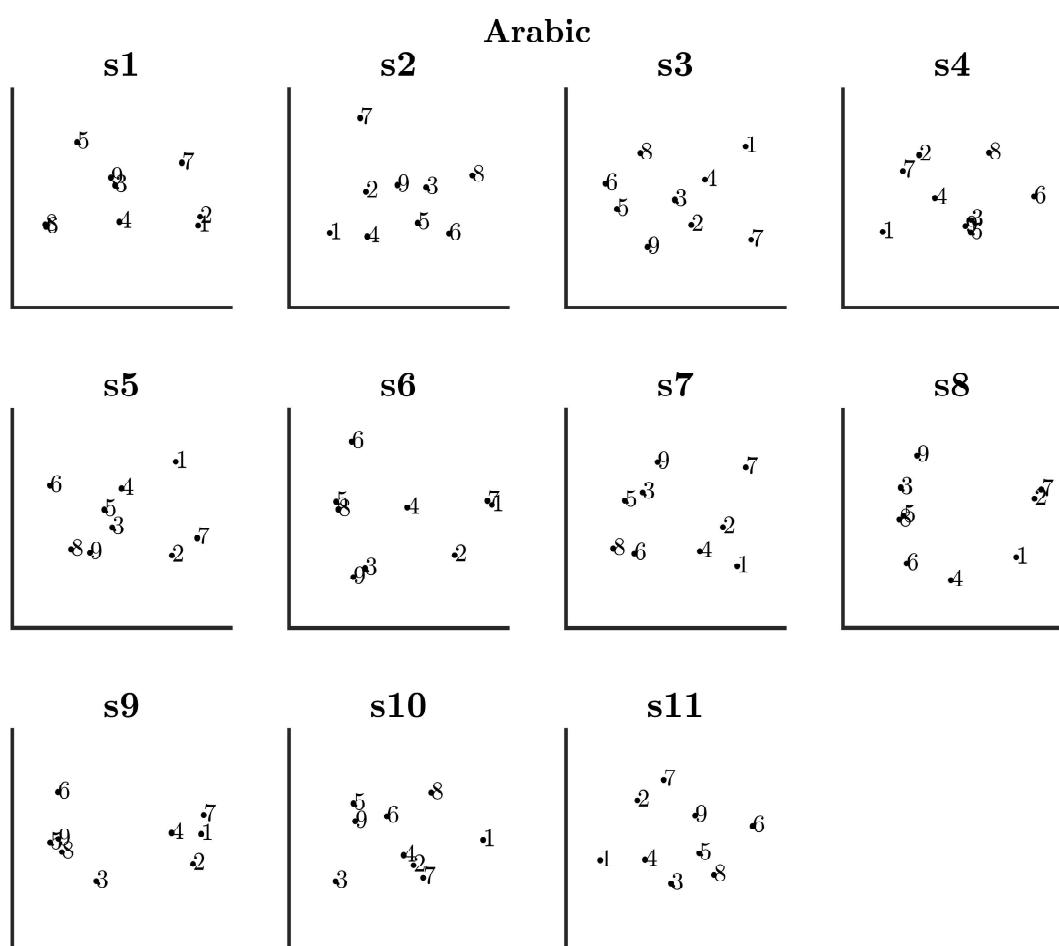


FIGURE S5.5.10: Individual biased MDS solutions for the Arabic digits.

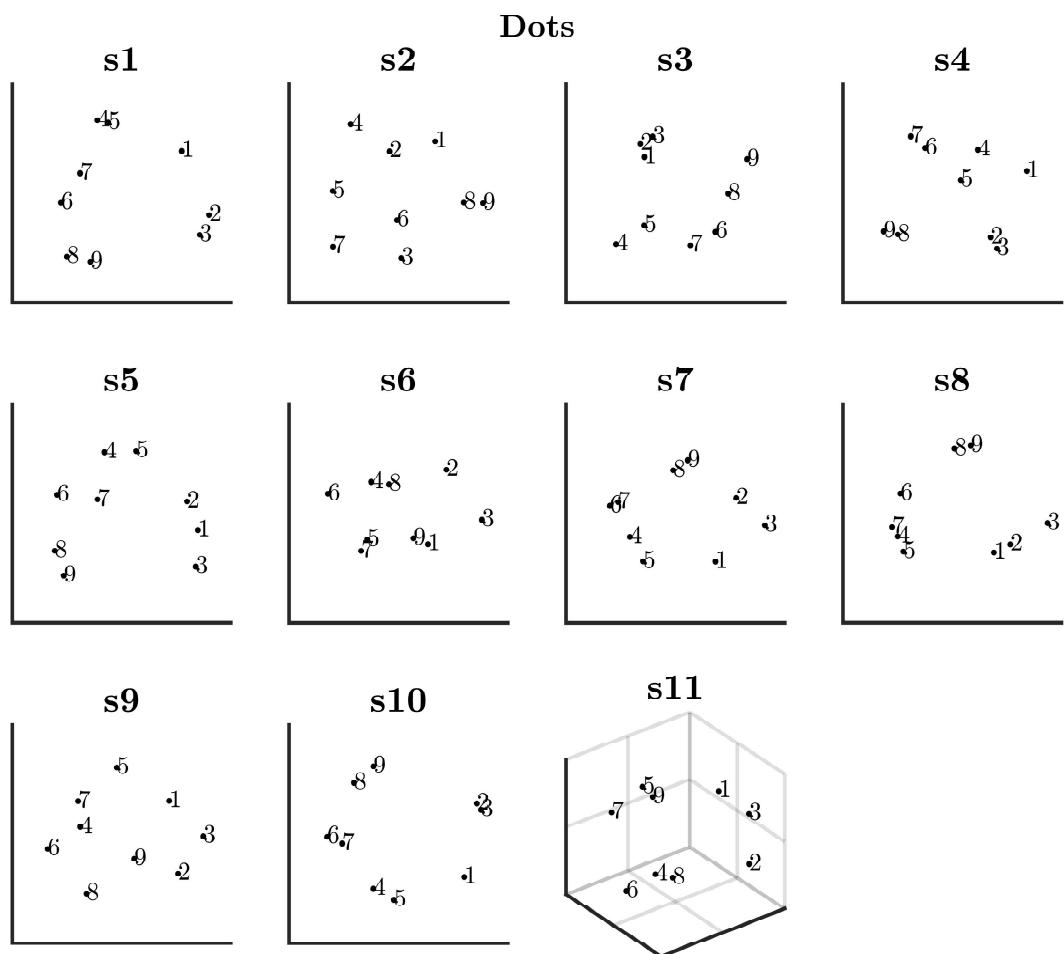


FIGURE S5.5.11: Individual biased MDS solutions for symbolic dots. Dots are represented by Arabic numbers for simplicity.

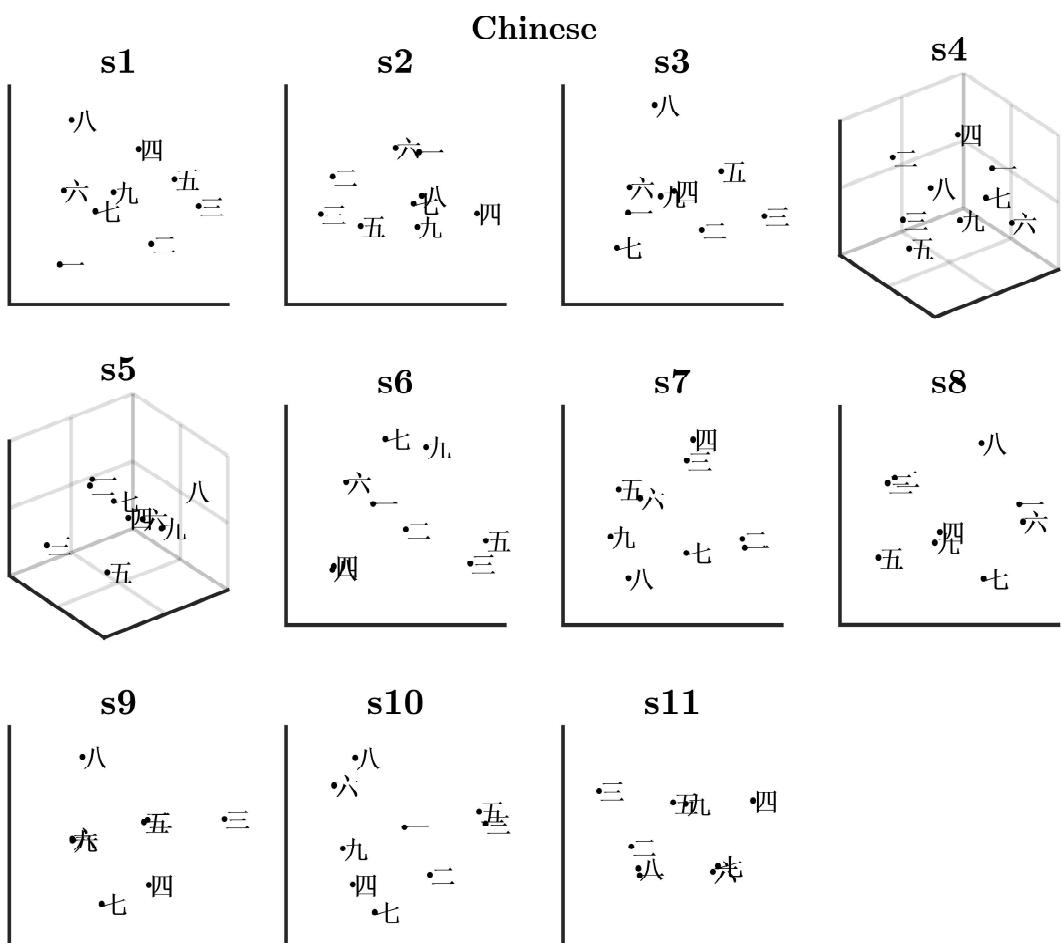


FIGURE S5.5.12: Individual bias-free MDS solutions for Chinese symbols.

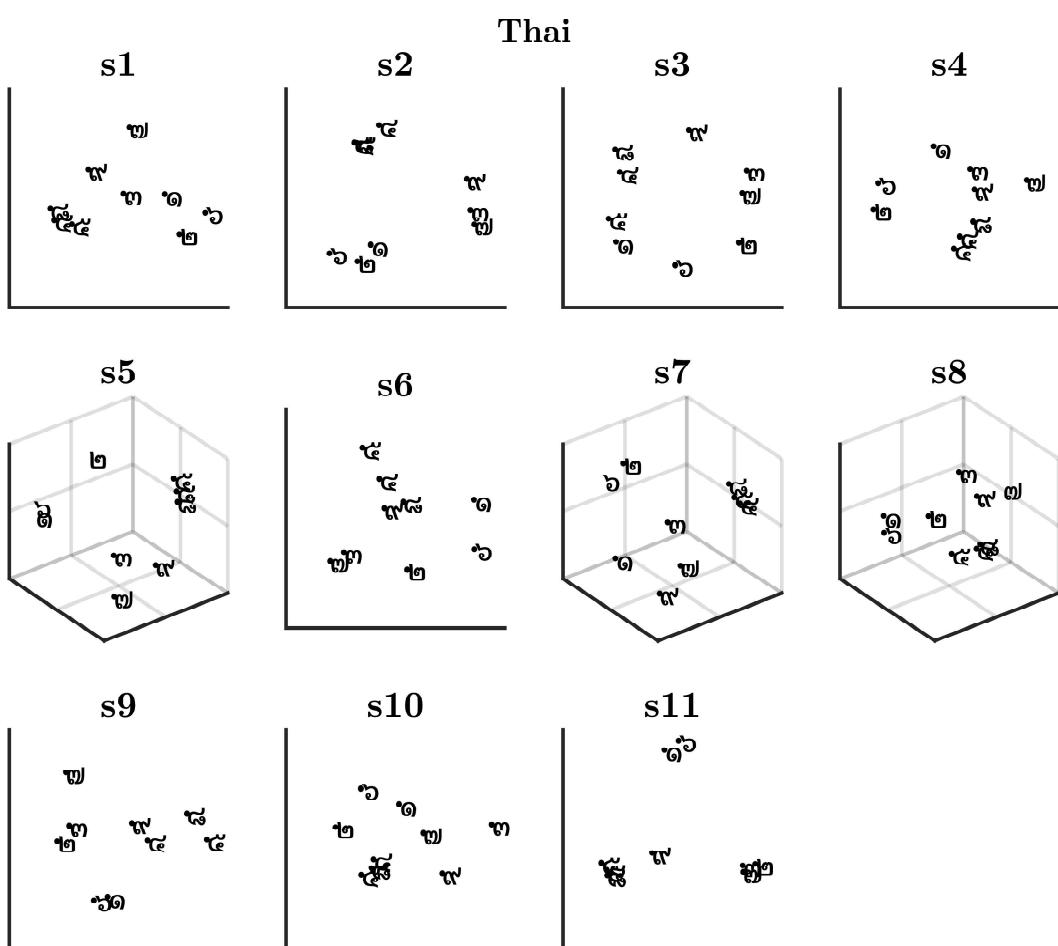


FIGURE S5.5.13: Individual bias-free MDS solutions for Thai symbols.

## .6 MDS cluster frequency heatmaps

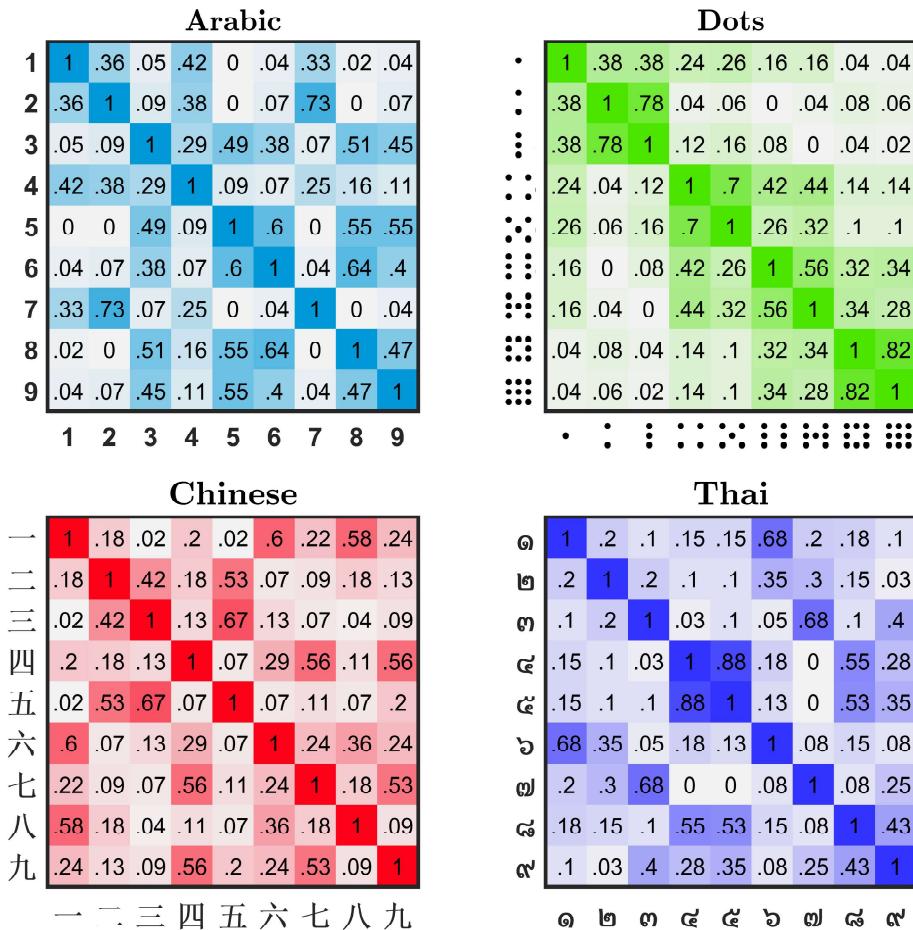


FIGURE S5.6.14: Proportional cluster-frequency heatmap for participants with two-dimensional MDS solutions, across 2–6 K-mean clusters. Larger proportions (darker colored squares) indicate items which most frequently cluster together. As opposed to the indscal heatmap in the main text, here, cluster frequencies are calculated separately for each participant, and an overall proportion presented.

## .7 Ideal observer comparison

The ideal observer analysis is a simple template matching process that compares numeric stimuli, pixel-by-pixel, to generate a confusion matrix. The ideal observer is not a model of human performance, but rather, a benchmark against which we may compare the performance of human observers (e.g., Gold et al., 1999; Eidels & Gold, 2014). The ‘ideal observer’ compares a noisy numeric stimulus to all possible templates, for example, comparing a noisy ‘1’ stimulus to the numerals ‘1–9’. The template with the best cross-correlational match over many iterations, with randomly sampled noise is selected as the ‘ideal observer response’. Normally distributed noise ( $\mu = 0$ ,  $\sigma = [1.065, .12, 1.127, 1.463]$  for Arabic, dot, Chinese and Thai numerals, respectively) is added to each numeric stimulus, until the ideal observer’s accuracy resembles the average accuracy of the participants. This process was repeated 10,000 times, per numeric-stimulus, per numeric-type, generating four confusion matrices. For direct comparison to the empirical data, Luce’s choice model was then applied to these confusion scores.

Figure S5.7.15.a displays the ideal observer MDS results for each numeric type, while Figure S5.7.15.b displays the corresponding K-mean cluster frequency heatmap. For Arabic numerals, the ideal observer and Chinese speaking cohort shared similar MDS and cluster patterns for the subset of numerals [2, 7], [5, 6], [5, 9] and [6, 9]. For non-symbolic dots, shared MDS and cluster patterns included numerals [1, 3] and [6, 7]. For Chinese numerals, shared MDS and cluster patterns included numerals [1, 6] and [2, 3]. Finally, for Thai numerals, shared MDS and cluster patterns included numerals [3, 7] and [4, 5].

The ideal observer does not share many commonalities with the empirical data of

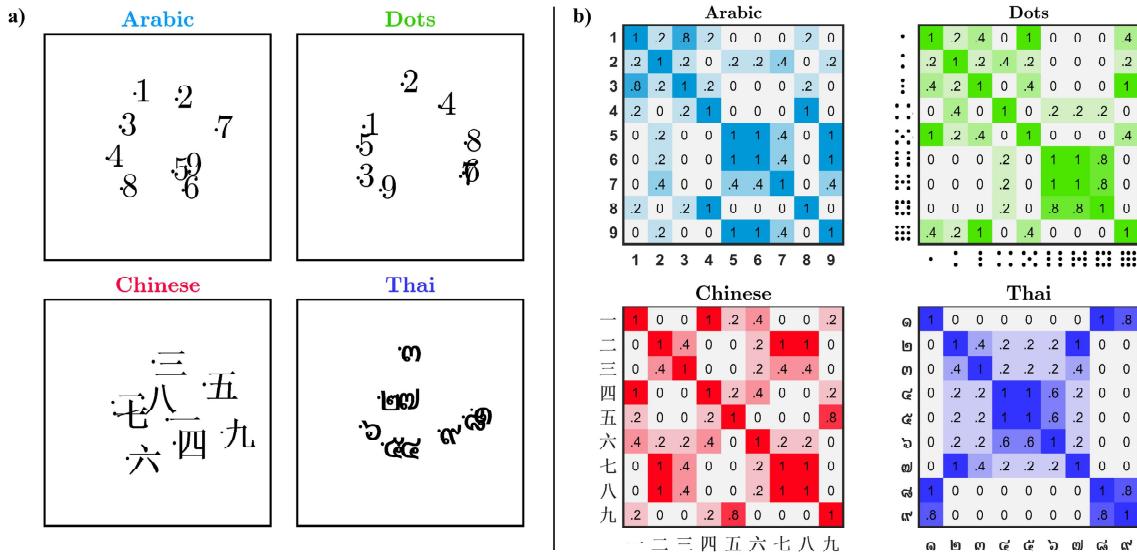


FIGURE S5.7.15: a) Ideal observer analysis bias-free MDS solutions, generated separately for each numeric-type. Non-symbolic dots are displayed as Arabic numerals in the MDS plot for clarity to the reader. b) Ideal observer K-mean cluster frequency heatmaps.

the Chinese speaking cohort. This observation may be caused by several factors. First, the ideal observer is a template matching process that only handles extant features. Arabic and Thai numerals appeared to be confused along a dimension of openness. Openness is the concave absence within a shape and it is not an extant feature, therefore it is poorly captured by our template matching process.

The Chinese speaking cohort appeared to confuse non-symbolic dots along dimensions of perceptual similarity and numerical proximity. The ideal observer holds no concept of ‘numerical proximity’, merely perceptual similarity. Lacking this extra dimensionality, it should be expected that the ideal observer results would differ from our empirical findings.

Finally, the Chinese speaking cohort appeared to confuse Chinese numerals based upon individual line-strokes, specifically the number of horizontal strokes and the

stroke curvature. These are highly specific features and would form only part of the template matching process carried out by the ideal observer. Expertise with the character set allows Chinese speakers to focus upon highly distinguishable character features — something not afforded to our simple observer. It is for this reason we speculate that a difference exists between the ideal observer and Chinese speaking cohort for Chinese numerals.

The ideal observer, as described here, is limited in a number of ways. Although a ‘better’ observer could be developed, for example, if one employed a machine learning algorithm; developing a ‘human-like’ observer was not the purpose of this exercise. Instead, this procedure was simply to provide a benchmark comparison with which to compare performance given items were confused only due to perceptual similarities. To this end, the ideal observer completed its task.